Serum metabolites and their association with risk of type 2 diabetes and cardiovascular diseases: a targeted metabolomics approach in EPIC-Potsdam

vorgelegt von

Master of Science, Nutritional Science,

Master of Public Health,

Anna Flögel

von der Fakultät VII – Wirtschaft und Management der Technischen Universität Berlin

zur Erlangung des akademischen Grades

Doktorin der Gesundheitswissenschaften/Public Health

-Dr. P.H.-

genehmigte Dissertation

Promotionsausschuss:

Vorsitzende: Prof. Dr. Jacqueline Müller-Nordhorn

Gutachter: Prof. Dr. Reinhard Busse

Gutachter: Prof. Dr. Heiner Boeing

Gutachter: Prof. Dr. Tobias Pischon

Tag der wissenschaftlichen Aussprache: 06. Mai 2013

Berlin 2013

D 83

Table of contents

List of tables

List of figures

List of abbreviations

ANOVA, analysis of variance

BMI, body mass index

CV, coefficient of variation

CVD, cardiovascular diseases

CI, confidence interval

CRP, C-reactive protein

EPIC, European Prospective Investigation into Cancer and Nutrition

ESI, electrospray ionization

FFQ, food frequency questionnaire

GC, gas chromatography

HbA_{1c}, glycated hemoglobin A1c

HDL, high density lipoprotein

HPLC, high performance liquid chromatography

HOMA, homeostasis model assessment

ICC, intraclass-correlation coefficient

KORA, COoperative health research in the Region of Augsburg

LC, liquid chromatography

LDL, low density lipoprotein

MS, mass spectrometry

NMR, nuclear magnetic resonance

ROC, receiver operating characteristic

RR, relative risk

SD, standard deviation

SE, standard error

TOF, time of flight

UPLC, ultra-performance liquid chromatography

VLDL, very low density lipoprotein

1. Introduction

1.1 Background and aim of the thesis

The increased burden of chronic diseases such as type 2 diabetes and cardiovascular diseases (CVD) is adversely affecting our well-being and health system budgets (1-4). It may be attributable to changes in human behavior and adaptation to a Western lifestyle, which is often characterized by imbalance of energy intake and expenditure, and leads to development of metabolic disorders including obesity, dyslipidemia, insulin resistance, and hypertension (5). Classical epidemiologic studies have identified many important risk factors that are linked to a higher risk of chronic diseases, e.g. dietary, lifestyle, and genetic factors, and also metabolic markers, some of them may further facilitate risk prediction (6-8). However, these classical approaches were to a certain degree limited, as they only focused on single risk factors, and it remains hard to grasp the complexity of biological mechanisms and the multifactorial origin that underlie chronic diseases (9-11). Rise of the 'omic'-sciences in analytical biochemistry offers more comprehensive approaches to study a whole system rather than a single risk factor, e.g. metabolomics describes the set of all low-weight molecular compounds present in a biological sample, and may better mirror the phenotype of disease (12-14). When metabolomics is combined with the strength of an epidemiologic study design, it may help unravel the black box of biological mechanisms and pathways that are linked to chronic diseases (9; 10). Furthermore, when applied to a prospective cohort study it may enable identification of predictive metabolites which can be used to identify high-risk individuals in advance of the disease onset to administer adequate and individualized prevention strategies. However, when metabolomics is adopted to large population-based studies, several challenges come up. First of all, the technical variation of biochemical methods used in metabolomics is usually higher compared to measurements of single biomarkers. This may introduce random measurement error, which can adversely affect validity of risk estimates (15-18). Second, the biological variation of metabolites has not yet been well examined, and particularly high within-person variation over time and low reliability may limit the use in epidemiologic studies that usually rely on a single blood sample collection (18-20). Last, the utility of metabolites needs to be evaluated, i.e. how well they reflect chronic diseases, e.g. in terms of strength, consistency and independency of risk association (18; 21; 22). Clinical usefulness of metabolites may further be characterized by their predictive value, e.g. by measures of discrimination and calibration (22; 23). Previously, metabolomics has rarely been adopted to epidemiologic studies, and most studies used a cross-sectional design which does not enable to identify predictive metabolites (22). A milestone paper was published in 2011 (24), when for the first time metabolomics was adopted to a prospective cohort study: thereby, in the Framingham Offspring cohort

and the Malmö Diet and Cancer study a set of five amino acids out of 61 metabolites was predictive for type 2 diabetes. This paper demonstrated the potential of metabolomics in an epidemiologic setting.

Against this background the objective of the present thesis was to adopt a targeted metabolomics approach, covering a broader set of metabolites, to the EPIC-Potsdam study with a large number of study participants. Thereby, different challenges of adopting metabolomics to an epidemiologic setting were addressed: first of all, technical variation of metabolites was investigated; second, biological variation and reliability of metabolites over time were studied; and finally, the utility of metabolites was evaluated by investigating their association with risk of type 2 diabetes and CVD, and their potential for risk prediction in the frame of a case-cohort study in EPIC-Potsdam. The present thesis represents an innovative approach in the fields of molecular and systems epidemiology, and noteworthy, it is one of the first available studies that used metabolomics in the frame of a large prospective cohort study.

1.2 Public health relevance

In line with prolonged life expectancy, incidence and prevalence of chronic diseases, among them CVD and diabetes mellitus, have increased in the past; thereby, adversely affecting life quality. Among the elderly German population, CVD, such as myocardial infarction and stroke, were the most common reasons for hospitalization in 2009 and the major burden for health care expenses (25). In 2008, the German health care system spent approximately 37 billion euro on CVD (4). Despite, CVD is the top cause of death in Germany and worldwide (2; 3). Furthermore, CVD represent a long-term complication of diabetes mellitus. About 50% of diabetics die of CVD (1). A recent survey from the Robert Koch Institute (DEGS 1) suggests that of all Germans between 18 and 79 years, 4.6 million have been diagnosed with diabetes mellitus, and an additional 1.3 million cases are estimated to be currently undiagnosed (26). This implies an age-adjusted increase of diabetes mellitus prevalence by 24% from 1998 to 2011. In 2008, health care expenses for diabetes mellitus were 6.3 billion euros in Germany (4). Worldwide, about 347 million people are estimated to have diabetes mellitus. In fact, it not only represents a major burden in high-income countries but also in low- and middle income countries (1). Particularly, type 2 diabetes (non-insulin dependent diabetes mellitus) is of public health interest, as it represents about 90% of all diabetes mellitus diagnoses (5).

The increased burden of chronic diseases may be partly attributable to prolonged life expectancy and be determined by non-modifiable factors, e.g. sex, ethnicity and genetic predisposition, to name a few; nevertheless, type 2 diabetes and CVD are also a result of changes in human behavior and lifestyle, such as physical inactivity, hyper-caloric diet, obesity, and smoking (5). Nevertheless, from a different perspective, these modifiable factors also offer a great chance to develop and implement prevention strategies to counteract the global burden of type 2 diabetes and CVD (6; 27; 28). In this context, it is of great interest to better understand the biological mechanisms that underlie these chronic diseases and to

identify individuals at high-risk of the diseases in advance of the actual disease onset. Thereby, intervention strategies could be applied early on, to enhance life quality and reduce health care expenses.

1.3 Proceeding from single risk factors to systems epidemiology

Several risk factors have been associated with higher incidence of type 2 diabetes and CVD. A risk factor is by definition of Brotman et al (29) "a variable with a significant statistical association with a clinical outcome". **Table 1** further summarizes different terms of risk factors. Risk factors may be causally or non-causally related to the outcome. In addition, their association with disease risk may be independent or non-independent of other risk factors. To study disease etiology, causal risk factors are of primary interest. However, non-causal risk factors may also be useful, e.g. for disease prediction.

Table 1. Different terms and definitions of risk factors (modified from Brotman et al. 2005 (29))

Risk factor	A variable with a significant statistical association with a clinical outcome.
Independent risk factor	A risk factor that retains its statistical association with the outcome when other established risk factors for the outcome are included in a statistical model (also referred to as an independent predictor.)
Non-independent risk factor	A risk factor that loses its statistical association with the outcome when other established risk factors for the outcome are included in a statistical model.
Direct causal risk factor	A causal factor that directly impacts the outcome (or the likelihood of the outcome). Causal factors are not defined statistically but are defined experimentally in that they are known to affect the outcome.
Non-direct causal risk factor	A causal factor that impacts the outcome (or affects its likelihood of occurrence) by changing a direct causal factor. If the direct causal factor is prevented from changing, then changes in the outcome will not be produced.
Non-causal risk factor	A variable that is associated with a causal factor but does not have a causal relationship with the outcome. However, by virtue of its association with a causal factor, a non-causal risk factor is indirectly associated with (and predictive for) the outcome.

Causal and non-causal risk factors that are used in current prediction models for type 2 diabetes and CVD are summarized in **Table 2** (22). These risk factors include non-modifiable characteristics, such as age, sex, family history, and genetic factors, as well as modifiable factors, such as obesity, low levels of physical activity, smoking and hyper-caloric diet. In addition, certain clinical parameters used for diagnosis and monitoring, may as well be useful to determine risk of developing the disease. In the prospective Whitehall II study, for example, it was shown, that clinical standard parameters of glucose metabolism, such as fasting glucose and homeostasis model assessment (HOMA)-insulin sensitivity, had been altered already three to six years before the actual onset of type 2 diabetes (30). Other biomarkers may reflect common pathophysiologic mechanisms involved in the etiology of type 2 diabetes and CVD, e.g. C-reactive protein (CRP), an acute phase protein, that serves as a marker of subclinical inflammation

(31; 32), or alterations of triglycerides and high density lipoprotein (HDL)-cholesterol as surrogate for dyslipidemia (33).

Table 2. Summary of common risk factors of type 2 diabetes and cardiovascular diseases
Age
Sex
Ethnicity
Family history
Socioeconomic factors
Lifestyle factors
Physical inactivity
Dietary components (red meat, fats, fruits, vegetables, fiber, coffee)
Smoking
Alcohol intake
Anthropometric factors (weight, BMI, waist circumference, waist-to-hip ratio, waist-to-height ratio)
Blood pressure (systolic, diastolic, use of medication)
Metabolic factors
Blood lipids (triglycerides, total-, HDL- and LDL-cholesterol, use of medication)
Glucose metabolism (fasting glucose, 2-h glucose, fasting insulin, HbA_{1c}, HOMA-IS, HOMA-BF)
Liver enzymes (Alanine-amino-transferase, Gamma-glutamyl-transferase)
Inflammatory markers (C-reactive protein, leukocyte count)
Other factors (adiponectin, fetuin-A, uric acid, natriuretic peptides)

Modified from Herder et al, 2011 (22). Abbreviations: HbA_{1c}, glycated hemoglobin A1c; HDL, high-density lipoprotein; HOMA, homeostasis model assessment; LDL, low density lipoprotein.

Although, type 2 diabetes and CVD are predictable to some extent through these anthropometric and lifestyle factors, and clinical parameters, their metabolic pathways are not yet fully understood (10). Classical epidemiologic studies have usually addressed single risk factors and related them to chronic disease endpoints. Thereby, a number of important risk factors of type 2 diabetes and CVD were identified. However, chronic diseases are very challenging in respect of their complexity of pathophysiology, their heterogeneity in phenotypes, and their multifactorial origin; which stands in contrast to the classical inherited diseases caused by single gene mutations (11). To better address the multifactorial origin of chronic diseases and to enhance the understanding of the biological mechanisms, a more comprehensive approach seems to be necessary that does not only focus on single risk factors.

The rise of the '–omic'-sciences in basic biology research (**Figure 1**), i.e. genomics (genetic predisposition), epi-genomics (gene-environment-interaction), transcriptomics (expression of genes), proteomics (enzymes), and metabolomics (substrates and products of enzymes), has enabled more systemic approaches to study biological mechanisms and pathways.

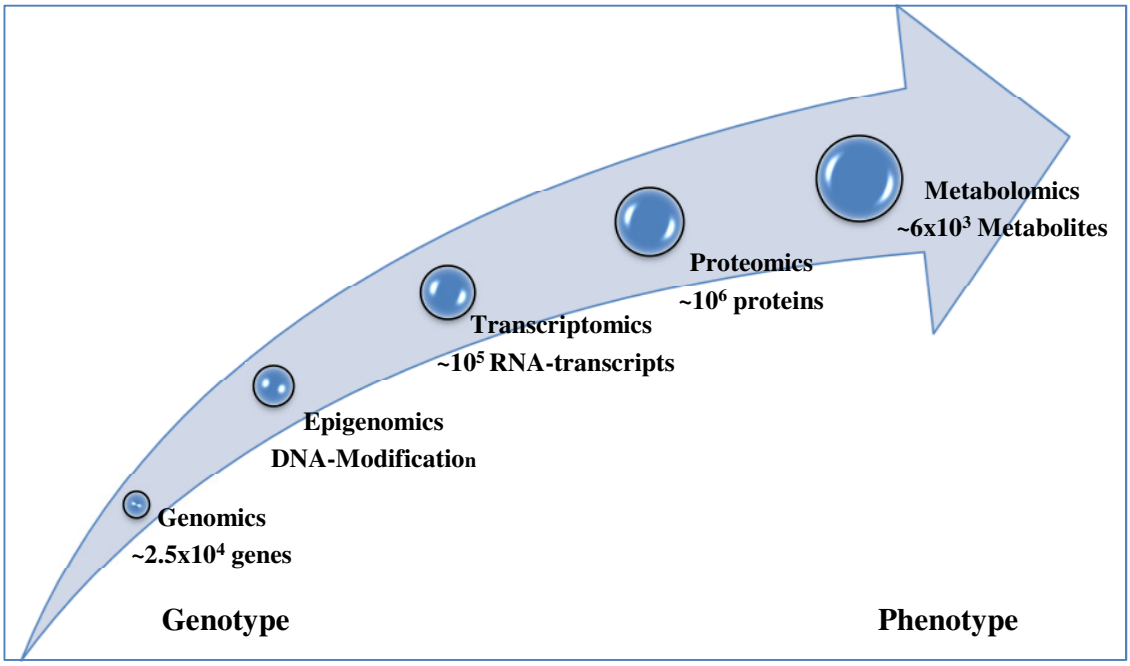


Figure 1. Hierarchy of the 'omic'-sciences.

In principal, the –'omic' sciences focus on entire systems rather than on single risk factors, e.g. genomics focuses on whole genome sequencing rather than on candidate genes, proteomics on the full protein set expressed in a system and not on single enzymes, and metabolomics studies the substrates and products of these enzymes that are present in a system. An important feature of these systemic approaches is their exploratory nature, covering a whole system with multiple factors. To date, however, they have mostly been applied in the context of experimental studies, e.g. animal studies or small scale clinical trials. Nevertheless, when combining observational study design with this innovative methodology from molecular biology, it might be possible to unravel the "black box" of biological mechanisms and pathways that underlie the observed associations between environmental and lifestyle exposures and chronic disease risk (10; 14; 34). To describe this concept the term 'systems epidemiology' has recently been proposed by Lund et al. (9), adopted from 'systems biology' which is by definition 'a discipline that seeks to determine how complex biological systems function by integrating experimentally derived information through mathematical and computational solutions' (35). Systems epidemiology has to be distinguished from systems biology though, as it further complements the strength of epidemiologic study design, e.g. nested case-control and case-cohort studies, for optimal application of these new molecular technologies in a population setting (9). In the frame of a prospective cohort study, for example, it may be possible to identify early and intermediary metabolic markers that mirror different stages of progression from a

healthy state up to disease onset. Ultimately, these markers may as well help to identify high-risk individuals, and facilitate to administer adequate and personalized prevention and treatment programs. Among the challenges of systems epidemiology, will be to integrate data from questionnaires with multiple level data of the biological scale (9). Moreover, systems epidemiology will require a shift in paradigm of research design from hypothesis-driven- towards exploratory approaches.

1.4 'Metabolomics meets epidemiology'

In this context, metabolomics has raised particular interest of clinicians and epidemiologists (12; 34; 36). Metabolomics is the simultaneous study of all (measurable) low-weight molecular compounds present in a biological system, e.g. a cell, biological fluid, tissue, organ or organism (37). Thereby, small compounds with less than 1 kDa of molecular weight are usually defined as metabolites (e.g. amino acids, mono- and di-saccharides, lipids, organic acids, steroids, nucleotides), whereas the entity of metabolites represents the metabolome. Metabolites can further be categorized based on their origin as endogenous (e.g. derived from de-novo synthesis) or exogenous (e.g. ingested with diet or medication) (38). In contrast to genomics, transcriptomics and proteomics, which are more reflective of the genotype and consequent enzyme expression; metabolomics focuses on substrates, intermediates and end products of metabolic pathways, and more directly mirrors cellular processes, such as enzyme activities and substrate fluxes. It provides a unique snapshot of metabolic phenotypes which may be reflective of genetic predisposition and modulation as well as environmental and lifestyle exposures. Ideally, it depicts metabolic perturbations linked to chronic disease risk (37; 39). An additional advantage of metabolomics compared to the other '-omic sciences' is that the human metabolome consists of approximately $6 \cdot 10^3$ metabolites (40), and thereby, it is considerably smaller than current estimates of the human genome (~$2.5 \cdot 10^4$ genes), transcriptome (~$1 \cdot 10^5$ RNA-transcripts) and proteome (~$1 \cdot 10^6$ proteins) (12).

1.4.1 Technical evolution of metabolomics

The rise of the rather "young science" metabolomics is closely linked to advancements in analytical technologies from clinical chemistry during the last century. Therefore, this section presents a little excursion into history. The key idea that a biological fluid with its ingredients may reflect health or disease states goes back to the ancient world or earlier (41). In the Middle Ages, the color, taste and smell of urine was frequently used for diagnosis of several medical conditions. In the 18th century, Dobson, a medical doctor from Liverpool, identified sugar as the sweet substance in urine of diabetes mellitus patients (42). Thus, the idea that chemical patterns are linked to disease risk has existed for longtime.

The technical basics for metabolomics, however, only go back to the beginning of the last century, when the principle of a mass spectrometer was discovered by Thomson in Cambridge. In the 1940s, nuclear magnetic resonance (NMR) was independently discovered by Bloch at Stanford and Purcell at Harvard with different techniques, and published simultaneously in the same issue of Physical Reviews in 1946 (43; 44). By then, the principals of the two detection techniques of metabolomics, mass spectrometry (MS) and NMR spectroscopy, had been discovered and ever since been applied. With these techniques it was made possible to distinguish between different atomic nuclei based on their magnetic properties (NMR) or mass to charge ratios of their ions (MS), and metabolites could now individually be characterized and identified in a sample. By the 1980s, NMR spectroscopy was able to identify metabolites even in unaltered biological fluids (41). In contrast to NMR spectroscopy, which does not necessarily require sample preparation, MS detection often involves prior separation of the different metabolite classes. In this context, the development and advancements of chromatographic separation methods in the 1960s, e.g. gas chromatography (GC) or liquid chromatography (LC), represented a big milestone for MS. In fact, these separation techniques are nowadays frequently used together with MS detection in metabolomics studies. The term “metabolic profile” was first introduced by Horning from Houston in 1971 (45) who used a GC separation coupled with a MS detection method to study different metabolites in urine samples. At the same time the laboratory of Pauling and Robinson studied variation of the human urine metabolome, although not yet called the latter, and measured 280 metabolites in response to a standardized diet (46). In his paper, Linus Pauling already pointed out the major challenge of metabolomics: “The problem is to make a quantitative determination of the amounts of each of several hundred substances present in a sample of urine, blood, spinal fluid, breath, saliva, or tissue.” (46). Until today, the simultaneous measurement of many compounds that are chemically diverse is one of the most difficult technical tasks of metabolomics.

Although metabolomics research was becoming more and more popular, the word itself was not introduced until 1998, when Oliver from Manchester suggested the term “metabolome” in addition to genome and proteome to describe the complete set of metabolites (47). Finally, the term “metabonomics” was used by Nicholson in 1999 to describe the study of the “metabolome” (48). Nowadays, the term “metabolomics” has become more frequently accepted than “metabonomics”. In 2004, the Human Metabolome Project was initiated by Prof. Wishart from the University of Alberta, which aims to characterize the complete human metabolome (40; 49). Today, the Human Metabolome database represents a comprehensive free access library of thousands of metabolites. During the last two decades, metabolomics research has rapidly increased, with particular focus on advancement of analytical technologies. Thus, it is about time to use these metabolomics technologies in epidemiologic studies.

1.4.2 Targeted and untargeted metabolomics

The discipline of metabolomics can be divided into two different approaches, which may go back to the two analytical techniques, NMR spectrometry and MS; however, they are also based on distinct theoretical concepts. The concept of untargeted (or non-targeted) metabolomics describes a 'top-down' approach, or as Nicholson stated a tactic "without prejudice" (48). Thereby, the sample is analyzed without major preparation and any compound present in the sample is being detected regardless of chemical class of metabolites (12). This leads to the simultaneous detection of a high number of peaks and consequent metabolites (order of magnitude: 10^3). NMR spectroscopy- sometimes also referred to as the "universal detector"- is widely applied in untargeted metabolomics (50), although MS methods are nowadays also frequently used (12); however, they could be affected by differential ionization efficiency, and therefore, may not detect completely unbiased (48). In contrast to untargeted metabolomics, targeted metabolomics describes a more focused 'bottom-up' approach. Thereby, a predefined set of metabolites is quantified within a sample, by monitoring specific chromatographic retention times of the analytes as well as use of predefined internal standards, often by MS (13). That is why targeted metabolomics involves a considerably lower amount of metabolites (order of magnitude: 10^2) than untargeted metabolomics. Metabolites can be selected based on their biological relevance, e.g. as part of certain pathways that may be involved in the development of chronic diseases, or based on the robustness of their measurements. Thus, the end results may be more confident, as the researchers a-priori define what should give rise to the signals (13). Both techniques have their strengths and limitations which are summarized in **Figure 2**. The advantages of untargeted metabolomics, on the one hand, include that it involves a large number of metabolites and it does not depend upon a priori assumptions; whereas targeted metabolomics, on the other hand, usually generates more robust data, particularly, in terms of sensitivity, and the use of internal standards enables absolute quantification of the metabolites. In terms of interpretation, targeted metabolomics data may be easier to handle, as it provides the identity and often also the quantity of metabolites. Bain et al. suggested that untargeted metabolomics approaches are better suited for in-vitro applications with a controlled setting than in studies with whole animals or humans (12). In the end, it should be decided based on the research question which approach is more appropriate.

When we look at the theoretical concepts that underlie these metabolomics techniques there is a fundamental difference. The untargeted metabolomics approach is of pure exploratory nature and seeks objectivity at its best in terms of selection of metabolites; whereas targeted metabolomics uses prior information, and thus, is closer to the classical hypothesis-driven studies (**Figure 2**). Nevertheless, both techniques involve a large number of metabolites, and therefore, should both be considered as exploratory compared to classical epidemiologic studies that look at single risk factors. This aspect has to be taken into mind, when analyzing metabolomics data with multiple statistical tests.

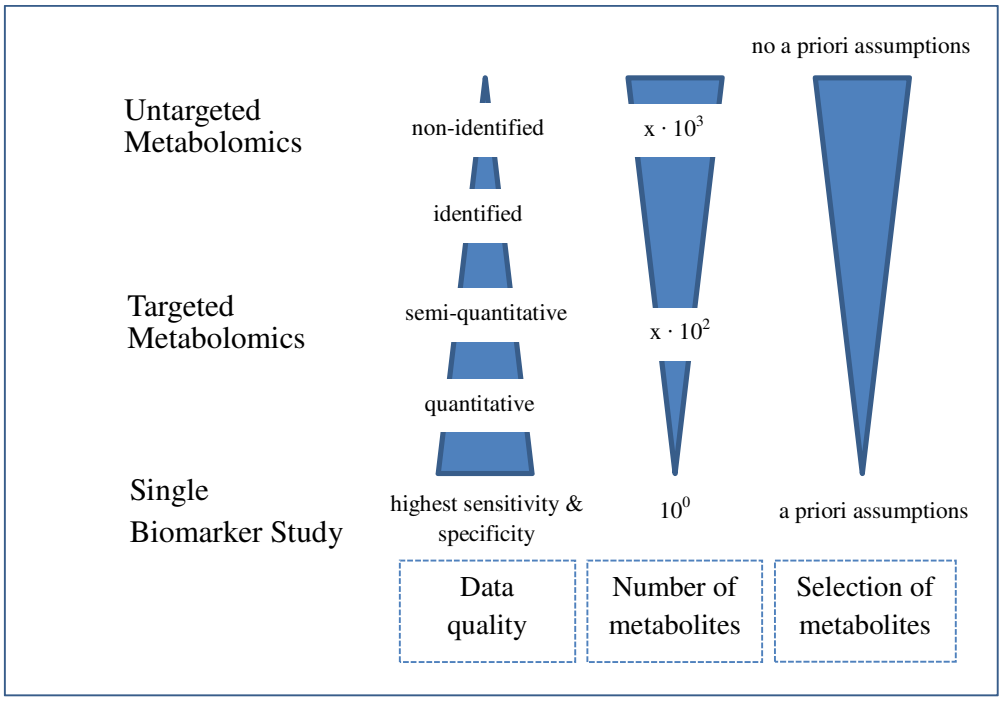


Figure 2. Comparison of untargeted and targeted metabolomics with single biomarker studies.

1.4.3 Challenges of adopting metabolomics to epidemiology

Ultimately, metabolomics may be adopted to epidemiology, to investigate associations between dietary, environmental and lifestyle exposures, and metabolites with disease risk, and may be useful for discovery of biomarkers of exposure or effect (21; 36; 51). Therefore, it seems appropriate to take similar steps in metabolomics studies as in classical biomarker studies. In the context of the present thesis, the definition of a biomarker by Paolo Vineis seems best as: "any substance, structure or process that can be measured in the human body or its products and may influence or predict the incidence or outcome of disease" (52). Before a biomarker can be applied in epidemiologic studies, it needs to be validated. Usually, biomarker validation studies, require different stages which were summarized by Shurubor (18) as follows: first, it is important to study technical variation, i.e. measurement error of the laboratory method; second, to investigate between- and within-person biological variation, and last, to confirm utility, which implies that the analyte corresponds to the phenotype of interest, e.g. disease risk. These stages should further be conducted in the proposed order and be independent of each other (18). This concept implies, that a broad understanding of the limitations of the measurements is necessary, before an analyte can be used to distinguish between different conditions of interest (18). When we adopt this theoretical concept of classical biomarker studies to metabolomics studies several issues arise.

Technical variation

As pointed out in **section 1.4.1** the evolution of metabolomics is closely linked to technical progress of NMR and MS technologies during the last decades. Advancement of these high-throughput technologies is, however, still ongoing. At the same time, these technologies are applied to study metabolites and their role in health and disease. Thus, in metabolomics studies it is particularly essential to understand the limitations of the measurements (18). As metabolomics aims to simultaneously study a high number of metabolites of very similar or different nature, the quality of the measurements usually represents a compromise between the individual metabolites. This applies to sensitivity (identification of "true positives") and specificity (identification of "true negatives") for identity of metabolites, as well as precision (i.e. reproducibility of the same results in unchanged conditions) and accuracy (i.e. obtaining the same mean value as a gold standard) for quantitative measurements. As a result, the technical variation, which consists of between-assay and within-assay variability (16), may vary across the metabolites and is generally expected to be higher compared to classical biomarker studies that focus on measurement of single candidates. High technical variation of the measurements introduces random error which in the end, reduces statistical power, and may as well adversely affect validity and reliability of the effect estimates (15; 16; 52; 53). While within-assay variability is inherent to the assay system itself, not much can be accounted for in the design or analysis of the study; except improvement of the laboratory assay, replicate measurements, or increasing the sample size for statistical power gain (16). Between-assay variability is introduced when there is variation between different plates or runs of the assay system. Ideally, there is not much additional variation beyond the within-assay variability. In a real life situation, however, the between-assay variability may sometimes even account for much larger proportions of the technical variability than the intrinsic variability (16). This becomes particularly relevant in the setting of high-throughput metabolomics measurements in epidemiologic studies with multiple samples and conducted over an extended period of time, where run-order or batch effects may occur, for example. Between-assay variation though, can be accounted for in the study design by some extent, e.g. by sample allocation before the run (16). Other sources of variation from technical issues which were not discussed in this paragraph but may also introduce measurement error include sample collection, processing and storage (54).

Biological variation and reliability

Another important source of variation in metabolomics studies represents the biological variation. As metabolites are substrates and products of enzymes, they may underlie very short term regulation which represents an advantage and a challenge at the same time. On the one hand, metabolites may be very reflective of exposures or effects. On the other hand, they may respond to multiple sources of variation and not only to the one of interest. In fact, many factors have been suggested to influence the human metabolome which are summarized in **Table 3** (38).

Table 3. Factors likely to influence the human metabolome (modified from Gibney et al. 2005 (38))	
Extrinsic factors	**Intrinsic factors**
Diet	Age
Medication	Sex
Physical activity	Genotype
Smoking	Body composition
Alcohol intake	Resting metabolic rate
Stress	Tissue turnover
Gastrointestinal microbiota	Health status
Environmental hazards	Reproductive status
	Diurnal cycle

When adopting metabolomics to epidemiologic studies this issue becomes even more challenging as there is not a controlled setting but a free-living population, which is exposed to many factors. On the one hand, high-between person variation is desirable as it represents the "true variance", which enables us to distinguish between people in respect of the factor of interest (53); however, on the other hand, it may be caused by other factors (confounders). These factors need as well to be accounted for, e.g. by adjustment or stratification in epidemiologic studies, which requires large sample sizes.

Besides the between-person variance component, the metabolites may vary over time within the same individual. High within-person variation is particularly challenging in epidemiologic studies that usually rely on a single blood sample collection, as the approximation that a single measurement represents the true value cannot be made (19). The within-person variance component is further not easy to grasp, as it may reflect true changes over time within one person (temporal variation) as well as technical variation (measurement error) (16; 54). Both reasons require restrain conditions and would limit the usefulness in epidemiologic studies. However, the within-person variation should always be evaluated in relation to the between-person variation, and should ideally be much smaller (53). The ratio of between-person variation to total variation (between- and within-person variation) is expressed by the interclass correlation coefficient (ICC) which is a measure of reliability. Reliability describes the consistency of a measurement, i.e. the ability to reproduce similar results under constant conditions (53). In the context of the present thesis, the term reliability is meant as test-retest reliability. Other forms of reliability, e.g. inter-rater reliability (53), were not studied in detail in the present thesis. The ICC as a measure of reliability further describes to what extent a single measurement relates to the "true average", and may range between zero and one (20). In general, high reliability is reflected by a higher ICC and achieved by high between-person variation and low within-person variation. However, as the ICC is a relative construct, by increasing between-person variation, reliable estimates can still be achieved with higher within-person

variation. A high degree of reliability of measurements is important in epidemiologic studies, as random measurement error may greatly attenuate the risk estimates and decrease the chance of finding true associations. **Figure 3** demonstrates the effect of random measurement error on relative risk (RR) estimates by applying the formula of Hankinson et al. (15). In case of an ICC of one (= perfect reliability) the observed RR will equal the expected RR; however, with smaller ICCs the observed RR may be considerably attenuated ("moved closer to 1.0").

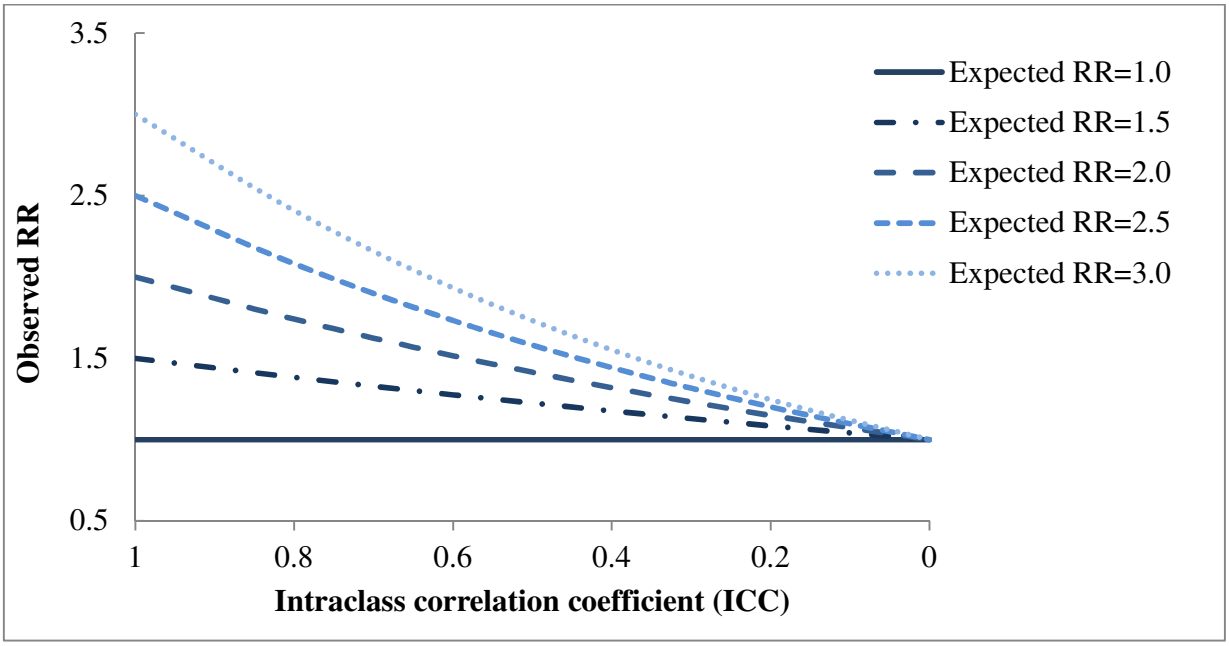


Figure 3. Impact of random measurement error of the exposure on relative risk (RR) estimates. Observed RR were calculated based on the following formula (15; 54): Observed RR= (expected RR)^{ICC}.

In summary, it should be a prerequisite in metabolomics studies to investigate technical variation a priori as it highly depends on the biochemical assay system. In the context of epidemiologic studies without controlled setting it is also important to investigate biological variation and reliability in order to better understand the quality of the measurements and the nature of the variation. If possible, metabolites that do not meet the criteria should be excluded or the error term should be considered, in order to achieve reliable and valid risk estimates.

Utility

Probably the most important criterion of a biomarker, in general, is its ability to reflect the phenotype of interest. Thereby, in the frame of a prospective study a biomarker of disease risk may be involved in (a) disease etiology and causally related to its effect, and/or (b) disease prediction, where it does not necessarily have to be causally related (55; 56). To evaluate causality of an association, the Bradford- Hill criteria may be used (**Table 4**) (57). These criteria may have implications for use of metabolomics in the frame of epidemiologic studies. Selected ideas are discussed in **Table 4**.

Table 4. Discussion of the Bradford-Hill criteria for a causal association in the context of epidemiologic studies with metabolomics data

1	Strength	Relative risk, odds ratio; effect size should not be attributable to possible confounding; exploratory approach: multiple testing correction
2	Consistency	Replication of results in independent study population; meta-analysis of multiple studies
3	Specificity	Difficult to apply, multifactorial origin of chronic diseases, multiple effects of one metabolite possible
4	Temporality	Change in metabolite concentration precedes disease occurrence, prospective design, time period sufficient for supposed mechanism?
5	Biological gradient	Dose-response-relation of metabolite concentration and disease risk, problem: does not consider possible non-linear relationships
6	Plausibility	Biological explanation of association, problem: high number of metabolites requires extensive literature review, how to treat novel candidates/non-identified metabolites?
7	Coherence	Agreement with previous studies, problem: limited comparability of studies with different metabolomics techniques/different metabolites
8	Experiment	Reversible process, usually requires dose-response-relation, problem: difficult to apply to human studies
9	Analogy	Similarity to other studies, similar metabolites have similar effect in different studies

Some of the Bradford-Hill criteria seem to have similar meanings in epidemiologic studies using metabolomics and classical epidemiologic studies with single risk factors, e.g. strength of association. The ‘strength’ of the association between metabolites and disease risk should be larger than the combined effects of all possible confounders (58). Thus, the effect size should still be relevant after multivariable adjustment. Nevertheless, it should be considered that in metabolomics studies with multiple statistical tests, multiple testing correction is often applied which leads to a lower alpha level of statistical significance. This implies that (when assuming identical sample size and precision of the effect estimates) larger effects are required to obtain significant results compared to hypothesis-driven studies with a single statistical test. The exploratory nature of metabolomics further indicates particular importance of the criterion of results ‘consistency’. In metabolomics studies it is frequently advised to replicate the results in an independent study population (as usually done in genome-wide association studies). Nevertheless, it should be debated whether a less conservative approach, e.g. a meta-analysis of multiple studies (as done when investigating single risk factors) may be a better approach, particularly when using targeted

metabolomics with a predefined selection of metabolites. Other Bradford-Hill criteria may not be readily applicable, e.g. the criterion of 'specificity'. This criterion has been suggested to be omitted (58) as it may not necessarily apply to chronic diseases with multifactorial origin, as well as metabolites that may be involved in common pathways of different chronic diseases.

To further evaluate the utility of a biomarker, its performance in disease prediction can be used (23). Thereby, risk factors are statistically modeled to predict an outcome. The prediction model can then be evaluated in terms of its performance. Thereby, the prediction model should be able to distinguish between people who develop a chronic disease in the future and those who do not, i.e. discrimination (59). In addition, the agreement between predicted and observed probabilities of the outcome should be well, i.e. calibration (23). General measures of model fit can also be taken as criteria, e.g. the explained variation R^2 in the outcome variable which quantifies the amount of information in a prediction model (23). When comparing prediction models in terms of measures of model fit, discrimination, and calibration, with and without a new predictor, the additional value of a predictor can be assessed (23). This may be a convenient approach to evaluate the utility of newly identified metabolites relative to classical risk factors, particularly when little is known about their biological role in etiology of chronic diseases.

Another aspect when evaluating the association between metabolites and risk of chronic disease is the question of the independency of the risk association from other risk factors. Independency in medical research was defined as follows by Brotman et al. (29): "a variable is called an independent risk factor if it has a significant contribution to an outcome in a statistical model that includes established risk factors." Thus, it is a statistical characteristic which does not necessarily imply causality. Knowledge about the independency or non-independency of an association may in the end facilitate the interpretation of the data. It may help to understand the role of metabolites, e.g. do they reflect a new aspect that was not covered by classical risk factors (independency)? If not, are they correlated to classical biomarkers or other metabolites? In this context, a "novel biomarker" that is uncorrelated to established risk factors will lead to greater improvement of risk prediction (22; 56); however, a dependent biomarker may also be valuable as it can add to the understanding of mechanisms and pathways of chronic disease risk. Another aspect is the dependency of metabolites with each other. Metabolomics analysis may generate highly correlated data, as lots of metabolites with similar structure are analyzed together. This problem called 'multicollinearity' represents a challenge for statistical data analysis. However, it also offers a chance. Statistical methods from systems biology, for example, may allow assigning statistically correlated metabolites to their biological pathways, or to identify unknown metabolites from untargeted metabolomics based on their correlations to other metabolites (60; 61).

In summary, much hope is put into exploratory metabolomics research in the frame of epidemiologic studies to study chronic disease risk, and in this respect, particularly for biomarker

discovery. However, many challenges appear, such as technical and biological variation, reliability, and utility of metabolites. These points need to be adequately addressed and evaluated when metabolomics is adopted to epidemiology.

1.4.4 Metabolites and risk of type 2 diabetes and cardiovascular diseases: first approaches

Although advancement of metabolomics technologies is still in progress, a number of studies have investigated the association between metabolites and risk of type 2 diabetes and CVD, the majority of them used case-control or cross-sectional designs. Of note, these designs are limited to estimated risks based on odds ratios. However, all previous studies are summarized in **Table 5A** and **Table 5B**. In total, 12 studies were found that investigated the association between metabolites and risk of type 2 diabetes, and 8 studies were identified that investigated the association between metabolites and risk of CVD.[1] Of them, four studies used a prospective design; however, three of them were patient cohorts with secondary complications or mortality as an outcome. Thus, only one study could be identified that followed-up healthy individuals prospectively. In this context it has to be considered that studies including prevalent cases may yield different results than studies with incident cases as metabolites may change considerably after disease onset, and it may not always be adequate to study disease etiology with prevalent cases (62). Metabolomics techniques that were previously used included targeted and untargeted approaches with measurement from about 30 up to ~2000 metabolites in serum, plasma or urine, whereas the latter was rarest analyzed. Sample size of the study populations was in a similar range, whereby generally in larger study populations smaller numbers of metabolites were measured.

Metabolites linked to a higher risk of type 2 diabetes in case-control and cross-sectional studies included acylcarnitines, sugar metabolites, and long-chain fatty acids (63-70). Metabolites that were reduced in type 2 diabetes cases compared to controls included the amino acids glycine and lysine, lysophosphatidylcholines and other lipids (64-68). Branched chain and aromatic amino acids and their derivates, ketone bodies and organic acids showed inconsistent results being associated with higher or lower risk of type 2 diabetes (65-67; 69; 71; 72). In 2011, Wang and colleagues (24) published a paper where for the first time a targeted metabolomics approach was adopted to a prospective study, namely the Framingham Offspring study. With a LC-MS/MS technique metabolites were profiled in baseline plasma samples in a nested case-control design including 189 incident cases of type 2 diabetes and an equal number of matched controls. Of the 61 metabolites measured, 5 branched chain and aromatic amino acids

[1] Studies that were published until November 14th 2012 in PubMed were included, except the paper that was published in the frame of this thesis. Search was conducted in title and abstract with the following keywords: "metabolomics", "metabonomics", "biomarker", "type 2 diabetes", "cardiovascular disease", "myocardial infarction", "stroke", "coronary heart disease", "coronary artery disease". Non-human studies were not included. Furthermore studies that measured less than 10 metabolites, and pre-disease conditions as an endpoint were not considered.

were found to be significantly positively associated with risk of incident type 2 diabetes. This result could further be replicated in the independent Malmö Diet and Cancer study.

In case-control and cross-sectional studies that investigated the association between metabolites and risk of CVD, sugar metabolites, methylated compounds, pyruvate and lactate were linked to higher risk (73-77) and free fatty acids and lipids with lower risk (73; 75; 77). In two prospective investigations within one cohort with patients that underwent a cardiac catheterization and/or bypass grafting and that had profiling of 69 metabolites, predictive metabolites for future complications, myocardial infarction, or mortality were acylcarnitines, ketone-related metabolites and free fatty acids (78; 79). More than 2000 plasma metabolites in patients with cardiac catheterization were measured in a nested case-control study within another patient cohort (GeneBank study). Of them, three metabolites from gut microbiota, namely choline, trimethylamine-N-oxide, and betaine, were linked to higher risk of myocardial infarction, stroke or death (80).

In summary, a number of metabolites have been linked to higher risk of type 2 diabetes and CVD in previous studies. Therefore, metabolomics seems to be promising for discovery of novel biomarkers of chronic disease risk. However, the consistency of results of previous studies was poor as some metabolites showed opposite risk associations in different studies. The inconsistency may be attributable to study limitations, e.g. a case-control or cross-sectional design, a small sample size or insufficient control of potential confounders, to name a few. Furthermore, it is not obvious whether technical and biological variation and reliability of these metabolites was always studied in detail, which could affect the results. Currently, prospective studies conducted on this topic are rare (22), as only one study could be identified which prospectively followed-up healthy individuals. For the endpoint CVD, only patient cohorts could be identified. Thus, more prospective studies investigating the association between (reliable) metabolites and risk of type 2 diabetes and CVD are very much required.

Table 1A. Previous studies investigating the association between metabolites and risk of type 2 diabetes

Author, year	Study design	Study population*	Metabolomics technique	Metabolites associated with higher risk of type 2 diabetes	Metabolites associated with lower risk of type 2 diabetes	Covariates
Adams et al. 2009 (63)	Case-control study	USA, Project SuGAR, overweight and obese African American women, cases n=44, controls n=12	Targeted, HPLC-MS, 42 plasma acylcarnitines	Total acylcarnitines, C2,C14, C18:1, C8DC	C3	Age- and BMI-matched
Bao et al. 2009 (64)	Case-control study	China, hospital-based, 74 cases (drug treated) and healthy controls	Targeted, GC-MS, serum metabolites	Valine, maltose, glutamate, urate, butanoate and long-chain fatty acid (C16:0, C18:1, C18:0, octadecanoate and C20:4),	Glucuronolactone, lysine, lactate	Unmatched
Fiehn et al. 2010 (65)	Case-control study	USA, Project SuGAR overweight and obese African American women, cases n=44, controls n=12	Untargeted, GC-MS, >350 plasma metabolites	Amino acids and derivates (leucine, 2-ketoisocaprorate, valine, cysteine, histidine), 2-hydroxybutanoate, long chain fatty acids (e.g. palmitic acid), carbohydrate derivates (e.g. glucose, fructose)	Glycine, lysine, C20:4, glycerol-3-phosphat, ethanolamine, benzylalcohol, benzoic acid	Age- and BMI-matched
Ha et al. 2012 (66)	Case-control study	Korea, hospital-based, men only, cases= 26, controls=27	Untargeted, UPLC-Q/TOF-MS, 382 plasma metabolites	Amino acids (leucine, lysine, phenylalanine), acylcarnitines (e.g. C3, C8), lyso-phosphatidylcholine (C14:0, C16:1, C18:1, C20:5, C22:6) lyso-phosphatidylethanolamine (C18:2, C22:6)	Serine, phosphatidyl-ethanolamine C18:1, lyso-phosphatidylcholine C18:1,	Age- and BMI-matched
Li et al. 2009 (70)	Case-control study	China, hospital-based, cases n=48, controls n=31	Untargeted, GCxGC-TOF-MS, plasma metabolites	Glucose, 2-hydroxyisobutyric acid, linoleic acid, palmitic acid and phosphate		Unmatched
Suhre et al. 2010 (67)	Cross-sectional	Germany, KORA study, men only, self-reported cases=40, controls=60	Targeted and untargeted, ESI-MS/MS (Biocrates), NMR (Chenomx), and UPLC-MS/MS and GC-MS (Metabolon), 482 plasma metabolites	Amino acids and derivates (phenylalanine, tyrosine, tryptophan, glutamyl-valine, glutamyl-isoleucine, phenyl-acetyl-glutamine, kynurinine, 3-indoxyl-sulfate, homocitrulline), sugars (glucose, mannose, desoxyhexose, uronic acid…), ketone body (hydroxybutyrate), creatinine	Lyso-phosphatidylcholine C20:4, diacyl-phosphatidylcholines C28:4 and C34:4, sphingomyelines C14:0 and C22:2, free fatty acids (C6:0, C7:0, C9:0)	Age-matched

Table 5A. Continued

Author, year	Study design	Study population*	Metabolomics technique	Metabolites associated with higher risk of type 2 diabetes	Metabolites associated with lower risk of type 2 diabetes	Covariates
Van Doorn et al. 2007 (71)	Case-control study	Netherlands, hospital-based, cases n=16, controls n=16, hospital-based, RCT for drug effect, baseline considered	Untargeted, ^{1}H-NMR spectroscopy, plasma and urine metabolites	Plasma: lactate, lipids, Urine: amino acids (alanine, phenylalanine, tyrosine), hippurate, citrate, phosphoenolpyruvat	Plasma: amino acids (leucine, isoleucine, valine, alanine, glutamine, tyrosine), 3D-hydroxybutyrate, citrate, formate Urine: amino acids (glutamate, glutamine), N-methyl-nicotinamid, uridine	Unmatched
Wang et al. 2011 (24)	Prospective cohort, nested case-control design	USA, Framingham Offspring study, cases n=189, controls n=189	Targeted, LC-MS/MS, 61 plasma metabolites	Amino acids (leucine, isoleucine, valine, phenylalanine, tryptophane)		Age-, sex-, BMI-, and fasting glucose-matched
Wang-Sattler et al. 2012 (68)	Cross-sectional	Germany, KORA study, cases n= 91, controls (normal and impaired glucose tolerance) n= 1206	Targeted, ESI-MS/MS, (Biocrates), 140 serum metabolites	Acetylcarnitine (C2)	Glycine, lyso-phosphatidylcholine C18:2	Age, sex, BMI, physical activity, alcohol intake, smoking, systolic blood pressure, HDL-cholesterol (and fasting glucose, insulin, HbA_{1c})
Zhang et al. 2009 (72)	Case-control study	China, hospital-based, cases n=33, controls n=25	Untargeted, UPLC-MS/MS, serum metabolites	Unidentified	Leucine, dihydrosphingosine and phytosphingosine	Unmatched
Zhang et al. 2009 (69)	Cross-sectional	China, nationwide survey on diabetes prevalence, cases n=74, controls n=157	Untargeted, ^{1}H-NMR spectroscopy, serum metabolites	Glucose	Amino acids (isoleucine, leucine, valine, alanine, methionine, glutamine, lysine, tyrosine, phenylalanine, histidine), citrate, lactate, choline	Age-, sex- and BMI-matched
Zhu et al. 2011 (81)	Case-control study	China, hospital-based, cases n=30, controls n=30	HPLC–TOF/MS, plasma metabolites	Lysophosphatidylcholines C18:0, C18:2, C20:4, phosphatidylcholine C34:2, phosphatidyl-ethanolamine C36:4, phosphatidylglycerol C36:2, sphingomyelin C20:2	Phosphatidylcholines C34:0, C38:4, phosphatidylinositols C34:0, C38:4, C40:6	Unmatched

*Cases are always cases of type 2 diabetes.

Table 5B. Previous studies investigating the association between metabolites and risk of cardiovascular diseases (CVD)

Author, year	Study design	Study population	Metabolomics technique	Metabolites associated with higher risk of CVD	Metabolites associated with lower risk of CVD	Covariates
Bodi et al. 2012 (73)	Case-control study	Spain, hospital-based, cases (myocardial ischemia) n=20, controls n=10	Untargeted, Hr-NMR spectroscopy, serum metabolites	Glucose, lactate, glutamine, glycine, glycerol, phenylalanine, tyrosine, and phosphoethanolamine	Choline-containing compounds and triacylglycerols	Unmatched
Jiang et al. 2011 (74)	Case-control study	China, hospital-based, cases (stroke) n=67, controls n=62	Untargeted, UPLC-TOF-MS, serum metabolites	Cysteine, S-adenosyl-homocysteine, oxidized glutathione, hydroxy-eicosatetraenoic acid, hydroxy-octa-decadienoic acid	Folic acid, tetrahydrofolate, adenosine, aldosterone, deoxocathasterone, sucrose-6-phosphat, betanine	Age- and sex-matched
Jung et al. 2011 (75)	Case-control study	Korea, hospital-based, cases (stroke) n (plasma/urine) =54/27, healthy controls (plasma/urine) n=47/27	Targeted, ^{1}H-NMR spectroscopy, plasma and urine metabolites	Plasma: Lactate, glycolate, pyruvate, formate Urine: -	Plasma: valine, lipids, methanol, glutamine Urine: citrate, dimethylamine, creatinine, glycine, hippurate	Unmatched
Kang et al. 2011 (76)	Case-control study	Korea, hospital-based, cases (systolic heart failure) n=15, controls n=20	Targeted, ^{1}H-NMR spectroscopy, urine metabolites	4-hydroxy-phenylacetate, acetate, acetoacetate, acetone, betaine, cytosine, methylmalonic acid, dimethylamine, glutarate, hippurate, methylguanidine, methylmalonate, phenylacetylglycine, taurine, trimethylamine-N-oxide, tyramine	1-methylnicotinamide, 2-hydroxy-isobutyrate, 2-oxoglutarate, adenine, citrate, glutamine, histidine, phenylalanine, pyruvate, succinate, cis-aconitate	Age- and sex-matched
Shah et al. 2012 (78)	Prospective patient cohort	USA, n=478 patients with bypass, cases (myocardial infarction, need for intervention, death) n=126, mean follow-up: 4.3y	Targeted LC-MS/MS, 69 plasma metabolites	Short-chain dicarboxyl-acylcarnitines, ketone-related metabolites, short-chain acylcarnitines		Multivariable adjustment (age, sex, race, smoking, BMI, clinic. parameters, metabolite factors)

Table 5B. Continued

Author, year	Study design	Study population	Metabolomics technique	Metabolites associated with higher risk of CVD	Metabolites associated with lower risk of CVD	Covariates
Shah et al. 2012 (79)	Prospective patient cohort	USA, n=2,023 patients with cardiac catheterization, cases (mortality) n=232, median follow-up 3.1y	Targeted LC-MS/MS, 69 plasma metabolites	Medium-chain acylcarnitines, short-chain dicarboxyl-acylcarnitines, long-chain dicarboxyl-acylcarnitines, free fatty acids	Branched-chain amino acids and related catabolites,	Multivariable adjustment (age, sex, smoking, clinic. parameters, metabolite factors)
Teul et al. 2011 (77)	Case-control study	Spain, hospital-based , cases (coronary artery syndromes) n=19, controls n=6	Untargeted GC-MS, and targeted (21 metabolites, fatty acids), plasma metabolites	Glucose, fructose, myoinositol, pyruvate, lactate, citrate, isocitrate, succinate, malate, valine, cysteine, 4-hydroxyproline, 2-hydroxybutyrate, 2,3,4-trihydroxybutyrate, 3-hydroxybutyrate	Oxalate, alanine, serine, glycine, threonine, aspartate, tryptophan, tyrosine, urea, 2-aminobutyrate, creatinine, aminomalonate, free fatty acids	Age-matched
Wang et al. 2011 (80)	Prospective patient cohort, nested case-control design	USA, GeneBank study, patients with cardiac catheterization, cases (myocardial infarction, stroke, death) n=50, controls n=50 Replication: BioBank study, CAD cohort	Untargeted, LC-MS, >2000 plasma metabolites	Choline, trimethylamine-N-oxide, betaine		Age- and sex-matched

1.5 Objective

Against this background and despite the challenges of adopting metabolomics to epidemiologic studies, the present thesis aimed to apply a targeted metabolomics approach with simultaneous measurement of 163 metabolites by flow injection analysis MS/MS technique to the prospective EPIC-Potsdam study. Thereby, a multistep approach was taken and the following objectives were of interest:

1.) Investigation of technical variation of metabolite measurements
2.) Investigation of biological variation and reliability of metabolites
3.) Investigation of utility of metabolites

First, technical variation was investigated by using data of replicate measures, and the final metabolite set was selected by excluding metabolites with high technical variation and those below their limit of detection. In the next step, biological variation and reliability of metabolites were studied among 100 healthy EPIC-Potsdam participants who had provided two blood samples four month apart in time. Last, utility of the metabolites was addressed by investigating their association with risk of type 2 diabetes and CVD in the frame of a case-cohort study including 849 incident cases of type 2 diabetes, 274 incident cases of myocardial infarction, 260 incident cases of stroke and a randomly drawn subcohort of 2500 participants. Thereby, strength and independency of risk associations as well as measures of discrimination and calibration for risk prediction were studied in detail. Overall, the present thesis addresses different challenges of adopting metabolomics to epidemiologic studies and represents a pioneering work in the emerging field of systems epidemiology.

2. Study population and methods

2.1 European Prospective Investigation into Cancer and Nutrition-Potsdam study

The European Prospective Investigation into Cancer and Nutrition (EPIC) is an ongoing multicenter prospective cohort study with the primary aim to study associations between diet and cancer, but considering other risk factors and chronic diseases as well (82). In total, the EPIC-study comprises about 520,000 participants across 23 centers in ten European countries, and is coordinated by the International Agency for Research on Cancer (IARC) of the World Health Organization (WHO) in Lyon, France. EPIC-Potsdam is the largest German study center that recruited 27548 participants mainly aged between 35 and 65 years from the city of Potsdam and surrounding communities based on general population registries (83). The participation rate in Potsdam was 22.7% and participants tended to be higher educated and more health conscious compared to the general population (83). Between 1994 and 1998, the participants were invited to the study center. They provided written informed consent before participating in the baseline examination which included anthropometric and blood pressure measurements, a lifestyle questionnaire, a computer-guided interview on diet, lifestyle and medical conditions, and a self-administered semi-quantitative food frequency questionnaire (84). In addition, a blood sample was collected (84).

Every two to three years, follow-up questionnaires were sent to the participants. Response rates were above 95% for the first and second follow-up period and above 90% for the third and fourth follow-up period (85). This was achieved by an effective reminder system with repeated written notices and phone calls and tracing of non-responders, which made high completeness of data possible (86). Incident cases of chronic diseases were identified by self-report and potential incident cases were further verified (86). The EPIC-Potsdam study was approved by the Ethics Committee of the Medical Society of the State of Brandenburg.

2.1.1 Blood sample collection at baseline

During the baseline examination, one blood sample was randomly collected from 95.7% of the EPIC-Potsdam participants, of them a small proportion (28%) had only consumed beverages prior to the blood withdrawal or fasted overnight (2). In particular, 30 mL of venous blood were drawn by qualified medical staff; 20 mL were inserted into monovettes containing citrate, and 10 mL into monovettes without anticoagulant. Blood was immediately fractionated into serum, plasma, buffy coat and erythrocytes, and aliquoted into straws of 0.5 mL each according to a strict and standardized protocol. During blood withdrawal and processing, time and room temperature were steadily documented. For each study participant 12 straws of plasma, 8 straws of serum, 4 straws of buffy coat, and 4 straws of erythrocytes were filled, summing up to 28 straws in total. The straws were stored in tanks of liquid nitrogen at -196°C.

As additional fractions remained, a second blood storage was established during the recruitment period, consisting of additional plasma, serum, serum with metaphosphoric acid, erythrocytes and for the last participants also whole blood (84). These additional blood fractions were stored in conventional tubes at -80°C.

2.1.2 Covariate assessment at baseline

Habitual diet of the participants was assessed by use of a self-administered food frequency questionnaire (FFQ) at baseline (84). The FFQ used in the German EPIC-centers had been validated by comparison to twelve repeated 24-hour dietary recalls administered monthly to a subset of participants (87; 88) and reproducibility was evaluated by repeated application of the FFQ after six month (87; 88). It was suitable to rank participants according to their usual food intake. The FFQ contained 148 items and asked about frequency and portion size of food and beverage consumption during the preceding 12 month. It also inquired about fat content of dairy products, fat quality for food preparation and dietary supplement use. The possible frequencies of food consumption ranged from never to 5 times per day or more. Photographs of different portion sizes and household measures were provided to facilitate the estimation of the quantity. Daily intake of an individual food item was calculated as the product of frequency and portion size of this item and expressed as grams per day.

During the physical examination at baseline, anthropometry (e.g. height, weight, waist and hip circumferences, skinfold measurement) and blood pressure were measured by trained staff following standardized procedures (84). Waist circumference was measured at midpoint between the lower rib and iliac crest. The participants wore no shoes and light closing during the examination. Body weight in kilograms was divided by the square of body height in meters to calculate body mass index (BMI). Blood pressure was measured on the right arm in a sitting position using automated devices (Boso oscillomat®). After a 15 to 30 minute resting period, each participant had three consecutive blood pressure measurements in two minute intervals. In the end, blood pressure was determined from the second and third reading (89).

A sociodemographic and lifestyle questionnaire was administered at baseline including items such as education, alcohol consumption and occupational activity, among others (84). Women filled in a supplementary questionnaire regarding women-specific questions such as menopausal status and parity. In addition, a PC-guided interview was used to assess other lifestyle factors, e.g. smoking habits and recreational activity, but also medical anamnesis of prevalent diseases and medication use.

2.1.3 Ascertainment of incident type 2 diabetes and cardiovascular diseases during follow-up

Since baseline, every two to three years the EPIC-Potsdam participants received a follow-up questionnaire by mail to identify potential incident cases of chronic diseases, including type 2 diabetes and CVD (86). The questionnaire included items on a new diagnosis of disease (**Figure 4**), disease-relevant medication use, or change in diet. Self-reported cases of incident type 2 diabetes and CVD were further verified by contacting the treating physician or hospital. Thereby, a standard inquiry form was sent to the respective physician or clinic to obtain the exact date and type of the diagnosis, the means of the confirmation and treatment information (**Figure 5**). In addition, local health offices were contacted to provide death certificates for confirmation of the diagnosis. This multistep approach contributed to high quality of follow-up data (86). Incident cases of type 2 diabetes and CVD were defined according to the international classification of diseases (ICD)-10 system: E11 for type 2 diabetes mellitus, I21 for myocardial infarction, I60 and I61 for haemorrhagic stroke, I63 for ischemic stroke and I64 for undetermined stroke.

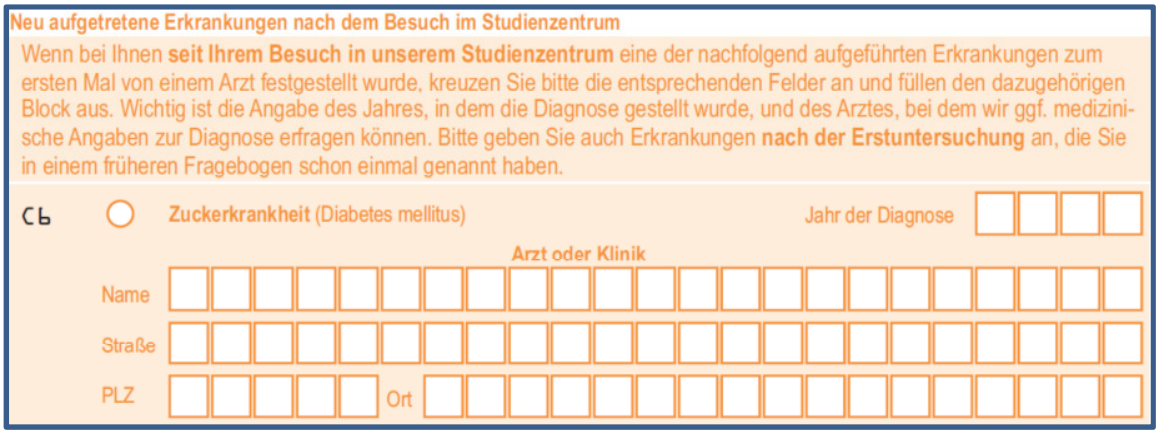
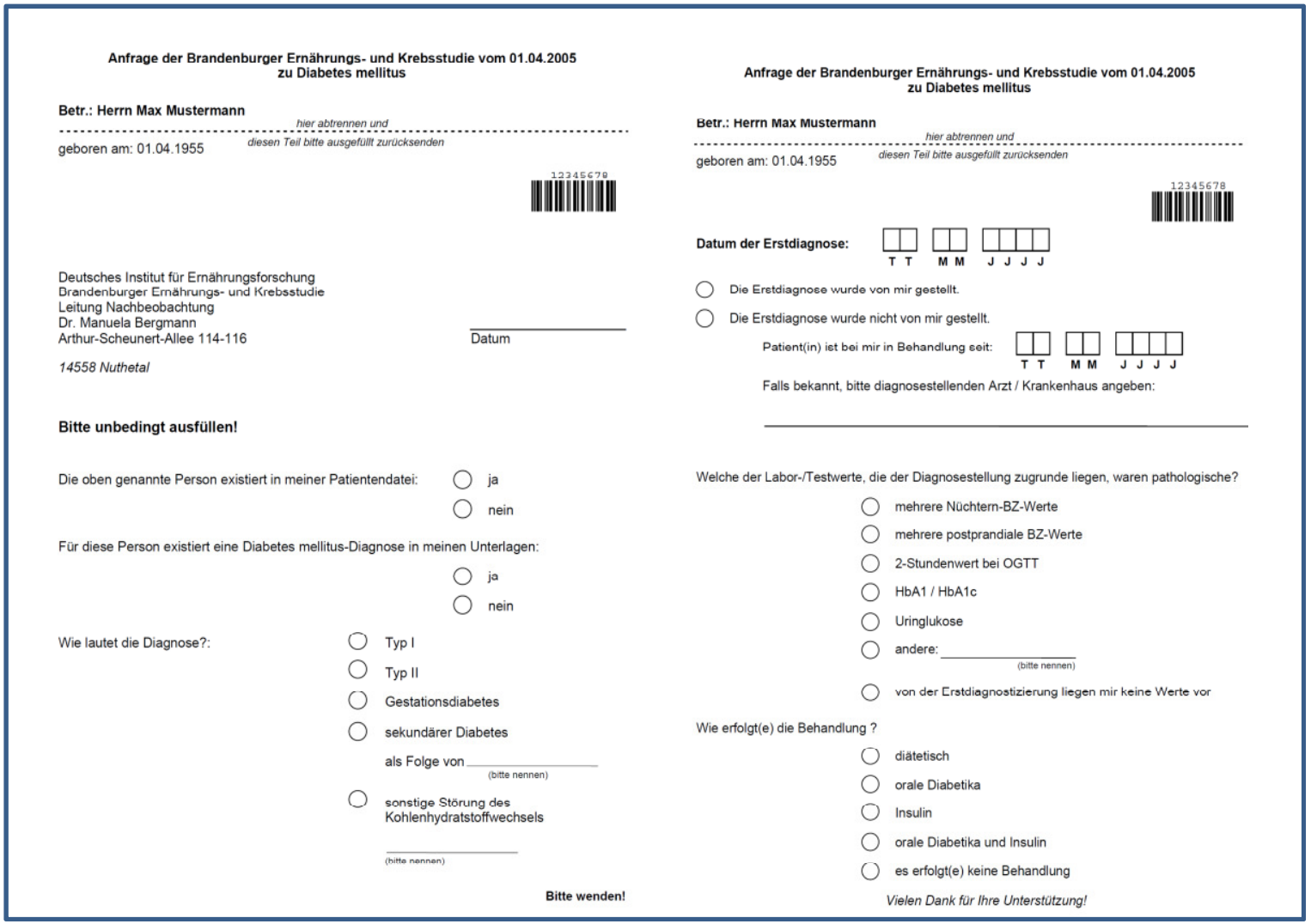


Figure 4. Identification form of incident type 2 diabetes in EPIC-Potsdam that was sent to participants.



Anfrage der Brandenburger Ernährungs- und Krebsstudie vom 01.04.2005
zu Diabetes mellitus

Betr.: Herrn Max Mustermann

hier abtrennen und
diesen Teil bitte ausgefüllt zurücksenden

geboren am: 01.04.1955

12345678

Deutsches Institut für Ernährungsforschung
Brandenburger Ernährungs- und Krebsstudie
Leitung Nachbeobachtung
Dr. Manuela Bergmann
Arthur-Scheunert-Allee 114-116

14558 Nuthetal

Datum

Bitte unbedingt ausfüllen!

Die oben genannte Person existiert in meiner Patientendatei:
- ja
- nein

Für diese Person existiert eine Diabetes mellitus-Diagnose in meinen Unterlagen:
- ja
- nein

Wie lautet die Diagnose?:
- Typ I
- Typ II
- Gestationsdiabetes
- sekundärer Diabetes
  als Folge von _____ (bitte nennen)
- sonstige Störung des Kohlenhydratstoffwechsels
  _____ (bitte nennen)

Bitte wenden!

Anfrage der Brandenburger Ernährungs- und Krebsstudie vom 01.04.2005
zu Diabetes mellitus

Betr.: Herrn Max Mustermann

hier abtrennen und
diesen Teil bitte ausgefüllt zurücksenden

geboren am: 01.04.1955

12345678

Datum der Erstdiagnose: T T M M J J J J

- Die Erstdiagnose wurde von mir gestellt.
- Die Erstdiagnose wurde nicht von mir gestellt.

  Patient(in) ist bei mir in Behandlung seit: T T M M J J J J

  Falls bekannt, bitte diagnosestellenden Arzt / Krankenhaus angeben:

  _____

Welche der Labor-/Testwerte, die der Diagnosestellung zugrunde liegen, waren pathologische?
- mehrere Nüchtern-BZ-Werte
- mehrere postprandiale BZ-Werte
- 2-Stundenwert bei OGTT
- HbA1 / HbA1c
- Uringlukose
- andere: _____ (bitte nennen)
- von der Erstdiagnostizierung liegen mir keine Werte vor

Wie erfolgt(e) die Behandlung ?
- diätetisch
- orale Diabetika
- Insulin
- orale Diabetika und Insulin
- es erfolgt(e) keine Behandlung

Vielen Dank für Ihre Unterstützung!

Figure 5. Verification form of incident type 2 diabetes in EPIC-Potsdam that was sent to the respective physician.

2.1.4 Reliability study

In 2007, a subgroup of 407 EPIC-Potsdam participants was re-invited to the study center to take part in a validation study on assessment methods for physical activity which involved a repeated blood sample collection. This subgroup was randomly selected from all EPIC-Potsdam participants who were younger than 64 years, had a systolic blood pressure <180 mmHg and a diastolic blood pressure <110 mmHg at baseline. For logistical reasons, they also needed to have a valid phone number and to live within 5 km radius of the study center. Exclusion criteria were immobility, a history of heart disease as well as the use of beta-blockers as documented at the time of invitation. Of the 407 invited participants, 176 declined participation, 11 did not respond, and 12 were excluded as it was revealed that they had used beta-blockers. Hence, a total of 208 individuals consented to participate in the validation study. From this sample, 100 participants (50% women) were randomly selected among those who had provided two fasting blood samples four month apart in the time period between October 2007 and July 2008. Blood was drawn after an overnight fast by qualified medical staff, inserted into monovettes with coagulation activator and centrifuged at 2,700g for 10 minutes. Samples were stored in a freezer at -80°C until metabolomics analysis.

2.1.5 Case-cohort study

A case-cohort study was previously designed within EPIC-Potsdam to study in more detail metabolic markers and their association with chronic disease risk without the need to measure these markers in the entire cohort (90; 91). Therefore, of all participants of the full cohort who had provided blood samples at baseline (n=26,444), a random sample of 2500 individuals was drawn as control group, referred to as the subcohort (8). In addition, all incident cases of the full cohort of type 2 diabetes that occurred until August 31st 2005 (n=849), and of myocardial infarction (n=274), and stroke (n=260) that occurred until December 31st 2006, were embedded (**Figure 6**). As the subcohort was representative of the full cohort, it also included incident cases of type 2 diabetes (n=85), myocardial infarction (n=24) and stroke (n=33) i.e. ‘internal cases’. The remaining ‘external cases’ were not part of the subcohort. With a case-cohort design the results are expected to be generalizable to the full cohort (90).

A number of clinical biomarkers were previously measured in the EPIC-Potsdam case-cohort study (92; 93). These biomarkers included, plasma glucose, glycated hemoglobin A1c (HbA_{1c}), total and HDL-cholesterol, triglycerides and CRP, among others. They were measured between 2007 and 2008 with the ADVIA 1650 chemistry system (Siemens Medical Solutions, Erlangen, Germany) at the Department of Internal Medicine, University of Tuebingen. A number of the above mentioned variables were considered as covariates in the present thesis.

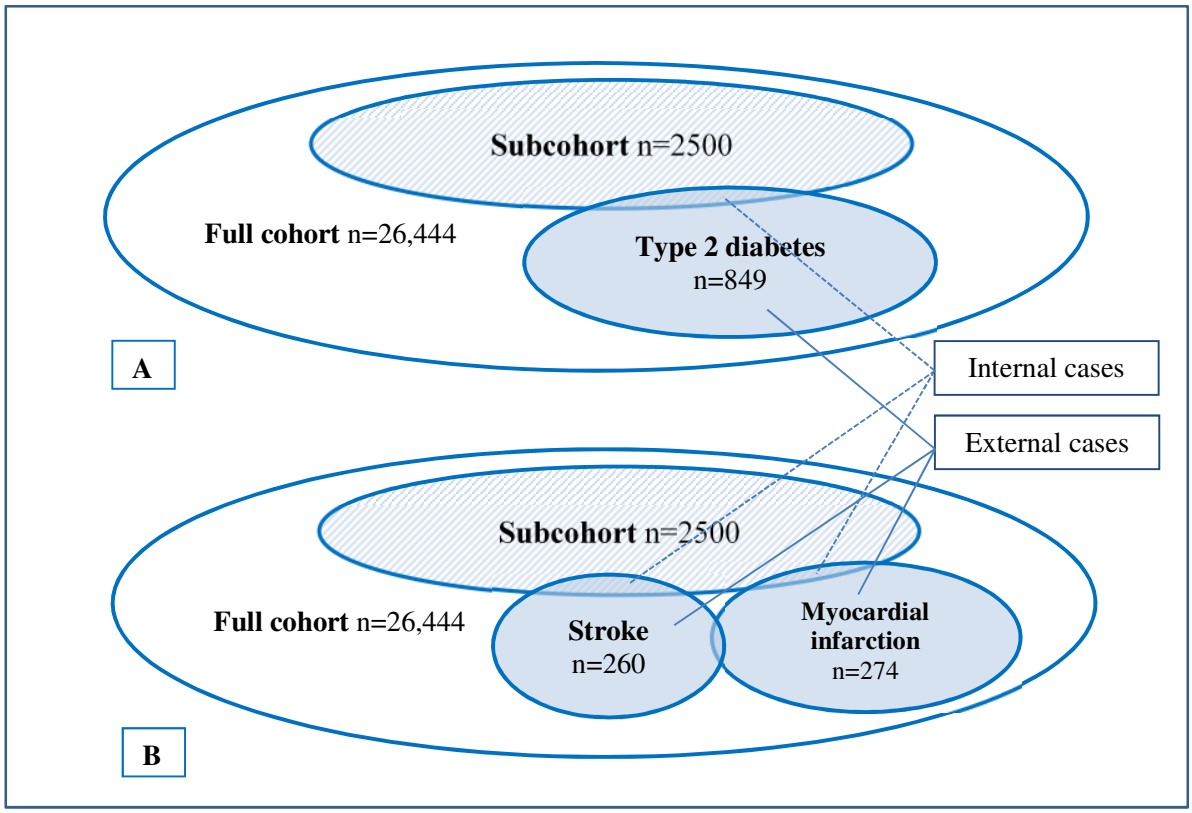


Figure 6. Case-cohort study for type 2 diabetes (A) and for cardiovascular diseases (B) in EPIC-Potsdam. The full cohort refers to all participants who had provided blood samples at baseline. Note: the subcohort is identical.

2.1.6 Measurement of serum metabolites

Concentrations of 163 serum metabolites were measured in blood samples of the reliability study and in baseline blood samples of the case-cohort study with a targeted metabolomics approach. Specifically, the Absolute*IDQ*™ p150 kit (BIOCRATES Life Sciences AG, Innsbruck, Austria) was applied, which uses flow injection analysis MS/MS technique. All samples were analyzed in the Genome Analysis Center of the Helmholtz Zentrum München between 2009 and 2010. Briefly, with a robotic system (Hamilton Bonaduz AG, Bonaduz, Switzerland) 10 µL of serum were pipetted onto a 96-well plate which included filters with isotope labeled internal standards. After incubation, the filters were dried in nitrogen, amino acids were derivated with 5% phenylisothiocyanate reagent and filters were dried again. The metabolites and internal standards were extracted with 5mM ammonium acetate in methanol, centrifuged through a filter membrane and diluted with the mass spectrometry (MS) running solvent. The final extracts were analyzed with an API 4000™ triple quadrupole mass spectrometer (ABSciex). For all measurements, two 20 µL injections (for positive and for negative electrospray ionization mode) were applied by the standard flow injection method. The metabolites were quantified by multiple reaction monitoring with the use of stable isotope labeled internal standards. Details were stated elsewhere (94).

According to the manufacturer, the analytical method was proven to meet the FDA-Guideline "Guidance for industry – bioanalytical method validation" (May 2001). Details on quantification ranges, accuracy, precision, reproducibility, specificity and possible interferences of the method were provided in the BIOCRATES manual AS-P150 (95). According to the manufacturer, the accuracy of the quantitative measurements was within 80-115%, and regarding precision, the uncertainty of the measurements was below 10% for most metabolites and between 10% and 20% for only few metabolites. Most metabolites were detected with high specificity and possible interferences may have occurred between few metabolites of the phospholipid class. It has to be noted, that the metabolite panel was selected by the manufacturer based on the robustness of the measurements. To investigate the technical variation of the assay system, 230 replicate measures (5 per plate on 46 plates) of one female serum sample were analyzed together with the EPIC-Potsdam samples. To account for run-order effects, the EPIC-Potsdam samples were randomly analyzed together regardless of the case status.

The targeted metabolomics approach simultaneously identified and quantified 163 metabolites, of them 92 phosphatidylcholines (diacyl- (PC aa Cx:y), acyl-alkyl- (PC ae Cx:y), and lyso-phosphatidylcholines (lysoPC a Cx:y)), 41 acylcarnitines (Cx:y), 15 sphingomyelins (SM Cx:y), 14 amino acids (standard three letter codes) and 1 hexose (sum of six carbon sugars without distinction of isomers). Fatty acid side chains were abbreviated x:y, where x represented the number of carbon atoms and y the number of double bonds. The prefix 'lyso' indicated a single fatty acid side chain. Of note, for the diacyl- and acyl-alkyl-phosphatidylcholines, which contained two fatty acid side chains, the method could neither differentiate the distribution of fatty acids across side chains nor the exact position of double bonds. All acylcarnitines were derivates of L-carnitine (C0) linked to one fatty acid. The individual biochemical names of metabolites and their limits of detection are provided in **Table S1**. Their biochemical structure is exemplified for each class of compounds in **Figure 7**.

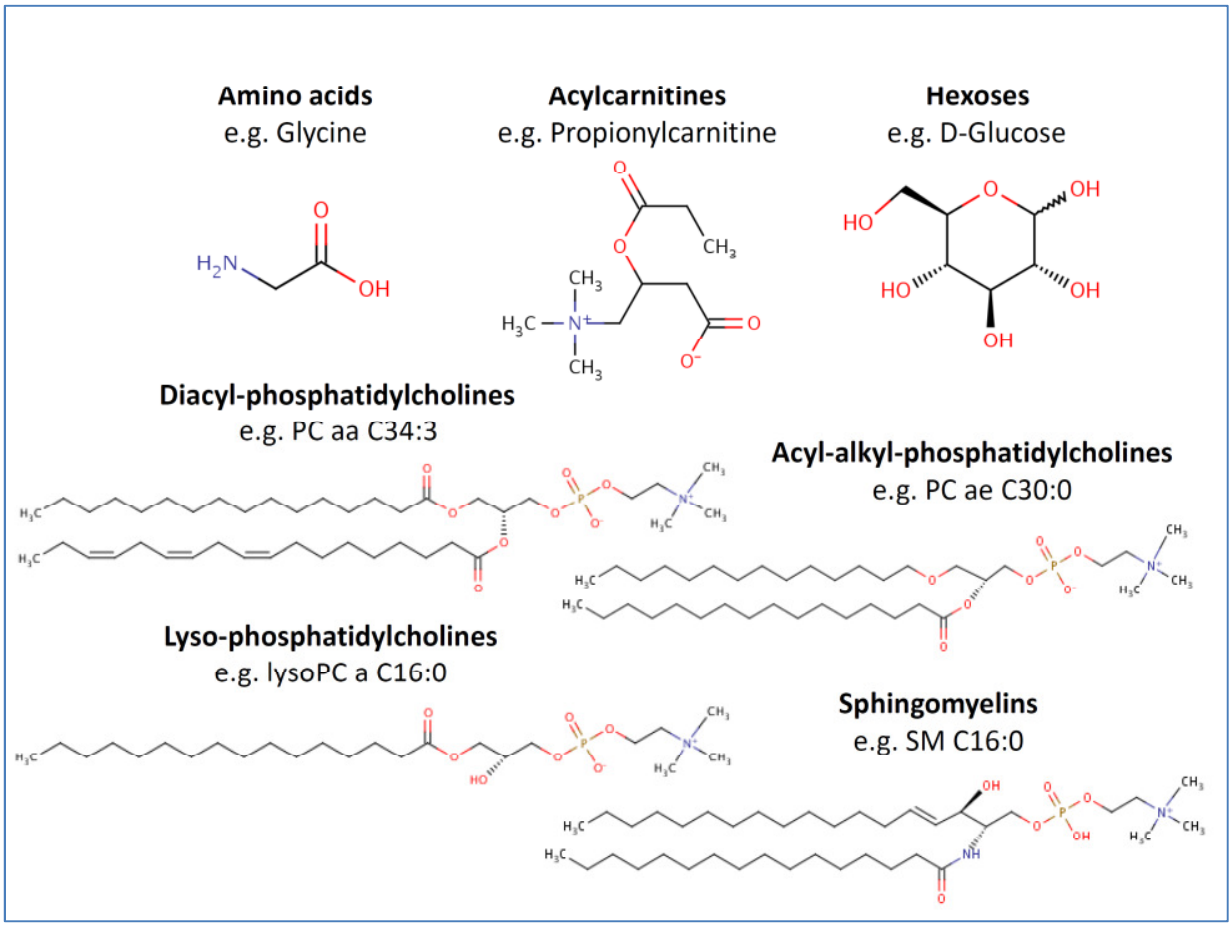


Figure 7. Biochemical structure of metabolites by subclass measured with the Absolute*IDQ*™ p150 kit (BIOCRATES Life Sciences AG, Innsbruck, Austria) in the EPIC-Potsdam study. Structural formulas were derived from the Human Metabolome Database (40; 49).

2.1.7 Analytical study sample

The present analysis involved three datasets: One to study reliability of metabolites and the other two datasets to study their association with risk of type 2 diabetes and CVD, respectively. The first dataset was based on the reliability study (n=100), and was complete for most of the metabolites. However, for 6 metabolites, the analysis was based on fewer observations because of missing metabolomics data: diacyl-phosphatidylcholines C30:2 (n=73) and C38:1 (n=82), acyl-alkyl-phosphatidylcholine C30:1 (n=97), lyso-phosphatidylcholine C6:0 (n=61), sphingomyelins C20:2(n=99) and C22:3 (n=99).

The other two datasets were based on the case-cohort study. Generation of the analytical study sample is illustrated in a flow diagram (**Figure 8**). To retrieve the final study sample for the endpoint type 2 diabetes, participants with prevalent type 2 diabetes or missing information on study variables were excluded. Thus, the analytical study sample for type 2 diabetes comprised 2282 participants of the

subcohort, and 800 incident cases of type 2 diabetes. For the endpoint CVD, prevalent cases of CVD or missing information of study variables were excluded. In addition, non-ischemic strokes, (ICD-10: I60, I61 and I64) were excluded. For individuals, who suffered from both myocardial infarction and stroke, only the first occurring event was counted. Therefore, the analytical study sample for CVD included 2309 participants of the subcohort and 235 incident cases of myocardial infarction and 178 incident cases of stroke.

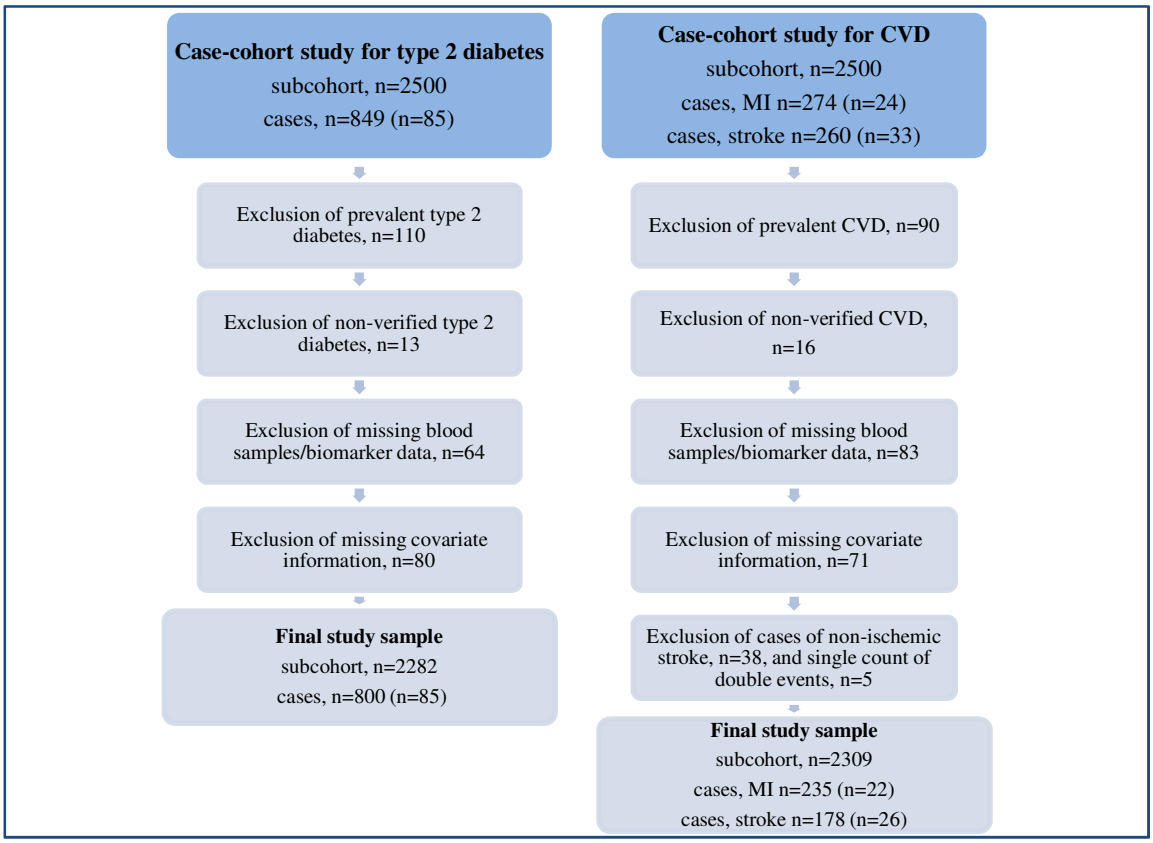


Figure 8. Retrieving the analytical study sample from the EPIC-Potsdam case-cohort study. Cases presented are incident cases. Internal cases, i.e. those that also belong to the subcohort, are shown in parenthesis. Abbreviations: CVD, cardiovascular disease; MI, myocardial infarction.

2.2 Statistical analysis

The analytical strategy of the present thesis consisted of a multi-step approach to address the different objectives, and is summarized in a flow-diagram at the end of this section (**Figure 10**). All statistical analysis were conducted with SAS (statistical analysis system; version 9.2 and SAS enterprise guide version 4.3; SAS Institute Inc, Cary, NC, USA). Correlation graphs were plotted with yED graph editor (yWorks GmbH, Tuebingen. www.yworks.com).

2.2.1 Descriptive statistics

Baseline characteristics of the EPIC-Potsdam subcohort and all incident cases of type 2 diabetes and CVD were calculated as age- and sex-adjusted mean and standard error (SE) for continuous variables, or age- and sex-adjusted percentages for categorical variables. Therefore, relevant socioeconomic, dietary and lifestyle factors, as well as obesity measures and medical conditions were considered. Serum metabolite concentrations were reported as unadjusted mean and SD. To evaluate the inter-correlation of metabolites, simple Spearman correlation coefficients were calculated between all metabolites, and illustrated for each class of metabolites as a circular network graph, where the color, shape and thickness of the lines connecting the individual metabolites indicated the strength of the correlation. In addition, Spearman correlation coefficients between metabolites and established biomarkers of type 2 diabetes and CVD were calculated. All correlations were analyzed within the random subcohort.

2.2.2 Selection of the final metabolite set

To select the final metabolite set, those metabolites with high technical variation and below their detection limit were excluded. The technical variation of the laboratory method for each metabolite was calculated with a one-way analysis of variance (ANOVA) (15; 96) from data of 230 replicate measures (5 per plate on 46 plates) of one sample and reported as between- and within-plate coefficient of variation (CV) in %, a common measure to describe technical error. A CV ≥50% for either between- or within-plate-variation was considered to be the cut-off point for exclusion of metabolites. The limit of detection (LOD) of the laboratory method for each metabolite was set as the three times median of the "zero" samples (buffer with internal standards only). Metabolites with concentrations below their LOD were excluded.

2.2.3 Biological variation and reliability of metabolites

Biological variation, i.e. between- and within-person variation (97), and reliability were estimated from data of repeated sample measurements (reliability study). Thereby, a one-way random effects model was used on a log-transformed scale with subject-ID as the random variable (96; 98). Between- and within-person CV were calculated as the square root of the between- and within-person variance components (96). In addition, geometric means of serum metabolite concentrations were calculated for both points in time and compared by paired t-test. To evaluate the reliability of serum metabolites, the ICC was calculated as the ratio of between-person variation and total variation (sum of between- and within-person variation) (**Equation 1**). The ICC was described as point estimate and 95% confidence interval (CI) (20; 99). In general, the ICC may range between zero and one, and a higher ICC indicates better reliability. Theoretically, the ICC is high when the within-person variation is low compared to the between-person variation (53). In accordance with previous studies (19; 20), an ICC <0.40 was considered poorly reliable, ICCs between 0.40 and 0.50 to indicate fair reliability, ICCs between 0.51 and 0.74 to show good reliability, and an ICC ≥0.75 was considered to indicate excellent reliability.

$$ICC = \frac{Between\ person\ variance}{Between\ person\ variance + within\ person\ variance}$$

Equation 1

To account for the reliability of metabolites that was estimated with the ICC, the observed RR estimates of metabolites associated with risk of type 2 diabetes and CVD in **section 2.2.4** were additionally corrected for the ICC to estimate the expected RR as done by Hankinson et al. (15) and originally suggested by Rosner et al. (17). Thereby, the equation used by Hankinson et al. (15) was solved for expected RR (**Equation 2**). Consequently, the expected RR was calculated as the ICC-th root of the observed RR. Next, differences in ranking of metabolites comparing expected and observed RR estimates were evaluated. Last, the theoretical effect of reliability of exposure assessment on statistical power in this thesis was estimated, using the same formula and assuming an ICC of 0.40 of all measurements, which has previously been used as cut-off for acceptable reliability (19; 20). In general, the method used for correction of reliability is suitable for a continuous exposure variable and assumes random measurement error and constant variance across metabolite levels.

$$observed\ RR = e^{ln(expected\ RR)\ \cdot\ ICC}$$

$$expected\ RR = e^{ln(observed\ RR)/\ ICC}$$

$$\boldsymbol{expected\ RR = \sqrt[ICC]{observed\ RR}}$$

Equation 2

2.2.4 Serum metabolites and risk of type 2 diabetes and cardiovascular diseases

To investigate the association between metabolites and risk of type 2 diabetes and CVD, a multi-step approach was taken, consisting of (1) identification of metabolites associated with risk of type 2 diabetes and CVD, (2) evaluation of strength and independency of risk associations and (3) usefulness for risk prediction.

Identification of metabolites associated with risk of type 2 diabetes and cardiovascular diseases

Cox proportional hazard regression (100) with weighting as suggested by Prentice (101) and robust sandwich covariance estimates to consider the case-cohort design was applied to calculate hazard rate ratios as a measure of RR and 95% CIs with age as the underlying time-scale from study entry (baseline) to exit time (diagnosis of type 2 diabetes or CVD, or censoring) for each participant. Thereby, the hazard of an individual for an event at a specific time (outcome) is given by the product of the baseline hazard and the exponential function of the linear combinations of explanatory variables (exposure and covariates), see **Equation 3**. The baseline hazard equals the hazard of an individual when the value of all explanatory variables is zero (102).

The Cox proportional hazard model is given as:

$$\boldsymbol{h(t,X) = h_0(t) \cdot exp \sum_{i=1}^{n} \beta_i \cdot x_i}$$

$where$:

$h(t,X) = \ hazard\ of\ event\ at\ time\ t\ for\ individual\ X$

$h_0(t) = baseline\ hazard$

$\beta_i = \beta_1 \dots \beta_n\ , regression\ coefficients\ of\ explanatory\ variables$

$x_i = x_1 \dots x_n\ , value\ of\ explanatory\ variables$

Equation 3

The Cox regression analysis offers the advantage over the logistic regression analysis that it considers the varying follow-up times of the participants (100). Before entering the metabolite concentrations into the Cox-model, they were transformed to standard normal distribution (mean=0, standard deviation (SD) =1) (**Equation 4**) and one by one considered as the exposure variable. With this transformation, it was accounted for the different absolute concentrations of the metabolites (e.g. concentration of sugar metabolites was in the mM range whereas acylcarnitines were concentrated in the µM range); thereby, the risk estimates were made comparable (hazard rate ratios per SD change of the metabolite concentrations).

The z-score (z_i) of a raw score (x_i) for each observation i is:

$$z_i = \frac{x_i - \bar{x}}{s}$$

where:

$\bar{x}$: mean of sample

s: standard deviation of sample.

Equation 4

To identify metabolites associated with risk of type 2 diabetes and CVD, a multivariable adjusted model was calculated that included common risk factors. This basic model was stratified by age (integer) and adjusted for sex, alcohol intake from beverages (non-consumers; women: >0-6 g/d, 6-12 g/d, >12 g/d; men: >0-12 g/d, 12-24 g/d, >24 g/d), smoking (never, former, current ≤20 cigarettes/d, current >20 cigarettes/d), cycling and sports (h/week), education (no degree/vocational training; trade/technical school; university degree), prevalent hypertension (yes/no), BMI (kg/m²), and waist circumference (cm). For the endpoint type 2 diabetes, this basic model was additionally adjusted for dietary factors including coffee intake (cups/d), red meat intake (g/d), and whole-grain bread intake (g/d). For the CVD analysis, the basic model was additionally adjusted for prevalent diabetes and fasting status. The latter was done as fasted participants were not equally distributed comparing the subcohort and incident CVD cases. As the metabolomics approach was rather exploratory and multiple statistical tests were conducted, the chance of false-positive findings was increased. Therefore, *p*-values from Cox regression were corrected for multiple testing with the Bonferroni-Holm method (103). Briefly, the Bonferroni-Holm correction is based on the principle of the classical Bonferroni test, i.e. that the significance level (α) is corrected by dividing α by the number of tests (n). However, the Bonferroni-Holm correction goes further as it considers the number of previous and remaining tests in each step, and thus uses multiple local significance levels ($\alpha_1 \ldots \alpha_n$), whereby the first significance level (α_1) equals the significance level of the classical Bonferroni test (α/n). The Bonferroni-Holm procedure is illustrated in **Figure 9**. The probability of rejecting false-hypotheses is similar comparing both correction methods; however, Bonferroni-Holm achieves a statistical power gain, and it was suggested that it should be preferred over the classical Bonferroni test (103).

In our study a corrected *p*-value <0.05 (Bonferroni-Holm) was considered as significant to identify metabolites associated with risk of type 2 diabetes and CVD. Only those metabolites that showed significant associations were considered for the next steps. To account for the inter-correlation of metabolites, an additional model was calculated that included all metabolites significantly associated with risk of type 2 diabetes and CVD and the respective covariates of the basic model. Then, stepwise Cox-regression was used to select the independent predictors. For the selection process, the covariates had to be included into the model.

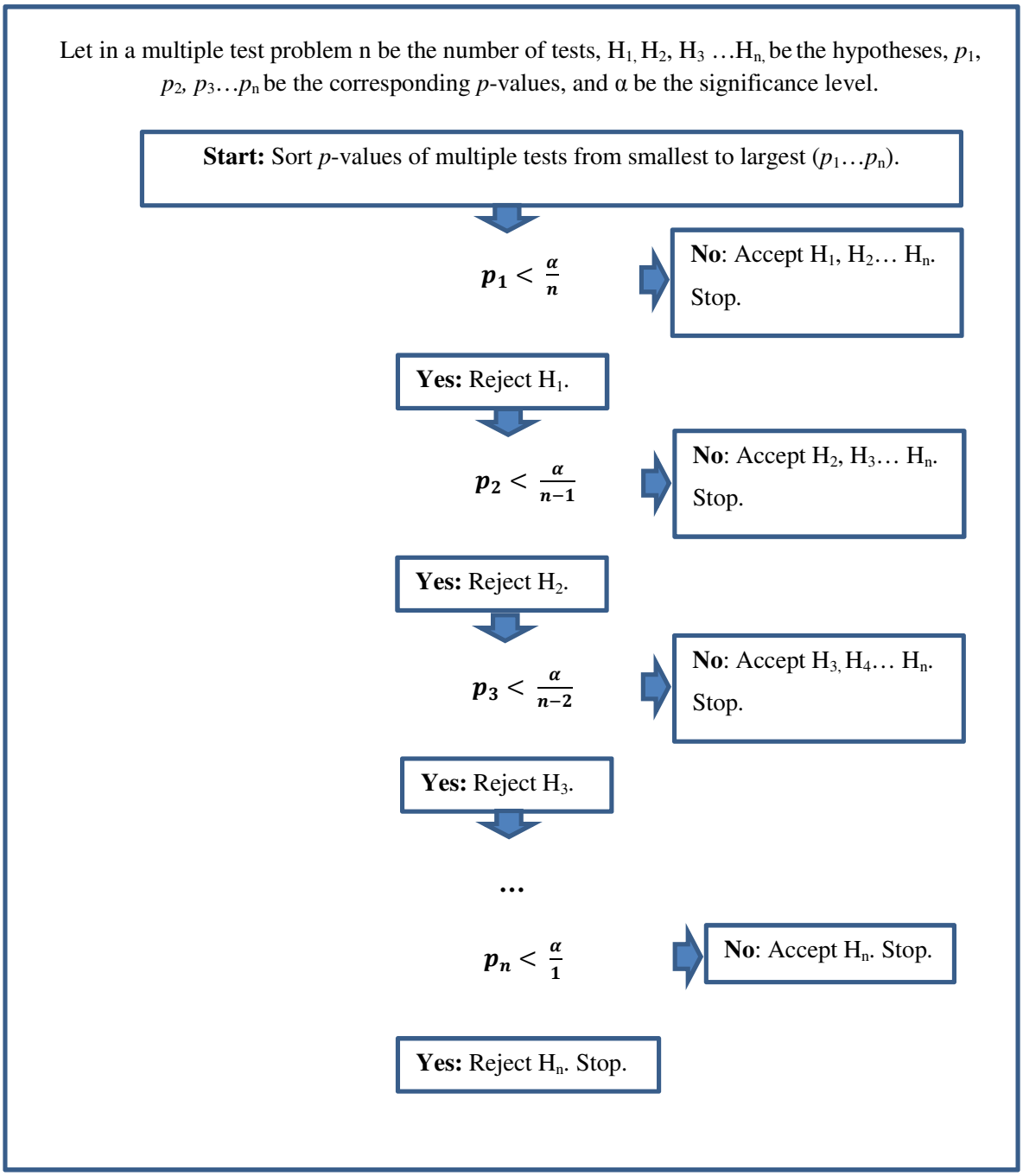


Figure 9. Overview of the procedure of the Bonferroni-Holm correction for multiple testing. Modified from S. Holm (103).

Evaluation of independency of risk associations

A principal component analysis (PCA) was conducted with all metabolites associated with risk of type 2 diabetes. In general, PCA is a data reduction strategy which aggregates multiple metabolites based on their degree of correlation with one another to obtain a smaller number of metabolite factors (principal components) which represent linear combinations of the individual metabolites (104). These metabolite factors are retrieved in a way that the explained variance between persons is maximized. As a result, a smaller number of metabolite factors which explains most of the variance of the individual metabolites is obtained. The PCA in this study was based on the correlation matrix of the standardized metabolites. The orthogonal varimax rotation method was applied to obtain uncorrelated metabolite factors, which facilitated the interpretation. Two metabolite factors were retained, as they explained the major fraction of the observed variance of metabolites (>50%). To investigate the association between metabolite factors and risk of type 2 diabetes and CVD, the population was divided into quintiles of metabolite factors and Cox regression analysis was used to calculate hazard rate ratios and 95% CIs across quintiles of metabolite factors with the first quintile as the reference group. Quintiles were particularly chosen to better interpret the results. *P*-values for a linear trend across all quintiles were calculated by assigning each participant the median value of the according quintile and modeling this value as a continuous variable.

To investigate whether the risk associations of the metabolites were independent of established biomarkers, additional Cox regression models were calculated that adjusted for other biomarkers. For the endpoint type 2 diabetes, the basic model was additionally adjusted for plasma glucose or HbA1c or HDL-cholesterol or triglycerides and a final model was adjusted for all of them. For the endpoint CVD, the basic model was additionally adjusted for HDL-cholesterol or total cholesterol or CRP and their combination.

Usefulness for risk prediction

Different measures of model fit, discrimination and calibration were calculated to evaluate whether the metabolites could be useful to improve prediction of type 2 diabetes and CVD. Therefore, different logistic regression models with and without metabolites were calculated; type 2 diabetes or CVD were considered as a dichotomous outcome. As the reference model for the endpoint type 2 diabetes, the German Diabetes Risk Score (7) was calculated according to **Equation 5**. Briefly, the German Diabetes Risk Score combines several risk factors for type 2 diabetes including age, diet, lifestyle and anthropometry, to estimate the individual risk of developing type 2 diabetes within the next 5 years. Thereby, individuals receive a point score based on their risk factors which is used to estimate the individual risk. It was derived within the EPIC-Potsdam study population and has been shown to accurately estimate the 5-year probability of developing type 2 diabetes (7).

$$\textbf{\textit{German Diabetes Risk Score}}$$

$$= 7.4 \times waist\ circumference\ (cm) - 2.4 \times height\ (cm) + 4.3 \times age\ (years)$$

$$+46 \times hypertension\ (y/n) + 49 \times red\ meat\left(150\frac{g}{d}\right) - 9 \times whole\ grain\ bread\left(50\frac{g}{d}\right)$$

$$-4 \times coffee\left(150\frac{g}{d}\right) - 20 \times moderate\ alcohol\ (y/n)\ - 2 \times physical\ activity\left(\frac{h}{week}\right)$$

$$+24 \times former\ smoker\ (y/n) + 64 \times current\ heavy\ smoker\ (y/n)$$

Equation 5

In the EPIC-Potsdam study population a CVD risk score has not yet been developed. Therefore, for the CVD analysis the basic model, which included the covariates (common risk factors of CVD), was used as the reference. In different steps metabolites and established biomarkers were added to the reference models. The model performance was evaluated by calculating -2Log-Likelihood and Nagelkerke's R^2 (105). In general, Nagelkerke's R^2 defines the fraction of -2Log-Likelihood which is explained by the predictors and may range from zero to one. Larger values for -2Log-Likelihood and Nagelkerke's R^2 indicate better model fit. To study discrimination, receiver operating characteristic curves (ROC) were plotted for models with and without metabolites. The ROC curve represents a plot of sensitivity versus false-positive rate of a continuous prediction model on a dichotomous outcome. The area under the ROC curve represents the probability that the prediction model assigns a true case with a higher risk compared to a control and may range from 0.5 (no discrimination) to 1.0 (perfect discrimination).The areas under the ROC curves and 95% CIs were then calculated and compared as suggested by DeLong and colleagues (106); thereby, a *p*-value <0.05 indicated significant difference of the ROC curves. Hosmer-Lemeshow-test (107) was used to study calibration of the different models. The Hosmer-Lemeshow statistic (χ^2) (107) compares predicted and observed probabilities of the dichotomous endpoint that are derived from deciles of predicted risk. A *p*-value <0.05 indicates significant differences between expected and observed probabilities and poor calibration. Smaller values of χ^2 and non-significant *p*-values specify better calibration.

Sensitivity analyses

Several sensitivity analyses were conducted to ensure the robustness of the results. Possible interactions were tested for the association between individual metabolites, metabolite factors and risk of type 2 diabetes and CVD, for the covariates sex and fasting status by including multiplicative interaction terms into the models. In addition, the PCA was repeated using data of only those participants who had provided fasting blood samples, to see whether similar metabolite factors could be reproduced compared to blood samples that were randomly collected. To rule out the chance of reverse causation, i.e. that the metabolite changes were due to overt and undiagnosed disease conditions and did not precede the onset of type 2 diabetes and CVD, the analysis was repeated excluding all incident cases that occurred shortly after the baseline examination during the first two years of follow-up.

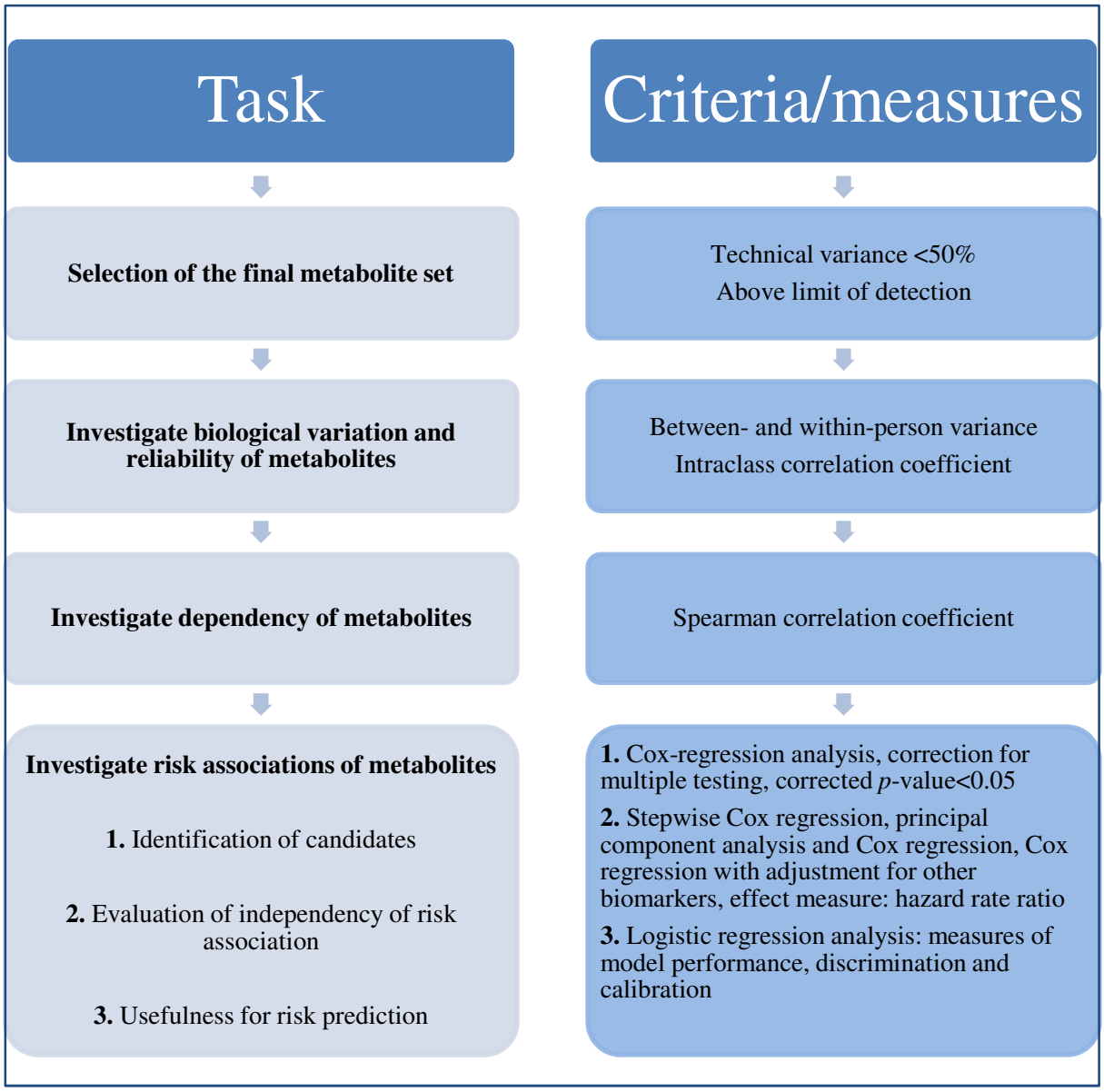


Figure 10. Summary of the analytical strategy to investigate 163 serum metabolites and their association with risk of type 2 diabetes and cardiovascular diseases in the frame of the EPIC-Potsdam study.

3. Results

3.1 Selection of the final metabolite set

In total 163 metabolites were measured with the BIOCRATES kit, including 41 acylcarnitines, 14 amino acids, 1 hexose, 38 diacyl-phosphatidylcholines, 39 acyl-alkyl-phosphatidylcholines, 15 lyso-phosphatidylcholines, and 15 sphingomyelins. Of them, 36 metabolites had to be excluded because of the poor quality of the measurements (**Table S2**). In particular, 29 metabolites were excluded because they were below the limit of detection of the assay system. An additional 7 metabolites were excluded, because they showed very high technical variance (>50%). In this respect, technical variance particularly consisted of between-assay variation. Thus, the final metabolite set for the present thesis comprised 127 metabolites (17 acylcarnitines, 14 amino acids; 1 hexose; 34 diacyl-phosphatidylcholines; 37 acyl-alkyl-phosphatidylcholines; 10 lyso-phosphatidylcholines, and 14 sphingomyelins). These 127 metabolites were used for the following analyses.

3.2 Biological variation and reliability of metabolites[2]

To investigate the biological variation and reliability of the metabolites, data from the reliability study with repeated metabolite measurements over a four month period was used. The reliability study conducted from 2007 to 2008 consisted of 100 participants evenly distributed by gender with mean age of 56.1 years (**Table 6**). Men were older and had higher anthropometric measures than women.

Mean serum concentrations of the metabolites ranged from 0.01 µmol/L for several acylcarnitines to 5207 µmol/L for hexose. In **Tables 7A-D** the detailed concentrations of each metabolite for the two time point measurements, their between- and within-person variation and reliability are listed according to the metabolite classes (**Table 7A**: acylcarnitines; **Table 7B**: amino acids, lyso-phosphatidylcholines, and sphingomyelins; **Table 7C**: diacyl-phosphatidylcholines; **Table 7D**: acyl-alkyl-phosphatidyl-cholines, and hexose). Overall, reliability of the 163 serum metabolites was good with a median ICC of 0.57. When excluding the 36 metabolites with low quality of measurements, the median ICC was 0.59. Excellent reliability (ICC≥0.75) was found for hexose, hydroxy-sphingomyelins C14:1, C22:2, and C16:1, sphingomyelins C18:1 and C16:1, acyl-alkyl-phosphatidylcholines C44:6, C42:5, C36:2, and C42:4, and butyryl-carnitine (C4). Metabolites that were poorly reliable (ICC<0.40) included 16 of 41 acylcarnitines, 14 of 92 glycerophospholipids and sphingomyelins C22:3, C20.2, and C26:0. Reliability of metabolites by

[2]In the framework of this thesis, parts of the results on biological variation and reliability of metabolites have already been published by the author: 108. Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, Joost HG, Boeing H, Pischon T: Reliability of Serum Metabolite Concentrations over a 4-Month Period Using a Targeted Metabolomic Approach. PLoS One 2011;6:e21103.

subclass is illustrated in **Figure 11**. The amino acids showed a fair to good reliability (median ICC: 0.58), sphingomyelins showed poor to excellent reliability (median ICC: 0.66) and glycerophospholipids showed poor to excellent reliability (median ICC: 0.58). Least reliable were the acylcarnitines with a median ICC of 0.41. Among this class particularly the hydroxy- and monounsaturated acylcarnitines showed poor to fair reliability, whereas short and medium chain saturated acylcarnitines showed good to excellent reliability.

Since reliability depends on both, within- and between-person variances, we also report these variance components to in detail show their contribution to the ICC. As an example, the acylcarnitines C4 and C16 showed very similar within-person variances (18.7% and 18.8%, respectively); however, their between-person variances differed (18.3% and 38.4%, respectively), which led to different magnitudes of their corresponding ICCs (0.49 and 0.85, respectively). Specifically, acylcarnitine C4 had a much lower within-person variance compared to its between-person variance leading to a high ICC and indicating excellent reliability. In contrast, acylcarnitine C16 showed similar within- and between-person variances, yielding to a lower ICC and suggesting only fair reliability.

Table 6. Characteristics of participants of the reliability study (2007-2008) in EPIC-Potsdam[3]

	All (n=100)	**Men (n=50)**	**Women (n=50)**
	Mean (SD)	Mean (SD)	Mean (SD)
Age (years)	56.1 (4.1)	57.9 (3.1)	54.4 (4.2)
BMI (kg/m²)	26.8 (4.1)	28.0 (3.8)	25.6 (4.0)
Waist (cm)	94.6 (13.5)	103 (10.9)	86.1 (10.1)
Waist-to-hip-ratio	0.89 (0.10)	0.96 (0.06)	0.81 (0.06)

[3] In the framework of this thesis, this table has already been published by the author: 108. Ibid. Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, Joost HG, Boeing H, Pischon T: Reliability of Serum Metabolite Concentrations over a 4-Month Period Using a Targeted Metabolomic Approach. PLoS One 2011;6:e21103. PLoS applies the Creative Commons Attribution License.

Table 7A. Geometric means, variances and intraclass-correlation coefficients (ICCs) of serum concentrations of acylcarnitines repeatedly measured 4 months apart among 100 healthy participants from the reliability study[4]

	1st Measurement		2nd Measurement			CV (%)			
Metabolite	**Mean** [µmol/L]	**95% CI**	**Mean** [µmol/L]	**95% CI**	***P-value*†**	**Within-person**	**Between-person**	**ICC‡**	**95% CI**
C0	35.8	(34.1-37.6)	37.0	(35.3-38.8)	0.09	13.4	19.7	0.69	(0.57-0.78)
C10	0.35	(0.32-0.38)	0.37	(0.33-0.41)	0.20	27.0	38.5	0.67	(0.55-0.77)
C10:1*	0.21	(0.20-0.22)	0.23	(0.21-0.24)	0.04	20.4	23.2	0.57	(0.42-0.68)
C10:2	0.04	(0.04-0.04)	0.04	(0.04-0.04)	0.44	21.0	19.2	0.46	(0.29-0.60)
C12*	0.13	(0.12-0.14)	0.14	(0.13-0.15)	0.09	23.1	23.0	0.50	(0.34-0.63)
C12-DC*	0.03	(0.03-0.04)	0.04	(0.03-0.04)	0.16	16.8	7.2	0.15	(0.00-0.34)
C12:1*	0.03	(0.03-0.04)	0.04	(0.03-0.04)	0.25	19.6	17.9	0.45	(0.28-0.60)
C14*	0.10	(0.09-0.10)	0.09	(0.09-0.10)	0.28	14.7	11.5	0.38	(0.20-0.53)
C14:1	0.26	(0.25-0.27)	0.27	(0.26-0.28)	0.05	15.5	16.2	0.52	(0.36-0.65)
C14:1-OH*	0.02	(0.02-0.02)	0.02	(0.02-0.02)	0.08	25.4	19.7	0.37	(0.19-0.53)
C14:2	0.04	(0.04-0.04)	0.04	(0.04-0.04)	0.39	30.7	28.4	0.46	(0.29-0.60)
C14:2-OH	0.01	(0.01-0.01)	0.01	(0.01-0.01)	0.17	37.8	20.5	0.23	(0.03-0.40)
C16	0.14	(0.13-0.14)	0.14	(0.14-0.15)	0.08	18.7	18.3	0.49	(0.33-0.63)
C16-OH*	0.01	(0.01-0.01)	0.01	(0.00-0.01)	0.20	44.8	15.4	0.11	(0.00-0.30)
C16:1*	0.05	(0.05-0.06)	0.05	(0.05-0.06)	0.99	17.6	17.1	0.49	(0.32-0.62)
C16:1-OH*	0.01	(0.01-0.01)	0.01	(0.01-0.01)	0.72	32.9	13.8	0.15	(0.00-0.34)
C16:2	0.01	(0.01-0.01)	0.01	(0.01-0.01	0.26	40.9	23.2	0.24	(0.05-0.42)
C16:2-OH*	0.01	(0.01-0.01)	0.01	(0.01-0.01)	0.02	31.6	12.4	0.13	(0.00-0.32)
C18	0.05	(0.05-0.06)	0.06	(0.05-0.06)	0.10	28.1	22.9	0.40	(0.22-0.55)
C18:1	0.18	(0.17-0.19)	0.19	(0.18-0.21)	0.00	19.2	18.7	0.49	(0.32-0.62)
C18:1-OH*	0.01	(0.01-0.01)	0.01	(0.01-0.01)	0.87	43.4	22.1	0.21	(0.01-0.39)
C18:2	0.06	(0.06-0.07)	0.07	(0.07-0.07)	0.00	21.9	23.9	0.54	(0.39-0.67)
C2	6.97	(6.55-7.41)	7.18	(6.74-7.65)	0.29	20.1	24.1	0.59	(0.45-0.70)
C3	0.39	(0.36-0.41)	0.39	(0.36-0.41)	0.97	18.7	27.9	0.69	(0.57-0.78)
C3-DC / C4-OH*	0.04	(0.04-0.04)	0.04	(0.04-0.05)	0.18	17.0	0.23	0.23	(0.04-0.41)
C3-DC-M / C5-OH	0.02	(0.02-0.02)	0.02	(0.02-0.02)	0.58	22.6	0.45	0.45	(0.28-0.59)
C3-OH*	0.01	(0.01-0.01)	0.01	(0.01-0.01)	0.08	36.7	15.6	0.15	(0.00-0.34)
C3:1*	0.01	(0.01-0.01)	0.01	(0.01-0.01)	0.27	57.6	0.00	0.00	-
C4	0.23	(0.21-0.25)	0.23	(0.21-0.25)	0.81	18.8	38.4	0.81	(0.73-0.87)
C4:1*	0.01	(0.01-0.01)	0.01	(0.01-0.01)	0.60	35.1	11.3	0.09	(0.00-0.28)
C4:1-DC / C6	0.09	(0.08-0.09)	0.09	(0.08-0.10)	0.13	22.0	30.6	0.66	(0.53-0.76)
C5	0.13	(0.12-0.14)	0.14	(0.13-0.15)	<.001	21.2	28.1	0.64	(0.50-0.74
C5-DC / C6-OH	0.02	(0.02-0.02)	0.02	(0.02-0.02)	0.80	25.4	0.41	0.41	(0.24-0.56)
C5-M-DC*	0.03	(0.02-0.03)	0.03	(0.03-0.03)	0.57	28.7	32.0	0.56	(0.40-0.68)
C5:1*	0.04	(0.03-0.04)	0.04	(0.03-0.04)	0.75	26.1	9.4	0.11	(0.00-0.30)
C5:1-DC*	0.01	(0.01-0.02)	0.01	(0.01-0.02)	0.68	31.8	11.3	0.11	(0.00-0.30)
C6:1*	0.01	(0.01-0.01)	0.01	(0.01-0.01)	0.57	30.1	9.8	0.10	(0.00-0.29)
C7-DC	0.04	(0.04-0.04)	0.04	(0.04-0.05)	0.26	26.1	21.7	0.41	(0.23-0.56)
C8*	0.26	(0.24-0.28)	0.28	(0.26-0.30)	0.02	20.1	33.5	0.73	(0.63-0.81)
C8:1	0.10	(0.10-0.11)	0.12	(0.11-0.13)	<.001	27.9	36.4	0.63	(0.50-0.73)
C9	0.04	(0.04-0.05)	0.04	(0.04-0.05)	0.46	26.1	30.8	0.58	(0.44-0.70)

*Metabolite concentration was below the assay's limit of detection.
†*P*-value for difference based on a paired t-test of log-transformed values to compare geometric means of metabolite concentrations over time.
‡The intraclass-correlation coefficient (ICC) was calculated as the ratio of between-person variation and total variation.

[4] In the framework of this thesis, this table has already been published by the author: 108. Ibid. Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, Joost HG, Boeing H, Pischon T: Reliability of Serum Metabolite Concentrations over a 4-Month Period Using a Targeted Metabolomic Approach. PLoS One 2011;6:e21103. PLoS applies the Creative Commons Attribution License.

Table 7B. Geometric means, variances and intraclass-correlation coefficients (ICCs) of serum concentrations of amino acids, lyso-phosphatidylcholines and sphingomyelins repeatedly measured 4 months apart among 100 healthy participants from the reliability study[5]

	1st Measurement		2nd Measurement			CV (%)			
Metabolite	**Mean** [µmol/L]	**95% CI**	**Mean** [µmol/L]	**95% CI**	***P-value*†**	**Within-person**	**Between-person**	**ICC‡**	**95% CI**
Amino acids									
Arg	135	(130-140)	136	(132-141)	0.49	11.4	13.7	0.59	(0.45-0.70)
Gln	726	(701-751)	741	(713-770)	0.22	11.7	14.3	0.60	(0.46-0.71)
Gly	313	(298-329)	329	(313-345)	0.01	14.1	20.4	0.68	(0.55-0.77)
His	93.7	(90.7-96.9)	96.4	(93.0-100)	0.07	11.0	13.5	0.60	(0.46-0.71)
Met	36.5	(35.1-38.0)	37.8	(36.2-39.4)	0.12	15.8	13.2	0.41	(0.24-0.56)
Orn	112	(107-118)	121	(116-127)	0.00	15.8	18.4	0.58	(0.43-0.69)
Phe	67.1	(64.6-69.7)	70.6	(68.1-73.3)	0.01	13.9	13.1	0.47	(0.30-0.61)
Pro	196	(185-208)	206	(194-219)	0.02	15.9	25.6	0.72	(0.62-0.80)
Ser	140	(134-146)	147	(141-154)	0.01	13.4	16.7	0.61	(0.47-0.72)
Thr	104	(98.7-109)	108	(103-113)	0.08	17.4	15.5	0.44	(0.27-0.59)
Trp	83.8	(81.2-86.5)	84.9	(82.2-87.8)	0.44	12.1	10.9	0.45	(0.28-0.59)
Tyr	80.1	(76.6-83.8)	84.2	(80.2-88.4)	0.03	15.9	17.4	0.54	(0.39-0.67)
Val	338	(324-353)	351	(333-369)	0.09	15.9	17.8	0.56	(0.41-0.68)
Ile	261	(250-274)	276	(263-289)	0.01	15.7	18.1	0.57	(0.42-0.69)
Lyso-phosphatidylcholines									
lysoPC a C14:0	3.28	(3.11-3.46)	3.36	(3.13-3.62)	0.47	24.3	21.6	0.44	(0.27-0.59)
lysoPC a C16:0	157	(151-163)	167	(160-175)	0.01	15.6	15.1	0.49	(0.32-0.62)
lysoPC a C16:1	4.51	(4.24-4.79)	4.69	(4.38-5.02)	0.20	21.7	24.8	0.57	(0.42-0.69)
lysoPC a C17:0	2.54	(2.39-2.69)	2.65	(2.49-2.82)	0.09	18.5	24.5	0.64	(0.50-0.74)
lysoPC a C18:0	49.8	(47.6-52.1)	52.6	(50.2-55.2)	0.02	16.7	16.5	0.50	(0.33-0.63)
lysoPC a C18:1	30.5	(28.9-32.1)	31.8	(30.0-33.6)	0.12	18.7	20.3	0.54	(0.39-0.66)
lysoPC a C18:2	36.33	(34.2-38.6)	37.52	(35.0-40.2)	0.29	21.2	25.1	0.58	(0.44-0.70)
lysoPC a C20:3	3.29	(3.08-3.50)	3.37	(3.18-3.58)	0.47	24.9	18.7	0.36	(0.18-0.52)
lysoPC a C20:4	8.50	(8.00-9.03)	9.11	(8.58-9.67)	0.01	19.8	23.2	0.58	(0.43-0.69)
lysoPC a C24:0*	0.37	(0.34-0.39)	0.35	(0.33-0.38)	0.32	25.2	19.3	0.37	(0.19-0.53)
lysoPC a C26:0*	0.51	(0.47-0.56)	0.51	(0.47-0.55)	0.91	31.0	32.0	0.52	(0.36-0.65)
lysoPC a C26:1*	3.09	(3.03-3.15)	3.05	(2.99-3.10)	0.28	9.1	3.7	0.14	(0.00-0.33)
lysoPC a C28:0*	0.48	(0.44-0.51)	0.47	(0.44-0.50)	0.62	24.7	24.1	0.49	(0.32-0.62)
lysoPC a C28:1	0.76	(0.71-0.81)	0.73	(0.68-0.78)	0.25	20.6	25.4	0.60	(0.46-0.71)
lysoPC a C6:0*	0.02	(0.02-0.02)	0.02	(0.02-0.02)	0.81	48.2	15.1	0.11	(0.00-0.32)
Sphingomyelins									
SM (OH) C14:1	7.08	(6.66-7.52)	7.18	(6.77-7.61)	0.41	11.6	27.7	0.85	(0.78-0.90)
SM (OH) C16:1	3.57	(3.37-3.78)	3.56	(3.35-3.79)	0.95	13.7	26.3	0.79	(0.70-0.85)
SM (OH) C22:1	13.7	(13.1-14.5)	13.9	(13.2-14.5)	0.69	13.7	20.7	0.70	(0.58-0.78)
SM (OH) C22:2	11.6	(10.9-12.3)	11.6	(11.0-12.3)	0.83	13.1	25.5	0.79	(0.70-0.85)
SM (OH) C24:1	1.26	(1.19-1.34)	1.22	(1.15-1.30	0.26	18.9	24.6	0.63	(0.49-0.73)
SM C16:0	125	(120-130)	127	(122-132)	0.42	11.3	15.5	0.66	(0.53-0.75)
SM C16:1	17.9	(17.0-18.7)	18.1	(17.4-19.0)	0.33	11.4	19.8	0.75	(0.65-0.83)
SM C18:0	26.4	(25.3-27.6)	26.7	(25.5-27.9)	0.60	12.8	19.0	0.69	(0.57-0.78)
SM C18:1	11.50	(10.9-12.1)	11.5	(10.9-12.1)	0.98	12.7	23.4	0.77	(0.68-0.84)
SM C20:2	0.46	(0.42-0.50)	0.44	(0.41-0.47)	0.39	32.3	19.1	0.26	(0.07-0.43)
SM C22:3	1.99	(1.81-2.19)	2.23	(2.08-2.38)	0.03	34.2	22.5	0.30	(0.11-0.47)
SM C24:0	24.0	(22.9-25.1)	24.1	(23.1-25.1)	0.78	14.2	16.3	0.57	(0.42-0.69)
SM C24:1	51.9	(49.8-54.2)	52.3	(50.0-54.6)	0.73	12.6	17.5	0.66	(0.54-0.76)
SM C26:0	0.21	(0.20-0.22)	0.22	(0.20-0.23)	0.48	29.0	16.2	0.24	(0.04-0.41)
SM C26:1	0.46	(0.44-0.49)	0.43	(0.40-0.46)	0.04	24.4	19.8	0.40	(0.22-0.55)

[5] In the framework of this thesis, this table has already been published by the author: 108. Ibid. Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, Joost HG, Boeing H, Pischon T: Reliability of Serum Metabolite Concentrations over a 4-Month Period Using a Targeted Metabolomic Approach. PLoS One 2011;6:e21103. PLoS applies the Creative Commons Attribution License.

Table 7C. Geometric means, variances and intraclass-correlation coefficients (ICCs) of serum concentrations of diacyl-phosphatidylcholines repeatedly measured 4 months apart among 100 healthy participants from the reliability study[6]

	1st Measurement		2nd Measurement			CV (%)			
Metabolite	**Mean** [µmol/L]	**95% CI**	**Mean** [µmol/L]	**95% CI**	***P-value*†**	**Within-person**	**Between-person**	**ICC‡**	**95% CI**
PC aa C24:0	0.12	(0.11-0.13)	0.12	(0.11-0.13)	0.56	32.3	25.8	0.39	(0.21-0.54)
PC aa C26:0*	0.84	(0.79-0.89)	0.80	(0.77-0.84)	0.04	16.6	20.4	0.60	(0.46-0.71)
PC aa C28:1	3.55	(3.37-3.74)	3.59	(3.43-3.77)	0.51	13.4	21.3	0.72	(0.61-0.80)
PC aa C30:0	4.89	(4.62-5.18)	4.86	(4.58-5.16)	0.81	19.6	21.8	0.55	(0.40-0.67)
PC aa C30:2	0.12	(0.10-0.16)	0.13	(0.10-0.16)	0.30	96.6	51.9	0.27	(0.00-0.43)
PC aa C32:0	16.2	(15.6-16.8)	16.4	(15.8-17.1)	0.38	13.2	14.3	0.54	(0.39-0.66)
PC aa C32:1	19.3	(17.6-21.2)	19.2	(17.4-21.2)	0.88	26.4	40.3	0.70	(0.58-0.79)
PC aa C32:2	4.71	(4.39-5.05)	4.59	(4.26-4.96)	0.44	23.6	28.2	0.59	(0.44-0.70)
PC aa C32:3	0.54	(0.51-0.57)	0.55	(0.52-0.58)	0.35	16.2	21.9	0.65	(0.52-0.75)
PC aa C34:1	245	(233-259)	254	(242-267)	0.13	16.2	20.3	0.61	(0.47-0.72)
PC aa C34:2	451	(433-469)	466	(450-483)	0.07	12.6	14.1	0.55	(0.40-0.68)
PC aa C34:3	17.8	(16.9-18.9)	18.0	(16.9-19.0)	0.80	18.8	21.8	0.57	(0.43-0.69)
PC aa C34:4	2.06	(1.92-2.21)	2.06	(1.92-2.22)	0.96	22.9	27.7	0.59	(0.45-0.71)
PC aa C36:0	2.72	(2.58-2.86)	2.79	(2.64-2.94)	0.30	17.8	20.1	0.56	(0.41-0.68)
PC aa C36:1	53.1	(50.4-55.8)	53.7	(51.3-56.1)	0.61	16.1	17.7	0.55	(0.39-0.67)
PC aa C36:2	267	(256-278)	272	(262-283)	0.32	14.3	13.2	0.46	(0.29-0.60)
PC aa C36:3	142	(135-148)	144	(138-150)	0.52	15.4	15.8	0.51	(0.35-0.64)
PC aa C36:4	200	(191-211)	209	(199-220)	0.02	13.5	20.8	0.70	(0.59-0.79)
PC aa C36:5	30.6	(27.9-33.6)	29.0	(26.5-31.8)	0.21	29.9	35.4	0.58	(0.44-0.70)
PC aa C36:6	1.16	(1.07-1.24)	1.09	(1.01-1.18)	0.11	24.6	28.9	0.58	(0.43-0.70)
PC aa C38:0	2.91	(2.75-3.08)	2.93	(2.78-3.10)	0.76	16.2	22.9	0.67	(0.540.76)
PC aa C38:1	0.43	(0.34-0.55)	0.43	(0.35-0.52)	0.94	100.5	15.3	0.03	(0.00-0.23)
PC aa C38:3	53.6	(50.9-56.4)	53.5	(51.1-56.0)	0.94	16.9	17.7	0.52	(0.37-0.65)
PC aa C38:4	118	(112-125)	123	(116-130)	0.08	14.6	22.6	0.70	(0.59-0.79)
PC aa C38:5	59.0	(56.0-62.2)	58.4	(55.4-61.6)	0.65	15.9	21.4	0.64	(0.51-0.75)
PC aa C38:6	86.0	(81.2-91.0)	84.9	(80.1-90.1)	0.61	16.8	23.9	0.67	(0.55-0.77)
PC aa C40:1*	0.42	(0.40-0.44)	0.40	(0.38-0.43)	0.26	18.3	17.2	0.47	(0.30-0.61)
PC aa C40:2	0.26	(0.24-0.28)	0.25	(0.23-0.27)	0.23	35.4	14.3	0.14	(0.00-0.33)
PC aa C40:3	0.48	(0.45-0.51)	0.46	(0.44-0.49)	0.28	26.9	16.5	0.28	(0.08-0.45)
PC aa C40:4	3.74	(3.54-3.94)	3.88	(3.68-4.09)	0.14	17.7	20.4	0.57	(0.42-0.69)
PC aa C40:5	10.7	(10.1-11.3)	10.7	(10.1-11.3)	0.90	17.4	22.4	0.62	(0.49-0.73)
PC aa C40:6	31.0	(29.1-33.1)	30.4	(28.5-32.4)	0.48	19.4	25.8	0.64	(0.51-0.74)
PC aa C42:0	0.54	(0.50-0.57)	0.53	(0.50-0.57)	0.98	17.2	28.7	0.73	(0.63-0.81)
PC aa C42:1	0.25	(0.24-0.27)	0.26	(0.24-0.28)	0.59	19.8	28.1	0.67	(0.54-0.76)
PC aa C42:2	0.21	(0.20-0.22)	0.21	(0.20-0.22)	0.87	22.8	18.8	0.40	(0.23-0.56)
PC aa C42:4	0.17	(0.16-0.17)	0.17	(0.16-0.18)	0.29	24.5	11.4	0.18	(0.00-0.36)
PC aa C42:5	0.39	(0.37-0.42)	0.37	(0.35-0.39)	0.13	23.2	19.7	0.42	(0.24-0.57)
PC aa C42:6	0.65	(0.63-0.68)	0.63	(0.60-0.66)	0.09	15.2	15.7	0.52	(0.36-0.65)

*Metabolite concentration was below the assay's limit of detection.
†*P*-value for difference based on a paired t-test of log-transformed values to compare geometric means of metabolite concentrations over time.
‡The intraclass-correlation coefficient (ICC) was calculated as the ratio of between-person variation and total variation.

[6] In the framework of this thesis, this table has already been published by the author: 108. Ibid. Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, Joost HG, Boeing H, Pischon T: Reliability of Serum Metabolite Concentrations over a 4-Month Period Using a Targeted Metabolomic Approach. PLoS One 2011;6:e21103. PLoS applies the Creative Commons Attribution License.

Table 7D. Geometric means, variances and intraclass-correlation coefficients (ICCs) of serum concentrations of acyl-alkyl-phosphatidylcholines and hexose repeatedly measured 4 months apart among 100 healthy participants from the reliability study[7]

	1st Measurement		2nd Measurement			CV (%)			
Metabolite	**Mean** [µmol/L]	**95% CI**	**Mean** [µmol/L]	**95% CI**	***P-value*†**	**Within-person**	**Between-person**	**ICC‡**	**95% CI**
PC ae C30:0	0.38	(0.36-0.40)	0.38	(0.36-0.41)	0.96	18.2	24.6	0.65	(0.52-0.75)
PC ae C30:1	0.25	(0.22-0.28)	0.21	(0.18-0.25)	0.12	67.9	17.7	0.06	(0.00-0.26)
PC ae C30:2	0.11	(0.11-0.12)	0.11	(0.10-0.12)	0.40	22.8	23.9	0.52	(0.37-0.65)
PC ae C32:1	3.07	(2.94-3.21)	3.10	(2.96-3.25)	0.64	13.3	18.4	0.66	(0.53-0.75)
PC ae C32:2	0.79	(0.75-0.83)	0.78	(0.75-0.82)	0.71	14.0	21.1	0.70	(0.58-0.78)
PC ae C34:0	1.55	(1.47-1.63)	1.57	(1.48-1.66)	0.70	18.0	20.9	0.57	(0.43-0.69)
PC ae C34:1	10.8	(10.4-11.3)	11.0	(10.5-11.5)	0.42	14.1	17.7	0.61	(0.47-0.72)
PC ae C34:2	12.8	(12.1-13.5)	13.2	(12.5-14.0)	0.14	17.1	22.3	0.63	(0.49-0.73)
PC ae C34:3	7.98	(7.54-8.44)	8.41	(7.93-8.91)	0.03	16.7	23.7	0.67	(0.54-0.76)
PC ae C36:0	0.68	(0.65-0.72)	0.71	(0.67-0.74)	0.18	17.4	20.1	0.57	(0.42-0.69)
PC ae C36:1	8.81	(8.40-9.23)	8.84	(8.44-9.27)	0.84	14.0	19.2	0.65	(0.53-0.75)
PC ae C36:2	14.1	(13.4-14.9)	14.1	(13.4-14.9)	1.00	13.0	23.0	0.76	(0.66-0.83)
PC ae C36:3	7.79	(7.39-8.21)	8.14	(7.74-8.55)	0.10	18.6	17.9	0.48	(0.32-0.62)
PC ae C36:4	18.8	(17.8-19.8)	19.9	(19.0-21.0)	0.02	17.8	19.4	0.54	(0.39-0.67)
PC ae C36:5	12.0	(11.3-12.6)	12.6	(11.9-13.3)	0.03	16.1	21.8	0.65	(0.52-0.75)
PC ae C38:0	1.69	(1.60-1.78)	1.66	(1.57-1.76)	0.49	17.3	21.6	0.61	(0.47-0.72)
PC ae C38:1	0.67	(0.60-0.74)	0.75	(0.69-0.82)	0.06	46.1	16.1	0.11	(0.00-0.30)
PC ae C38:2	1.48	(1.40-1.56)	1.44	(1.35-1.53)	0.38	24.5	15.7	0.29	(0.10-0.46)
PC ae C38:3	4.23	(4.03-4.45)	4.24	(4.04-4.45)	0.95	14.9	19.4	0.63	(0.50-0.73)
PC ae C38:4	13.5	(12.9-14.1)	14.0	(13.4-14.7)	0.04	13.5	18.6	0.65	(0.53-0.75)
PC ae C38:5	17.6	(16.8-18.4)	18.4	(17.6-19.2)	0.03	14.4	18.1	0.61	(0.47-0.72)
PC ae C38:6	7.82	(7.40-8.26)	7.87	(7.47-8.29)	0.76	16.2	21.6	0.64	(0.51-0.74)
PC ae C40:0*	8.62	(8.36-8.89)	8.59	(8.33-8.85)	0.80	10.3	11.5	0.56	(0.40-0.68)
PC ae C40:1	1.04	(0.98-1.09)	1.03	(0.98-1.09)	0.95	17.3	20.7	0.59	(0.44-0.70)
PC ae C40:2	2.12	(2.02-2.22)	2.11	(2.00-2.24)	0.94	14.0	22.5	0.72	(0.61-0.80)
PC ae C40:3	1.05	(1.00-1.10)	1.01	(0.96-1.06)	0.07	15.2	18.4	0.60	(0.45-0.71)
PC ae C40:4	2.11	(2.02-2.21)	2.18	(2.07-2.29	0.12	13.6	20.6	0.70	(0.58-0.79)
PC ae C40:5	3.66	(3.52-3.81)	3.71	(3.54-3.89)	0.51	14.0	17.1	0.60	(0.46-0.71)
PC ae C40:6	4.81	(4.59-5.05)	4.79	(4.55-5.05)	0.80	13.9	20.8	0.69	(0.57-0.78)
PC ae C42:0*	0.31	(0.30-0.32)	0.30	(0.29-0.32)	0.25	18.4	5.5	0.08	(0.00-0.27)
PC ae C42:1	0.36	(0.34-0.38)	0.36	(0.34-0.38)	0.93	19.0	18.7	0.49	(0.33-0.63)
PC ae C42:2	0.64	(0.61-0.67)	0.63	(0.60-0.66)	0.35	16.1	18.4	0.56	(0.42-0.68)
PC ae C42:3	0.68	(0.65-0.72)	0.65	(0.61-0.69)	0.10	18.8	22.3	0.59	(0.44-0.70)
PC ae C42:4	0.93	(0.88-0.98)	0.95	(0.90-1.01)	0.26	13.8	24.2	0.75	(0.65-0.83)
PC ae C42:5	2.10	(2.00-2.20)	2.12	(2.01-2.22)	0.66	11.7	21.5	0.77	(0.68-0.84)
PC ae C44:3	0.11	(0.10-0.12)	0.10	(0.10-0.11)	0.06	25.0	23.5	0.47	(0.30-0.61)
PC ae C44:4	0.38	(0.36-0.41)	0.40	(0.37-0.43)	0.10	18.8	29.1	0.71	(0.59-0.79)
PC ae C44:5	1.91	(1.81-2.02)	1.94	(1.83-2.06)	0.45	15.0	25.3	0.74	(0.64-0.82)
PC ae C44:6	1.27	(1.20-1.35)	1.30	(1.22-1.38)	0.38	13.3	27.3	0.81	(0.73-0.87)
Hexose	5059	(4811-5320)	5207	(4940-5488)	0.11	12.7	22.6	0.76	(0.66-0.83)

*Metabolite concentration was below the assay's limit of detection.

†*P*-value for difference based on a paired t-test of log-transformed values to compare geometric means of metabolite concentrations over time.

‡The intraclass-correlation coefficient (ICC) was calculated as the ratio of between-person variation and total variation.

[7] In the framework of this thesis, this table has already been published by the author: 108. Ibid. Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, Joost HG, Boeing H, Pischon T: Reliability of Serum Metabolite Concentrations over a 4-Month Period Using a Targeted Metabolomic Approach. PLoS One 2011;6:e21103. PLoS applies the Creative Commons Attribution License.

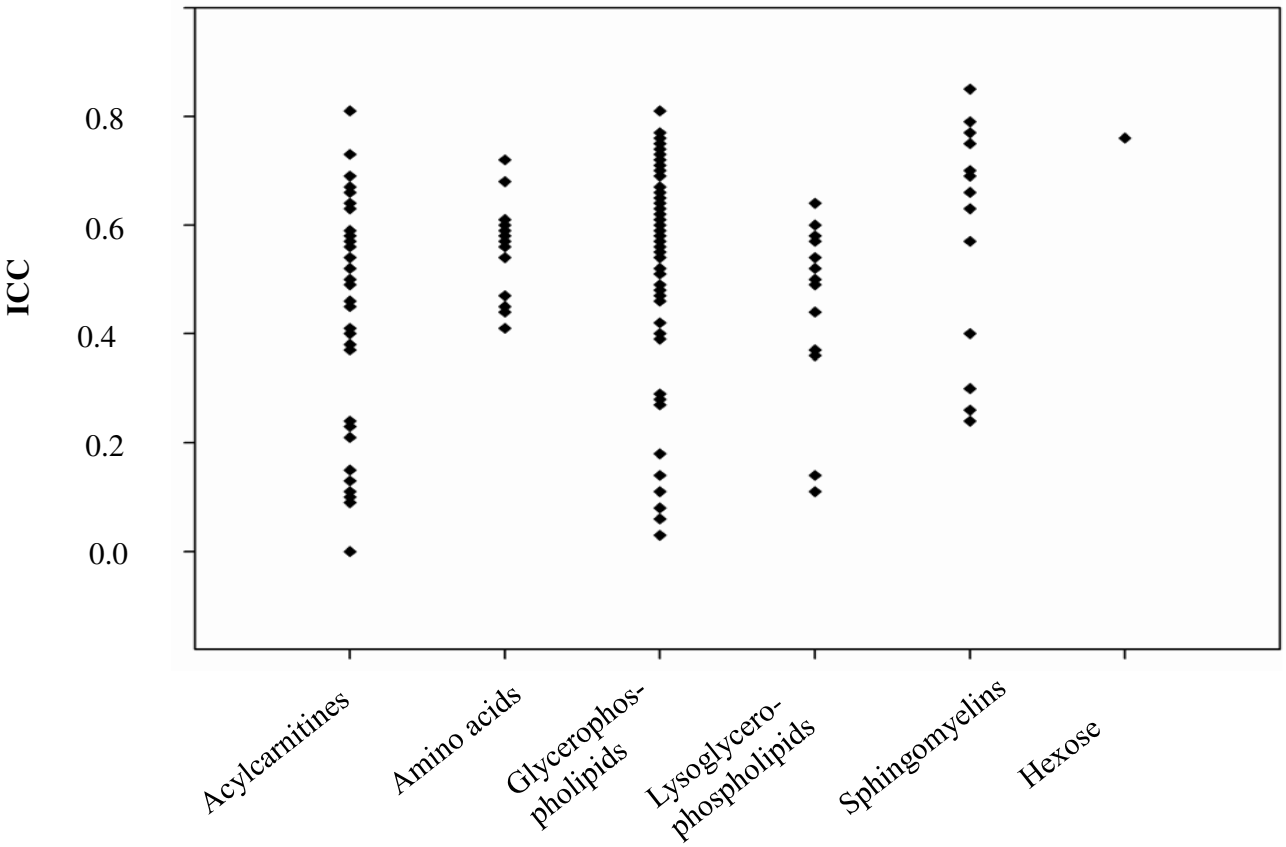


Figure 11. Intraclass correlation coefficients (ICC) of 163 metabolites repeatedly measured 4 month apart in 100 healthy EPIC-Potsdam participants.[8] Each data point represents one metabolite.

[8] In the framework of this thesis, this figure has already been published by the author: 108. Ibid. Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, Joost HG, Boeing H, Pischon T: Reliability of Serum Metabolite Concentrations over a 4-Month Period Using a Targeted Metabolomic Approach. PLoS One 2011;6:e21103. PLoS applies the Creative Commons Attribution License.

3.3 Dependency of metabolites

To investigate the dependency of metabolites, Spearman correlation coefficients were calculated between metabolites. The correlation coefficients of metabolites of the same subclasses are illustrated in **Figure 12** (**12A**: acylcarnitines, **12B**: amino acids and hexose, **12C**: lyso-phosphatidylcholines, **12D**: sphingomyelins, **12E**: diacyl-phosphatidylcholines, and **12F**: acyl-alkyl-phospatidylcholines).

In general, metabolites were positively correlated with other metabolites of the same class; however, the strength of the correlation differed. Strong positive correlations (Spearman correlation coefficient >0.70) could be found between acylcarnitines C14:1 and C14:2, and between C14:2 and C7-DC, and very strong positive correlations (Spearman correlation coefficient >0.80) between acylcarnitines C16:0 and C18:1 (**Figure 12A**). Among amino acids, there was a strong positive correlation between methionine, phenylalanine and tryptophan, and between phenylalanine and isoleucine. A very strong correlation was found between valine and isoleucine (**Figure 12B**). Strong positive correlations were found between lyso-phosphatidylcholines C16:0, C18:0 and C18:1, between lysophosphatidylcholines C18:1 and C20:3, and lysophosphatidylcholines C20:3 and C20:4 (**Figure 12C**). Strong to very strong positive correlations were observed between several C14, C16, C18, C22 and C24 sphingomyelins and hydroxy-sphingomyelins (**Figure 12D**). Strong to very strong correlations were also observed for several diacyl-phosphatidylcholines, particularly between the C34 and C36 diacyl-phosphatidylcholines, the C38 and C40 diacyl-phosphatidylcholines, and different C42 species (**Figure 12E**). For the acyl-alkyl-phosphatidylcholines, particularly neighboring species with similar number of carbon atoms were strongly to very strongly positively correlated, e.g. acyl-alkyl-phosphatidylcholines C34 and C36, C36 and C38, C38 and C40, and C40 and C42 (**Figure 12F**).

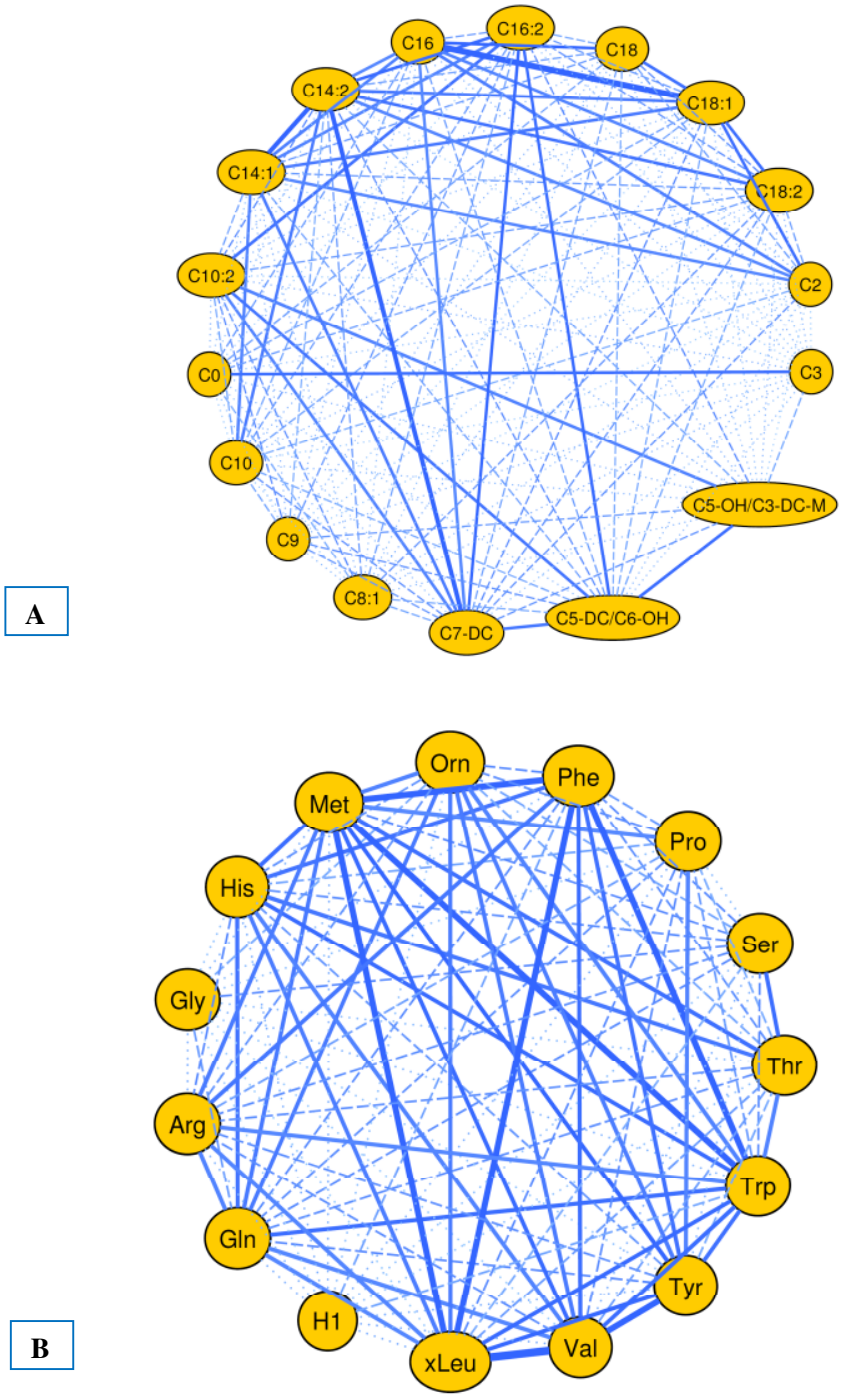
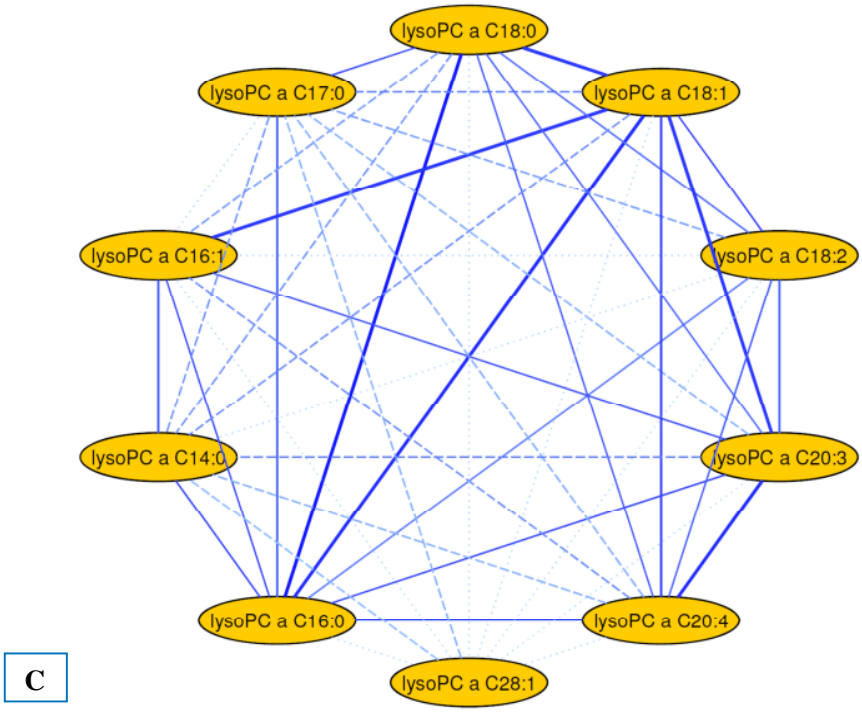


Figure 12 A/B. Correlation between different metabolites of acylcarnitines (A), and amino acids and hexose (B) in the EPIC-Potsdam subcohort. Solid, darker and thicker lines indicate a stronger positive correlation (Spearman correlation coefficients: solid line >0.50, solid thicker lines >0.70, >0.80, and >0.90; respectively).



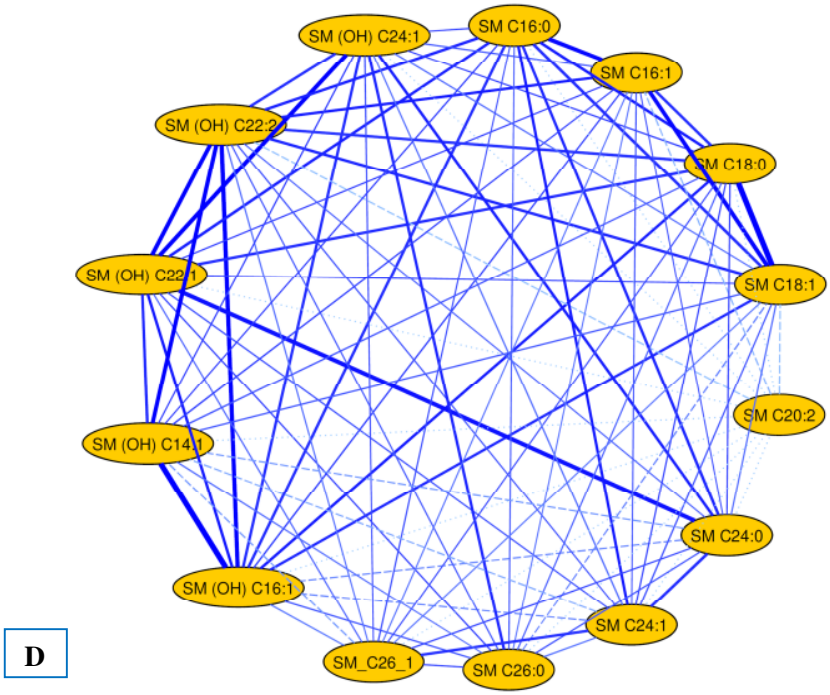


Figure 12 C/D. Correlation between different metabolites of lysophosphatidylcholines (C), and sphingomyelins (D) in the EPIC-Potsdam subcohort. Solid, darker and thicker lines indicate a stronger positive correlation (Spearman correlation coefficients: solid line >0.50, solid thicker lines >0.70, >0.80, and >0.90; respectively).

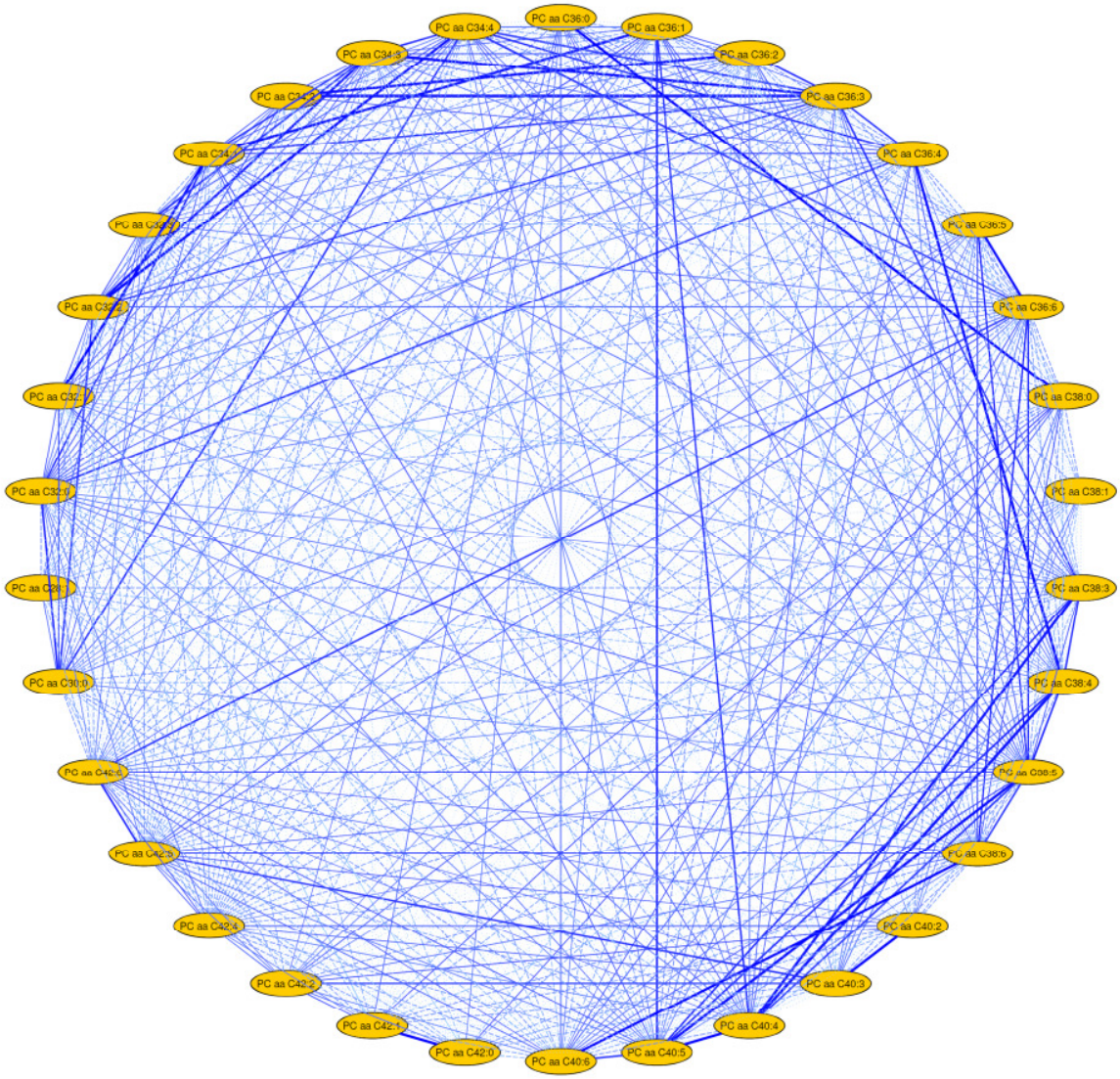


Figure 12 E. Correlation between different metabolites of diacyl-phosphatidylcholines in the EPIC-Potsdam subcohort. Solid, darker and thicker lines indicate a stronger positive correlation (Spearman correlation coefficients: solid line >0.50, solid thicker lines >0.70, >0.80, and >0.90; respectively).

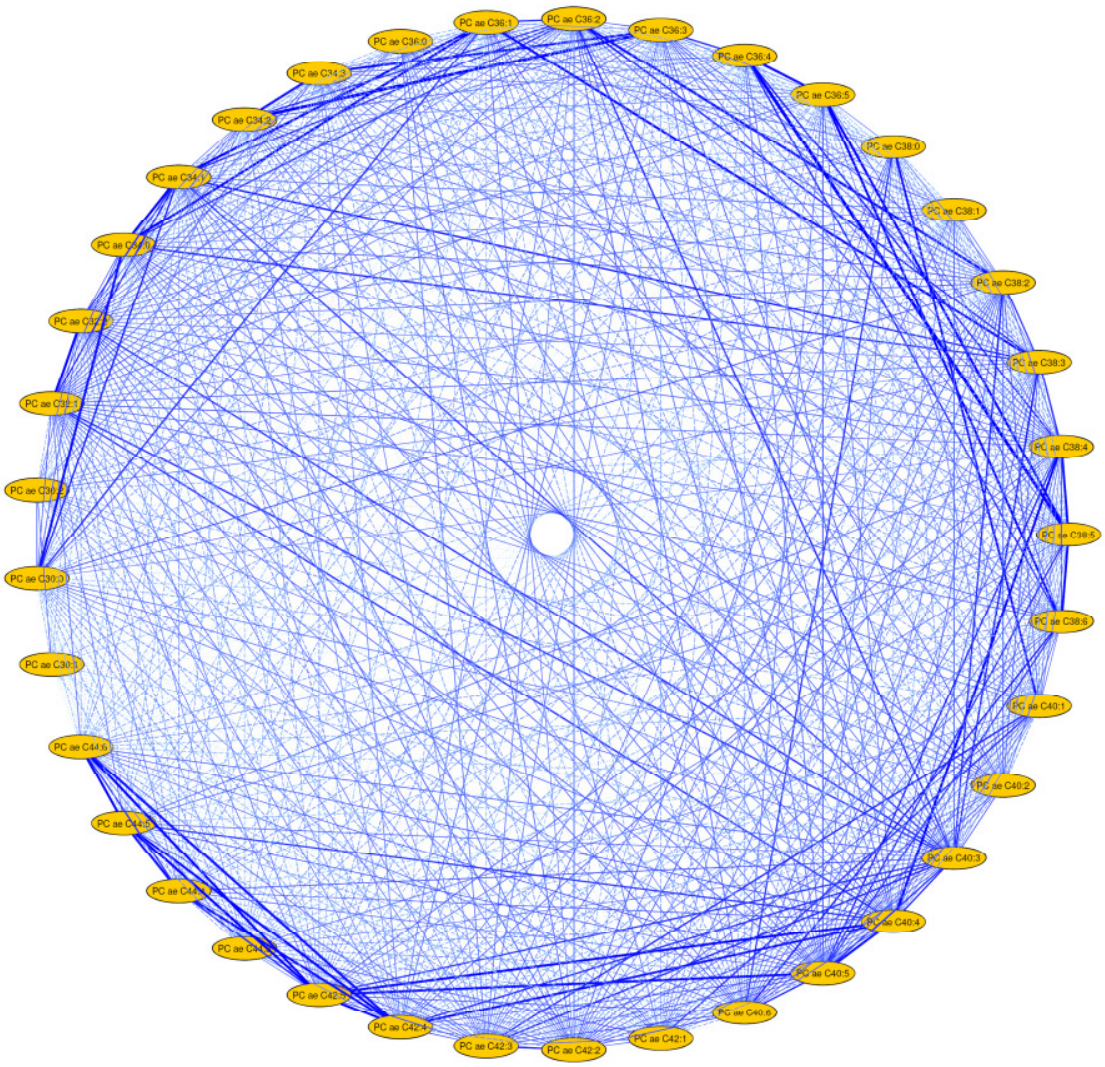


Figure 12 F. Correlation between different metabolites of acyl-alkyl-phosphatidylcholines in the EPIC-Potsdam subcohort. Solid, darker and thicker lines indicate a stronger positive correlation (Spearman correlation coefficients: solid line >0.50, solid thicker lines >0.70, >0.80, and >0.90; respectively).

3.4 Serum metabolites and risk of type 2 diabetes and cardiovascular diseases[9]

In the next step, the utility of the 127 metabolites, i.e. their association with risk of type 2 diabetes and CVD, was investigated within the EPIC-Potsdam case-cohort study. **Section 3.4.1** represents the results for type 2 diabetes, and **section 3.4.2** includes the results for CVD.

3.4.1 Serum metabolites and risk of type 2 diabetes

Baseline characteristics of the case-cohort study sample for type 2 diabetes

The baseline characteristics of the EPIC-Potsdam subcohort and all incident cases of type 2 diabetes are presented in **Table 8**. In general, participants who developed type 2 diabetes were older and more likely to be male. After adjusting for age and sex, their lifestyle habits and biomarker profiles at baseline were still disadvantageous compared to the subcohort. People who developed type 2 diabetes particularly had a higher BMI and waist circumference, were more frequently hypertensive and had lower HDL-cholesterol concentrations.

[9] In the framework of this thesis, parts of the results (type 2 diabetes) have already been published by the author: 109. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, Fritsche A, Haring HU, Hrabe de Angelis M, Peters A, Roden M, Prehn C, Wang-Sattler R, Illig T, Schulze MB, Adamski J, Boeing H, Pischon T: Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. Diabetes 2012; Oct.4th (Epub ahead).

Table 8. Age- and sex-adjusted baseline characteristics* (1994-1998) of the EPIC-Potsdam case-cohort sample for type 2 diabetes[10]		
	Subcohort (n=2282)	**Type 2 Diabetes Cases† (n=800)**
Age (years) ‡	49.5 (8.9)	54.7 (7.3)
Women (%)‡	62.0	42.2
BMI (kg/m²)	26.1 (0.09)	30.1 (0.15)
Waist circumference, men (cm) §	93.7 (0.34)	103.6 (0.46)
Waist circumference, women (cm) §	80.6 (0.30)	93.4 (0.62)
Prevalent Hypertension (%)	49.5	70.8
Education		
No degree/vocational training (%)	37.1	45.6
Trade/technical school (%)	24.0	25.4
University degree (%)	39.0	29.0
Smoking status		
Never (%)	46.9	36.2
Former (%)	33.0	42.3
Current (%)	20.1	21.5
Among smokers: number of cigarettes/d	12.6 (0.43)	16.0 (0.74)
Physical activity (h/week)¶	2.8 (0.07)	2.2 (0.13)
Alcohol intake from beverages (g/d)	14.8 (0.41)	14.5 (0.71)
Coffee consumption (cups/d)	2.8 (0.04)	2.7 (0.08)
Whole grain bread intake (g/d)	45.9 (1.11)	38.2 (1.91)
Red meat intake (g/d)	43.3 (0.61)	48.9 (1.06)
Biomarkers		
Glucose (mg/dL)	88.1 (0.53)	107.0 (0.92)
HbA_{1c} (%)‖	5.42 (0.01)	6.30 (0.03)
HDL-cholesterol (mg/dL)	47.5 (0.25)	40.9 (0.43)
Triglycerides (mg/dL)	114.8 (2.12)	177.2 (3.65)

*Presented are age- and sex-adjusted mean (standard error) for continuous variables or percentages for categorical variables.
†Includes only incident cases that occurred after the baseline examination.
‡Unadjusted mean (standard deviation) or percent.
§Age-adjusted mean (standard error).
¶Average of cycling and sports during summer and winter season.
‖Data was only available for n= 2900 (HbA1c).
Abbreviations: BMI, body mass index; HbA_{1c}, glycated hemoglobin A1C; HDL, high density lipoprotein.

[10] In the framework of this thesis, this table has already been published by the author: 109. Ibid. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, Fritsche A, Haring HU, Hrabe de Angelis M, Peters A, Roden M, Prehn C, Wang-Sattler R, Illig T, Schulze MB, Adamski J, Boeing H, Pischon T: Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. Diabetes 2012. Oct.4th (Epub ahead). Copyright: American Diabetes Association (ADA). http://diabetes.diabetesjournals.org

Identification of candidates associated with risk of type 2 diabetes

Of the 127 metabolites validly measured with the targeted metabolomics technology, 69 metabolites were associated with risk of type 2 diabetes independent of relevant dietary and lifestyle factors, anthropometry and hypertension (**Table S3**). After correction for multiple testing, 34 metabolites remained significantly associated with risk of type 2 diabetes. Of them, 14 metabolites were identified with stepwise regression to be associated with risk of type 2 diabetes independent of the others (**Table 9**). On the one hand, hexose, phenylalanine, and diacyl-phosphatidylcholines C32:1, C36:1, C38:3 and C40:5 were significantly positively associated with risk of type 2 diabetes. On the other hand, glycine, sphingomyelin C16:1, lysophosphatidylcholine C18:2 and acyl-alkyl-phosphatidylcholines C34:3, C40:6, C42:5, C44:4 and C44:5 were significantly inversely associated with risk of type 2 diabetes.

The impact of reliability of the metabolite concentrations was evaluated by correcting observed RR by the ICC. For metabolites significantly associated with risk of type 2 diabetes the ICC ranged between 0.47 and 0.77. The expected RR estimates were always higher in magnitude compared to observed RR estimates. For some metabolites expected RR was not in the 95% CI of the observed RR, particularly for metabolites with higher observed RR (e.g. hexose) or those metabolites with smaller ICCs (e.g. phenylalanine), **Table 9**. For most metabolites, ranking of their risk estimates was not changed comparing expected and observed RR (**Table S4**). In the present study, statistical power enabled detection of RR per SD of about 1.18 for type 2 diabetes considering multiple testing correction (as seen in **Table 9**). Theoretically assuming that all metabolite measurements were done with an ICC of 0.4, which has been used as the cut-off point of acceptable reliability in previous studies (19; 20), in the present study statistical power could be increased when improving reliability up to a detectable RR per SD of 1.07.

Table 9. Mean serum concentrations of identified metabolites and their association with risk of type 2 diabetes and possible impact of reliability of metabolites in the EPIC-Potsdam case-cohort study[11]

	Subcohort n=2282 µmol/L*	Cases n=800 µmol/L*	Relative risk per SD (95% CI), basic model†	ICC	Expected relative risk per SD‡
Hexose	4698 (984)	5783 (2167)	2.36 (2.06-2.71)	0.76	3.10
Phenylalanine	56.4 (12.0)	61.0 (12.9)	1.35 (1.22-1.49)	0.47	1.89
Glycine	256 (77.2)	227 (59.6)	0.73 (0.64-0.83)	0.68	0.63
SM C16:1	17.6 (3.86)	17.4 (4.02)	0.80 (0.72-0.88)	0.75	0.74
PC ae C34:3	8.51 (2.48)	7.01 (2.02)	0.64 (0.56-0.72)	0.67	0.51
PC ae C40:6	5.52 (1.43)	5.03 (1.34)	0.77 (0.69-0.86)	0.69	0.68
PC ae C42:5	2.35 (0.54)	2.06 (0.46)	0.71 (0.63-0.79)	0.77	0.64
PC ae C44:4	0.38 (0.10)	0.33 (0.09)	0.76 (0.67-0.85)	0.71	0.68
PC ae C44:5	1.77 (0.48)	1.53 (0.43)	0.71 (0.63-0.80)	0.74	0.63
PC aa C32:1	16.9 (9.76)	20.4 (11.3)	1.18 (1.08-1.29)	0.70	1.27
PC aa C36:1	57.3 (15.6)	62.1 (17.2)	1.18 (1.08-1.30)	0.55	1.35
PC aa C38:3	55.4 (14.8)	65.3 (18.2)	1.37 (1.24-1.51)	0.52	1.83
PC aa C40:5	11.4 (3.49)	12.8 (4.17)	1.18 (1.08-1.29)	0.62	1.31
LysoPC a C18:2	35.2 (13.7)	29.6 (11.1)	0.74 (0.66-0.84)	0.58	0.60

*Presented are metabolite concentrations as mean (standard deviation).
†Adjusted for age, sex, alcohol intake from beverages (non-consumers; women: >0-6 g/d, 6-12 g/d, >12 g/d; men: >0-12 g/d, 12-24 g/d, >24 g/d), smoking (never, former, current ≤20 cigarettes/d, current >20 cigarettes/d), physical activity (cycling and sports in h/week), education (low, medium, high), coffee intake (cups/d), red meat intake (g/d) and whole grain bread intake (g/d), prevalent hypertension (yes/no), BMI (kg/m²), and waist circumference (cm).
‡Expected relative risk was additionally calculated to account for reliability of metabolites, based on the observed relative risk and ICCs for the individual metabolites (15).
Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; ICC, intraclass-correlation coefficient, PC, phosphatidylcholine; SM, sphingomyelin.

[11] In the framework of this thesis, parts of this table have already been published by the author: 109. Ibid. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, Fritsche A, Haring HU, Hrabe de Angelis M, Peters A, Roden M, Prehn C, Wang-Sattler R, Illig T, Schulze MB, Adamski J, Boeing H, Pischon T: Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. Diabetes 2012. Oct.4th (Epub ahead). Copyright: American Diabetes Association (ADA). http://diabetes.diabetesjournals.org

Evaluation of independency of risk associations for type 2 diabetes

As many metabolites were associated with risk of type 2 diabetes, additionally a PCA was conducted to account for their dependency. The PCA was based on the correlation matrix of the 34 metabolites associated with risk of type 2 diabetes. The scree plot of all the metabolite factors is shown in **Figure S1**. In general, these factors included metabolites of the same classes. Of all factors, 2 metabolite factors were identified that explained more than 50% of the variation of the individual metabolites. Metabolite factor 1 included acyl-alkyl-phosphatidylcholines, sphingomyelins, and lyso-phosphatidylcholines, whereas metabolite factor 2 contained diacyl-phosphatidylcholines, branched chain and aromatic amino acids, propionyl-carnitine and hexose (**Figure 13**). The distribution of metabolite factors in this population is shown in **Figure S2**. The metabolite factors were linked to established biomarkers of risk of chronic diseases (**Table 10**). Metabolite factor 1 was positively correlated with HDL-cholesterol and total cholesterol and inversely correlated with triglycerides, glucose and HbA_{1c}. Metabolite factor 2 was positively correlated with triglycerides, total cholesterol, glucose and HbA_{1c}. The metabolite factors showed significant and opposing associations with risk of type 2 diabetes. When comparing extreme quintiles (**Table 11**) metabolite factor 1 was associated with a 69% lower risk of type 2 diabetes; opposite to metabolite factor 2, which was associated with a 3.82-fold increased risk of type 2 diabetes.

To investigate whether the association was independent of established biomarkers the analysis was additionally adjusted for glucose, HbA1c, HDL-cholesterol and triglycerides. The results are presented in **Table 12**. The association between acyl-alkyl-phosphatidylcholines, and lyso-phosphatidylcholine and type 2 diabetes risk was slightly weakened after adjustment for HDL-cholesterol; whereas the association of diacyl-phosphatidylcholines was weakened after adjustment for triglycerides. Glucose adjustment particularly weakened the associations between hexose, phenylalanine and glycine and type 2 diabetes risk. In the full model, adjusted for glucose, HbA1c, HDL-cholesterol and triglycerides, 7 out of 14 metabolites (i.e. hexose, phenylalanine, diacyl-phosphatidylcholines C32:1, C36:1, C38:3 and C40:5 and lyso-phosphatidylcholine C18:2) and metabolite factor 2 remained significantly associated with risk of type 2 diabetes.

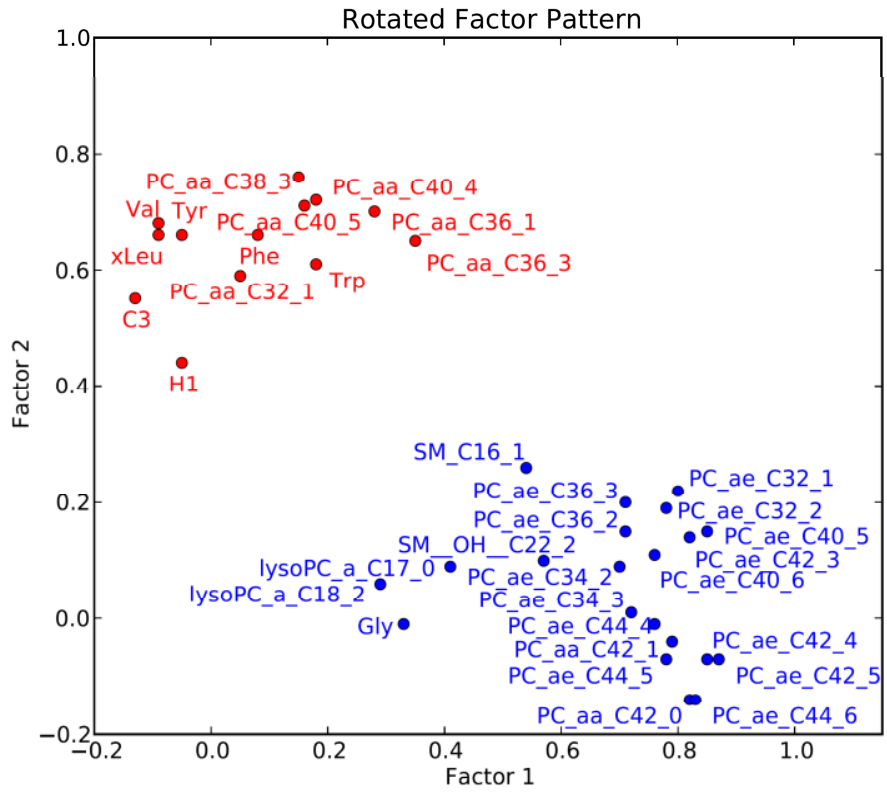


Figure 13. Two metabolite factors associated with risk of type 2 diabetes.[12] Presented is a two-dimensional factor loadings plot obtained from principal component analysis (PCA). Metabolites colored in blue are associated with decreased risk of type 2 diabetes, whereas metabolites in red are associated with increased risk of type 2 diabetes. The factor loadings represent the correlation coefficients of individual metabolites with corresponding metabolite factors, and may range between -1 and 1. The PCA was based on the correlation matrix of all metabolites significantly associated with risk of type 2 diabetes in the EPIC-Potsdam case-cohort study. An orthogonal varimax rotation was used, and two factors were retained. Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; C3, propionyl-carnitine; Gly, glycine; H1, hexose; PC, phosphatidylcholine; Phe, phenylalanine; SM, sphingomyelin; Trp, tryptophan; Tyr, tyrosine; Val; valine; xLeu, isoleucine.

Table 10. Correlation* of metabolite factors with established biomarkers of risk of type 2 diabetes and cardiovascular diseases in the EPIC-Potsdam subcohort (n=2282)

	Glucose	HbA1c	HDL-Chol	Total-Chol	Triglycerides	CRP
Factor 1	-0.13	-0.16	0.46	0.21	-0.27	-0.12
Factor 2	0.17	0.12	-0.02	0.30	0.46	0.14

*Depicted are Spearman correlation coefficients. Blue color indicates positive correlation and red color indicates negative correlation. All correlations were significant $p<0.0001$, except factor 2 and HDL-Chol, $p=0.307$. Abbreviations: Chol, cholesterol; CRP, C-reactive protein; HbA1c, glycated hemoglobin A1c; HDL, high density lipoprotein.

[12] In the framework of this thesis, this figure has already been published by the author: 109. Ibid. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, Fritsche A, Haring HU, Hrabe de Angelis M, Peters A, Roden M, Prehn C, Wang-Sattler R, Illig T, Schulze MB, Adamski J, Boeing H, Pischon T: Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. Diabetes 2012. Oct.4th (Epub ahead). Copyright: American Diabetes Association (ADA). http://diabetes.diabetesjournals.org

Table 11. Relative risk of type 2 diabetes by quintiles of metabolite factors[13]

	N (Cases)	Relative risk (95% CI), basic model*
Factor 1†		
Quintile 1	296	1.00
Quintile 2	194	0.61 (0.46-0.80)
Quintile 3	148	0.50 (0.37-0.67)
Quintile 4	95	0.37 (0.27-0.51)
Quintile 5	67	0.31 (0.21-0.44)
***P*-Trend**		<0.0001
	N (Cases)	**Relative risk (95% CI), basic model***
Factor 2†		
Quintile 1	51	1.00
Quintile 2	102	1.23 (0.82-1.85)
Quintile 3	149	1.72 (1.17-2.51)
Quintile 4	191	1.80 (1.24-2.63)
Quintile 5	307	3.82 (2.64-5.52)
***P*-Trend**		<0.0001

*Adjusted for age, sex, alcohol intake from beverages (non-consumers; women: >0-6 g/d, 6-12 g/d, >12 g/d; men: >0-12 g/d, 12-24 g/d, >24 g/d), smoking (never, former, current ≤20 cigarettes/d, current >20 cigarettes/d), physical activity (cycling and sports in h/week), education (low, medium, high), coffee intake (cups/d), red meat intake (g/d) and whole grain bread intake (g/d), prevalent hypertension (yes/no), BMI (kg/m²), and waist circumference (cm).

†Factor 1= 0.80 x PC ae C32:1 + 0.78 x PC ae C32:2 + 0.70 x PC ae C34:2 + 0.72 x PC ae C34:3 + 0.71 x PC ae C36:2 + 0.71 x PC ae C36:3 + 0.85 x PC ae C40:5 + 0.76 x PC ae C40:6 + 0.82 x PC ae C42:3 + 0.85 x PC ae C42:4 + 0.87 x PC ae C42:5 + 0.76 x PC ae C44:4 + 0.78 x PC ae C44:5 + 0.83 x PC ae C44:6 + 0.82 x PC aa C42:0 + 0.79 x PC aa C42:1 + 0.54 x SM C16:1 + 0.57 x SM OH C22:2 + 0.41 x lysoPC a C17:0. Factor 2= 0.55 x propionyl-carnitine + 0.66 x phenylalanine + 0.61 x tryptophan + 0.66 x tyrosine + 0.68 x valine + 0.66 x isoleucine + 0.59 x PC aa C32:1 + 0.70 x PC aa C36:1 + 0.65 x PC aa C36:3 + 0.76 x PC aa C38:3 + 0.72 x PC aa C40:4 + 0.71 x PC aa C40:5 + 0.44 x hexose.

Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; PC, phosphatidylcholine; SM, sphingomyelin.

[13] In the framework of this thesis, this table has already been published by the author: 109. Ibid. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, Fritsche A, Haring HU, Hrabe de Angelis M, Peters A, Roden M, Prehn C, Wang-Sattler R, Illig T, Schulze MB, Adamski J, Boeing H, Pischon T: Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. Diabetes 2012. Oct.4th (Epub ahead). Copyright: American Diabetes Association (ADA). http://diabetes.diabetesjournals.org

Table 12. Metabolites and metabolite factors and their association with risk of type 2 diabetes in the EPIC-Potsdam case-cohort sample with additional adjustment for established biomarkers[14]

	Relative Risk per SD (95% CI)					
	Basic model*	**+ Glucose†**	**+ HbA_{1c}†**	**+ HDL-Chol†**	**+ Triglycerides†**	**+ Glucose + HbA_{1c} + HDL-Chol + Triglycerides†**
Hexose	2.36 (2.06-2.71)	1.62 (1.37-1.91)	2.11 (1.82-2.44)	2.43 (2.11-2.79)	2.34 (2.04-2.69)	1.51 (1.28-1.80)
Phenylalanine	1.35 (1.22-1.49)	1.24 (1.12-1.37)	1.37 (1.23-1.52)	1.34 (1.21-1.49)	1.30 (1.16-1.44)	1.25 (1.12-1.39)
Glycine	0.73 (0.64-0.83)	0.81 (0.71-0.92)	0.81 (0.71-0.93)	0.74 (0.65-0.84)	0.74 (0.65-0.84)	0.89 (0.78-1.01)
SM C16:1	0.80 (0.72-0.88)	0.82 (0.73-0.92)	0.74 (0.67-0.83)	0.87 (0.78-0.97)	0.79 (0.71-0.88)	0.91 (0.80-1.03)
PC ae C34:3	0.64 (0.56-0.72)	0.70 (0.61-0.80)	0.67 (0.58-0.77)	0.72 (0.63-0.82)	0.66 (0.58-0.74)	0.92 (0.78-1.07)
PC ae C40:6	0.77 (0.69-0.86)	0.81 (0.72-0.92)	0.83 (0.74-0.94)	0.82 (0.73-0.92)	0.76 (0.67-0.85)	0.94 (0.82-1.08)
PC ae C42:5	0.71 (0.63-0.79)	0.76 (0.67-0.86)	0.82 (0.72-0.92)	0.76 (0.68-0.86)	0.73 (0.64-0.82)	0.96 (0.83-1.12)
PC ae C44:4	0.76 (0.67-0.85)	0.77 (0.68-0.87)	0.84 (0.75-0.96)	0.82 (0.72-0.93)	0.77 (0.68-0.87)	0.97 (0.84-1.12)
PC ae C44:5	0.71 (0.63-0.80)	0.74 (0.65-0.84)	0.81 (0.71-0.91)	0.77 (0.68-0.87)	0.74 (0.65-0.83)	0.95 (0.82-1.09)
PC aa C32:1	1.18 (1.08-1.29)	1.12 (1.02-1.24)	1.12 (1.02-1.23)	1.23 (1.11-1.36)	1.12 (1.02-1.22)	1.15 (1.03-1.29)
PC aa C36:1	1.18 (1.08-1.30)	1.18 (1.07-1.30)	1.09 (0.98-1.21)	1.26 (1.14-1.39)	1.07 (0.96-1.19)	1.25 (1.10-1.41)
PC aa C38:3	1.37 (1.24-1.51)	1.35 (1.22-1.50)	1.31 (1.18-1.45)	1.38 (1.25-1.52)	1.24 (1.11-1.38)	1.38 (1.22-1.55)
PC aa C40:5	1.18 (1.08-1.29)	1.21 (1.10-1.33)	1.12 (1.02-1.24)	1.21 (1.10-1.32)	1.07 (0.96-1.18)	1.19 (1.06-1.32)
LysoPC a C18:2	0.74 (0.66-0.84)	0.72 (0.64-0.81)	0.77 (0.68-0.88)	0.79 (0.70-0.90)	0.72 (0.63-0.81)	0.84 (0.73-0.96)
Factor 1	0.69 (0.62-0.77)	0.72 (0.64-0.81)	0.74 (0.66-0.83)	0.75 (0.67-0.85)	0.69 (0.62-0.77)	0.93 (0.80-1.07)
Factor 2	1.50 (1.34-1.67)	1.36 (1.22-1.51)	1.43 (1.28-1.60)	1.51 (1.36-1.69)	1.37 (1.21-1.54)	1.38 (1.23-1.56)

*Adjusted for age, sex, alcohol intake from beverages (non-consumers; women: >0-6 g/d, 6-12 g/d, >12 g/d; men: >0-12 g/d, 12-24 g/d, >24 g/d), smoking (never, former, current ≤20 cigarettes/d, current >20 cigarettes/d), physical activity (cycling and sports in h/week), education (low, medium, high), coffee intake (cups/d), red meat intake (g/d) and whole grain bread intake (g/d), prevalent hypertension (yes/no), BMI (kg/m^2), and waist circumference (cm).
†Adjustment according to the basic model with additional adjustment for specified biomarkers.
Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; HbA_{1c}, glycated hemoglobin A1C; HDL-Chol, high-density lipoprotein cholesterol; PC, phosphatidylcholine; SM, sphingomyelin.

[14] In the framework of this thesis, parts of this table have already been published by the author: 109. Ibid. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, Fritsche A, Haring HU, Hrabe de Angelis M, Peters A, Roden M, Prehn C, Wang-Sattler R, Illig T, Schulze MB, Adamski J, Boeing H, Pischon T: Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. Diabetes 2012. Oct.4th (Epub ahead). Copyright: American Diabetes Association (ADA). http://diabetes.diabetesjournals.org

Usefulness for prediction of type 2 diabetes

To evaluate the usefulness of identified metabolites for prediction of type 2 diabetes models with and without metabolites were compared regarding measures of discrimination and calibration. The ability of metabolites alone to discriminate between cases and non-cases of type 2 diabetes was similar to that of the German Diabetes Risk Score, which includes dietary and lifestyle factors, anthropometry and hypertension (area under the ROC curve: 0.849 and 0.847, respectively), **Figure 14**. When the metabolites were added to established risk prediction models, prediction of type 2 diabetes could further be improved up to an area under the ROC curve of 0.912. Models including the metabolites were well calibrated as indicated by Hosmer-Lemeshow test (**Table 13**).

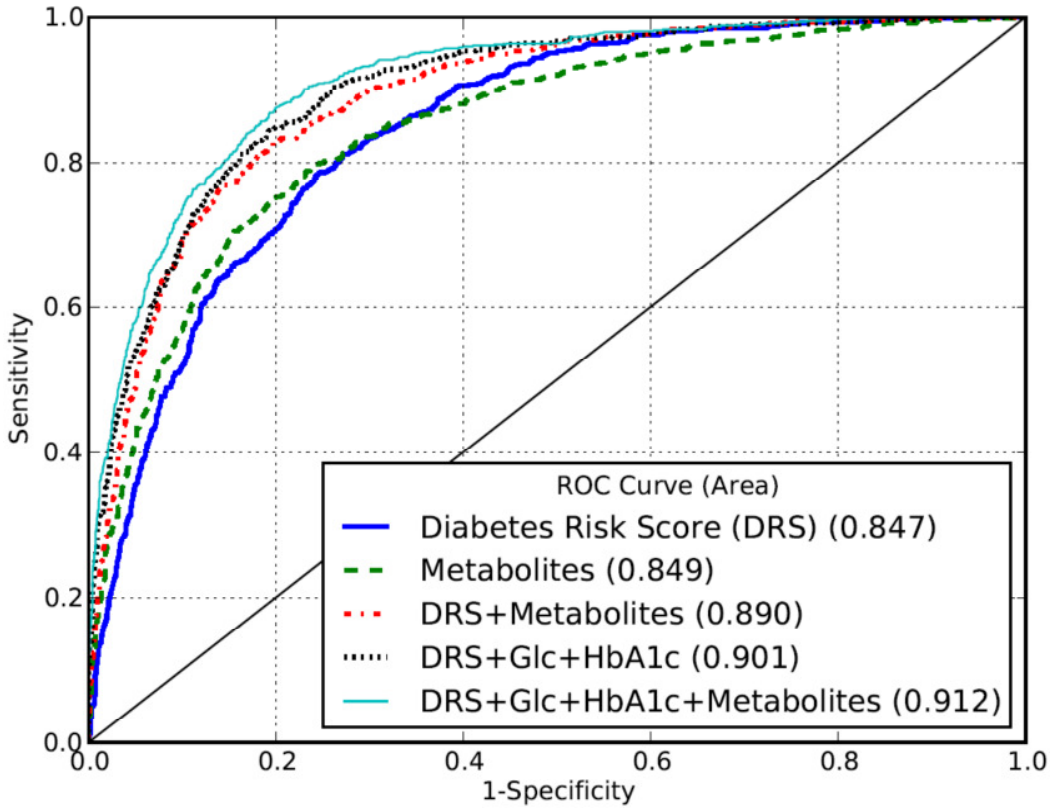


Figure 14. Relative contribution of metabolites to predict type 2 diabetes in the EPIC-Potsdam case-cohort sample.[15] Presented are receivers operating characteristic (ROC) curves comparing different multivariable adjusted models to predict type 2 diabetes, including the German Diabetes Risk Score (which includes classical risk factors), the identified metabolites (hexose, phenylalanine, glycine, SM C16:1, PC aa C32:1, PC aa C36:1, PC aa C38:3, PC aa C40:5, PC ae C34:3, PC ae C40:6, PC ae C42:5, PC ae C44:4, PC ae C44:5, lysoPC a C18:2) and established biomarkers (glucose, and HbA_{1c}). Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; HbA1c, glycated hemoglobin A1c; PC, phosphatidylcholine; SM, sphingomyelin.

[15] In the framework of this thesis, this figure has already been published by the author: 109. Ibid. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, Fritsche A, Haring HU, Hrabe de Angelis M, Peters A, Roden M, Prehn C, Wang-Sattler R, Illig T, Schulze MB, Adamski J, Boeing H, Pischon T: Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. Diabetes 2012. Oct.4th (Epub ahead). Copyright: American Diabetes Association (ADA). http://diabetes.diabetesjournals.org

Table 13. Measures* of discrimination and calibration comparing multivariable adjusted models to predict type 2 diabetes in the EPIC-Potsdam case-cohort sample[16]

	-2LOG-Likelihood	Nagelkerke R^2†	Hosmer-Lemeshow‡		Receiver Operating Characteristic (ROC)§			
Model			χ^2	***P*-Value**	**Area under the curve (95% CI)**	***P* for Difference**		
Diabetes Risk Score (DRS)¶	2413	0.40	15.89	0.044	0.847 (0.832-0.862)	Ref.		
Metabolites‖	2387	0.41	13.53	0.095	0.849 (0.833-0.865)	0.838		
DRS + Glucose	2210	0.47	18.41	0.018	0.873 (0.859-0.887)	<0.0001	Ref.	
DRS + Metabolites‖	2083	0.52	9.39	0.310	0.890 (0.877-0.903)	<0.0001	<0.0001	
DRS + Glucose + HbA_{1c}	1978	0.55	15.25	0.054	0.901 (0.888-0.913)	<0.0001	<0.0001	Ref.
DRS + Glucose + HbA_{1c} + Metabolites‖	1864	0.58	15.04	0.058	0.912 (0.901-0.924)	<0.0001	<0.0001	<0.0001

*Better model performance is indicated by smaller -2LOG-Likelihood Ratio, larger Nagelkerke's R^2, good calibration with a Homer-Lemeshow *P*-Value ≥0.05, and better discrimination mirrored by larger area under the ROC curve with significant improvement of model performance, *P* for difference <0.05.
†Specifically, Nagelkerke's R^2 defines the fraction of -2 log likelihood explained by the predictors ranging from 0 to 1. Larger values indicate better model fit (105).
‡As a measure of model calibration, the Hosmer-Lemeshow statistic compares predicted and observed probabilities of type 2 diabetes derived from deciles of predicted risk. Smaller $\chi 2$ values and larger *p*-values specify better model fit. *P*-values <0.05 indicate significant difference between expected and observed probabilities (107).
§The ROC curve is a plot of sensitivity versus false-positive rate of a continuous prediction model on a binary outcome. The area under the ROC curve, a measure of discrimination that mirrors the probability the model assigns a higher risk to future cases compared to controls. It may range from 0.5 (no discrimination) to 1.0 (perfect discrimination) (110). The method of DeLong (106) was used to test for difference between the areas under the ROC curve, $p<0.05$ indicate significant model improvement.
¶The German Diabetes Risk Score (7) combines information on several diabetes risk factors, such as diet, lifestyle and anthropometry, to estimate risk of developing type 2 diabetes.
‖Metabolites linked to type 2 diabetes risk: hexose, phenylalanine, glycine, SM C16:1, PC aa C32:1, PC aa C36:1, PC aa C38:3, PC aa C40:5, PC ae C34:3, PC ae C40:6, PC ae C42:5, PC ae C44:4, PC ae C44:5, lysoPC a C18:2.
Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; CI, confidence interval; DRS, Diabetes Risk Score; PC, phosphatidylcholine; ROC, receiver operating characteristic; SM, sphingomyelin.

[16] In the framework of this thesis, parts of this table have already been published by the author: 109. Ibid. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, Fritsche A, Haring HU, Hrabe de Angelis M, Peters A, Roden M, Prehn C, Wang-Sattler R, Illig T, Schulze MB, Adamski J, Boeing H, Pischon T: Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. Diabetes 2012. Oct.4th (Epub ahead). Copyright: American Diabetes Association (ADA). http://diabetes.diabetesjournals.org

3.4.2 Serum metabolites and risk of cardiovascular diseases

Baseline characteristics of the case-cohort study sample for cardiovascular diseases

Next, the associations between metabolites and risk of CVD were investigated. The baseline characteristics of the EPIC-Potsdam subcohort and all incident cases of CVD are presented in **Table 14**. Individuals who developed CVD were older and more likely to be male. After adjusting for age and sex, the baseline characteristics of incident cases of CVD remained unfavorable. Incident cases of myocardial infarction particularly had the highest proportion of smokers and lowest HDL-cholesterol together with highest total cholesterol. People who developed stroke were oldest and had the highest CRP.

Table 14. Age- and sex-adjusted baseline characteristics* (1994-1998) of the EPIC-Potsdam case-cohort sample for cardiovascular diseases

	Subcohort (n=2309)	MI cases† (n=235)	Stroke cases† (n=178)
Age (years)‡	49.5 (8.9)	55.3 (7.1)	56.1 (7.8)
Women (%)‡	62.3	28.1	46.1
BMI (kg/m²)	26.2 (0.09)	26.9 (0.28)	26.5 (0.32)
Waist circumference, men (cm)§	94.2 (0.34)	96.8 (0.77)	95.4 (1.0)
Waist circumference, women (cm)§	80.9 (0.29)	82.8 (1.4)	81.8 (1.2)
Prevalent Hypertension (%)	48.8	58.7	64.4
Education			
No degree/vocational training (%)	36.8	40.3	41.5
Trade/technical school (%)	23.8	23.9	29.2
University degree (%)	39.4	35.8	29.3
Smoking status			
Never (%)	46.9	31.3	41.6
Former (%)	32.4	22.3	34.3
Current (%)	20.6	46.4	24.1
Among smokers: number of cigarettes/d	13.1 (0.45)	16.2 (0.97)	16.5 (1.6)
Physical activity (h/week) ¶	2.8 (0.07)	2.0 (0.23)	2.6 (0.3)
Alcohol intake from beverages (g/d)	14.9 (0.38)	10.3 (1.2)	14.8 (1.4)
Biomarkers			
C-reactive protein (mg/L)	0.17 (0.01)	0.23 (0.03)	0.31 (0.03)
HDL-cholesterol (mg/dL)	47.5 (0.3)	44.1 (0.8)	46.0 (0.94)
Total cholesterol (mg/dL)	174.6 (0.77)	184.1 (2.5)	171.5 (2.8)

*Presented are age- and sex-adjusted mean (SE) for continuous variables or percentages for categorical variables.
†Includes only incident cases that occurred after the baseline examination.
‡Unadjusted mean (SD) or percent.
§Age-adjusted mean (SE).
¶Average of cycling and sports during summer and winter season.
Abbreviations: BMI, body mass index; HDL, high density lipoprotein.

Association between metabolite factors linked to type 2 diabetes and risk of cardiovascular diseases

In a first step, the association between metabolite factors –that were linked to risk of type 2 diabetes- and risk of CVD was evaluated, as it was assumed that common risk factors may be involved in development of these chronic diseases. As type 2 diabetes is an established risk factor for CVD, the association was studied with and without adjustment for prevalent type 2 diabetes. There was a positive association between metabolite factor 2 and risk of stroke (RR: 1.87; 95% CI: 1.05-3.36, when comparing extreme quintiles) which was slightly weakened after adjustment for prevalent type 2 diabetes, but the trend remained significant (**Table 15**). No significant association was found for metabolite factor 1 and risk of stroke. In addition, both metabolite factors were associated with a higher risk of myocardial infarction. When comparing extreme quintiles metabolite factor 2 was associated with a 2.2-fold increased risk of myocardial infarction independently of established risk factors and prevalent diabetes (**Table 16**). Metabolite factor 1, which was associated with a lower risk of type 2 diabetes, was associated with an 87% higher risk of myocardial infarction after multivariable adjustment and when comparing extreme quintiles. The association was non-significant without adjustment for anthropometric measures and prevalent diabetes (data not shown).

Table 15. Relative risk of stroke by quintiles of metabolite factors linked to type 2 diabetes			
	N (Cases)	**Relative risk (95% CI), basic model***	**Relative risk (95% CI), basic model* – without adjustment for prevalent diabetes**
Factor 1†			
Quintile 1	46	1.00	1.00
Quintile 2	30	0.73 (0.44-1.21)	0.65 (0.40-1.08)
Quintile 3	38	0.96 (0.59-1.54)	0.92 (0.57-1.47)
Quintile 4	31	0.72 (0.42-1.23)	0.69 (0.41-1.16)
Quintile 5	33	0.95 (0.57-1.60)	0.89 (0.53-1.49)
***P*-Trend**		0.820	0.691
	N (Cases)	**Relative risk (95% CI), basic model***	**Relative risk (95% CI), basic model* – without adjustment for prevalent diabetes**
Factor 2†			
Quintile 1	20	1.00	1.00
Quintile 2	24	1.05 (0.56-1.98)	1.02 (0.54-1.92)
Quintile 3	40	1.51 (0.84-2.73)	1.53 (0.86-2.75)
Quintile 4	42	1.56 (0.86-2.82)	1.57 (0.87-2.82)
Quintile 5	52	1.76 (0.98-3.19)	1.87 (1.05-3.36)
***P*-Trend**		0.024	0.010

*Adjusted for age, sex, alcohol intake from beverages (non-consumers; women: >0-6 g/d, 6-12 g/d, >12 g/d; men: >0-12 g/d, 12-24 g/d, >24 g/d), smoking (never, former, current ≤20 cigarettes/d, current >20 cigarettes/d), physical activity (cycling and sports in h/week), education (low, medium, high), fasting status, BMI (kg/m²), waist circumference (cm), prevalent hypertension, and prevalent diabetes.

†Factor 1= 0.80 x PC ae C32:1 + 0.78 x PC ae C32:2 + 0.70 x PC ae C34:2 + 0.72 x PC ae C34:3 + 0.71 x PC ae C36:2 + 0.71 x PC ae C36:3 + 0.85 x PC ae C40:5 + 0.76 x PC ae C40:6 + 0.82 x PC ae C42:3 + 0.85 x PC ae C42:4 + 0.87 x PC ae C42:5 + 0.76 x PC ae C44:4 + 0.78 x PC ae C44:5 + 0.83 x PC ae C44:6 + 0.82 x PC aa C42:0 + 0.79 x PC aa C42:1 + 0.54 x SM C16:1 + 0.57 x SM OH C22:2 + 0.41 x lysoPC a C17:0. Factor 2= 0.55 x propionyl-carnitine + 0.66 x phenylalanine + 0.61 x tryptophan + 0.66 x tyrosine + 0.68 x valine + 0.66 x isoleucine + 0.59 x PC aa C32:1 + 0.70 x PC aa C36:1 + 0.65 x PC aa C36:3 + 0.76 x PC aa C38:3 + 0.72 x PC aa C40:4 + 0.71 x PC aa C40:5 + 0.44 x hexose. Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; PC, phosphatidylcholine; SM, sphingomyelin.

Table 16. Relative risk of myocardial infarction by quintiles of metabolite factors linked to type 2 diabetes			
	N (Cases)	**Relative risk (95% CI), basic model***	**Relative risk (95% CI), basic model* – without adjustment for prevalent diabetes**
Factor 1†			
Quintile 1	49	1.00	1.00
Quintile 2	51	1.39 (0.89-2.16)	1.40 (0.90-2.19)
Quintile 3	48	1.52 (0.96-2.40)	1.52 (0.96-2.40)
Quintile 4	44	1.17 (0.73-1.88)	1.18 (0.73-1.89)
Quintile 5	43	1.84 (1.12-3.02)	1.87 (1.14-3.07)
***P*-Trend**		0.048	0.043
	N (Cases)	**Relative risk (95% CI), basic model***	**Relative risk (95% CI), basic model* – without adjustment for prevalent diabetes**
Factor 2†			
Quintile 1	25	1.00	1.00
Quintile 2	31	1.06 (0.60-1.87)	1.06 (0.60-1.88)
Quintile 3	50	1.26 (0.73-2.16)	1.26 (0.74-2.17)
Quintile 4	52	1.44 (0.85-2.44)	1.43 (0.84-2.44)
Quintile 5	77	2.22 (1.33-3.72)	2.20 (1.31-3.69)
***P*-Trend**		<0.001	<0.001

*Adjusted for age, sex, alcohol intake from beverages (non-consumers; women: >0-6 g/d, 6-12 g/d, >12 g/d; men: >0-12 g/d, 12-24 g/d, >24 g/d), smoking (never, former, current ≤20 cigarettes/d, current >20 cigarettes/d), physical activity (cycling and sports in h/week), education (low, medium, high), fasting status, BMI (kg/m²), waist circumference (cm), prevalent hypertension, and prevalent diabetes..

†Factor 1= 0.80 x PC ae C32:1 + 0.78 x PC ae C32:2 + 0.70 x PC ae C34:2 + 0.72 x PC ae C34:3 + 0.71 x PC ae C36:2 + 0.71 x PC ae C36:3 + 0.85 x PC ae C40:5 + 0.76 x PC ae C40:6 + 0.82 x PC ae C42:3 + 0.85 x PC ae C42:4 + 0.87 x PC ae C42:5 + 0.76 x PC ae C44:4 + 0.78 x PC ae C44:5 + 0.83 x PC ae C44:6 + 0.82 x PC aa C42:0 + 0.79 x PC aa C42:1 + 0.54 x SM C16:1 + 0.57 x SM OH C22:2 + 0.41 x lysoPC a C17:0. Factor 2= 0.55 x propionylcarnitine + 0.66 x phenylalanine + 0.61 x tryptophan + 0.66 x tyrosine + 0.68 x valine + 0.66 x isoleucine + 0.59 x PC aa C32:1 + 0.70 x PC aa C36:1 + 0.65 x PC aa C36:3 + 0.76 x PC aa C38:3 + 0.72 x PC aa C40:4 + 0.71 x PC aa C40:5 + 0.44 x hexose. Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; PC, phosphatidylcholine; SM, sphingomyelin.

Identification of candidates associated with risk of cardiovascular diseases

In the next step, individual metabolites were identified that are specifically linked to risk of CVD. Of all 127 metabolites analyzed, three metabolites were associated with risk of stroke (diacyl-phosphatidylcholines C 42:0, C36:3 and acyl-alkyl-phosphatidylcholine C 44:6) when adjusting for lifestyle factors, anthropometry, prevalent hypertension and type 2 diabetes. However, the association was non-significant after correction for multiple testing (**Table S5**). Therefore, the endpoint stroke was not considered for further analysis. Of all the metabolites, 42 were associated with risk of myocardial infarction independent of lifestyle factors, anthropometry, prevalent hypertension and type 2 diabetes (**Table S6**). Of them, 11 remained significant after correction for multiple testing (**Table 17**). These metabolites included sphingomyelins C24:0, C18:0 and C16:0, hydroxy-sphingomyelins C22:1, C24:1 and C14:1, diacyl-phosphatidylcholines C38:3, C40:4 and C40:5 and acyl-alkyl-phosphatidylcholines C38:3 and C36:1. Noteworthy, higher concentrations of all of these metabolites were associated with an increased risk of myocardial infarction. In stepwise regression, two metabolites, namely hydroxy-sphingomyelin C22:1 and diacyl-phosphatidylcholine C40:4, were selected to be associated with risk of myocardial infarction independently of the other metabolites.

The impact of reliability of metabolites on RR estimates was also evaluated for the endpoint CVD. ICCs of metabolites associated with risk of myocardial infarction ranged from 0.52 to 0.85. For two metabolites, namely hydroxy-sphingomyelin C14:1 and diacyl-phosphatidylcholine C38:3, there was a considerably different ranking of their expected and observed risk estimates (difference of 6 and 5 ranks, respectively, see **Table S7**). Statistical power enabled detection of RR per SD of about 1.25 for myocardial infarction considering multiple testing corrections (**Table 17**). Assuming that all metabolite measurements were conducted with acceptable reliability (ICC of 0.4) (19; 20), in the present study statistical power could be increased up to a detectable RR per SD of 1.09, if reliability was improved.

Table 17. Mean serum concentrations of identified metabolites and their association with risk of myocardial infarction and possible impact of reliability of metabolites in the EPIC-Potsdam case-cohort study

	Subcohort n=2309 µmol/L*	Cases n=235 µmol/L*	Relative risk per SD (95% CI), basic model†	ICC	Expected relative risk per SD‡
SM (OH) C22:1	14.7 (3.98)	16.0 (4.58)	1.45 (1.24-1.70)	0.70	1.70
SM C24:0	26.0 (6.78)	29.0 (7.85)	1.41 (1.22-1.64)	0.57	1.83
SM (OH) C24:1	1.59 (0.49)	1.75 (0.54)	1.39 (1.21-1.61)	0.63	1.69
PC aa C38:3	55.5 (14.7)	61.3 (16.1)	1.37 (1.18-1.59)	0.52	1.83
SM C18:0	25.2 (6.24)	27.6 (7.55)	1.36 (1.17-1.59)	0.69	1.56
PC ae C38:3	4.76 (1.14)	4.93 (1.23)	1.43 (1.20-1.70)	0.63	1.76
PC aa C40:4	4.00 (1.12)	4.37 (1.22)	1.30 (1.14-1.48)	0.57	1.58
SM (OH) C14:1	7.44 (2.12)	7.95 (2.52)	1.40 (1.18-1.65)	0.85	1.49
SM C16:0	115 (24.0)	124 (30.1)	1.35 (1.15-1.59)	0.66	1.58
PC aa C40:5	11.4 (3.44)	12.6 (3.57)	1.25 (1.11-1.41)	0.62	1.43
PC ae C36:1	9.47 (2.33)	9.93 (2.58)	1.38 (1.16-1.64)	0.65	1.64

*Presented are metabolite concentrations as mean (standard deviation).
†Adjusted for age, sex, alcohol intake from beverages (non-consumers; women: >0-6 g/d, 6-12 g/d, >12 g/d; men: >0-12 g/d, 12-24 g/d, >24 g/d), smoking (never, former, current ≤20 cigarettes/d, current >20 cigarettes/d), physical activity (cycling and sports in h/week), education (low, medium, high), fasting status, BMI (kg/m²), waist circumference (cm), prevalent hypertension, and prevalent diabetes.
‡Expected relative risk was additionally calculated to account for reliability of metabolites, based on the observed relative risk and ICCs for the individual metabolites (15).
Abbreviations: aa, diacyl; ae, acyl-alkyl; ICC, intraclass-correlation coefficient; PC, phosphatidylcholine; SM, sphingomyelin.

Evaluation of independency of risk associations for cardiovascular diseases

The analysis of metabolites and risk of myocardial infarction was additionally adjusted for established biomarkers of CVD, namely CRP, HDL- and total cholesterol. The results are presented in **Table 18**. The association between serum metabolites and risk of myocardial infarction was marginally affected by adjustment for established biomarkers and remained significant for all of the metabolites and metabolite factors. Adjustment for total cholesterol slightly weakened the associations of all of the metabolites and attenuated the association for metabolite factor 1; whereas adjustment for HDL-cholesterol strengthened the association for all of the metabolites and metabolite factors. CRP-adjustment had no impact on the association between metabolites and risk of myocardial infarction.

Table 18. Metabolites and metabolite factors, and their association with risk of myocardial infarction in the EPIC-Potsdam case-cohort sample with additional adjustment for established biomarkers

	Relative Risk per SD (95% CI)				
	Basic model*	**+ CRP†**	**+ HDL-Chol†**	**+ Total-Chol†**	**+ CRP + HDL-Chol + Total-Chol†**
SM (OH) C22:1	1.45 (1.24-1.70)	1.46 (1.24-1.70)	1.53 (1.30-1.79)	1.40 (1.18-1.67)	1.40 (1.19-1.66)
SM C24:0	1.41 (1.22-1.64)	1.41 (1.22-1.64)	1.49 (1.28-1.72)	1.37 (1.16-1.61)	1.37 (1.17-1.6)
SM (OH) C24:1	1.39 (1.21-1.61)	1.39 (1.21-1.61)	1.45 (1.25-1.67)	1.34 (1.15-1.56)	1.34 (1.15-1.55)
PC aa C38:3	1.37 (1.18-1.59)	1.37 (1.18-1.59)	1.40 (1.21-1.62)	1.30 (1.11-1.54)	1.26 (1.08-1.47)
SM C18:0	1.36 (1.17-1.59)	1.36 (1.17-1.58)	1.43 (1.23-1.67)	1.31 (1.11-1.54)	1.31 (1.11-1.54)
PC ae C38:3	1.43 (1.20-1.70)	1.44 (1.21-1.72)	1.52 (1.27-1.82)	1.37 (1.14-1.64)	1.42 (1.18-1.71)
PC aa C40:4	1.30 (1.14-1.48)	1.30 (1.14-1.48)	1.33 (1.17-1.51)	1.25 (1.09-1.43)	1.24 (1.09-1.41)
SM (OH) C14:1	1.40 (1.18-1.65)	1.40 (1.19-1.65)	1.49 (1.25-1.77)	1.34 (1.13-1.59)	1.39 (1.17-1.65)
SM C16:0	1.35 (1.15-1.59)	1.35 (1.15-1.59)	1.47 (1.24-1.74)	1.29 (1.08-1.55)	1.34 (1.12-1.61)
PC aa C40:5	1.25 (1.11-1.41)	1.26 (1.11-1.42)	1.27 (1.13-1.44)	1.20 (1.06-1.36)	1.18 (1.04-1.34)
PC ae C36:1	1.38 (1.16-1.64)	1.39 (1.17-1.66)	1.49 (1.24-1.79)	1.31 (1.09-1.58)	1.38 (1.14-1.66)
Factor 1	1.18 (0.99-1.41)	1.19 (1.00-1.42)	1.37 (1.13-1.68)	1.10 (0.92-1.33)	1.32 (1.07-1.62)
Factor 2	1.25 (1.07-1.45)	1.25 (1.07-1.46)	1.27 (1.09-1.48)	1.19 (1.01-1.40)	1.17 (1.00-1.37)

*Adjusted for age, sex, alcohol intake from beverages (non-consumers; women: >0-6 g/d, 6-12 g/d, >12 g/d; men: >0-12 g/d, 12-24 g/d, >24 g/d), smoking (never, former, current ≤20 cigarettes/d, current >20 cigarettes/d), physical activity (cycling and sports in h/week), education (low, medium, high), fasting status, BMI (kg/m²), and waist circumference (cm), prevalent hypertension, and prevalent diabetes.

†Adjustment according to the basic model with additional adjustment for specified biomarkers.

Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; Chol, cholesterol; HDL, high-density lipoprotein; PC, phosphatidylcholine; SM, sphingomyelin.

Usefulness for prediction of cardiovascular diseases

The potential to discriminate between cases and non-cases of myocardial infarction was lower for metabolites alone compared to the basic model which included known risk factors such as age, sex, lifestyle factors, anthropometry and medical condition (area under the ROC curve: 0.657 and 0.818; respectively), see **Figure 15**. When adding the metabolites to established risk factors, the area under the ROC curve could slightly be improved up to 0.837. As a risk score to predict CVD has not yet been developed in EPIC-Potsdam, the area under the ROC curve of individual predictors included in the basic model and the established biomarkers are presented in more detail in **Table S8**. All of the models to predict myocardial infarction with and without metabolites, were well calibrated (**Table 19**).

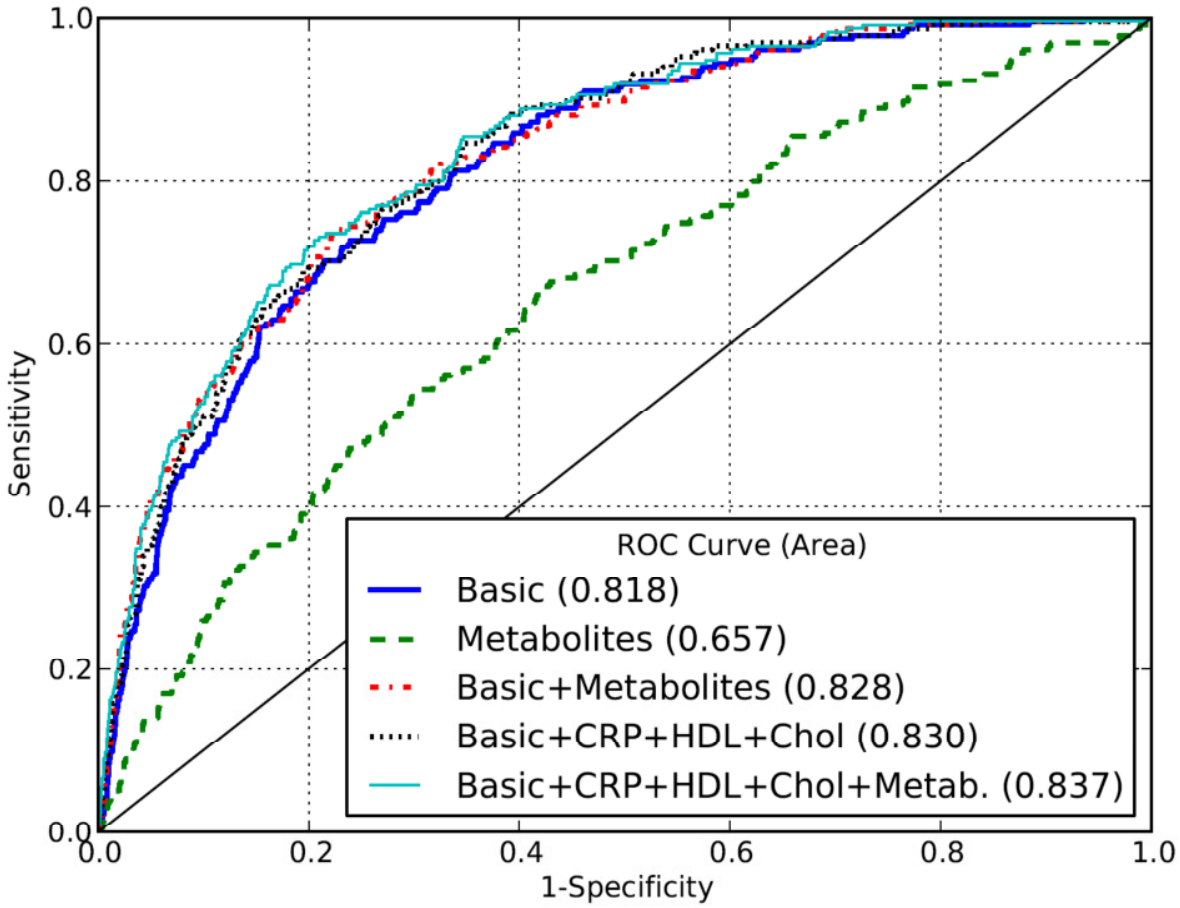


Figure 15. Relative contribution of metabolites to predict myocardial infarction in the EPIC-Potsdam case-cohort sample. Presented are receivers operating characteristic (ROC) curves comparing different multivariable adjusted models to predict myocardial infarction, including a basic model (age, sex, alcohol intake from beverages, smoking, physical activity, education, fasting status, prevalent hypertension, prevalent diabetes, BMI, and waist circumference), the identified metabolites (SM (OH) C22:1, SM C24:0, SM (OH) C24:1, PC aa C38:3, PC ae C38:3, SM C18:0, PC aa C40:4, SM (OH) C14:1, SM C16:0, PC ae C36:1, PC aa C40:5) and established biomarkers (CRP, HDL-Chol, Chol). Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; CRP, C-reactive protein; Chol, cholesterol; HDL, high density lipoprotein; PC, phosphatidylcholine; SM, sphingomyelin.

Table 19. Measures* of discrimination and calibration comparing multivariable adjusted models to predict myocardial infarction in the EPIC-Potsdam case-cohort sample

	-2LOG-Likelihood	Nagelkerke R^2†	Hosmer-Lemeshow‡		Receiver Operating Characteristic (ROC)§			
Model			χ^2	***P*-Value**	**Area under the curve (95% CI)**	***P* for Difference**		
Basic¶	1266	0.24	9.30	0.318	0.818 (0.792-0.845)	Ref.		
Metabolites‖	1501	0.05	4.63	0.796	0.657 (0.620-0.694)	<0.0001		
Basic + CRP	1264	0.24	4.07	0.851	0.819 (0.792-0.845)	0.516	Ref.	
Basic + Metabolites‖	1232	0.27	3.41	0.906	0.828 (0.801-0.855)	0.073	0.096	
Basic + CRP + HDL + Chol	1233	0.27	3.67	0.886	0.830 (0.804-0.856)	0.023	0.026	Ref.
Basic + CRP + HDL + Chol + Metabolites‖	1207	0.29	6.48	0.594	0.837 (0.811-0.863)	0.006	0.007	0.084

*Better model performance is indicated by smaller -2LOG-Likelihood Ratio, larger Nagelkerke's R^2, good calibration with a Homer-Lemeshow *P*-Value ≥0.05, and better discrimination mirrored by larger area under the ROC curve with significant improvement of model performance, *P* for difference <0.05.

†Specifically, Nagelkerke's R^2 defines the fraction of -2 log likelihood explained by the predictors ranging from 0 to 1. Larger values indicate better model fit (105).

‡As a measure of model calibration, the Hosmer-Lemeshow statistic compares predicted and observed probabilities of type 2 diabetes derived from deciles of predicted risk. Smaller $\chi 2$ values and larger *p*-values specify better model fit. *P*-values <0.05 indicate significant difference between expected and observed probabilities (107).

§The ROC curve is a plot of sensitivity versus false-positive rate of a continuous prediction model on a binary outcome. The area under the ROC curve, a measure of discrimination that mirrors the probability the model assigns a higher risk to future cases compared to controls. It may range from 0.5 (no discrimination) to 1.0 (perfect discrimination) (110). The method of DeLong (106) was used to test for difference between the areas under the ROC curve, $p<0.05$ indicate significant model improvement.

¶Basic model includes age, sex, alcohol intake from beverages (non-consumers; women: >0-6 g/d, 6-12 g/d, >12 g/d; men: >0-12 g/d, 12-24 g/d, >24 g/d), smoking (never, former, current ≤20 cigarettes/d, current >20 cigarettes/d), physical activity (cycling and sports in h/week), education (low, medium, high), fasting status, prevalent hypertension, prevalent diabetes, BMI (kg/m²), and waist circumference (cm).

‖Metabolites linked to risk of myocardial infarction: SM (OH) C22:1, SM C24:0, SM (OH) C24:1, PC aa C38:3, PC ae C38:3, SM C18:0, PC aa C40:4, SM (OH) C14:1, SM C16:0, PC ae C36:1, PC aa C40:5.

Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; CI, confidence interval; DRS, Diabetes Risk Score; PC, phosphatidylcholine; ROC, receiver operating characteristic; SM, sphingomyelin.

3.4.3 Sensitivity analyses

To ensure the robustness of the results different sensitivity analyses were conducted. The results are presented in the following paragraphs.

Fasting status

In EPIC-Potsdam, a small proportion had fasted prior to the blood sample collection. The proportion of cases of type 2 diabetes was equally distributed across participants who had fasted and those who had not (26.8% and 26.7%; respectively) and accordingly, adjustment for fasting status did not change the results. For the endpoint CVD, incident cases were not equally distributed across fasting and non-fasting participants (12.1% and 10.1%; respectively); therefore, the main analysis was adjusted for fasting status. However, the results were very similar without adjustment for fasting status. In addition, the PCA was re-conducted when restricting the analysis to fasting samples only (n=429). Thereby, very similar metabolite factors as in the main analysis could be reproduced (**Table S9**).

A possible effect modification of fasting status on the association between metabolite factors, individual metabolites and risk of type 2 diabetes and myocardial infarction was also evaluated by including multiplicative interaction terms into the models. No significant interaction of fasting status was observed on the association between metabolite factors and risk of type 2 diabetes and CVD. There was a significant interaction of fasting status on the association between hexose and risk of type 2 diabetes, indicating that in people who were fasted the positive association was 3-fold stronger compared to their counterparts. The association between acyl-alkyl-phosphatidylcholine C42:5 and risk of type 2 diabetes was only seen in fasted individuals and was non-significant for their counterparts. A significant interaction was found for fasting status on the association between hydroxy-sphingomyelin C 14:1 and sphingomyelin C16:0 and risk of myocardial infarction (*p* for interaction: 0.004 and 0.01; respectively). When looking at fasting and non-fasting participants separately, the strength of association was approximately 2-folds higher in individuals who were fasted compared to non-fasted individuals.

Interaction with sex

A possible effect modification of sex on the association between metabolite factors, individual metabolites and risk of type 2 diabetes and myocardial infarction was also evaluated. No significant interactions of sex were observed on the association between metabolite factors and risk of type 2 diabetes and CVD. There was a significant interaction of sex on the association between hexose and risk of type 2

diabetes, indicating that in men the positive association was much stronger compared to women. The positive association between diacyl-phosphatidylcholines C38:3 and C40:5 and risk of myocardial infarction was only found in men and was non-significant in women.

Reverse causation

To ensure that the changes in metabolites preceded the onset of type 2 diabetes and CVD and were not caused by undiagnosed or pre-disease conditions, the main analysis was repeated when excluding all cases of type 2 diabetes (n=208) and CVD (n=54) that occurred shortly after the baseline examination (two years into follow-up). The associations for metabolite factors, individual metabolites and risk of type 2 diabetes were not markedly different and all remained significant. For diacyl-phosphatidylcholine C38:3 the association was slightly stronger (RR per SD: 1.39, 95% CI: 1.25-1.55). The associations for metabolite factors and risk of myocardial infarction were non-significant (RR per SD, factor 1: 1.18, 95% CI: 0.97-1.44; factor 2: RR: 1.14, 95% CI: 0.96-1.36; respectively). For individual metabolites linked to risk of myocardial infarction, the associations were slightly weakened but all remained significant.

4. Discussion

4.1 Discussion of methods

4.1.1 Study design

The overall strength and challenge of this study was, that in the frame of a large prospective cohort study a comparably high number of metabolites were investigated, representing a very large dataset in this field. Additionally, a systematic and multistep approach was taken to address the different challenges of metabolomics in an epidemiologic setting. It was started with investigation of technical variation, biological variation and reliability, before utility of metabolites was investigated in terms of their association with risk of type 2 diabetes and CVD and their usefulness for risk prediction.

The present thesis was conducted in the frame of a prospective cohort study. A general advantage of prospective cohort studies is that several exposures and outcomes can be studied at the same time (111), which was also done in this thesis. The prospective design of this thesis enabled to show temporality of exposure and outcome; reverse causation was additionally evaluated in a sensitivity analysis. That is among the reasons why this design is superior to previous case-control studies with measurement of metabolites in prevalent cases (62). The main analysis was conducted in the frame of a nested case-cohort design. Thereby, the results are expected to be generalizable to the full cohort, whereas the metabolite measurements only need to be done in a randomly drawn subcohort and all incident cases; which is convenient in terms of cost particularly when considering comprehensive and expensive metabolomics measurements (90). This design additionally offered some advantages over nested case-control studies (62). That is a single control group can be used to study multiple disease endpoints, which was also done in the present thesis, which again offers cost benefits. Furthermore, the chance of selection bias for the control group is reduced as each participant has the same chance to be selected as a control regardless of the case status or follow-up time and matching is not necessary (62). However, in the case-cohort design compared to the full cohort it is difficult to estimate the absolute risk as the incident cases are oversampled. Therefore, in the present thesis the measures of discrimination for risk prediction should only be seen as relative constructs compared to other prediction models and not be interpreted in absolute numbers (8).

4.1.2 Exposure assessment

The present study is based on data retrieved in the frame of a large prospective cohort study. Therefore, compromises had to be taken in biological specimen collection because of logistic reasons and cost constrains. Blood samples of the full cohort were only collected at baseline, and thus, the exposure of this thesis (serum metabolites) was only measured once (84). However, in the frame of this thesis, reliability of the metabolites was also investigated in a reliability study, which was based on two blood sample collections 4 month apart in time. This investigation suggested that the majority of metabolites showed acceptable reliability. Long-term storage and blood processing may as well have affected concentrations of serum metabolites (16); unfortunately, data was not available to investigate this issue. Only participants of the morning appointments had provided fasting blood samples and the majority of participants were non-fasted. In the main analysis, data of fasted and non-fasted participants were combined. However, a sensitivity analysis suggested that fasting status did not modify the association between metabolites and risk of type 2 diabetes and CVD; an exception being hexose and three phospholipids, where the association was much stronger for fasted individuals. However, the association also had the same direction for non-fasted individuals. In this context, the results of the interaction tests of the individual metabolites have to be interpreted with caution as multiple statistical tests were conducted, which increased the cumulative chance of false positive findings. Measurement of serum metabolites were done with a commercially available kit and technical variation, which was evaluated in the frame of this thesis, was quite high for some metabolites particularly those with lower serum concentrations. These metabolites were, however, excluded for the main analysis of this thesis.

4.1.3 Outcome assessment

The primary outcomes of this thesis were incident type 2 diabetes, myocardial infarction and stroke. The outcomes were identified by self-reports; nevertheless, they were also verified by contacting the treating physician and retrieving details about the diagnosis. For the endpoint type 2 diabetes, it may be particularly difficult to assess the exact onset of the disease, as development of type 2 diabetes is a rather continuous process (112) and many people may have undiagnosed but overt diabetic conditions. In fact, in this thesis it was observed that incident cases of type 2 diabetes also had higher levels of serum hexose at baseline. This may as well be caused by undiagnosed type 2 diabetes cases. To address this issue the main analysis was repeated when excluding all cases that occurred shortly after the baseline examination. In fact, the associations were not markedly different, suggesting that this problem may not have majorly affected the results.

4.1.4 Methodological and statistical approaches

Technical and biological variation and reliability of metabolites

In the present thesis, it was aimed to study technical variation and biological variation. However, the biological variation that was measured over a 4-month period cannot exclusively be assessed, as it also depends on the measurement error of the laboratory method and always includes the technical variation (16). The ICC was calculated as a measure of reliability in the present thesis. An advantage of the ICC is that it accounts for both variance components: between- and within-person variation and their relation (19; 20). In addition, the single variance components were also reported for each metabolite in the present thesis, to better understand the origin of variation (16).

RR risk estimates of metabolites significantly associated with risk of type 2 diabetes and CVD were projected towards expected RR; the values hypothesized an ICC of one. However, the expected RR has to be interpreted with caution, because there are many assumptions made: it assumes equal variation across the exposure levels and only random error; systematic error, however, is not corrected (15; 54). In addition, it only accounts for measurement error of the exposure, but assumes no measurement error of the outcome or covariates. This assumption cannot be true (113). Furthermore, the case of different directions of expected and observed RR is not covered by the correction method. As fulfillment of all of these assumptions cannot be met, the expected RR was suggested to be used as an indicator of the degree of bias in the risk estimates rather than the estimate of "true association" (54).

Identification of metabolites associated with risk of type 2 diabetes and cardiovascular diseases

In general, identification of metabolites was accomplished with a rather conservative approach, in terms of model adjustment and multiple testing corrections. The basic adjustment model for selection of metabolites was multivariable adjusted for age, sex, dietary and lifestyle factors, anthropometry and medical conditions. Thus, only metabolites that were significant after multivariable adjustment were selected. It may be assumed that the number of metabolites related to risk of type 2 diabetes and CVD is different when calculating crude or less-adjusted models. However, only the full model was considered as calculation of different models would further increase the number of statistical tests, which is critical.

To account for the exploratory design of the study with multiple metabolites as exposure multiple testing correction was done according to Bonferroni-Holm method (103). This method represents a rather strict and conservative approach similar to the classic Bonferroni correction, which is frequently used in

exploratory studies, and has been previously applied in metabolomics studies (24; 114). However, it was suggested to be preferred over the classic Bonferroni method, as it achieves a statistical power gain with similar chance of detecting false positives (103). Less conservative methods, e.g. correction based on the false discovery rate (115), have also been used in previous metabolomics studies (67), e.g. if they focused on metabolites of specific pathways. In the present thesis, though, the metabolites were not a-priori selected by the researchers; that is why it was corrected conservatively.

Dependency of metabolites

The main analysis was done for each metabolite separately as in classic epidemiologic studies with single risk factors. Dependency of metabolites was accounted for by stepwise models and conduction of a PCA. However, these methods may not necessarily best account for the multicollinearity of metabolites. In long-term, more sophisticated statistical methods, e.g. classification methods from systems biology, should be adopted to epidemiologic study design to approach these highly correlated data.

Risk prediction

To assess the ability of metabolites to discriminate between cases and non-cases, the area under the ROC curve was calculated in models with and without metabolites. Thereby, an independent risk factor which is uncorrelated to other risk factors will yield greater improvement in discrimination compared to a non-independent risk factors (116). The ROC curve has been discussed previously, as on the one hand, it was shown that it may overestimate the effect of weak predictors (117). However, on the other hand, in risk prediction models that already contain strong predictors only little improvement can be achieved when adding another strong predictor (116-118). Furthermore, the area under the ROC curve depends on the study population, e.g. it is easier to discriminate people in a more heterogeneous population (117). This has to be kept in mind when interpreting the results. The DeLong-test was used to compare areas under the ROC curve, although it has recently been suggested that it may not be the best choice to use in nested models (119). Last, calibration was evaluated by Hosmer-Lemeshow test (107). A limitation of the Hosmer-Lemeshow test is that the result is depending on the sample size. With large sample size it will be significant when there are only small differences between expected and observed probabilities. Any significant difference, however, implies poor calibration. Reclassification, e.g. by net reclassification index, is another measure to evaluate performance of risk prediction (118). Unfortunately, it was not analyzed in the present thesis as it requires estimation of the absolute risk. However, methods for assessment of reclassification indices in case-cohort studies are currently lacking.

4.2 Discussion of results

4.2.1 Technical and biological variation and reliability of metabolites[17]

The present thesis comprehensively investigated different variance components of 163 serum metabolites, including technical variation (between- and within-assay variation), biological variation (between- and within-person variation) and reliability. In addition, the RR estimates of metabolites significantly associated with risk of type 2 diabetes and CVD were corrected for the ICC. In this thesis the technical variation was comparable to that in other studies with measurement of multiple biomarkers (15; 19). Few metabolites showed technical variation above 50% particularly those with low serum concentrations in this healthy population, e.g. hydroxy-acylcarnitines. These metabolites were excluded from the present analysis as the validity of measurements is questionable. It was also observed that median between-assay variation was higher compared to within-assay variation. Ideally, the between-assay variation should not substantially add to the within-assay variation (16); however, it has previously been reported that it may even account for much larger proportions of the variation (15; 120). In the present study it may be due to the extended time period that was necessary to measure the large number of samples. This may be a general problem when conducting metabolomic measurements in epidemiologic studies. However, the between-assay variation can be considered in epidemiologic study design, e.g. by allocation of the samples before the metabolomic measurements (16). Studies that investigated biological variation and reliability of metabolites over time are rarely found in the literature (19; 121; 122). These studies looked at different biological specimen, e.g. plasma or cerebral spinal fluid, used untargeted metabolomics, or investigated different metabolites. Most previous studies focused on technical variation, improvement of laboratory methods and conditions of sample handling and storage, but did not study biological variation (123-125).

In the present thesis amino acids showed good reliability and the ICCs were of similar magnitude as in a previous study, were plasma amino acids were determined in different nutritional states (126). Metabolism of amino acids is tightly regulated and underlies homeostatic control mechanisms (126). Thus, in healthy individuals the within-person variation should be small compared to between-person variation, which was also observed in this thesis. Low within-person variation and excellent reliability was also observed for hexose. Hexose represents the sum of all monosaccharides with 6 carbon atoms. Thereby, it consists of glucose (68) but also fructose and galactose among other C6-sugars. Hormonal

[17] In the framework of this thesis, parts of the discussion on biological variation and reliability of metabolites have already been published by the author in similar text: 108. Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, Joost HG, Boeing H, Pischon T: Reliability of Serum Metabolite Concentrations over a 4-Month Period Using a Targeted Metabolomic Approach. PLoS One 2011;6:e21103

control mechanisms immediately respond to different nutritional states to ensure that blood sugar concentrations are retained within a narrow range over time. These results suggest that amino acids and hexose are reliable over time, and thus, may be suitable for a single measurement in epidemiologic studies. Acylcarnitines are esterified fatty acid derivates of carnitine, which are involved in fatty acid translocation into the inner mitochondrial membrane. This represents the limiting step for fatty acid degradation by β-oxidation. Acylcarnitines can further pass into the cytosol and subsequently into the circulation. Blood concentrations of acylcarnitines may reflect substrate flux through β-oxidation; however, it has also been suggested that higher blood concentrations of acylcarnitines may point towards altered fatty acid β-oxidation and impaired mitochondrial function (68). It was previously reported that blood acylcarnitine concentrations were higher in people with obesity and type 2 diabetes (63; 127). Nevertheless, information on reliability of acylcarnitines is rare. The present thesis found that acylcarnitines showed low serum concentrations in healthy individuals and that short and medium chain acylcarnitines were more reliable than longer chain hydroxy- and monounsaturated acylcarnitines. However, many of the acylcarnitines, that showed high within-person variation and poor reliability, were also below their limit of detection and showed high technical variation; therefore, they were excluded for the main analysis. This fact may demonstrate how technical variation is linked to biological variation and reliability. Phosphatidylcholines belong to a class of membrane phospholipids; however, in the circulation they are most abundant in lipoproteins, although they may also be derived from redistribution of plasma membranes (128; 129). They consist of a glycerol chore with a choline head group and two fatty acid residues (diacyl-phosphatidylcholines). Lysophosphatidylcholines originate from hydrolysis of one fatty acid, usually, in the sn2 position, and trans-esterification by phospholipase A2. Acyl-alkyl-phosphatidylcholines represent the ether derivates of diacyl-phosphatidylcholines, where one fatty acid is linked by a vinyl-ether bond (130). Sphingomyelins contain a ceramide core instead of glycerol linked to a fatty acid and a polar head group. Their function remains relatively unknown until today which gave rise to their name (riddle of the “sphinx” from greek mythology). They may be involved in cellular signal transduction, e.g. the nuclear factor-κB-pathway, and may play a role in the pathophysiology of CVD (131). In a previous study, reliability of total phospholipid measures in blood was investigated over a three week period in 12 subjects (132). The authors reported slightly lower ICCs as observed in the present thesis, e.g. an ICC of 0.50 for total phosphatidylcholines and an ICC of 0.54 for total sphingomyelins as compared to a median ICC of 0.58 for phosphatidylcholines and 0.66 for sphingomyelins, respectively, in the present thesis. The results suggest the usefulness of these phospholipids as reliable biomarker candidates in epidemiologic studies.

In the present thesis, the RR estimates of metabolites significantly associated with risk of type 2 diabetes and CVD were corrected for reliability of individual metabolites. This correction always increased the effect size, which is to be expected with the correction method used (15). This was particularly relevant for metabolites with lower reliability, e.g. phenylalanine, or with a stronger association, e.g. hexose and type 2 diabetes risk. These results suggest that the "true effect" may be stronger when reliability of metabolite measurements is improved, which would consequently also increase statistical power as demonstrated in the present thesis. Thereby, it has to be considered that the chance of false negative findings is increased for metabolites with lower reliability as the "true effect" could be considerably attenuated and underestimated (19; 20). Ranking of the RR estimates before and after reliability correction did not change for most of the metabolites; however, it changed considerably for few metabolites. Thus, when many metabolites with different reliability are analyzed together the results in epidemiologic studies could potentially be biased (113). In terms of practical relevance metabolites with poor reliability would not be useful as biomarkers to be measured in epidemiologic or clinical settings. Nevertheless, the findings of the present thesis highlight the necessity of improvement of reliability of the metabolite measurements and to reduce technical variation to obtain valid risk estimates in epidemiologic studies (54). However, to account for true temporal variation of metabolites within one individual (within-person variation), which is also included in the construct of reliability, multiple biological sample collections and metabolite measurements would be required, which are not feasible to conduct in large-scale epidemiologic studies.

4.2.2 Serum metabolites and risk of type 2 diabetes and cardiovascular diseases[18]

In the present thesis using a targeted metabolomics approach in the prospective EPIC-Potsdam study a number of metabolites were associated with higher risk of type 2 diabetes and myocardial infarction, but not with stroke risk. The biological plausibility of these results is discussed in the following paragraphs.

[18] In the framework of this thesis, parts of the discussion (type 2 diabetes) have already been published by the author in similar text: 109. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, Fritsche A, Haring HU, Hrabe de Angelis M, Peters A, Roden M, Prehn C, Wang-Sattler R, Illig T, Schulze MB, Adamski J, Boeing H, Pischon T: Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. Diabetes 2012; Oct.4th (Epub ahead).

Serum metabolites and risk of type 2 diabetes

Metabolites linked to higher risk of type 2 diabetes included higher concentrations of hexose, branched chain- and aromatic amino acids and diacyl-phosphatidylcholines, and lower concentrations of glycine, lyso-phosphatidylcholines, sphingomyelins, and acyl-alkyl-phosphatidylcholines. These results are in line with recent results from case-control and cross-sectional studies, where type 2 diabetes patients showed higher concentrations of sugar metabolites (64; 67; 69; 70), branched chain and aromatic amino acids (65-67), and lower concentrations of glycine (65; 68). However, despite propionyl-carnitine acylcarnitines were not related to risk of type 2 diabetes in the present thesis, which is in contrast to findings from previous studies (63; 68). In this context it needs to be considered that previous studies investigated prevalent cases of type 2 diabetes and may not be directly comparable to the present thesis. In fact, metabolite levels may be very different after the onset of type 2 diabetes and consequent disruption of metabolism. The only prospective study which was previously conducted on this topic found a signature of five branched chain and aromatic amino acids, that predicted type 2 diabetes in the Framingham Offspring study and the Malmö Diet and Cancer study (24). In agreement, the present thesis also found higher concentrations of phenylalanine, isoleucine, tyrosine and valine to be associated with higher risk of type 2 diabetes and glycine with lower risk (which was borderline significant in the Framingham Offspring study). In the present thesis it was further observed that branched chain and aromatic amino acids were linked to each other and only phenylalanine was significantly positively associated with risk of type 2 diabetes once accounting for the other metabolites. Branched chain amino acids are potential substrates for the glucose-alanine cycle in skeletal muscle. They may undergo transamination reactions catalyzed by alanine-aminotransferase, and thereby, increase substrate availability for hepatic gluconeogenesis, and finally, increase hepatic glucose production (133). Reduced serum glycine may as well reflect increased hepatic gluconeogenesis as glycine is a gluconeogenic amino acid itself. Other theories argue that glycine depletion is caused by elevated glutathione consumption as a consequence of increased oxidative stress in type 2 diabetes (134) or high abundance of incompletely oxidized fatty acids, that are excreted as acyl-glycine conjugates by the kidney (135-137). Higher concentrations of branched chain amino acids in people with insulin resistance have also been suggested to result from reduced activities of key catabolic enzymes in amino acid metabolism in liver and adipose tissue (138). In addition, branched chain amino acids may directly cause muscular insulin resistant by disturbing insulin signaling (139).

Increased hexose was associated with higher risk of type 2 diabetes. Elevated fasting glucose is used as a criterion for type 2 diabetes diagnosis (112). In addition, in the prospective Whitehall II study it was shown that plasma glucose was increased already years before diagnosis of type 2 diabetes (30).

However, in the present thesis the association remained significant after glucose adjustment. This observation could be triggered by the different methods used to measure hexose and glucose. Nevertheless, hexose did not only represent glucose but all 6-carbon monosaccharides, e.g. fructose and galactose, which are also well absorbed in the intestine (140). A previous studies observed that fructose levels were increased in people with type 2 diabetes (65). In addition, fructose intake has been linked to higher risk of type 2 diabetes in prospective cohort studies (141; 142).

In the present thesis, choline-containing phospholipids (diacyl-, acyl-alkyl-, and lyso-phosphatidylcholines and sphingomyelins) were linked to risk of type 2 diabetes; unfortunately, these were not investigated in the Framingham Offspring study (24). Thus, this thesis is probably the first published study that investigated their prospective association. Cross-sectional and case-control studies did often not include choline-containing phospholipids; and the few studies reported inconsistent results. In accordance to our study, Wang-Sattler et al. (68) reported decreased lyso-phosphatidylcholine C18:2 in patients with prevalent type 2 diabetes. However, other studies found increased lyso-phosphatidylcholine C18:2 in patients with type 2 diabetes, or no association (66; 81). Diacyl-phosphatidylcholines and sphingomyelins were previously found to be associated with lower or higher risk of type 2 diabetes (67; 81). Acyl-alkyl-phosphatidylcholines have hardly been measured in previous studies with the endpoint type 2 diabetes. However, previous studies have reported that acyl-alkyl-phosphatidylcholines were lower in obese people and individuals with insulin resistance (143; 144). In general, choline-containing phospholipids make up a main constituent of cellular membranes and are involved in signal transduction (145). In the circulation, they are most common as part of lipoproteins (128). Diacyl-phosphatidylcholines are particularly required for hepatic secretion of triglyceride-rich very low density lipoprotein (VLDL)-particles as well as HDL (145). In contrast, acyl-alkyl-phosphatidylcholines may act as blood antioxidants to protect lipoproteins from oxidation, which is possible through the unique structure of the vinyl-ether bond found in the plasmalogen fraction of acyl-alkyl-phosphatidylcholines (143). This may highlight the importance of biochemical structure of metabolites in disease pathophysiology and particularly the type of binding between phospholipid core and fatty acid residue. Synthesis of phosphatidylcholines requires dietary choline and is partly restricted to the liver (145). Choline-deficient mice on a high-fat diet previously showed reduced phosphatidylcholine biosynthesis, and thereby, accumulated hepatic fat but at the same time they also improved glucose tolerance and reduced fasting insulin (146). Furthermore, impaired hepatic phosphatidylcholine biosynthesis reduced levels of plasma triglycerides and HDL-cholesterol in-vivo (147). In this thesis, on the one hand, HDL-adjustment weakened the inverse association between acyl-alkyl-phosphatidylcholines, lysophosphatidylcholines, sphingomyelins and risk of type 2 diabetes. On the other hand, the positive association between diacyl-phosphatidylcholines and

risk of type 2 diabetes was particularly weakened by adjustment for plasma triglycerides. Thus, these phospholipid species seem to be differentially linked to plasma lipoproteins, which may partly account for their opposite association with risk of type 2 diabetes. Furthermore, the present thesis observed that, in general, those phosphatidylcholines containing fatty acids with lower number of carbons and double bonds were positively associated with type 2 diabetes risk, contrary to those with higher number of carbons and double bonds. Similar observations have recently been reported for fatty acid compositions of erythrocyte membrane phospholipids (148) and of triglycerides (149) suggesting that lipids with shorter chain length and saturated fatty acid residues may trigger development of type 2 diabetes while those containing longer chain and unsaturated fatty acids may offer protection. Furthermore, phosphatidylcholines have previously been linked to a genetic variant of fatty acid desaturase 1 (114). This enzyme which is required for long chain omega-6 and omega-3 polyunsaturated fatty acid biosynthesis may as well be involved in pathogenesis of type 2 diabetes (148).

In general, metabolomic studies may highly depend on the dataset, and thus, external validation of the results is important. The findings from the present thesis were replicated in the independent Cooperative Health Research in the Region of Augsburg (KORA)-study, which is a cohort study in Southern Germany with prospective follow-up of participants (150-152). The metabolite factors found in the present thesis could partly be replicated in KORA (109). In particular, metabolite factor 2 was also significantly positively associated with risk of type 2 diabetes in KORA (RR (95% CI) in extreme quintiles, KORA: 4.95 (1.96-12.48), *p*-trend $1.10 \cdot 10^{-5}$; EPIC-Potsdam: 3.82 (2.64-5.52), *p*-trend $6.64 \cdot 10^{-18}$). Metabolite factor 1 was also inversely associated with risk of type 2 diabetes in KORA but the association was non-significant (RR (95% CI) in extreme quintiles, KORA: 0.49 (0.17-1.25), *p*-trend $1.00 \cdot 10^{-1}$; EPIC-Potsdam: 0.31 (0.21-0.44), *p*-trend $2.95 \cdot 10^{-13}$). However, there was no significant heterogeneity between studies and KORA had a considerably lower number of incident type 2 diabetes cases (n=91) compared to EPIC-Potsdam (n=800) (109). In addition, the metabolite factors found in EPIC-Potsdam were investigated in 76 participants of the Tübingen Family study, who participated in an oral glucose tolerance test and had data on these metabolites available (109). Metabolite factor 1, which was inversely associated with type 2 diabetes risk in EPIC-Potsdam, was positively correlated to insulin sensitivity and inversely correlated to insulin secretion in the Tübingen Family study. Metabolite factor 2 positively associated with risk of type 2 diabetes in EPIC-Potsdam was inversely correlated to insulin sensitivity and not correlated to insulin secretion. Thus, the findings in these two independent study populations further underline the validity of observed associations between metabolite factors and risk of type 2 diabetes.

To sum up, the metabolites that were found to be associated with risk of type 2 diabetes in the present thesis could be involved in different pathways of type 2 diabetes and may also add to the current understanding of disease pathophysiology. In addition, risk prediction of type 2 diabetes with the metabolites alone was similar to that of the German Diabetes Risk Score when comparing the areas under the ROC curve. When combining metabolites with established risk prediction models the discrimination could further be improved. Thus, metabolite measurement may also be a useful approach to predict type 2 diabetes risk for individuals in the future. The metabolites offer potential to serve as markers for deranged pathways in type 2 diabetes, and as an outlook, they could allow the implementation of individualized preventive and therapeutic strategies. Nevertheless, future studies are required to study individual risks in detail and to better understand the biological mechanisms and pathways of these metabolites.

Serum metabolites and risk of cardiovascular diseases

In the present thesis, the number of metabolites associated with risk of CVD was considerably smaller than the number of metabolites linked to risk of type 2 diabetes. This finding may be expected as type 2 diabetes represents an endocrinologic disease contrary to CVD. However, metabolite factor 2, which was associated with higher risk of type 2 diabetes, was also significantly positively associated with risk of stroke and myocardial infarction. Adjustment for prevalent type 2 diabetes slightly weakened this association. Similar pathophysiologic mechanisms may be involved in etiology of different chronic diseases and particularly type 2 diabetes and CVD (22). Obesity, and insulin resistance, and consequent hyperglycemia, and dyslipidemia, for example, represent strong risk factors for both CVD and type 2 diabetes (22). Furthermore, CVD is a long-term complication of type 2 diabetes. This theory is also supported by the finding that in the present thesis there was a remarkable overlap of metabolites linked to a higher risk of type 2 diabetes and those metabolites, which were ranked highest for the association with stroke risk. However, none of the individual metabolites were significantly linked to risk of stroke, after multiple testing corrections. This may be due to the fact, that the number of stroke cases (n=178) was considerably smaller than the number of cases with type 2 diabetes (n=800). Thus, it could be a statistical power problem. However, the fact, that a couple of metabolites were linked to higher risk of myocardial infarction and the number of cases of myocardial infarction and stroke were of similar magnitude, suggests that these metabolites are not strongly linked to stroke risk. Therefore, these metabolites may play a minor direct role in pathophysiology of stroke. Two case-control studies have previously investigated metabolite levels of patients with an acute stroke event (74; 75). The untargeted metabolomics study reported a number of endogenous one-carbon-cycle metabolites to be altered in patients with stroke (74) and the targeted metabolomics study found metabolites linked to anaerobic

glycolysis, folic acid deficiency and hyper-homocysteinemia (75). Thus, it is possible, that the present thesis did not find association between metabolites and risk of stroke as inappropriate metabolites were studied. An alternative explanation is that metabolite concentrations could change rapidly in response to an acute stroke event but are not altered years before the onset of stroke. Future prospective investigations and particularly untargeted metabolomics approaches seem necessary to better understand the association between metabolites and risk of incident stroke.

A number of sphingomyelins and phosphatidylcholines were linked to higher risk of myocardial infarction in the present thesis. Three of the diacyl-phosphatidylcholines, namely C38:3, C40:4, and C40:5, were also linked to higher risk of type 2 diabetes, which support the theory of common pathophysiologic processes in development of chronic diseases (22). A previous study reported glucose, amino acids and choline-containing compounds to be altered in patients with myocardial infarction (73). Prospective cohorts of patients with cardiac catheterization or bypass grafting found acylcarnitines and metabolites linked to gastrointestinal microbiota to be associated with risk of myocardial infarction or mortality (78-80). The findings from the present thesis do not directly compare to these previous studies. Unfortunately, no prospective cohort study could be found that investigated the association between metabolites and CVD risk in originally healthy individuals. Future prospective investigations are, therefore, very much required. The present thesis found sphingomyelins and phosphatidylcholines to be associated with risk of myocardial infarction. Sphingomyelins and phosphatidylcholines have previously been suggested to be involved in the pathophysiological process of atherosclerosis, which often precedes the onset of hard CVD events. A proposed mechanism of atherosclerosis (153-156) includes low density-lipoprotein (LDL)-accumulation in the arterial wall, enzyme action of sphingomyelinase and secretory phospholipase A2, which liberate free fatty acids, lyso-phosphatidylcholines and ceramides. The free lipid species may further enhance atherogenic processes at the vessel wall. Additionally, LDL-particles are substantially modified by the enzyme actions. These modified LDL-particles may promote inflammatory processes at the vessel wall and cause monocyte emigration, and consequently, differentiation into macrophages and foam cell formation. Thereby, vessel walls may be ruptured and the plaque is worsened, which in a late stage may lead to thrombosis. In this process, sphingomyelins are particularly linked to atherogenic LDL-particles (156). Previous studies found that LDL-particles in atherogenic plaques were substantially enriched with sphingomyelins compared to plasma LDL particles (157; 158). An in-vivo study reported that oral administration of an inhibitor of sphingomyelin de-novo biosynthesis prevented the development of atherosclerotic lesions in apo-E knockout mice (154). Treatment with statins also resulted in lower plasma sphingomyelin levels among patients with CVD in a randomized trial (153). Individual sphingomyelins that were affected in the study included sphingomyelins C16:0, C18:0 and C24:0, which were also positively associated with risk of myocardial infarction in the present thesis.

In cross-sectional studies, higher concentrations of total sphingomyelins were linked to subclinical atherosclerosis (159) and coronary artery disease (160). Additionally, in a small patient cohort with acute coronary syndromes a worse prognosis was found for those with higher plasma sphingomyelins (161). In the prospective Multi Ethnic study that followed healthy individuals (162), total sphingomyelins were not associated with risk of coronary heart disease. The present thesis, however, found several sphingomyelins to be associated with higher risk of myocardial infarction and provides first evidence for a prospective association. These results support the hypothesis that serum sphingomyelins could be a marker of atherosclerosis, which precedes the onset of myocardial infarction. In terms of risk prediction, the sphingomyelins and phosphatidylcholines discriminated less efficient between incident cases of myocardial infarction and non-cases than a combination of established CVD risk factors. However, their single contribution was comparable to that of other CVD risk factors and they slightly added to the information of established risk factors.

In summary, the present thesis found that some metabolites and metabolite factors linked to type 2 diabetes were also related to CVD risk underlining the concept of common pathways in chronic disease development. Not a single metabolite could be found that was strongly and exclusively linked to risk of stroke. The class of sphingomyelins was particularly identified to be strongly associated with risk of incident myocardial infarction and the biological plausibility is supported by findings from experimental and cross-sectional studies. Thus, sphingomyelins could be potential candidates for risk of myocardial infarction in the future. The present thesis was one of the first prospective studies on this topic, and ideally, these findings need external validation. In fact, these results are aimed to be replicated in the EPIC-Heidelberg cohort, where the metabolite measurements are currently being conducted.

4.3 Public health implications

Chronic diseases, including type 2 diabetes and CVD, are diminishing our life expectancy and quality, and bear severe social and financial implications for the Western society as well as for developing countries (1; 2). More systematic approaches than the study of single risk factors may help to better address and counteract the multifactorial genesis of chronic diseases (9-11). With a targeted metabolomics approach in the EPIC-Potsdam study the present thesis identified a number of metabolites that were associated with higher risk of type 2 diabetes and CVD; these metabolites also contributed to risk prediction. The discrimination by the metabolites alone was similar to that by common risk factors. This observation suggests that serum metabolite measurements could be an alternative approach to assessment of classical risk factors for risk prediction, which may not always be possible or adequate. For instance, metabolite measurements in blood may be more objective than assessment of traditional risk factors by

questionnaires. In the present thesis discrimination was only slightly improved when the metabolites were added to established risk prediction models. However, the concept of the area under the ROC curve has previously been criticized, as it may underestimate the impact of strong predictors that are included into risk prediction models that already show good discrimination (116-118). In fact, it was previously shown that even a modest improvement of the area under the ROC curve may require very strong associations between a novel predictor and the outcome (e.g. odds ratios of about 20) (116; 163). Thus, the metabolites suggested in the present thesis may facilitate early identification of individuals at high risk of chronic diseases. This offers a great chance as risk of chronic diseases can be reduced by changes in behavior and lifestyle (6; 28). Once people at high-risk are identified targeted intervention strategies could be administered to delay or even prevent the onset of chronic diseases. Thereby, health could be improved and quality-adjusted life years extended. This is particularly relevant considering the increasing life expectancy of Western society in the era post the infectious diseases. It may be argued that existing risk models may already be sufficient to identify high-risk individuals, e.g. the German Diabetes Risk Score for type 2 diabetes (7; 8), or the Framingham risk score for CVD (164; 165). Nevertheless, it was discussed previously that current risk prediction models may have a "detection gap" (116). For instance, it was shown that up to 20% of people with CVD have no traditional risk factors and up to 40% have only one (166; 167). Thus, the metabolites studied here could be informative about specific pathways that are disturbed, and thereby, allow individualized prevention or therapeutic strategies. In fact, predictive biomarkers have recently been suggested as the "holy grail" of personalized medicine (168). In the future, there is potential that these metabolites could be useful in stratified and personalized medicine.

4.4 Conclusion and outlook

The present thesis represents one of the first available studies that successfully adopted targeted metabolomics to a large prospective cohort study. In particular, 163 metabolites measured in EPIC-Potsdam were systematically studied regarding their technical and biological variation, reliability and utility in terms of their association with risk of type 2 diabetes and CVD. Thereby, most of the metabolites showed acceptable technical and biological variation and reliability, although few metabolites were identified with high technical variation and low reliability. These metabolites may not be useful in epidemiologic studies with a single measurement of metabolites as the validity of risk estimates may be negatively affected. A number of individual metabolites were identified, that were associated with risk of type 2 diabetes and myocardial infarction but not with stroke risk. In specific, amino acids, hexose and choline-containing phospholipids were linked to risk of type 2 diabetes and phosphatidylcholines and sphingomyelins were associated with risk of myocardial infarction. These associations were independent

of established risk factors and biomarkers and the metabolites further added to risk prediction. Thus, these metabolites may represent promising candidates of chronic disease risk. In addition, a metabolite factor, that was strongly associated with risk type 2 diabetes in the present thesis and also replicated in the KORA-study, was also linked to risk of myocardial infarction and stroke. This finding supports the theory of common risk factors and mechanisms involved in the etiology of chronic diseases.

In the future it is aimed to also validate the results for CVD in another cohort study, namely the EPIC-Heidelberg study. For biomarker discovery of chronic disease risk untargeted metabolomics approaches should be adopted to prospective cohort studies as they may provide a broader perspective than targeted metabolomics. This may be particularly interesting for the endpoint stroke as in the present thesis none of the individual metabolites were linked to risk of stroke. Furthermore, determinants of the metabolites identified in the present thesis should be investigated in detail with particular focus on modifiable risk factors, e.g. diet, physical activity and obesity, and their association with serum metabolites. Once modifying factors of individual serum metabolites are identified individualized recommendations for the prevention of chronic disease risk could be generated in long-term.

Summary

The global burden of chronic diseases, such as type 2 diabetes and cardiovascular diseases (CVD), may originate from multiple factors. To grasp the complexity of their pathophysiology more comprehensive and systematic approaches than the study of single risk factors seem appropriate. Analytical biochemistry offers sophisticated technology, e.g. with a targeted metabolomics approach hundreds of low weight molecular compounds (metabolites) can be measured simultaneously. When metabolomics is combined with epidemiologic study design it may be possible to better approach the biological mechanisms and pathways involved in etiology of chronic diseases. Furthermore, in the frame of a prospective cohort study predictive metabolites can be studied, which may facilitate identification of individuals at high risk of chronic diseases. However, before metabolomics can be adopted to epidemiologic studies several challenges appear, such as technical and biological variation and reliability of metabolites. As metabolites may rapidly respond to stimuli a single measurement, which epidemiologic studies usually rely on, may not always be adequate.

Considering the challenges the present thesis aimed to adopt a targeted metabolomics approach with measurement of 163 serum metabolites to the prospective European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study. Thereby, first, technical and biological variation of metabolites was studied in detail with data from repeated metabolite measurements and reliability of metabolites was expressed by the intraclass-correlation coefficient (ICC). Later on, a quantitative estimate of the decrease in relative risk (RR) because of the limited reliability was conducted. The association between metabolites and risk of type 2 diabetes and CVD was investigated in a case-cohort study including all incident cases of type 2 diabetes (n=849), myocardial infarction (n=274) and stroke (n=260) of the full cohort and a randomly drawn subcohort (n=2500). Multivariable-adjusted Cox regression analysis combined with correction for multiple testing was used to identify metabolites associated with risk of type 2 diabetes and CVD. Independency of the risk associations was further investigated by adjustment for established biomarkers. In addition, a principal component analysis was used to retrieve metabolite factors. The usefulness of metabolites for risk prediction was evaluated with measures of discrimination and calibration in comparison to established risk factors.

Of the 163 metabolites measured 127 were included into the present analysis based on their technical variation and limit of detection of the laboratory method. Most of the metabolites showed acceptable reliability; however, particularly metabolites with low serum concentrations were poorly reliable. This may cause relevant attenuation of the risk estimates and loss of statistical power. Of the 127 metabolites 34 metabolites were associated with risk of type 2 diabetes after multiple testing correction.

Particularly, increased branched chain and aromatic amino acids, hexose and diacyl-phosphatidylcholines, and decreased glycine, lyso-phosphatidylcholines, sphingomyelins and acyl-alkyl-phosphatidylcholines were associated with higher risk of type 2 diabetes. In addition, a metabolite factor associated with higher risk of type 2 diabetes was also linked to higher risk of myocardial infarction and stroke (RR (95% confidence interval) comparing extreme quintiles, *p*-trend across quintiles, type 2 diabetes: 3.82 (2.64-5.52), $p<0.0001$; myocardial infarction: 2.20 (1.31-3.69), $p<0.001$, stroke: 1.87 (1.05-3.36), $p=0.01$; respectively). Furthermore, 11 individual metabolites specifically increased phosphatidylcholines and sphingomyelins were linked to higher risk of myocardial infarction but none of the individual metabolites was associated with stroke risk after multiple testing correction. The observed associations were independent of established risk factors and partly remained significant after adjustment for classical biomarkers of chronic diseases. The metabolites contributed to risk prediction in similar magnitude as established risk factors and when combined discrimination and calibration could further be improved.

In the present thesis it was possible to adopt targeted metabolomics to the prospective EPIC-Potsdam study with a systematic and multistep approach. On the one hand, promising candidates of risk of type 2 diabetes and myocardial infarction were identified. On the other hand, the present thesis highlights the importance to understand the limitations of metabolite measurements to obtain valid risk estimates and sufficient statistical power in epidemiologic studies. In conclusion, although metabolomics combined with epidemiologic study design bears many challenges it offers great potential to study chronic disease risk. In the future, determinants of the metabolites studied here should be investigated. In addition, untargeted metabolomics which may provide a broader insight into risk factors and pathways of chronic diseases should be adopted to epidemiologic studies.

Zusammenfassung

Die weltweit gestiegene Inzidenz und Prävalenz chronischer Erkrankungen wie Typ-2-Diabetes und Herz-Kreislauf-Erkrankungen ist durch viele Faktoren begründet. Um die Komplexität ihrer Pathophysiologie besser abzubilden, scheint die Anwendung umfassender und systematischer Ansätze der Untersuchung einzelner Risikofaktoren überlegen. Die analytische Biochemie bietet entsprechende Technologien. So ermöglicht der *Targeted Metabolomics*-Ansatz beispielsweise die gleichzeitige Messung hunderter niedermolekularer Verbindungen (Metabolite). Die Kombination von *Metabolomics* mit einem epidemiologischen Studiendesign bietet dabei vielversprechende Möglichkeiten, sich den biologischen Mechanismen und Stoffwechselwegen, die bei der Entstehung chronischer Erkrankungen eine Rolle spielen könnten, zu nähern. Zudem können im Rahmen einer prospektiven Kohortenstudie prädiktive Metabolite untersucht werden, die zur frühzeitigen Identifizierung von Hochrisikopersonen chronischer Erkrankungen beitragen könnten. Bevor *Metabolomics* jedoch im Rahmen epidemiologischer Studien eingesetzt werden kann, müssen Fragen zu technischer und biologischer Variabilität sowie der Reliabilität der Metabolite geklärt werden. Da Metabolite sehr kurzfristig auf Reize reagieren können, ist eine einzelne Messung, auf der epidemiologische Studien häufig beruhen, nicht immer ausreichend.

In Anbetracht dieser Herausforderungen war es das Ziel dieser Arbeit, einen *Targeted Metabolomics*-Ansatz anzuwenden, der die Messung von 163 Serum-Metaboliten im Rahmen der *European Prospective Investigation into Cancer and Nutrition* (EPIC)-Potsdam-Studie beinhaltete. Dabei wurde zuerst die technische und biologische Variabilität der Metabolite mit Daten von wiederholten Messungen untersucht und die Reliabilität der Metabolite als Intraklassen-Korrelationskoeffizient ausgedrückt. Im Anschluss daran wurde eine quantitative Schätzung vorgenommen, inwieweit das Relative Risiko (RR) durch die limitierte Reliabilität abgeschwächt wird. Der Zusammenhang zwischen Metaboliten und dem Risiko für Typ-2-Diabetes und Herz-Kreislauf-Erkrankungen wurde in einer Fall-Kohortenstudie untersucht, die alle inzidenten Fälle von Typ-2-Diabetes (n=849), Herzinfarkt (n=274) und Schlaganfall (n=260) der gesamten Kohorte beinhaltete sowie eine zufällig gezogene Subkohorte von 2500 Personen. Mittels multivariabel-adjustierter Cox-Regressionsanalyse in Kombination mit Korrektur für multiples Testen wurden Metabolite identifiziert, die mit dem Risiko für Typ-2-Diabetes und Herz-Kreislauf-Erkrankungen assoziiert waren. Um Metaboliten-Faktoren basierend auf deren Korrelation zu generieren, wurde zusätzlich eine Hauptkomponentenanalyse durchgeführt. Der Zusatznutzen der Metabolite zur Risikoprädiktion wurde anhand von Maßzahlen der Diskriminierung und Kalibrierung im Vergleich zu etablierten Risikofaktoren abgeschätzt.

Von den 163 gemessenen Metaboliten wurden basierend auf ihrer technischen Variabilität und der Detektionsgrenze des Messinstruments 127 Metabolite in die Analyse einbezogen. Die meisten Metabolite wiesen eine ausreichende Reliabilität auf, dennoch zeigten insbesondere Metabolite mit niedriger Serumkonzentration eine geringe Reliabilität. Dieses könnte eine relevante Abschwächung des RR und damit einhergehend einen Verlust der statistischen Power bewirken. Von den 127 Metaboliten waren nach Korrektur für multiples Testen 34 Metabolite mit dem Typ-2-Diabetes-Risiko assoziiert. Insbesondere waren erhöhte Konzentrationen von verzweigtkettigen und aromatischen Aminosäuren, Hexose und Diacyl-Phosphatidylcholinen sowie geringere Konzentrationen von Glycin, Lyso-Phosphatidylcholinen, Sphingomyelinen und Acyl-Alkyl-Phosphatidylcholinen mit einem erhöhten Typ-2-Diabetes-Risiko assoziiert. Zudem war ein Metaboliten-Faktor, der mit einem höheren Typ-2-Diabetes-Risiko zusammenhing, auch mit einem erhöhten Risiko für Herzinfarkt und Schlaganfall assoziiert (RR (95% Konfidenzintervall) bei Vergleich der extremen Quintile, *p*-trend über die Quintile: entsprechend, Typ-2-Diabetes: 3.82 (2.64-5.52), $p<0.0001$; Herzinfarkt: 2.20 (1.31-3.69), $p<0.001$; Schlaganfall: 1.87 (1.05-3.36), $p=0.01$). Außerdem hingen 11 einzelne Metabolite aus der Gruppe der Sphingomyeline und Phosphatidylcholine nach Korrektur für multiples Testen positiv mit dem Herzinfarktrisiko zusammen. Keiner der individuellen 127 Metabolite war jedoch mit dem Schlaganfallrisiko assoziiert. Die beobachteten Zusammenhänge waren unabhängig von etablierten Risikofaktoren und blieben nach weiterer Adjustierung für klassische Biomarker chronischer Erkrankungen teilweise signifikant. Ferner trugen die Metabolite zur Risikoprädiktion von Typ-2-Diabetes und Herzinfarkt bei. Durch Kombination mit etablierten Risikofaktoren konnten Diskriminierung und Kalibrierung weiter verbessert werden.

In der vorliegenden Arbeit ist es gelungen einen *Targeted-Metabolomics*-Ansatz in systematischer Vorgehensweise auf die prospektive EPIC-Potsdam-Studie anzuwenden. Einerseits konnten vielversprechende Kandidaten für Typ-2-Diabetes und Herzinfarkt identifiziert werden, andererseits unterstreichen die Ergebnisse der Arbeit die Bedeutung, die Grenzen der Metaboliten-Messungen zu verstehen, um valide Risikoschätzer und ausreichend statistische Power in epidemiologischen Studien zu erreichen. Fazit ist, dass *Metabolomics* im Rahmen einer epidemiologischen Studie trotz der damit einhergehenden Herausforderungen viel Potential zur Untersuchung des Risikos chronischer Erkrankungen bietet. Zukünftige Studien sollten Determinanten der hier untersuchten Metabolite näher beleuchten. Zusätzlich könnte ein *Untargeted-Metabolomics*-Ansatz, der eine noch unvoreingenommenere Betrachtungsweise auf Risikofaktoren und Stoffwechselwegen chronischer Erkrankungen ermöglicht, in epidemiologischen Studien angewandt werden.

References

1. World Health Organisation (WHO): Diabetes Fact Sheet No 312. Septermber 2012. Webpage: http://www.who.int/mediacentre/factsheets/fs312/en/index.html. Accessed October 16th 2012.
2. World Health Organisation (WHO):Cardiovascular diseases Fact Sheet No 317. September 2012. Webpage:http://www.who.int/mediacentre/factsheets/fs317/en/index.html. Accessed October 16th 2012.
3. Statistisches Bundesamt Deutschland; Todesursachen: Anzahl der Gestorbenen nach Kapiteln der ICD-10. Website: https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Gesundheit/Todesursachen/Tabellen/GestorbeneAnzahl.html; Accessed October 15th 2012
4. Statistisches Bundesamt Deutschland; Gesundheit: Krankheitskosten. Website: https://www-genesis.destatis.de/genesis/online/data;jsessionid=F631DCF8A7B884BB8595FA15D5DD9AA2.tomcat_GO_2_1?operation=abruftabelleBearbeiten&levelindex=1&levelid=1350308500991&auswahloperation=abruftabelleAuspraegungAuswaehlen&auswahlverzeichnis=ordnungsstruktur&auswahlziel=werteabruf&selectionname=23631-0001&auswahltext=&werteabruf=Werteabruf Accessed October 15th, 2012.
5. Zimmet P, Alberti KG, Shaw J: Global and societal implications of the diabetes epidemic. Nature 2001;414:782-787
6. Ford ES, Bergmann MM, Kroger J, Schienkiewitz A, Weikert C, Boeing H: Healthy living is the best revenge: findings from the European Prospective Investigation Into Cancer and Nutrition-Potsdam study. Arch Intern Med 2009;169:1355-1362
7. Schulze MB, Hoffmann K, Boeing H, Linseisen J, Rohrmann S, Mohlig M, Pfeiffer AF, Spranger J, Thamer C, Haring HU, Fritsche A, Joost HG: An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. Diabetes Care 2007;30:510-515
8. Schulze MB, Weikert C, Pischon T, Bergmann MM, Al-Hasani H, Schleicher E, Fritsche A, Haring HU, Boeing H, Joost HG: Use of multiple metabolic and genetic markers to improve the prediction of type 2 diabetes: the EPIC-Potsdam Study. Diabetes Care 2009;32:2116-2119
9. Lund E, Dumeaux V: Systems epidemiology in cancer. Cancer Epidemiol Biomarkers Prev 2008;17:2954-2957
10. Hu FB: Metabolic profiling of diabetes: from black-box epidemiology to systems epidemiology. Clin Chem 2011;57:1224-1226
11. Muller G: Personalized prognosis and diagnosis of type 2 diabetes--vision or fiction? Pharmacology 2010;85:168-187
12. Bain JR, Stevens RD, Wenner BR, Ilkayeva O, Muoio DM, Newgard CB: Metabolomics applied to diabetes research: moving from information to knowledge. Diabetes 2009;58:2429-2443
13. Lewis GD, Asnani A, Gerszten RE: Application of metabolomics to cardiovascular biomarker and pathway discovery. J Am Coll Cardiol 2008;52:117-123
14. Gjesing AP, Pedersen O: 'Omics'-driven discoveries in prevention and treatment of type 2 diabetes. Eur J Clin Invest 2012;42:579-588
15. Hankinson SE, Manson JE, London SJ, Willett WC, Speizer FE: Laboratory reproducibility of endogenous hormone levels in postmenopausal women. Cancer Epidemiol Biomarkers Prev 1994;3:51-56
16. Tworoger SS, Hankinson SE: Use of biomarkers in epidemiologic studies: minimizing the influence of measurement error in the study design and analysis. Cancer Causes Control 2006;17:889-899
17. Rosner B, Willett WC, Spiegelman D: Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Stat Med 1989;8:1051-1069; discussion 1071-1053
18. Shurubor YI, Matson WR, Willett WC, Hankinson SE, Kristal BS: Biological variability dominates and influences analytical variance in HPLC-ECD studies of the human plasma metabolome. BMC Clin Pathol 2007;7:9

19. Kotsopoulos J, Tworoger SS, Campos H, Chung FL, Clevenger CV, Franke AA, Mantzoros CS, Ricchiuti V, Willett WC, Hankinson SE, Eliassen AH: Reproducibility of plasma and urine biomarkers among premenopausal and postmenopausal women from the Nurses' Health Studies. Cancer Epidemiol Biomarkers Prev 2010;19:938-946
20. Rosner B: The intraclass correlation coefficient. In: Rosner, B. ed. Fundamentals of biostatistics. BelmontCA: Duxbury. pp 613-8. 2006.
21. Gerszten RE, Wang TJ: The search for new cardiovascular biomarkers. Nature 2008;451:949-952
22. Herder C, Karakas M, Koenig W: Biomarkers for the prediction of type 2 diabetes and cardiovascular disease. Clin Pharmacol Ther 2011;90:52-66
23. "Chapter 15: Evaluation of performance" in "Clinical Prediction Models". E.W. Steyerberg, Springer, New York, 2009.
24. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, O'Donnell CJ, Carr SA, Mootha VK, Florez JC, Souza A, Melander O, Clish CB, Gerszten RE: Metabolite profiles and the risk of developing diabetes. Nat Med 2011;17:448-453
25. Statistisches Bundesamt Deutschland; Gesundheit im Alter, 2011. Website: https://www.destatis.de/DE/Publikationen/Thematisch/Gesundheit/Gesundheitszustand/GesundheitimAlter5230002119004.pdf?__blob=publicationFile; Accessed October 15th 2012
26. "Diabetes mellitus: How many adults in Germany have diabetes?" C. Heidemann, Y. Du, C. Scheidt-Nave. In: Bundesgesundheitsblatt 2012, 55:980-990. Robert Koch Institut Berlin. Websites: http://www.rki.de/DE/Content/Gesundheitsmonitoring/Studien/Degs/BGBL_2012_55_BM_Kurth.pdf?__blob=publicationFile and http://www.rki.de/DE/Content/Gesundheitsmonitoring/Studien/Degs/degs_w1/Symposium/degs_diabetes.pdf;jsessionid=BA4661DD5B01B009F0FA5324F8958C18.2_cid241?__blob=publicationFile Accessed October 15th 2012.
27. Laaksonen MA, Knekt P, Rissanen H, Harkanen T, Virtala E, Marniemi J, Aromaa A, Heliovaara M, Reunanen A: The relative importance of modifiable potential risk factors of type 2 diabetes: a meta-analysis of two cohorts. Eur J Epidemiol 2010;25:115-124
28. World Health Organization, Food and Agriculture Organization (2003) Diet, Nutrition and the Prevention of Chronic Diseases. Report of a Joint WHO/FAO Expert Consultation. WHO, Geneva, Switzerland. 2003;
29. Brotman DJ, Walker E, Lauer MS, O'Brien RG: In search of fewer independent risk factors. Arch Intern Med 2005;165:138-145
30. Tabak AG, Jokela M, Akbaraly TN, Brunner EJ, Kivimaki M, Witte DR: Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. Lancet 2009;373:2215-2221
31. Lee CC, Adler AI, Sandhu MS, Sharp SJ, Forouhi NG, Erqou S, Luben R, Bingham S, Khaw KT, Wareham NJ: Association of C-reactive protein with type 2 diabetes: prospective analysis and meta-analysis. Diabetologia 2009;52:1040-1047
32. Kaptoge S, Di Angelantonio E, Lowe G, Pepys MB, Thompson SG, Collins R, Danesh J: C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis. Lancet 2010;375:132-140
33. Law MR, Wald NJ, Thompson SG: By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? BMJ 1994;308:367-372
34. Bictash M, Ebbels TM, Chan Q, Loo RL, Yap IK, Brown IJ, de Iorio M, Daviglus ML, Holmes E, Stamler J, Nicholson JK, Elliott P: Opening up the "Black Box": metabolic phenotyping and metabolome-wide association studies in epidemiology. J Clin Epidemiol 2010;63:970-979
35. Institute of Systems and Synthetic Biology, Imperial College London. Website: http://www3.imperial.ac.uk/systemsbiology. Accessed October 18th, 2012. .
36. Mayr M: Metabolomics: ready for the prime time? Circ Cardiovasc Genet 2008;1:58-65

37. Definition by the Metabolomics society. Website: http://www.metabolomicssociety.org/metabolomics. Accessed October 22nd 2012.
38. Gibney MJ, Walsh M, Brennan L, Roche HM, German B, van Ommen B: Metabolomics in human nutrition: opportunities and challenges. Am J Clin Nutr 2005;82:497-503
39. Weinberger KM: [Metabolomics in diagnosing metabolic diseases]. Ther Umsch 2008;65:487-491
40. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly MA, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, MacInnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L: HMDB: the human metabolome database. Nucleic Acids Research 2007;35:D521-D526
41. Nicholson JK, Lindon JC: Systems biology: Metabonomics. Nature 2008;455:1054-1056
42. "Diabetes timeline". In: "The illustrated timeline of medicine." Davies, G. Page 152. Rosen Publishing Group, New York, 2012.
43. Bloch F, Hansen WW, Packard M: Nuclear Induction. Phys. Rev. 1946;69:127-127
44. Purcell EU, Torrey HC, Pound RV: Resonance Absorption by Nuclear Magnetic Moments in a Solid. Phys. Rev. 1946;69:37-38
45. Horning EC, Horning MG: Metabolic profiles: gas-phase methods for analysis of metabolites. Clin Chem 1971;17:802-809
46. Pauling L, Robinson AB, Teranishi R, Cary P: Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. Proc Natl Acad Sci U S A 1971;68:2374-2376
47. Oliver SG, Winson MK, Kell DB, Baganz F: Systematic functional analysis of the yeast genome. Trends Biotechnol 1998;16:373-378
48. Nicholson JK, Lindon JC, Holmes E: 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 1999;29:1181-1189
49. Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, Mandal R, Sinelnikov I, Xia JG, Jia L, Cruz JA, Lim E, Sobsey CA, Shrivastava S, Huang P, Liu P, Fang L, Peng J, Fradette R, Cheng D, Tzur D, Clements M, Lewis A, De Souza A, Zuniga A, Dawe M, Xiong YP, Clive D, Greiner R, Nazyrova A, Shaykhutdinov R, Li L, Vogel HJ, Forsythe I: HMDB: a knowledgebase for the human metabolome. Nucleic Acids Research 2009;37:D603-D610
50. Smolinska A, Blanchet L, Buydens LM, Wijmenga SS: NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. Anal Chim Acta 2012;750:82-97
51. Primrose S, Draper J, Elsom R, Kirkpatrick V, Mathers JC, Seal C, Beckmann M, Haldar S, Beattie JH, Lodge JK, Jenab M, Keun H, Scalbert A: Metabolomics and human nutrition. Br J Nutr 2011;105:1277-1283
52. Vineis P: Sources of variation in biomarkers. IARC Sci Publ 1997:59-71
53. Streiner DL, Norman GR: "Precision" and "accuracy": two terms that are neither. J Clin Epidemiol 2006;59:327-330
54. "Application of Biomarkers in Cancer Epidemiology". Toniolo, P., Boffeta, P., Shuker, D.E.G., Rotthman, N., et al. IARC Scientific Publications, No 142. International Agency for Research on Cancer, Lyon, 1997.
55. "Chapter 3: Study design for prediction models" in "Clinical Prediction Models". E.W. Steyerberg, Springer, New York, 2009.
56. Pencina MJ, D'Agostino RB, Vasan RS: Statistical methods for assessment of added usefulness of new biomarkers. Clin Chem Lab Med 2010;48:1703-1711
57. Hill AB: The Environment and Disease: Association or Causation? Proc R Soc Med 1965;58:295-300
58. Howick J, Glasziou P, Aronson JK: The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation' contribute? J R Soc Med 2009;102:186-194

59. "Chapter 5: Assessing the validity and reliability of diagnostic and screening tests". In: "Epidemiology". Gordis, L. Page 85-108. Saunders Elsevier, 4th ed, Philadelphia, 2009.
60. Krumsiek J, Suhre K, Illig T, Adamski J, Theis FJ: Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. BMC Syst Biol 2011;5:21
61. Krumsiek J, Suhre K, Evans AM, Mitchell MW, Mohney RP, Milburn MV, Wagele B, Romisch-Margl W, Illig T, Adamski J, Gieger C, Theis FJ, Kastenmuller G: Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. PLoS Genet 2012;8:e1003005
62. Rothman KJ, Greenland S, Lash TL: Case-control studies. In: Rothman KJ, Greenland S, eds. Modern epidemiology. 3rd ed. Philadelphia; Lippincott-Williams & Wilkins; 2008: 111-127.
63. Adams SH, Hoppel CL, Lok KH, Zhao L, Wong SW, Minkler PE, Hwang DH, Newman JW, Garvey WT: Plasma acylcarnitine profiles suggest incomplete long-chain fatty acid beta-oxidation and altered tricarboxylic acid cycle activity in type 2 diabetic African-American women. J Nutr 2009;139:1073-1081
64. Bao Y, Zhao T, Wang X, Qiu Y, Su M, Jia W: Metabonomic variations in the drug-treated type 2 diabetes mellitus patients and healthy volunteers. J Proteome Res 2009;8:1623-1630
65. Fiehn O, Garvey WT, Newman JW, Lok KH, Hoppel CL, Adams SH: Plasma metabolomic profiles reflective of glucose homeostasis in non-diabetic and type 2 diabetic obese African-American women. PLoS One 2010;5:e15234
66. Ha CY, Kim JY, Paik JK, Kim OY, Paik YH, Lee EJ, Lee JH: The association of specific metabolites of lipid metabolism with markers of oxidative stress, inflammation and arterial stiffness in men with newly diagnosed type 2 diabetes. Clin Endocrinol (Oxf) 2012;76:674-682
67. Suhre K, Meisinger C, Doring A, Altmaier E, Belcredi P, Gieger C, Chang D, Milburn MV, Gall WE, Weinberger KM, Mewes HW, Hrabe de Angelis M, Wichmann HE, Kronenberg F, Adamski J, Illig T: Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. PLoS One 2010;5:e13953
68. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, Heim K, Campillos M, Holzapfel C, Thorand B, Grallert H, Xu T, Bader E, Huth C, Mittelstrass K, Doring A, Meisinger C, Gieger C, Prehn C, Roemisch-Margl W, Carstensen M, Xie L, Yamanaka-Okumura H, Xing G, Ceglarek U, Thiery J, Giani G, Lickert H, Lin X, Li Y, Boeing H, Joost HG, de Angelis MH, Rathmann W, Suhre K, Prokisch H, Peters A, Meitinger T, Roden M, Wichmann HE, Pischon T, Adamski J, Illig T: Novel biomarkers for pre-diabetes identified by metabolomics. Mol Syst Biol 2012;8:615
69. Zhang X, Wang Y, Hao F, Zhou X, Han X, Tang H, Ji L: Human serum metabonomic analysis reveals progression axes for glucose intolerance and insulin resistance statuses. J Proteome Res 2009;8:5188-5195
70. Li X, Xu Z, Lu X, Yang X, Yin P, Kong H, Yu Y, Xu G: Comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry for metabonomics: Biomarker discovery for diabetes mellitus. Anal Chim Acta 2009;633:257-262
71. van Doorn M, Vogels J, Tas A, van Hoogdalem EJ, Burggraaf J, Cohen A, van der Greef J: Evaluation of metabolite profiles as biomarkers for the pharmacological effects of thiazolidinediones in Type 2 diabetes mellitus patients and healthy volunteers. Br J Clin Pharmacol 2007;63:562-574
72. Zhang J, Yan L, Chen W, Lin L, Song X, Yan X, Hang W, Huang B: Metabonomics research of diabetic nephropathy and type 2 diabetes mellitus based on UPLC-oaTOF-MS system. Anal Chim Acta 2009;650:16-22
73. Bodi V, Sanchis J, Morales JM, Marrachelli VG, Nunez J, Forteza MJ, Chaustre F, Gomez C, Mainar L, Minana G, Rumiz E, Husser O, Noguera I, Diaz A, Moratal D, Carratala A, Bosch X, Llacer A, Chorro FJ, Vina JR, Monleon D: Metabolomic profile of human myocardial ischemia by nuclear magnetic resonance spectroscopy of peripheral blood serum: a translational study based on transient coronary occlusion models. J Am Coll Cardiol 2012;59:1629-1641
74. Jiang Z, Sun J, Liang Q, Cai Y, Li S, Huang Y, Wang Y, Luo G: A metabonomic approach applied to predict patients with cerebral infarction. Talanta 2011;84:298-304

75. Jung JY, Lee HS, Kang DG, Kim NS, Cha MH, Bang OS, Ryu do H, Hwang GS: 1H-NMR-based metabolomics study of cerebral infarction. Stroke 2011;42:1282-1288
76. Kang SM, Park JC, Shin MJ, Lee H, Oh J, Ryu do H, Hwang GS, Chung JH: (1)H nuclear magnetic resonance based metabolic urinary profiling of patients with ischemic heart failure. Clin Biochem 2011;44:293-299
77. Teul J, Garcia A, Tunon J, Martin-Ventura JL, Tarin N, Bescos LL, Egido J, Barbas C, Ruperez FJ: Targeted and non-targeted metabolic time trajectory in plasma of patients after acute coronary syndrome. J Pharm Biomed Anal 2011;56:343-351
78. Shah AA, Craig DM, Sebek JK, Haynes C, Stevens RC, Muehlbauer MJ, Granger CB, Hauser ER, Newby LK, Newgard CB, Kraus WE, Hughes GC, Shah SH: Metabolic profiles predict adverse events after coronary artery bypass grafting. J Thorac Cardiovasc Surg 2012;143:873-878
79. Shah SH, Sun JL, Stevens RD, Bain JR, Muehlbauer MJ, Pieper KS, Haynes C, Hauser ER, Kraus WE, Granger CB, Newgard CB, Califf RM, Newby LK: Baseline metabolomic profiles predict cardiovascular events in patients at risk for coronary artery disease. Am Heart J 2012;163:844-850 e841
80. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung YM, Wu Y, Schauer P, Smith JD, Allayee H, Tang WH, DiDonato JA, Lusis AJ, Hazen SL: Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 2011;472:57-63
81. Zhu C, Liang QL, Hu P, Wang YM, Luo GA: Phospholipidomic identification of potential plasma biomarkers associated with type 2 diabetes mellitus and diabetic nephropathy. Talanta 2011;85:1711-1720
82. Bingham S, Riboli E: Diet and cancer--the European Prospective Investigation into Cancer and Nutrition. Nat Rev Cancer 2004;4:206-215
83. Boeing H, Korfmann A, Bergmann MM: Recruitment procedures of EPIC-Germany. European Investigation into Cancer and Nutrition. Ann Nutr Metab 1999;43:205-215
84. Boeing H, Wahrendorf J, Becker N: EPIC-Germany--A source for studies into diet and risk of chronic diseases. European Investigation into Cancer and Nutrition. Ann Nutr Metab 1999;43:195-204
85. Schienkiewitz A, Schulze MB, Hoffmann K, Kroke A, Boeing H: Body mass index history and risk of type 2 diabetes: results from the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. Am J Clin Nutr 2006;84:427-433
86. Bergmann MM, Bussas U, Boeing H: Follow-up procedures in EPIC-Germany--data quality aspects. European Prospective Investigation into Cancer and Nutrition. Ann Nutr Metab 1999;43:225-234
87. Bohlscheid-Thomas S, Hoting I, Boeing H, Wahrendorf J: Reproducibility and relative validity of energy and macronutrient intake of a food frequency questionnaire developed for the German part of the EPIC project. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol 1997;26 Suppl 1:S71-81
88. Bohlscheid-Thomas S, Hoting I, Boeing H, Wahrendorf J: Reproducibility and relative validity of food group intake in a food frequency questionnaire developed for the German part of the EPIC project. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol 1997;26 Suppl 1:S59-70
89. Schulze MB, Kroke A, Bergmann MM, Boeing H: Differences of blood pressure estimates between consecutive measurements on one occasion: implications for inter-study comparability of epidemiologic studies. Eur J Epidemiol 2000;16:891-898
90. Ford ES, Schulze MB, Bergmann MM, Thamer C, Joost HG, Boeing H: Liver enzymes and incident diabetes: findings from the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. Diabetes Care 2008;31:1138-1143
91. Weikert C, Stefan N, Schulze MB, Pischon T, Berger K, Joost HG, Haring HU, Boeing H, Fritsche A: Plasma fetuin-a levels and the risk of myocardial infarction and ischemic stroke. Circulation 2008;118:2555-2562
92. Montonen J, Drogan D, Joost HG, Boeing H, Fritsche A, Schleicher E, Schulze MB, Pischon T: Estimation of the contribution of biomarkers of different metabolic pathways to risk of type 2 diabetes. Eur J Epidemiol 2011;26:29-38

93. Montonen J, Boeing H, Steffen A, Lehmann R, Fritsche A, Joost HG, Schulze MB, Pischon T: Body iron stores and risk of type 2 diabetes: results from the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study. Diabetologia 2012;55:2613-2621
94. Romisch-Margl W, Prehn C, Bogumil R, Rohring C, Suhre K, Adamski J: Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. Metabolomics 2012;8:133-142
95. Absolute*IDQ*™ Kit. Analytical Specifications p150. BIOCRATES Life Sciences, Innsbruck, Austria. 2006.
96. Rosner B: One-way ANOVA - the random-effects model. In: Rosner, B. ed. Fundamentals of biostatistics. BelmontCA: Duxbury. pp 606-13. 2006.
97. Lacher DA, Hughes JP, Carroll MD: Estimate of biological variation of laboratory analytes based on the third national health and nutrition examination survey. Clin Chem 2005;51:450-452
98. Fleiss JL: Reliability of measurement. In: Fleiss J.L., ed. The design and analysis of clinical experiments. Wiley and Sons: New York. pp p.1-32. 1986.
99. Rosner B, Willett WC: Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. Am J Epidemiol 1988;127:377-386
100. Cox DR, Oakes D: Analysis of survival data. London: Chapman and Hall. 1984;
101. Prentice RL: Design issues in cohort studies. Stat Methods Med Res 1995;4:273-292
102. "Chapter 3: Survival Data", in: "Modelling survival data in medical research", David Collett, p56-58, 2nd ed, Chapman & Hall/CRC. Boca Raton, 2003.
103. Holm S: A simple sequentially rejective multiple test procedure. Scand J Statist 1979;6:65-70
104. Schulze MB, Hoffmann K: Methodological approaches to study dietary patterns in relation to risk of coronary heart disease and stroke. Br J Nutr 2006;95:860-869
105. Nagelkerke NJD: A Note on a General Definition of the Coefficient of Determination. Biometrika 1991;78:691-692
106. DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-845
107. Hosmer DW, S. L: Applied Logistic Regression. 2 ed. New York: John Wiley & Sons, Inc.; 2000.
108. Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, Joost HG, Boeing H, Pischon T: Reliability of Serum Metabolite Concentrations over a 4-Month Period Using a Targeted Metabolomic Approach. PLoS One 2011;6:e21103
109. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, Fritsche A, Haring HU, Hrabe de Angelis M, Peters A, Roden M, Prehn C, Wang-Sattler R, Illig T, Schulze MB, Adamski J, Boeing H, Pischon T: Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. Diabetes 2012;
110. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36
111. Müller, M.J. and E. Trautwein, Gesundheit und Ernährung. Public Health Nutrition. 2005, Stuttgart: Ulmer UTB. 304.
112. "Selected lifestyle factors and mortality risk in individuals with diabetes mellitus: Are the associations different from individuals without diabetes?" Diewertje Sluik, doctoral thesis Technical University of Berlin, p. 2. , Berlin, 2012.
113. Schatzkin A, Kipnis V: Could exposure assessment problems give us wrong answers to nutrition and cancer questions? J Natl Cancer Inst 2004;96:1564-1565
114. Illig T, Gieger C, Zhai G, Romisch-Margl W, Wang-Sattler R, Prehn C, Altmaier E, Kastenmuller G, Kato BS, Mewes HW, Meitinger T, de Angelis MH, Kronenberg F, Soranzo N, Wichmann HE, Spector TD, Adamski J, Suhre K: A genome-wide perspective of genetic variation in human metabolism. Nat Genet 2010;42:137-141

115. Benjamini Y, Hochberg Y: Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 1995;57:289-300
116. Ge Y, Wang TJ: Identifying novel biomarkers for cardiovascular disease risk prediction. J Intern Med 2012;272 430-439
117. Kowall B, Rathmann W, Strassburger K: Use of areas under the receiver operating curve (AROCs) and some caveats. Int J Public Health 2012;
118. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS: Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157-172; discussion 207-112
119. Demler OV, Pencina MJ, D'Agostino RB, Sr.: Misuse of DeLong test to compare AUCs for nested models. Stat Med 2012;31:2577-2587
120. Falk RT, Gail MH, Fears TR, Rossi SC, Stanczyk F, Adlercreutz H, Kiura P, Wahala K, Donaldson JL, Vaught JB, Fillmore CM, Hoover RN, Ziegler RG: Reproducibility and validity of radioimmunoassays for urinary hormones and metabolites in pre- and postmenopausal women. Cancer Epidemiol Biomarkers Prev 1999;8:567-577
121. Kaplan RC, Ho GY, Xue X, Rajpathak S, Cushman M, Rohan TE, Strickler HD, Scherer PE, Anastos K: Within-individual stability of obesity-related biomarkers among women. Cancer Epidemiol Biomarkers Prev 2007;16:1291-1293
122. Crews B, Wikoff WR, Patti GJ, Woo HK, Kalisiak E, Heideker J, Siuzdak G: Variability analysis of human plasma and cerebral spinal fluid reveals statistical significance of changes in mass spectrometry-based metabolomics data. Anal Chem 2009;81:8538-8544
123. Strnadova KA, Holub M, Muhl A, Heinze G, Ratschmann R, Mascher H, Stockler-Ipsiroglu S, Waldhauser F, Votava F, Lebl J, Bodamer OA: Long-term stability of amino acids and acylcarnitines in dried blood spots. Clin Chem 2007;53:717-722
124. Zivkovic AM, Wiest MM, Nguyen UT, Davis R, Watkins SM, German JB: Effects of sample handling and storage on quantitative lipid analysis in human serum. Metabolomics 2009;5:507-516
125. Zelena E, Dunn WB, Broadhurst D, Francis-McIntyre S, Carroll KM, Begley P, O'Hagan S, Knowles JD, Halsall A, Wilson ID, Kell DB: Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. Anal Chem 2009;81:1357-1364
126. McBride KL, Belmont JW, O'Brien WE, Amin TJ, Carter S, Lee BH: Heritability of plasma amino acid levels in different nutritional states. Mol Genet Metab 2007;90:217-220
127. Mihalik SJ, Goodpaster BH, Kelley DE, Chace DH, Vockley J, Toledo FG, DeLany JP: Increased levels of plasma acylcarnitines in obesity and type 2 diabetes and identification of a marker of glucolipotoxicity. Obesity (Silver Spring) 2010;18:1695-1700
128. Quehenberger O, Dennis EA: The human plasma lipidome. N Engl J Med 2011;365:1812-1823
129. Zhao X, Peter A, Fritsche J, Elcnerova M, Fritsche A, Haring HU, Schleicher ED, Xu G, Lehmann R: Changes of the plasma metabolome during an oral glucose tolerance test: is there more than glucose to look at? Am J Physiol Endocrinol Metab 2009;296:E384-393
130. Magnusson CD, Haraldsson GG: Ether lipids. Chem Phys Lipids 2011;164:315-340
131. Pavoine C, Pecker F: Sphingomyelinases: their regulation and roles in cardiovascular pathophysiology. Cardiovasc Res 2009;82:175-183
132. Soares JC, Dippold CS, Wells KF, Houck P, Mallinger AG: Reproducibility of in vivo measures of platelet membrane phospholipids in human subjects. Psychiatry Res 1999;86:107-112
133. Ruderman NB: Muscle amino acid metabolism and gluconeogenesis. Annu Rev Med 1975;26:245-258
134. Sekhar RV, McKay SV, Patel SG, Guthikonda AP, Reddy VT, Balasubramanyam A, Jahoor F: Glutathione synthesis is diminished in patients with uncontrolled diabetes and restored by dietary supplementation with cysteine and glycine. Diabetes Care 2011;34:162-167

135. Koves TR, Ussher JR, Noland RC, Slentz D, Mosedale M, Ilkayeva O, Bain J, Stevens R, Dyck JR, Newgard CB, Lopaschuk GD, Muoio DM: Mitochondrial overload and incomplete fatty acid oxidation contribute to skeletal muscle insulin resistance. Cell Metab 2008;7:45-56
136. Muoio DM, Neufer PD: Lipid-induced mitochondrial stress and insulin action in muscle. Cell Metab 2012;15:595-605
137. Vianey-Liaud C, Divry P, Gregersen N, Mathieu M: The inborn errors of mitochondrial fatty acid oxidation. J Inherit Metab Dis 1987;10 Suppl 1:159-200
138. She P, Van Horn C, Reid T, Hutson SM, Cooney RN, Lynch CJ: Obesity-related elevations in plasma leucine are associated with alterations in enzymes involved in branched-chain amino acid metabolism. Am J Physiol Endocrinol Metab 2007;293:E1552-1563
139. Tremblay F, Brule S, Hee Um S, Li Y, Masuda K, Roden M, Sun XJ, Krebs M, Polakiewicz RD, Thomas G, Marette A: Identification of IRS-1 Ser-1101 as a target of S6K1 in nutrient- and obesity-induced insulin resistance. Proc Natl Acad Sci U S A 2007;104:14056-14061
140. Nuttall FQ, Gannon MC: Plasma glucose and insulin response to macronutrients in nondiabetic and NIDDM subjects. Diabetes Care 1991;14:824-838
141. Montonen J, Jarvinen R, Knekt P, Heliovaara M, Reunanen A: Consumption of sweetened beverages and intakes of fructose and glucose predict type 2 diabetes occurrence. J Nutr 2007;137:1447-1454
142. Hu FB, Malik VS: Sugar-sweetened beverages and risk of obesity and type 2 diabetes: epidemiologic evidence. Physiol Behav 2010;100:47-54
143. Wallner S, Schmitz G: Plasmalogens the neglected regulatory and scavenging lipid species. Chem Phys Lipids 2011;164:573-589
144. Pietilainen KH, Sysi-Aho M, Rissanen A, Seppanen-Laakso T, Yki-Jarvinen H, Kaprio J, Oresic M: Acquired obesity is associated with changes in the serum lipidomic profile independent of genetic effects--a monozygotic twin study. PLoS One 2007;2:e218
145. Cole LK, Vance JE, Vance DE: Phosphatidylcholine biosynthesis and lipoprotein metabolism. Biochim Biophys Acta 2011;
146. Raubenheimer PJ, Nyirenda MJ, Walker BR: A choline-deficient diet exacerbates fatty liver but attenuates insulin resistance and glucose intolerance in mice fed a high-fat diet. Diabetes 2006;55:2015-2020
147. Jacobs RL, Devlin C, Tabas I, Vance DE: Targeted deletion of hepatic CTP:phosphocholine cytidylyltransferase alpha in mice decreases plasma high density and very low density lipoproteins. J Biol Chem 2004;279:47402-47410
148. Kroger J, Zietemann V, Enzenbach C, Weikert C, Jansen EH, Doring F, Joost HG, Boeing H, Schulze MB: Erythrocyte membrane phospholipid fatty acids, desaturase activity, and dietary fatty acids in relation to risk of type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. Am J Clin Nutr 2011;93:127-142
149. Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, Yang E, Farrell L, Fox CS, O'Donnell CJ, Carr SA, Vasan RS, Florez JC, Clish CB, Wang TJ, Gerszten RE: Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. J Clin Invest 2011;121:1402-1411
150. Meisinger C, Strassburger K, Heier M, Thorand B, Baumeister SE, Giani G, Rathmann W: Prevalence of undiagnosed diabetes and impaired glucose regulation in 35-59-year-old individuals in Southern Germany: the KORA F4 Study. Diabet Med 2010;27:360-362
151. Rathmann W, Strassburger K, Heier M, Holle R, Thorand B, Giani G, Meisinger C: Incidence of Type 2 diabetes in the elderly German population and the effect of clinical and lifestyle risk factors: KORA S4/F4 cohort study. Diabet Med 2009;26:1212-1219
152. Kowall B, Rathmann W, Strassburger K, Meisinger C, Holle R, Mielck A: Socioeconomic status is not associated with type 2 diabetes incidence in an elderly population in Germany: KORA S4/F4 Cohort Study. J Epidemiol Community Health 2011;65:606-612

153. Bergheanu SC, Reijmers T, Zwinderman AH, Bobeldijk I, Ramaker R, Liem AH, van der Greef J, Hankemeier T, Jukema JW: Lipidomic approach to evaluate rosuvastatin and atorvastatin at various dosages: investigating differential effects among statins. Curr Med Res Opin 2008;24:2477-2487
154. Park TS, Panek RL, Rekhter MD, Mueller SB, Rosebury WS, Robertson A, Hanselman JC, Kindt E, Homan R, Karathanasis SK: Modulation of lipoprotein metabolism by inhibition of sphingomyelin synthesis in ApoE knockout mice. Atherosclerosis 2006;189:264-272
155. Oestvang J, Bonnefont-Rousselot D, Ninio E, Hakala JK, Johansen B, Anthonsen MW: Modification of LDL with human secretory phospholipase A(2) or sphingomyelinase promotes its arachidonic acid-releasing propensity. J Lipid Res 2004;45:831-838
156. Bismuth J, Lin P, Yao Q, Chen C: Ceramide: a common pathway for atherosclerosis? Atherosclerosis 2008;196:497-504
157. Hoff HF, Morton RE: Lipoproteins containing apo B extracted from human aortas. Structure and function. Ann N Y Acad Sci 1985;454:183-194
158. Jeong T, Schissel SL, Tabas I, Pownall HJ, Tall AR, Jiang X: Increased sphingomyelin content of plasma lipoproteins in apolipoprotein E knockout mice reflects combined production and catabolic defects and enhances reactivity with mammalian sphingomyelinase. J Clin Invest 1998;101:905-912
159. Nelson JC, Jiang XC, Tabas I, Tall A, Shea S: Plasma sphingomyelin and subclinical atherosclerosis: findings from the multi-ethnic study of atherosclerosis. Am J Epidemiol 2006;163:903-912
160. Jiang XC, Paultre F, Pearson TA, Reed RG, Francis CK, Lin M, Berglund L, Tall AR: Plasma sphingomyelin level as a risk factor for coronary artery disease. Arterioscler Thromb Vasc Biol 2000;20:2614-2618
161. Schlitt A, Blankenberg S, Yan D, von Gizycki H, Buerke M, Werdan K, Bickel C, Lackner KJ, Meyer J, Rupprecht HJ, Jiang XC: Further evaluation of plasma sphingomyelin levels as a risk factor for coronary artery disease. Nutr Metab (Lond) 2006;3:5
162. Yeboah J, McNamara C, Jiang XC, Tabas I, Herrington DM, Burke GL, Shea S: Association of plasma sphingomyelin levels and incident coronary heart disease events in an adult population: Multi-Ethnic Study of Atherosclerosis. Arterioscler Thromb Vasc Biol 2010;30:628-633
163. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P: Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol 2004;159:882-890
164. Kannel WB, McGee D, Gordon T: A general cardiovascular risk profile: the Framingham Study. Am J Cardiol 1976;38:46-51
165. D'Agostino RB, Sr., Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, Kannel WB: General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008;117:743-753
166. Khot UN, Khot MB, Bajzer CT, Sapp SK, Ohman EM, Brener SJ, Ellis SG, Lincoff AM, Topol EJ: Prevalence of conventional risk factors in patients with coronary heart disease. JAMA 2003;290:898-904
167. Hozawa A, Folsom AR, Sharrett AR, Chambless LE: Absolute and attributable risks of cardiovascular disease incidence in relation to optimal and borderline risk factors: comparison of African American with white subjects--Atherosclerosis Risk in Communities Study. Arch Intern Med 2007;167:573-579
168. Quote from Prof. Dr. Tomasz Burzykowski from the International Drug Development Institute, Louvain-la-Neuve, Belgium, in "Panel discussion: biomarker-driven clinical studies" at 2nd Munich Biomarker Conference, Hilton Munich Park, Munich, November 22nd, 2012

Appendix[19]

Table S1: Biochemical names and limit of detection of 163 metabolites measured with the BIOCRATES Absolute IDQ targeted metabolomics technology		
Abbreviation	***Biochemical name***	***Limit of detection (µM)****
Acylcarnitines		
C0	DL-Carnitine	4
C10	Decanoyl-L-carnitine	0.16
C10:1	Decenoyl-L-carnitine	0.12
C10:2	Decadienyl-L-carnitine	0.04
C12	Dodecanoyl-L-carnitine	0.057
C12-DC	Dodecanedioyl-L-carnitine	0.2
C12:1	Dodecenoyl-L-carnitine	0.2
C14	Tetradecanoyl-L-carnitine	0.03
C14:1	Tetradecenoyl-L-carnitine	0.015
C14:1-OH	Hydroxytetradecenoyl-L-carnitine	0.015
C14:2	Tetradecadienyl-L-carnitine	0.012
C14:2-OH	Hydroxytetradecadienyl-L-carnitine	0.015
C16	Hexadecanoyl-L-carnitine	0.018
C16-OH	Hydroxyhexadecanoyl-L-carnitine	0.015
C16:1	Hexadecenoyl-L-carnitine	0.06
C16:1-OH	Hydroxyhexadecenoyl-L-carnitine	0.02
C16:2	Hexadecadienyl-L-carnitine	0.008
C16:2-OH	Hydroxyhexadecadienyl-L-carnitine	0.03
C18	Octadecanoyl-L-carnitine	0.02
C18:1	Octadecenoyl-L-carnitine	0.04
C18:1-OH	Hydroxyoctadecenoyl-L-carnitine	0.023
C18:2	Octadecadienyl-L-carnitine	0.009
C2	Acetyl-L-carnitine	0.15
C3	Propionyl-L-carnitine	0.08
C3-DC / C4-OH	Malonyl-L-carnitine / Hydroxybutyryl-L-carnitine	0.09
C3-DC-M / C5-OH	Methylmalonyl-L-carnitine / Hydroxyvaleryl-L-carnitine	0.1
C3-OH	Hydroxypropionyl-L-carnitine	0.05
C3:1	Propenyl-L-carnitine	0.03
C4	Butyryl-L-carnitine	0.03
C4:1	Butenyl-L-carnitine	0.03
C4:1-DC / C6	Fumaryl-L-carnitine / Hexanoyl-L-carnitine	0.08
C5	Valeryl-L-carnitine	0.04
C5-DC / C6-OH	Glutaryl-L-carnitine / Hydroxyhexanoyl-L-carnitine	0.035
C5-M-DC	Methylglutaryl-L-carnitine	0.06

[19] In the framework of this thesis, parts of the appendix have already been published by the author. Parts of Tables S1 and S2 were published in: 108. Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, Joost HG, Boeing H, Pischon T: Reliability of Serum Metabolite Concentrations over a 4-Month Period Using a Targeted Metabolomic Approach. PLoS One 2011;6:e21103 (PLoS applies the Creative Commons Attribution License). Parts of Tables S3 and S9 were published in: 109. Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, Fritsche A, Haring HU, Hrabe de Angelis M, Peters A, Roden M, Prehn C, Wang-Sattler R, Illig T, Schulze MB, Adamski J, Boeing H, Pischon T: Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. Diabetes 2012; Oct.4th (Epub ahead). (Copyright: American Diabetes Association (ADA). http://diabetes.diabetesjournals.org

C5:1	Tiglyl-L-carnitine	0.04
C5:1-DC	Glutaconyl-L-carnitine	0.015
C6:1	Hexenoyl-L-carnitine	0.035
C7-DC	Pimelyl-L-carnitine	0.035
C8	Octanoyl-L-carnitine	0.17
C8:1	Octenoyl-L-carnitine	0.025
C9	Nonayl-L-carnitine	0.04
Amino acids		
Arg	Arginine	4
Gln	Glutamine	3
Gly	Glycine	6
His	Histidine	4
Met	Methionine	1
Orn	Ornithine	5
Phe	Phenylalanine	2
Pro	Proline	2
Ser	Serine	3
Thr	Threonine	3
Trp	Tryptophan	10
Tyr	Tyrosine	3
Val	Valine	4
xLeu	Leucine/Isoleucin	2
Glycerophospholipids		
PC aa C24:0	Phosphatidylcholine diacyl C 24:0	0.1
PC aa C26:0	Phosphatidylcholine diacyl C 26:0	1.4
PC aa C28:1	Phosphatidylcholine diacyl C 28:1	0.04
PC aa C30:0	Phosphatidylcholine diacyl C 30:0	0.2
PC aa C30:2	Phosphatidylcholine diacyl C 30:2	0.006
PC aa C32:0	Phosphatidylcholine diacyl C 32:0	0.04
PC aa C32:1	Phosphatidylcholine diacyl C 32:1	0.06
PC aa C32:2	Phosphatidylcholine diacyl C 32:2	0.03
PC aa C32:3	Phosphatidylcholine diacyl C 32:3	0.008
PC aa C34:1	Phosphatidylcholine diacyl C 34:1	0.06
PC aa C34:2	Phosphatidylcholine diacyl C 34:2	0.1
PC aa C34:3	Phosphatidylcholine diacyl C 34:3	0.01
PC aa C34:4	Phosphatidylcholine diacyl C 34:4	0.006
PC aa C36:0	Phosphatidylcholine diacyl C 36:0	0.2
PC aa C36:1	Phosphatidylcholine diacyl C 36:1	0.03
PC aa C36:2	Phosphatidylcholine diacyl C 36:2	0.15
PC aa C36:3	Phosphatidylcholine diacyl C 36:3	0.04
PC aa C36:4	Phosphatidylcholine diacyl C 36:4	0.04
PC aa C36:5	Phosphatidylcholine diacyl C 36:5	0.01
PC aa C36:6	Phosphatidylcholine diacyl C 36:6	0.015
PC aa C38:0	Phosphatidylcholine diacyl C 38:0	0.2
PC aa C38:1	Phosphatidylcholine diacyl C 38:1	0.08
PC aa C38:3	Phosphatidylcholine diacyl C 38:3	0.04
PC aa C38:4	Phosphatidylcholine diacyl C 38:4	0.03
PC aa C38:5	Phosphatidylcholine diacyl C 38:5	0.015
PC aa C38:6	Phosphatidylcholine diacyl C 38:6	0.02
PC aa C40:1	Phosphatidylcholine diacyl C 40:1	0.4
PC aa C40:2	Phosphatidylcholine diacyl C 40:2	0.02

PC aa C40:3	Phosphatidylcholine diacyl C 40:3	0.006
PC aa C40:4	Phosphatidylcholine diacyl C 40:4	0.01
PC aa C40:5	Phosphatidylcholine diacyl C 40:5	0.04
PC aa C40:6	Phosphatidylcholine diacyl C 40:6	1.2
PC aa C42:0	Phosphatidylcholine diacyl C 42:0	0.05
PC aa C42:1	Phosphatidylcholine diacyl C 42:1	0.008
PC aa C42:2	Phosphatidylcholine diacyl C 42:2	0.006
PC aa C42:4	Phosphatidylcholine diacyl C 42:4	0.006
PC aa C42:5	Phosphatidylcholine diacyl C 42:5	0.05
PC aa C42:6	Phosphatidylcholine diacyl C 42:6	0.3
PC ae C30:0	Phosphatidylcholine acyl-alkyl C 30:0	0.15
PC ae C30:1	Phosphatidylcholine acyl-alkyl C 30:1	0.02
PC ae C30:2	Phosphatidylcholine acyl-alkyl C 30:2	0.57
PC ae C32:1	Phosphatidylcholine acyl-alkyl C 32:1	0.009
PC ae C32:2	Phosphatidylcholine acyl-alkyl C 32:2	0.02
PC ae C34:0	Phosphatidylcholine acyl-alkyl C 34:0	0.017
PC ae C34:1	Phosphatidylcholine acyl-alkyl C 34:1	0.012
PC ae C34:2	Phosphatidylcholine acyl-alkyl C 34:2	0.01
PC ae C34:3	Phosphatidylcholine acyl-alkyl C 34:3	0.015
PC ae C36:0	Phosphatidylcholine acyl-alkyl C 36:0	0.12
PC ae C36:1	Phosphatidylcholine acyl-alkyl C 36:1	0.03
PC ae C36:2	Phosphatidylcholine acyl-alkyl C 36:2	0.01
PC ae C36:3	Phosphatidylcholine acyl-alkyl C 36:3	0.007
PC ae C36:4	Phosphatidylcholine acyl-alkyl C 36:4	0.013
PC ae C36:5	Phosphatidylcholine acyl-alkyl C 36:5	0.012
PC ae C38:0	Phosphatidylcholine acyl-alkyl C 38:0	0.066
PC ae C38:1	Phosphatidylcholine acyl-alkyl C 38:1	0.015
PC ae C38:2	Phosphatidylcholine acyl-alkyl C 38:2	0.018
PC ae C38:3	Phosphatidylcholine acyl-alkyl C 38:3	0.01
PC ae C38:4	Phosphatidylcholine acyl-alkyl C 38:4	0.015
PC ae C38:5	Phosphatidylcholine acyl-alkyl C 38:5	0.01
PC ae C38:6	Phosphatidylcholine acyl-alkyl C 38:6	0.03
PC ae C40:0	Phosphatidylcholine acyl-alkyl C 40:0	12
PC ae C40:1	Phosphatidylcholine acyl-alkyl C 40:1	0.06
PC ae C40:2	Phosphatidylcholine acyl-alkyl C 40:2	0.01
PC ae C40:3	Phosphatidylcholine acyl-alkyl C 40:3	0.015
PC ae C40:4	Phosphatidylcholine acyl-alkyl C 40:4	0.1
PC ae C40:5	Phosphatidylcholine acyl-alkyl C 40:5	0.006
PC ae C40:6	Phosphatidylcholine acyl-alkyl C 40:6	0.025
PC ae C42:0	Phosphatidylcholine acyl-alkyl C 42:0	0.4
PC ae C42:1	Phosphatidylcholine acyl-alkyl C 42:1	0.03
PC ae C42:2	Phosphatidylcholine acyl-alkyl C 42:2	0.006
PC ae C42:3	Phosphatidylcholine acyl-alkyl C 42:3	0.006
PC ae C42:4	Phosphatidylcholine acyl-alkyl C 42:4	0.3
PC ae C42:5	Phosphatidylcholine acyl-alkyl C 42:5	1.3
PC ae C44:3	Phosphatidylcholine acyl-alkyl C 44:3	0.006
PC ae C44:4	Phosphatidylcholine acyl-alkyl C 44:4	0.01
PC ae C44:5	Phosphatidylcholine acyl-alkyl C 44:5	0.02
PC ae C44:6	Phosphatidylcholine acyl-alkyl C 44:6	0.09
lysoPC a C14:0	lysoPhosphatidylcholine acyl C14:0	5
lysoPC a C16:0	lysoPhosphatidylcholine acyl C16:0	0.12

lysoPC a C16:1	lysoPhosphatidylcholine acyl C16:1	0.07
lysoPC a C17:0	lysoPhosphatidylcholine acyl C17:0	0.05
lysoPC a C18:0	lysoPhosphatidylcholine acyl C18:0	0.05
lysoPC a C18:1	lysoPhosphatidylcholine acyl C18:1	0.1
lysoPC a C18:2	lysoPhosphatidylcholine acyl C18:2	0.1
lysoPC a C20:3	lysoPhosphatidylcholine acyl C20:3	0.2
lysoPC a C20:4	lysoPhosphatidylcholine acyl C20:4	0.02
lysoPC a C24:0	lysoPhosphatidylcholine acyl C24:0	1.3
lysoPC a C26:0	lysoPhosphatidylcholine acyl C26:0	0.5
lysoPC a C26:1	lysoPhosphatidylcholine acyl C26:1	4
lysoPC a C28:0	lysoPhosphatidylcholine acyl C28:0	0.3
lysoPC a C28:1	lysoPhosphatidylcholine acyl C28:1	0.15
lysoPC a C6:0	lysoPhosphatidylcholine acyl C6:0	0.04
Sphingolipids		
SM (OH) C14:1	Hydroxysphingomyelin C 14:1	0.025
SM (OH) C16:1	Hydroxysphingomyelin C 16:1	0.012
SM (OH) C22:1	Hydroxysphingomyelin C 22:1	0.015
SM (OH) C22:2	Hydroxysphingomyelin C 22:2	0.01
SM (OH) C24:1	Hydroxysphingomyelin C 24:1	0.01
SM C16:0	Sphingomyelin C 16:0	0.03
SM C16:1	Sphingomyelin C 16:1	0.01
SM C18:0	Sphingomyelin C 18:0	0.07
SM C18:1	Sphingomyelin C 18:1	0.01
SM C20:2	Sphingomyelin C 20:2	0.005
SM C22:3	Sphingomyelin C 22:3	0.01
SM C24:0	Sphingomyelin C 24:0	0.13
SM C24:1	Sphingomyelin C 24:1	0.035
SM C26:0	Sphingomyelin C 26:0	0.015
SM C26:1	Sphingomyelin C 26:1	0.006
Sugars		
H1	Hexose	20

*The limit of detection was determined by the manufacturer of the kit (BIOCRATES Life Sciences AG, Innsbruck, Austria)(95).

Table S2: Selection of the final metabolite set based on technical variance* and limit of detection of 163 metabolites measured with the BIOCRATES Absolute IDQ targeted metabolomics technology				
Metabolite	**Within-plate CV (%)**	**Between-plate CV (%)**	**LOD**	
Acylcarnitines				
C0	5.1	1.6		
C10	5.6	11.4		
C10:1	8.9	11.7	below	**Ex**
C10:2	11.4	30.2		
C12	6.8	10.4	below	**Ex**
C12-DC	10.2	13.0	below	**Ex**
C12:1	11.9	7.7	below	**Ex**
C14	8.5	15.3	below	**Ex**
C14:1	7.2	18.7		
C14:1-OH	14.1	30.2	below	**Ex**
C14:2	11.2	17.4		
C14:2-OH	18.4	50.5		**Ex**
C16	7.7	9.0		
C16-OH	26.3	55.0	below	**Ex**
C16:1	9.8	17.5	below	**Ex**
C16:1-OH	18.5	39.7	below	**Ex**
C16:2	22.3	48.0		
C16:2-OH	18.7	44.4	below	**Ex**
C18	13.8	12.8		
C18:1	7.1	3.9		
C18:1-OH	23.2	46.9	below	**Ex**
C18:2	10.0	10.2		
C2	4.5	4.0		
C3	7.4	12.5		
C3-DC / C4-OH	20.2	31.3	below	**Ex**
C3-DC-M / C5-OH	17.1	47.3		
C3-OH	15.1	37.4	below	**Ex**
C3:1	20.7	86.1	below	**Ex**
C4	29.3	94.8		**Ex**
C4:1	8.8	25.2	below	**Ex**
C4:1-DC / C6	16.6	83.3		**Ex**
C5	23.3	95.9		**Ex**
C5-DC / C6-OH	8.4	17.6		
C5-M-DC	10.2	60.4	below	**Ex**
C5:1	14.9	36.7	below	**Ex**
C5:1-DC	15.8	49.1	below	**Ex**
C6:1	17.0	76.4	below	**Ex**
C7-DC	10.7	26.1		
C8	6.2	12.4	below	**Ex**
C8:1	8.5	22.3		
C9	8.5	24.4		
Amino acids				
Arg	6.1	3.4		

Gln	6.7	6.3		
Gly	6.7	4.4		
His	7.0	6.7		
Met	7.2	6.5		
Orn	8.7	4.8		
Phe	6.7	6.2		
Pro	5.3	3.7		
Ser	8.0	15.3		
Thr	8.2	16.6		
Trp	5.8	6.6		
Tyr	8.3	5.4		
Val	7.8	4.3		
Ile	7.5	4.5		
Diacyl-phosphatidylcholines				
PC aa C24:0	35.9	65.2		**Ex**
PC aa C26:0	27.9	56.2	below	**Ex**
PC aa C28:1	5.4	8.2		
PC aa C30:0	5.1	4.3		
PC aa C30:2	69.9	88.6		**Ex**
PC aa C32:0	5.7	6.2		
PC aa C32:1	6.5	13.7		
PC aa C32:2	7.2	6.4		
PC aa C32:3	7.4	6.5		
PC aa C34:1	5.4	9.5		
PC aa C34:2	6.4	15.9		
PC aa C34:3	5.5	9.1		
PC aa C34:4	6.8	6.9		
PC aa C36:0	8.5	15.9		
PC aa C36:1	5.2	2.6		
PC aa C36:2	5.4	11.2		
PC aa C36:3	5.0	7.7		
PC aa C36:4	5.9	7.9		
PC aa C36:5	6.5	6.1		
PC aa C36:6	6.9	6.2		
PC aa C38:0	6.5	11.4		
PC aa C38:1	27.2	25.1		
PC aa C38:3	5.2	3.0		
PC aa C38:4	5.1	4.8		
PC aa C38:5	5.5	6.3		
PC aa C38:6	4.9	5.2		
PC aa C40:1	10.3	6.0	below	**Ex**
PC aa C40:2	12.8	13.4		
PC aa C40:3	8.4	9.5		
PC aa C40:4	5.5	3.5		
PC aa C40:5	4.8	3.5		
PC aa C40:6	4.9	3.5		
PC aa C42:0	6.9	3.7		
PC aa C42:1	9.2	7.8		
PC aa C42:2	10.7	10.5		
PC aa C42:4	11.5	8.2		
PC aa C42:5	8.2	3.9		

PC aa C42:6	6.5	6.3		
Acyl-alkyl-phosphatidylcholines				
PC ae C30:0	9.2	11.6		
PC ae C30:1	23.8	42.4		
PC ae C30:2	18.1	24.8		
PC ae C32:1	7.4	6.1		
PC ae C32:2	6.8	7.5		
PC ae C34:0	7.1	5.5		
PC ae C34:1	5.4	6.6		
PC ae C34:2	5.4	7.6		
PC ae C34:3	5.5	7.0		
PC ae C36:0	13.6	19.0		
PC ae C36:1	5.9	7.3		
PC ae C36:2	6.3	6.0		
PC ae C36:3	7.0	6.2		
PC ae C36:4	5.6	7.7		
PC ae C36:5	5.7	8.4		
PC ae C38:0	6.1	11.4		
PC ae C38:1	14.8	17.6		
PC ae C38:2	9.3	7.1		
PC ae C38:3	5.7	5.4		
PC ae C38:4	5.2	6.5		
PC ae C38:5	5.2	5.8		
PC ae C38:6	5.5	5.2		
PC ae C40:0	3.8	11.5	below	**Ex**
PC ae C40:1	6.8	12.4		
PC ae C40:2	6.4	4.6		
PC ae C40:3	8.9	4.2		
PC ae C40:4	5.8	9.3		
PC ae C40:5	5.4	4.0		
PC ae C40:6	5.1	4.9		
PC ae C42:0	9.1	14.1	below	**Ex**
PC ae C42:1	13.7	15.3		
PC ae C42:2	7.6	5.8		
PC ae C42:3	8.5	6.7		
PC ae C42:4	7.1	2.8		
PC ae C42:5	5.3	1.3		
PC ae C44:3	21.1	12.5		
PC ae C44:4	9.1	4.6		
PC ae C44:5	6.0	2.6		
PC ae C44:6	5.9	1.3		
Lyso-phosphatidylcholines				
lysoPC a C14:0	4.0	12.0		
lysoPC a C16:0	5.0	12.6		
lysoPC a C16:1	6.6	11.7		
lysoPC a C17:0	7.3	13.9		
lysoPC a C18:0	5.3	13.7		
lysoPC a C18:1	5.3	11.1		
lysoPC a C18:2	5.4	8.4		
lysoPC a C20:3	7.9	9.5		
lysoPC a C20:4	7.4	10.0		

lysoPC a C24:0	26.6	43.7	below	**Ex**
lysoPC a C26:0	39.1	65.9	below	**Ex**
lysoPC a C26:1	11.4	14.8	below	**Ex**
lysoPC a C28:0	28.1	45.6	below	**Ex**
lysoPC a C28:1	20.5	34.7		
lysoPC a C6:0	52.3	68.3	below	**Ex**
Sphingomyelins				
SM (OH) C14:1	6.8	14.3		
SM (OH) C16:1	6.8	17.4		
SM (OH) C22:1	6.8	21.9		
SM (OH) C22:2	7.0	19.2		
SM (OH) C24:1	8.3	22.6		
SM C16:0	6.3	16.4		
SM C16:1	8.9	14.3		
SM C18:0	7.2	16.7		
SM C18:1	6.5	15.9		
SM C20:2	21.0	23.0		
SM C22:3	89.8	53.8		**Ex**
SM C24:0	6.8	22.2		
SM C24:1	7.2	23.0		
SM C26:0	20.1	25.9		
SM C26:1	11.4	20.9		
Monosaccharides				
Hexose	4.1	2.0		
Median	**7.3**	**11.4**		

Abbreviations: CV, coefficient of variation; Ex, excluded; LOD, limit of detection.
*Analytical variance was determined by measuring 5 replicates on each of the 46 plates containing the EPIC-samples (total of 230 replicates) and reported as CV%. Note: higher CVs are mainly observed for metabolites that show very low concentrations and are below the LOD of the assay.

Table S3. Selection of metabolites according to their association with type 2 diabetes risk in the EPIC-Potsdam case-cohort sample

Rank*	Metabolite	β-Coefficent (SE)†	*P*-Value†	Corrected *P*-Value‡
1	Hexose	0.86 (0.07)	1.00×10^{-34}	1.28×10^{-32}§
2	PC ae C34:3	-0.45 (0.06)	1.81×10^{-12}	2.28×10^{-10}§
3	PC ae C44:6	-0.40 (0.06)	5.70×10^{-11}	7.13×10^{-09}§
4	PC aa C42:0	-0.36 (0.06)	1.01×10^{-09}	1.25×10^{-07}§
5	PC aa C38:3	0.31 (0.05)	1.13×10^{-09}	1.39×10^{-07}§
6	PC aa C42:1	-0.34 (0.06)	2.69×10^{-09}	3.29×10^{-07}§
7	PC ae C42:5	-0.35 (0.06)	5.83×10^{-09}	7.06×10^{-07}§
8	Phenylalanine	0.30 (0.05)	7.50×10^{-09}	9.01×10^{-07}§
9	PC ae C44:5	-0.35 (0.06)	1.82×10^{-08}	2.16×10^{-06}§
10	PC ae C42:4	-0.34 (0.06)	3.88×10^{-08}	4.58×10^{-06}§
11	Isoleucine	0.26 (0.05)	2.60×10^{-07}	3.04×10^{-05}§
12	Tyrosine	0.27 (0.05)	4.26×10^{-07}	4.94×10^{-05}§
13	PC ae C40:5	-0.28 (0.06)	4.45×10^{-07}	5.12×10^{-05}§
14	Valine	0.24 (0.05)	5.17×10^{-07}	5.89×10^{-05}§
15	Glycine	-0.32 (0.06)	9.62×10^{-07}	1.09×10^{-04}§
16	PC ae C32:2	-0.28 (0.06)	9.62×10^{-07}	1.08×10^{-04}§
17	lysoPC a C18:2	-0.30 (0.06)	9.97×10^{-07}	1.11×10^{-04}§
18	C3	0.25 (0.05)	1.35×10^{-06}	1.49×10^{-04}§
19	PC ae C44:4	-0.28 (0.06)	2.70×10^{-06}	2.94×10^{-04}§
20	PC ae C34:2	-0.26 (0.06)	4.43×10^{-06}	4.79×10^{-04}§
21	PC ae C40:6	-0.26 (0.06)	5.45×10^{-06}	5.83×10^{-04}§
22	SM C16:1	-0.23 (0.05)	1.61×10^{-05}	1.71×10^{-03}§
23	PC ae C32:1	-0.23 (0.05)	2.28×10^{-05}	2.40×10^{-03}§
24	PC ae C42:3	-0.23 (0.06)	3.30×10^{-05}	3.44×10^{-03}§
25	PC aa C40:4	0.19 (0.05)	6.05×10^{-05}	6.24×10^{-03}§
26	SM (OH) C22:2	-0.21 (0.05)	9.50×10^{-05}	9.69×10^{-03}§
27	lysoPC a C17:0	-0.22 (0.06)	1.73×10^{-04}	1.75×10^{-02}§
28	Tryptophan	0.22 (0.06)	1.73×10^{-04}	1.73×10^{-02}§
29	PC ae C36:3	-0.21 (0.06)	2.01×10^{-04}	1.99×10^{-02}§
30	PC aa C36:3	0.19 (0.05)	2.24×10^{-04}	2.20×10^{-02}§
31	PC aa C32:1	0.17 (0.04)	2.56×10^{-04}	2.48×10^{-02}§
32	PC ae C36:2	-0.21 (0.06)	2.87×10^{-04}	2.76×10^{-02}§
33	PC aa C40:5	0.16 (0.05)	3.60×10^{-04}	3.42×10^{-02}§
34	PC aa C36:1	0.17 (0.05)	4.04×10^{-04}	3.79×10^{-02}§
35	Glutamine	-0.16 (0.05)	5.59×10^{-04}	5.20×10^{-02}
36	PC aa C42:2	-0.19 (0.05)	6.32×10^{-04}	5.81×10^{-02}
37	C0	0.17 (0.05)	1.01×10^{-03}	9.23×10^{-02}

38	PC ae C40:4	-0.19 (0.06)	1.04×10^{-03}	9.40×10^{-02}
39	PC aa C38:4	0.15 (0.05)	1.50×10^{-03}	1.33×10^{-01}
40	PC ae C38:6	-0.16 (0.05)	1.50×10^{-03}	1.32×10^{-01}
41	lysoPC a C18:1	-0.18 (0.06)	1.60×10^{-03}	1.39×10^{-01}
42	Arginine	0.17 (0.05)	1.69×10^{-03}	1.46×10^{-01}
43	PC aa C34:1	0.16 (0.05)	2.07×10^{-03}	1.76×10^{-01}
44	PC ae C36:5	-0.16 (0.05)	2.10×10^{-03}	1.77×10^{-01}
45	PC aa C32:0	0.16 (0.05)	2.24×10^{-03}	1.86×10^{-01}
46	PC ae C38:5	-0.16 (0.05)	2.36×10^{-03}	1.94×10^{-01}
47	SM C16:0	-0.15 (0.05)	3.19×10^{-03}	2.58×10^{-01}
48	PC ae C40:3	-0.16 (0.06)	3.48×10^{-03}	2.78×10^{-01}
49	SM (OH) C14:1	-0.16 (0.06)	3.79×10^{-03}	3.00×10^{-01}
50	C5-DC / C6-OH	0.14 (0.05)	4.58×10^{-03}	3.57×10^{-01}
51	Methionine	0.14 (0.05)	4.78×10^{-03}	3.68×10^{-01}
52	PC aa C30:0	0.15 (0.05)	5.48×10^{-03}	4.17×10^{-01}
53	C3-DC-M / C5-OH	0.13 (0.05)	6.59×10^{-03}	4.94×10^{-01}
54	SM C24:1	-0.14 (0.05)	8.01×10^{-03}	5.93×10^{-01}
55	PC ae C40:1	-0.14 (0.05)	8.85×10^{-03}	6.46×10^{-01}
56	PC aa C40:6	0.12 (0.05)	9.31×10^{-03}	6.71×10^{-01}
57	PC aa C38:0	-0.14 (0.05)	9.49×10^{-03}	6.74×10^{-01}
58	lysoPC a C14:0	0.13 (0.05)	1.16×10^{-02}	8.15×10^{-01}
59	PC aa C36:2	0.12 (0.05)	1.43×10^{-02}	9.89×10^{-01}
60	C8:1	0.13 (0.05)	1.48×10^{-02}	1
61	PC ae C38:4	-0.13 (0.05)	1.66×10^{-02}	1
62	PC aa C42:5	0.12 (0.05)	1.67×10^{-02}	1
63	SM C18:1	-0.13 (0.06)	1.79×10^{-02}	1
64	PC aa C36:4	0.13 (0.05)	2.22×10^{-02}	1
65	Ornithine	0.12 (0.05)	2.40×10^{-02}	1
66	PC aa C34:2	0.12 (0.05)	2.44×10^{-02}	1
67	SM (OH) C16:1	-0.12 (0.05)	3.12×10^{-02}	1
68	PC ae C38:0	-0.11 (0.05)	3.29×10^{-02}	1
69	C16	0.11 (0.05)	3.72×10^{-02}	1
70	PC aa C32:3	-0.11 (0.06)	5.06×10^{-02}	1
71	PC ae C30:1	-0.12 (0.06)	5.39×10^{-02}	1
72	C10	0.07 (0.04)	6.13×10^{-02}	1
73	SM C26:0	-0.09 (0.05)	7.23×10^{-02}	1
74	PC aa C36:0	-0.09 (0.05)	7.77×10^{-02}	1
75	SM C26:1	-0.09 (0.05)	8.82×10^{-02}	1
76	PC aa C36:6	-0.08 (0.05)	1.08×10^{-01}	1
77	C16:2	0.08 (0.05)	1.17×10^{-01}	1

78	lysoPC a C16:0	0.08 (0.05)	1.24×10^{-01}	1
79	C9	0.07 (0.05)	1.53×10^{-01}	1
80	SM C20:2	-0.08 (0.06)	1.54×10^{-01}	1
81	PC ae C30:0	-0.08 (0.06)	1.72×10^{-01}	1
82	PC ae C38:2	-0.07 (0.05)	1.79×10^{-01}	1
83	lysoPC a C16:1	0.06 (0.05)	1.80×10^{-01}	1
84	lysoPC a C20:4	-0.07 (0.06)	1.87×10^{-01}	1
85	lysoPC a C20:3	0.07 (0.05)	1.94×10^{-01}	1
86	PC ae C40:2	-0.06 (0.05)	2.31×10^{-01}	1
87	PC aa C40:3	-0.06 (0.05)	2.55×10^{-01}	1
88	PC aa C32:2	0.06 (0.05)	2.60×10^{-01}	1
89	C10:2	0.06 (0.05)	2.61×10^{-01}	1
90	PC ae C30:2	-0.07 (0.06)	2.77×10^{-01}	1
91	PC aa C28:1	-0.06 (0.05)	2.83×10^{-01}	1
92	PC aa C42:6	0.05 (0.05)	2.92×10^{-01}	1
93	PC aa C34:3	0.06 (0.05)	3.10×10^{-01}	1
94	PC ae C34:1	-0.06 (0.06)	3.13×10^{-01}	1
95	PC ae C42:2	-0.05 (0.05)	3.22×10^{-01}	1
96	Threonine	-0.05 (0.06)	3.26×10^{-01}	1
97	PC aa C38:1	-0.05 (0.06)	3.48×10^{-01}	1
98	PC ae C36:4	-0.05 (0.05)	3.50×10^{-01}	1
99	PC ae C44:3	-0.05 (0.05)	3.62×10^{-01}	1
100	Proline	0.04 (0.05)	3.88×10^{-01}	1
101	SM C24:0	0.04 (0.05)	4.03×10^{-01}	1
102	C18:2	-0.04 (0.05)	4.17×10^{-01}	1
103	PC aa C34:4	0.05 (0.06)	4.20×10^{-01}	1
104	PC ae C36:0	0.04 (0.05)	4.41×10^{-01}	1
105	PC aa C42:4	0.04 (0.05)	4.43×10^{-01}	1
106	SM (OH) C24:1	-0.03 (0.05)	4.78×10^{-01}	1
107	PC ae C36:1	-0.04 (0.06)	4.93×10^{-01}	1
108	C14:1	-0.03 (0.05)	5.00×10^{-01}	1
109	C18:1	0.03 (0.05)	5.07×10^{-01}	1
110	PC ae C42:1	-0.03 (0.05)	5.80×10^{-01}	1
111	PC aa C36:5	-0.03 (0.05)	6.08×10^{-01}	1
112	PC ae C38:1	-0.03 (0.05)	6.08×10^{-01}	1
113	PC aa C38:5	0.02 (0.05)	6.20×10^{-01}	1
114	C14:2	0.02 (0.05)	6.64×10^{-01}	1
115	SM (OH) C22:1	0.02 (0.05)	6.90×10^{-01}	1
116	PC ae C38:3	-0.02 (0.06)	6.91×10^{-01}	1
117	Serine	-0.02 (0.06)	6.94×10^{-01}	1

118	SM C18:0	0.02 (0.05)	7.13×10^{-01}	1
119	lysoPC a C18:0	0.02 (0.05)	7.37×10^{-01}	1
120	Histidine	-0.02 (0.06)	7.80×10^{-01}	1
121	PC aa C40:2	-0.02 (0.06)	7.85×10^{-01}	1
122	C7-DC	0.01 (0.05)	8.01×10^{-01}	1
123	C2	0.01 (0.05)	8.32×10^{-01}	1
124	PC ae C34:0	0.01 (0.06)	8.32×10^{-01}	1
125	C18	0.01 (0.05)	8.67×10^{-01}	1
126	PC aa C38:6	0.00 (0.05)	9.58×10^{-01}	1
127	lysoPC a C28:1	0.00 (0.06)	9.67×10^{-01}	1

*Ranking from lowest to highest uncorrected *p*-value.
†β-coefficients and *P*-values were derived from multivariate Cox regression analysis modified using the prentice method to account for the case-cohort design in a continuous model after standardizing metabolite concentrations to mean of zero and standard deviation of one. The model was adjusted for age, sex, alcohol intake from beverages (non-consumers; women: >0-6 g/d, 6-12 g/d, >12 g/d; men: >0-12 g/d, 12-24 g/d, >24 g/d), smoking (never, former, current ≤20 cigarettes/d, current >20 cigarettes/d), physical activity (cycling and sports in h/week), education (no degree/vocational training; trade/technical school; university degree), coffee intake (cups/d), red meat intake (g/d), whole grain bread intake (g/d), prevalent hypertension (yes/no), BMI (kg/m²), and waist circumference (cm).
‡*P*-values were corrected to account for multiple testing (n=127) using the Bonferroni-Holm procedure (103).
§Corrected *p*-value <0.05.

Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; DC, decarboxyl; PC, phosphatidylcholine; SE, standard error; SM, sphingomyelin.

Table S4. Possible impact of reliability of metabolite s on the association between metabolites and risk of type 2 diabetes in the EPIC-Potsdam case-cohort study

	Observed RR per SD*	Observed Rank	Expected RR per SD†	Expected Rank	|Δ Expected-Observed Rank|
Hexose	2.36	1	3.10	1	0
PC aa C38:3	1.37	2	1.83	3	1
Phenylalanine	1.35	3	1.89	2	1
PC aa C32:1	1.18	4	1.27	4	0
PC aa C36:1	1.18	5	1.35	5	0
PC aa C40:5	1.18	6	1.31	6	0
SM C16:1	0.80	7	0.74	7	0
PC ae C40:6	0.77	8	0.68	8	0
PC ae C44:4	0.76	9	0.68	9	0
LysoPC a C18:2	0.74	10	0.60	13	3
Glycine	0.73	11	0.63	11	0
PC ae C42:5	0.71	12	0.64	10	2
PC ae C44:5	0.71	13	0.63	12	1
PC ae C34:3	0.64	14	0.51	14	0

*Basic model as in Table 9.

†Expected relative risk was additionally calculated to account for reliability of metabolites, based on the observed relative risk and intraclass-correlation coefficient for the individual metabolites (15).

Abbreviations: aa, diacyl; ae, acyl-alkyl; PC, phosphatidylcholine; SM, sphingomyelin.

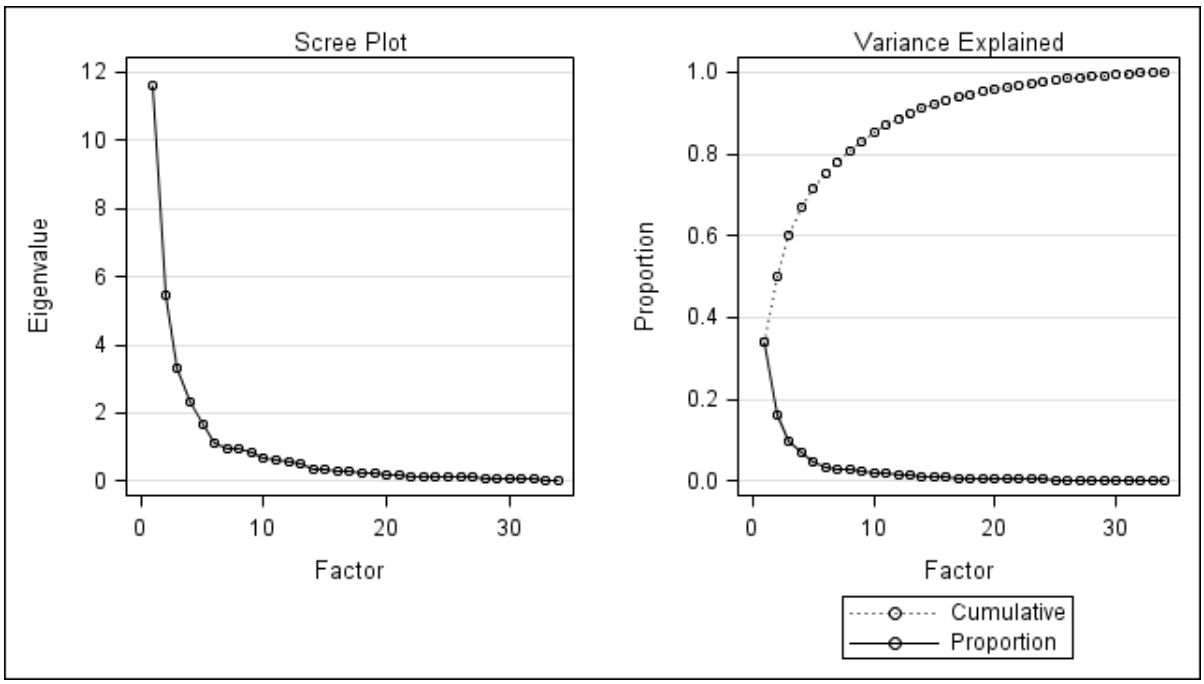


Figure S1: Scree plot and proportion of variance explained of all factors derived from principal component analysis.

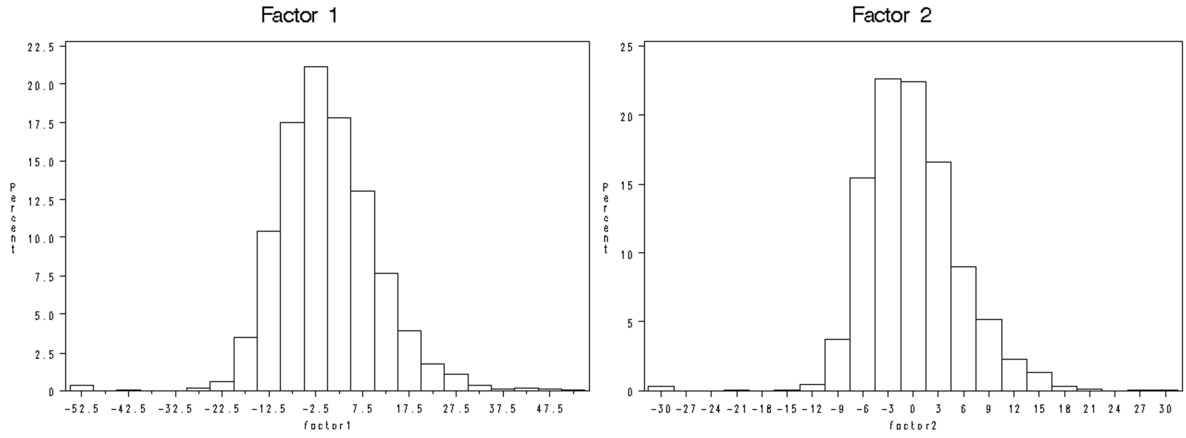


Figure S2. Histogram of two metabolite factors associated with risk of type 2 diabetes in the EPIC-Potsdam case-cohort sample.

Table S5: Selection of metabolites according to their association with stroke risk in the EPIC-Potsdam case-cohort sample

Rank*	Metabolite	β-Coefficent (SE)†	*P*-value†	Corrected *p*-value‡
1	PC aa C42:0	-0.24 (0.09)	8.43×10^{-03}	1
2	PC aa C36:3	0.17 (0.08)	3.69×10^{-02}	1
3	PC ae C44:6	-0.19 (0.09)	4.55×10^{-02}	1
4	C8:1	0.16 (0.08)	5.57×10^{-02}	1
5	PC aa C34:1	0.15 (0.08)	6.14×10^{-02}	1
6	PC aa C34:3	0.13 (0.07)	6.55×10^{-02}	1
7	PC ae C36:3	0.17 (0.10)	7.79×10^{-02}	1
8	Isoleucine	0.14 (0.08)	7.96×10^{-02}	1
9	PC ae C40:6	-0.15 (0.09)	8.05×10^{-02}	1
10	PC ae C42:5	-0.16 (0.09)	8.39×10^{-02}	1
11	Valine	0.15 (0.09)	9.17×10^{-02}	1
12	Glycine	-0.17 (0.10)	9.90×10^{-02}	1
13	Proline	-0.14 (0.09)	1.18×10^{-01}	1
14	PC ae C44:5	-0.14 (0.09)	1.24×10^{-01}	1
15	PC aa C42:1	-0.14 (0.09)	1.35×10^{-01}	1
16	PC aa C34:4	0.13 (0.09)	1.38×10^{-01}	1
17	PC aa C36:4	0.13 (0.09)	1.42×10^{-01}	1
18	C16:2	0.11 (0.07)	1.42×10^{-01}	1
19	lysoPC a C18:2	-0.15 (0.10)	1.45×10^{-01}	1
20	PC ae C34:2	0.14 (0.10)	1.46×10^{-01}	1
21	lysoPC a C17:0	-0.12 (0.09)	1.82×10^{-01}	1
22	Glutamine	-0.10 (0.07)	1.87×10^{-01}	1
23	C14:1	0.10 (0.08)	1.89×10^{-01}	1
24	PC aa C36:0	-0.11 (0.08)	1.90×10^{-01}	1
25	lysoPC a C16:1	0.08 (0.06)	1.94×10^{-01}	1
26	PC aa C32:1	0.07 (0.06)	2.00×10^{-01}	1
27	PC aa C32:2	0.09 (0.07)	2.12×10^{-01}	1
28	SM (OH) C16:1	-0.10 (0.08)	2.14×10^{-01}	1
29	PC ae C36:4	0.09 (0.08)	2.46×10^{-01}	1
30	PC aa C42:2	-0.10 (0.09)	2.52×10^{-01}	1
31	PC ae C42:4	-0.11 (0.09)	2.58×10^{-01}	1
32	PC ae C42:3	-0.09 (0.08)	2.60×10^{-01}	1
33	PC aa C38:3	0.10 (0.09)	2.70×10^{-01}	1
34	Ornitine	-0.09 (0.08)	2.73×10^{-01}	1
35	lysoPC a C20:3	0.09 (0.08)	2.85×10^{-01}	1
36	Threonine	-0.10 (0.09)	2.87×10^{-01}	1
37	Hexose	0.07 (0.07)	2.87×10^{-01}	1
38	PC aa C40:6	-0.08 (0.08)	3.05×10^{-01}	1
39	PC ae C40:2	-0.08 (0.08)	3.09×10^{-01}	1
40	PC aa C32:3	0.08 (0.08)	3.35×10^{-01}	1
41	Serine	-0.07 (0.08)	3.49×10^{-01}	1
42	PC aa C34:2	0.09 (0.09)	3.50×10^{-01}	1
43	PC aa C36:1	0.08 (0.08)	3.51×10^{-01}	1
44	C3-DC-M/C5-OH	0.08 (0.08)	3.51×10^{-01}	1

45	PC ae C44:4	-0.09 (0.10)	3.54×10^{-01}	1
46	PC aa C38:1	-0.08 (0.08)	3.60×10^{-01}	1
47	PC aa C38:0	-0.07 (0.08)	3.65×10^{-01}	1
48	PC ae C32:2	-0.07 (0.08)	3.67×10^{-01}	1
49	SM (OH) C14:1	-0.08 (0.09)	3.93×10^{-01}	1
50	PC ae C42:1	0.07 (0.08)	3.97×10^{-01}	1
51	C10:2	0.07 (0.08)	3.98×10^{-01}	1
52	PC ae C40:1	-0.07 (0.08)	4.00×10^{-01}	1
53	C18	-0.08 (0.09)	4.07×10^{-01}	1
54	SM (OH) C22:2	-0.07 (0.09)	4.07×10^{-01}	1
55	PC ae C40:4	-0.07 (0.09)	4.13×10^{-01}	1
56	SM C26:0	-0.07 (0.08)	4.33×10^{-01}	1
57	C16	-0.07 (0.09)	4.39×10^{-01}	1
58	PC ae C40:5	-0.07 (0.09)	4.43×10^{-01}	1
59	PC ae C36:1	0.07 (0.09)	4.45×10^{-01}	1
60	PC ae C44:3	-0.06 (0.08)	4.57×10^{-01}	1
61	SM C26:1	-0.06 (0.08)	4.62×10^{-01}	1
62	lysoPC a C18:0	-0.06 (0.09)	4.75×10^{-01}	1
63	PC ae C38:3	0.07 (0.09)	4.79×10^{-01}	1
64	Histidine	-0.06 (0.08)	4.82×10^{-01}	1
65	Methionine	-0.06 (0.08)	4.87×10^{-01}	1
66	Tryptophan	-0.05 (0.08)	4.96×10^{-01}	1
67	PC ae C30:0	-0.06 (0.09)	5.00×10^{-01}	1
68	PC aa C42:4	0.05 (0.08)	5.01×10^{-01}	1
69	PC ae C36:0	0.05 (0.07)	5.06×10^{-01}	1
70	PC aa C40:5	-0.05 (0.08)	5.07×10^{-01}	1
71	lysoPC a C16:0	0.05 (0.08)	5.10×10^{-01}	1
72	C0	-0.05 (0.09)	5.31×10^{-01}	1
73	PC ae C38:6	-0.05 (0.08)	5.39×10^{-01}	1
74	PC ae C34:1	0.05 (0.09)	5.53×10^{-01}	1
75	PC ae C32:1	-0.05 (0.08)	5.62×10^{-01}	1
76	PC aa C40:4	0.04 (0.08)	5.64×10^{-01}	1
77	C10	-0.07 (0.12)	5.76×10^{-01}	1
78	SM (OH) C24:1	-0.05 (0.08)	5.76×10^{-01}	1
79	PC aa C38:6	-0.04 (0.08)	5.79×10^{-01}	1
80	PC ae C34:3	0.05 (0.09)	5.88×10^{-01}	1
81	Tyrosine	0.04 (0.08)	5.99×10^{-01}	1
82	PC ae C42:2	-0.04 (0.08)	6.01×10^{-01}	1
83	PC ae C40:3	-0.04 (0.08)	6.07×10^{-01}	1
84	PC aa C38:4	0.04 (0.09)	6.20×10^{-01}	1
85	PC aa C32:0	0.03 (0.07)	6.24×10^{-01}	1
86	SM C16:1	0.04 (0.09)	6.42×10^{-01}	1
87	PC aa C30:0	0.03 (0.07)	6.48×10^{-01}	1
88	PC ae C38:0	-0.04 (0.08)	6.50×10^{-01}	1
89	SM C24:1	-0.03 (0.08)	6.54×10^{-01}	1
90	SM C24:0	0.03 (0.08)	6.94×10^{-01}	1

91	PC aa C40:2	-0.03 (0.08)	6.99×10^{-01}	1
92	C7-DC	-0.03 (0.09)	7.21×10^{-01}	1
93	PC aa C28:1	0.03 (0.08)	7.23×10^{-01}	1
94	PC aa C42:6	-0.02 (0.08)	7.81×10^{-01}	1
95	Arginine	-0.02 (0.08)	7.88×10^{-01}	1
96	PC aa C40:3	-0.02 (0.07)	7.89×10^{-01}	1
97	PC aa C42:5	0.02 (0.07)	7.93×10^{-01}	1
98	C18:1	0.02 (0.09)	8.09×10^{-01}	1
99	PC aa C36:2	0.02 (0.08)	8.34×10^{-01}	1
100	Phenylalanine	0.02 (0.07)	8.34×10^{-01}	1
101	PC ae C30:2	-0.02 (0.08)	8.37×10^{-01}	1
102	PC ae C36:5	-0.02 (0.08)	8.39×10^{-01}	1
103	PC ae C30:1	-0.02 (0.10)	8.42×10^{-01}	1
104	PC aa C36:6	-0.02 (0.09)	8.50×10^{-01}	1
105	SM C18:1	0.02 (0.09)	8.51×10^{-01}	1
106	PC ae C36:2	-0.02 (0.09)	8.55×10^{-01}	1
107	C18:2	-0.02 (0.09)	8.57×10^{-01}	1
108	PC aa C38:5	-0.01 (0.08)	8.57×10^{-01}	1
109	lysoPC a C28:1	0.01 (0.09)	8.63×10^{-01}	1
110	lysoPC a C14:0	0.01 (0.08)	8.83×10^{-01}	1
111	PC ae C38:1	0.01 (0.08)	8.88×10^{-01}	1
112	SM C18:0	-0.01 (0.09)	9.11×10^{-01}	1
113	PC ae C38:5	-0.01 (0.08)	9.27×10^{-01}	1
114	PC ae C38:2	0.01 (0.08)	9.35×10^{-01}	1
115	lysoPC a C18:1	-0.01 (0.09)	9.39×10^{-01}	1
116	C2	0.01 (0.08)	9.39×10^{-01}	1
117	PC ae C34:0	-0.01 (0.09)	9.43×10^{-01}	1
118	C9	0.01 (0.09)	9.49×10^{-01}	1
119	C3	0.00 (0.08)	9.54×10^{-01}	1
120	SM C20:2	0.00 (0.07)	9.55×10^{-01}	1
121	C14:2	0.00 (0.09)	9.69×10^{-01}	1
122	C5-DC/C6-OH	0.00 (0.09)	9.70×10^{-01}	1
123	SM (OH) C22:1	0.00 (0.09)	9.72×10^{-01}	1
124	PC aa C36:5	0.00 (0.08)	9.72×10^{-01}	1
125	lysoPC a C20:4	0.00 (0.09)	9.75×10^{-01}	1
126	PC ae C38:4	0.00 (0.09)	9.78×10^{-01}	1
127	SM C16:0	0.00 (0.08)	9.78×10^{-01}	1

*Ranking from lowest to highest uncorrected *p*-value.

†β-coefficients and *P*-values were derived from multivariate Cox regression analysis modified using the prentice method to account for the case-cohort design in a continuous model after standardizing metabolite concentrations to mean of zero and standard deviation of one. The model was adjusted for age, sex, alcohol intake from beverages (non-consumers; women: >0-6 g/d, 6-12 g/d, >12 g/d; men: >0-12 g/d, 12-24 g/d, >24 g/d), smoking (never, former, current ≤20 cigarettes/d, current >20 cigarettes/d), physical activity (cycling and sports in h/week), education (no degree/vocational training; trade/technical school; university degree), fasting status, prevalent hypertension, prevalent type 2 diabetes, BMI (kg/m²), and waist circumference (cm).

‡*P*-values were corrected to account for multiple testing (n=127) using the Bonferroni-Holm procedure (103).

§Corrected *p*-value <0.05.

Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; DC, decarboxyl; PC, phosphatidylcholine; SE, standard error; SM, sphingomyelin.

Table S6: Selection of metabolites according to their association with risk of myocardial infarction in the EPIC-Potsdam case-cohort sample

Rank*	Metabolite	β-Coefficent (SE)†	*P*-value†	Corrected *p*-value‡
1	SM (OH) C22:1	0.37 (0.08)	3.98×10^{-06}	5.06×10^{-04}§
2	SM C24:0	0.35 (0.08)	4.74×10^{-06}	5.97×10^{-04}§
3	SM (OH) C24:1	0.33 (0.07)	6.48×10^{-06}	8.10×10^{-04}§
4	PC aa C38:3	0.31 (0.08)	3.91×10^{-05}	4.84×10^{-03}§
5	SM C18:0	0.31 (0.08)	5.40×10^{-05}	6.64×10^{-03}§
6	PC ae C38:3	0.36 (0.09)	5.43×10^{-05}	6.64×10^{-03}§
7	PC aa C40:4	0.26 (0.07)	7.46×10^{-05}	9.03×10^{-03}§
8	SM (OH) C14:1	0.33 (0.08)	7.78×10^{-05}	9.33×10^{-03}§
9	SM C16:0	0.30 (0.08)	2.65×10^{-04}	3.16×10^{-02}§
10	PC aa C40:5	0.22 (0.06)	3.16×10^{-04}	3.73×10^{-02}§
11	PC ae C36:1	0.32 (0.09)	3.35×10^{-04}	3.92×10^{-02}§
12	SM (OH) C16:1	0.29 (0.08)	4.39×10^{-04}	5.10×10^{-02}
13	PC aa C28:1	0.29 (0.08)	5.92×10^{-04}	6.80×10^{-02}
14	PC aa C36:2	0.25 (0.08)	1.14×10^{-03}	1.29×10^{-01}
15	PC ae C40:3	0.28 (0.09)	1.20×10^{-03}	1.35×10^{-01}
16	PC aa C36:1	0.20 (0.06)	1.46×10^{-03}	1.64×10^{-01}
17	C10	0.13 (0.04)	1.49×10^{-03}	1.65×10^{-01}
18	SM C16:1	0.27 (0.09)	1.78×10^{-03}	1.95×10^{-01}
19	SM (OH) C22:2	0.25 (0.08)	3.12×10^{-03}	3.40×10^{-01}
20	SM C26:1	0.20 (0.07)	5.58×10^{-03}	6.02×10^{-01}
21	PC ae C40:2	0.21 (0.08)	6.81×10^{-03}	7.29×10^{-01}
22	PC ae C38:4	0.21 (0.08)	7.14×10^{-03}	7.57×10^{-01}
23	PC ae C38:2	0.22 (0.08)	7.31×10^{-03}	7.68×10^{-01}
24	PC ae C40:5	0.18 (0.07)	7.47×10^{-03}	7.77×10^{-01}
25	PC aa C36:3	0.20 (0.08)	7.98×10^{-03}	8.22×10^{-01}
26	PC aa C38:4	0.19 (0.07)	8.31×10^{-03}	8.48×10^{-01}
27	PC ae C40:4	0.21 (0.08)	9.97×10^{-03}	1
28	SM C26:0	0.18 (0.07)	1.32×10^{-02}	1
29	PC ae C36:2	0.20 (0.08)	1.58×10^{-02}	1
30	C9	0.16 (0.07)	1.73×10^{-02}	1
31	PC ae C38:1	0.19 (0.08)	1.84×10^{-02}	1
32	PC aa C30:0	0.15 (0.06)	2.08×10^{-02}	1
33	PC ae C42:4	0.20 (0.09)	2.09×10^{-02}	1
34	PC aa C34:2	0.19 (0.08)	2.14×10^{-02}	1
35	SM C18:1	0.18 (0.08)	2.24×10^{-02}	1
36	PC ae C34:0	0.19 (0.08)	2.25×10^{-02}	1
37	PC aa C40:6	0.15 (0.07)	2.53×10^{-02}	1
38	Arginine	0.16 (0.07)	2.73×10^{-02}	1
39	PC ae 34:1	0.19 (0.09)	2.98×10^{-02}	1
40	PC aa C42:4	0.15 (0.07)	4.12×10^{-02}	1
41	PC ae C36:3	0.17 (0.08)	4.15×10^{-02}	1
42	C18	0.15 (0.08)	4.26×10^{-02}	1
43	PC aa C34:1	0.13 (0.07)	5.41×10^{-02}	1
44	C14:1	0.13 (0.07)	5.79×10^{-02}	1

45	Tyrosine	0.16 (0.08)	5.96×10^{-02}	1
46	PC aa C32:0	0.12 (0.06)	6.06×10^{-02}	1
47	PC aa C32:3	0.14 (0.08)	6.26×10^{-02}	1
48	C14:2	0.12 (0.07)	6.88×10^{-02}	1
49	PC ae C30:0	0.15 (0.08)	7.19×10^{-02}	1
50	SM C20:2	0.11 (0.06)	8.59×10^{-02}	1
51	SM C24:1	0.13 (0.07)	8.69×10^{-02}	1
52	lysoPC a C18:2	-0.14 (0.08)	9.04×10^{-02}	1
53	PC ae C42:3	0.13 (0.08)	9.18×10^{-02}	1
54	PC ae C34:2	0.14 (0.08)	9.44×10^{-02}	1
55	PC ae C42:2	0.12 (0.08)	9.84×10^{-02}	1
56	lysoPC a C20:4	-0.12 (0.07)	1.06×10^{-01}	1
57	lysoPC a C18:1	-0.12 (0.08)	1.08×10^{-01}	1
58	PC ae C42:5	0.14 (0.09)	1.10×10^{-01}	1
59	PC aa C32:2	0.10 (0.07)	1.23×10^{-01}	1
60	PC ae C32:1	0.12 (0.08)	1.36×10^{-01}	1
61	PC ae C40:6	0.12 (0.08)	1.39×10^{-01}	1
62	PC ae C36:4	0.10 (0.07)	1.45×10^{-01}	1
63	PC aa C34:3	0.10 (0.07)	1.47×10^{-01}	1
64	lysoPC a C16:1	-0.09 (0.07)	1.90×10^{-01}	1
65	C16	0.09 (0.07)	2.08×10^{-01}	1
66	Glutamine	-0.09 (0.07)	2.10×10^{-01}	1
67	Serine	-0.10 (0.08)	2.31×10^{-01}	1
68	PC ae C44:4	0.10 (0.09)	2.33×10^{-01}	1
69	PC aa C42:5	0.07 (0.06)	2.36×10^{-01}	1
70	Phenylalanine	0.09 (0.07)	2.38×10^{-01}	1
71	Glycine	-0.11 (0.09)	2.48×10^{-01}	1
72	Hexose	0.08 (0.07)	2.57×10^{-01}	1
73	Valine	0.09 (0.08)	2.63×10^{-01}	1
74	PC ae C30:1	0.08 (0.08)	2.94×10^{-01}	1
75	PC aa C38:5	0.07 (0.07)	3.12×10^{-01}	1
76	PC aa C36:4	0.07 (0.08)	3.38×10^{-01}	1
77	PC aa C36:0	-0.07 (0.07)	3.39×10^{-01}	1
78	Proline	-0.07 (0.07)	3.41×10^{-01}	1
79	Ornithine	0.07 (0.08)	3.58×10^{-01}	1
80	Isoleucine	0.07 (0.08)	3.83×10^{-01}	1
81	PC ae C30:2	0.07 (0.08)	3.90×10^{-01}	1
82	PC ae C44:6	0.07 (0.09)	3.95×10^{-01}	1
83	PC ae C38:5	0.06 (0.07)	3.96×10^{-01}	1
84	C16:2	0.06 (0.07)	4.00×10^{-01}	1
85	C2	-0.06 (0.08)	4.11×10^{-01}	1
86	PC aa C36:5	-0.06 (0.08)	4.33×10^{-01}	1
87	PC ae C42:1	0.05 (0.07)	4.52×10^{-01}	1
88	C10:2	0.05 (0.07)	4.54×10^{-01}	1
89	PC aa C42:2	-0.06 (0.08)	4.57×10^{-01}	1
90	PC aa C38:6	0.05 (0.07)	4.67×10^{-01}	1

91	C8:1	0.05 (0.07)	4.71×10^{-01}	1
92	lysoPC a C28:1	0.05 (0.08)	4.92×10^{-01}	1
93	C5-DC/C6-OH	0.05 (0.07)	4.93×10^{-01}	1
94	PC aa C34:4	0.05 (0.08)	5.01×10^{-01}	1
95	C18:1	0.04 (0.08)	6.10×10^{-01}	1
96	lysoPC a C17:0	0.04 (0.08)	6.19×10^{-01}	1
97	PC aa C42:0	0.04 (0.09)	6.20×10^{-01}	1
98	PC ae C38:0	-0.03 (0.07)	6.26×10^{-01}	1
99	PC aa C42:6	0.03 (0.07)	6.32×10^{-01}	1
100	PC aa C38:1	0.04 (0.08)	6.32×10^{-01}	1
101	Threonine	-0.04 (0.09)	6.33×10^{-01}	1
102	PC aa C42:1	0.04 (0.09)	6.49×10^{-01}	1
103	PC ae C44:5	0.04 (0.08)	6.54×10^{-01}	1
104	lysoPC a C16:0	-0.03 (0.07)	6.56×10^{-01}	1
105	Tryptophan	-0.03 (0.07)	6.61×10^{-01}	1
106	PC ae C38:6	0.03 (0.08)	6.75×10^{-01}	1
107	PC aa C40:2	0.03 (0.07)	6.81×10^{-01}	1
108	lysoPC a C18:0	0.03 (0.07)	7.18×10^{-01}	1
109	lysoPC a C20:3	0.02 (0.07)	7.25×10^{-01}	1
110	Histidine	-0.02 (0.08)	7.58×10^{-01}	1
111	PC aa C36:6	-0.02 (0.07)	7.67×10^{-01}	1
112	PC ae C36:0	0.02 (0.07)	7.67×10^{-01}	1
113	C18:2	0.02 (0.08)	7.68×10^{-01}	1
114	PC ae C34:3	-0.02 (0.08)	7.83×10^{-01}	1
115	PC ae C32:2	0.02 (0.08)	7.96×10^{-01}	1
116	PC ae C44:3	0.02 (0.08)	8.17×10^{-01}	1
117	PC ae C36:5	0.01 (0.07)	8.24×10^{-01}	1
118	Methionine	-0.02 (0.08)	8.27×10^{-01}	1
119	C0	0.01 (0.08)	8.67×10^{-01}	1
120	lysoPC a C14:0	0.01 (0.07)	8.83×10^{-01}	1
121	PC aa C40:3	-0.01 (0.07)	9.14×10^{-01}	1
122	C5-OH/C3-DC-M	0.01 (0.07)	9.16×10^{-01}	1
123	C3	-0.01 (0.08)	9.17×10^{-01}	1
124	PC aa C32:1	0.00 (0.05)	9.42×10^{-01}	1
125	C7-DC	0.00 (0.07)	9.44×10^{-01}	1
126	PC ae C40:1	0.00 (0.07)	9.59×10^{-01}	1
127	PC aa C38:0	0.00 (0.08)	9.74×10^{-01}	1

*Ranking from lowest to highest uncorrected *p*-value.

†β-coefficients and *P*-values were derived from multivariate Cox regression analysis modified using the prentice method to account for the case-cohort design in a continuous model after standardizing metabolite concentrations to mean of zero and standard deviation of one. The model was adjusted for age, sex, alcohol intake from beverages (non-consumers; women: >0-6 g/d, 6-12 g/d, >12 g/d; men: >0-12 g/d, 12-24 g/d, >24 g/d), smoking (never, former, current ≤20 cigarettes/d, current >20 cigarettes/d), physical activity (cycling and sports in h/week), education (no degree/vocational training; trade/technical school; university degree), fasting status, prevalent hypertension, prevalent type 2 diabetes, BMI (kg/m²), and waist circumference (cm).

‡*P*-values were corrected to account for multiple testing (n=127) using the Bonferroni-Holm procedure (103).

§Corrected *p*-value <0.05.

Abbreviations: a, acyl; aa, diacyl; ae, acyl-alkyl; DC, decarboxyl; PC, phosphatidylcholine; SE, standard error; SM, sphingomyelin.

Table S7. Possible impact of reliability of metabolite measurements on the association between metabolites and risk of cardiovascular diseases in the EPIC-Potsdam case-cohort study

	Observed RR per SD*	Observed Rank	Expected RR per SD†	Expected Rank	|Δ Expected-Observed Rank|
SM (OH) C22:1	1.45	1	1.70	4	3
PC ae C38:3	1.43	2	1.76	3	1
SM C24:0	1.41	3	1.83	1	2
SM (OH) C14:1	1.4	4	1.49	10	6
SM (OH) C24:1	1.39	5	1.69	5	0
PC ae C36:1	1.38	6	1.64	6	0
PC aa C38:3	1.37	7	1.83	2	5
SM C18:0	1.36	8	1.56	9	1
SM C16:0	1.35	9	1.58	7	2
PC aa C40:4	1.30	10	1.58	8	2
PC aa C40:5	1.25	11	1.43	11	0

*Basic model as in Table 17.
†Expected relative risk was additionally calculated to account for reliability of metabolites, based on the observed relative risk and intraclass-correlation coefficient for the individual metabolites (15).
Abbreviations: aa, diacyl; ae, acyl-alkyl; PC, phosphatidylcholine; SM, sphingomyelin.

Table S8: Relative contribution of individual risk factors to predict myocardial infarction in the EPIC-Potsdam case-cohort sample	
Risk factor	**Area under the ROC curve (95% CI)**
Sex	0.674 (0.643-0.704)
Age	0.687 (0.655-0.718)
Smoking	0.664 (0.628-0.700)
Alcohol from beverages	0.557 (0.521-0.594)
Physical activity	0.585 (0.548-0.622)
Education	0.518 (0.482-0.553)
BMI & waist circumference	0.708 (0.673-0.742)
Prevalent hypertension	0.613 (0.582-0.644)
Prevalent diabetes	0.545 (0.532-0.568)
C-reactive protein	0.616 (0.580-0.652)
HDL-cholesterol	0.658 (0.621-0.696)
Total cholesterol	0.622 (0.019-0.585)
C-reactive protein & HDL-cholesterol & total cholesterol	0.734 (0.700-0.767)

Abbreviations: CI, confidence interval; HDL, high-density lipoprotein; ROC, receiver operating characteristic.

Table S9. Metabolites and their corresponding factor loadings as obtained from principal component analysis (PCA) comparing random and fasting blood sampling in the EPIC-Potsdam case-cohort sample

	Factor Loading			
	Factor 1†		Factor 2†	
	Random‡	Fasting‡	Random‡	Fasting‡
Propionyl-L-Carnitine	-13	-16	55*	46*
Glycine	33	42*	-1	0
Phenylalanine	8	3	66*	67*
Tryptophane	18	21	61*	62*
Tyrosine	-5	-10	66*	67*
Valine	-9	-18	68*	67*
Isoleucine	-9	-20	66*	69*
PC aa C32:1	5	9	59*	64*
PC aa C36:1	28	31	70*	71*
PC aa C36:3	35	39*	65*	66*
PC aa C38:3	15	17	76*	79*
PC aa C40:4	18	23	72*	74*
PC aa C40:5	16	19	71*	75*
PC aa C42:0	82*	81*	-14	-14
PC aa C42:1	79*	79*	-4	-3
PC ae C32:1	80*	83*	22	19
PC ae C32:2	78*	83*	19	14
PC ae C34:2	70*	76*	9	5
PC ae C34:3	72*	76*	1	3
PC ae C36:2	71*	78*	15	8
PC ae C36:3	71*	76*	20	17
PC ae C40:5	85*	87*	15	18
PC ae C40:6	76*	76*	11	12
PC ae C42:3	82*	82*	14	14
PC ae C42:4	85*	85*	-7	-4
PC ae C42:5	87*	88*	-7	-5
PC ae C44:4	76*	78*	-1	-1
PC ae C44:5	78*	77*	-7	-4
PC ae C44:6	83*	82*	-14	-12
LysoPC a C17:0	41*	48*	9	5
LysoPC a C18:2	29	32	6	7
SM C16:1	54*	57*	26	32
SM OH C22:2	57*	61*	10	13
Hexose	-5	2	44*	62*

*factor loading>0.35.

†Metabolite factors were identified by PCA, based on the correlation matrix, of all metabolites significantly associated with risk of type 2 diabetes in EPIC-Potsdam. An orthogonal varimax rotation was used, and two factors were retained as they accounted for more than 50% of the observed variance.

‡Random blood: n= 2997, of them fasting blood: n=429

Danksagung

Die vorliegende Arbeit wurde in der Abteilung Epidemiologie am Deutschen Institut für Ernährungsforschung Potsdam-Rehbrücke erstellt. In diesem Zusammenhang möchte ich allen Personen danken, die mich auf diesem Weg unterstützt haben.

Besonderer Dank gilt meinen beiden Betreuern Prof. Dr. Heiner Boeing und Prof. Dr. Tobias Pischon, die mich in den letzten Jahren in meiner wissenschaftlichen Entwicklung begleitet, motiviert und unterstützt haben. Außerdem möchte ich mich bei Prof. Dr. Rainhard Busse für die Möglichkeit der Promotion an der TU-Berlin und das konstruktive Gespräch bedanken.

Vielen Dank an meine lieben Kolleginnen, insbesondere Anne von Rüsten, Angelika Wientzek, Janine Wirth, Silke Feller und Jana Förster, die stets ein offenes Ohr für mich hatten in wissenschaftlichen wie in privaten Angelegenheiten. Ich danke auch Wolfgang Bernigau und Sven Knüppel für die Bereitschaft, Ihre statistische Expertise mit mir zu teilen.

Ein besonderer Dank gilt meiner Familie, insbesondere meinen Eltern und meinen Geschwistern, die mir auf meinem bisherigen Lebensweg stets die bestmögliche Unterstützung geboten haben und damit maßgeblich zum Gelingen dieser Arbeit beitrugen. Ich möchte meinem Mann Martin herzlich danken, für seinen liebevollen Support und nicht zuletzt dafür, dass er gerne die Mathematik mit mir diskutiert und die Arbeit Korrektur gelesen hat.

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die im Fachbereich Management im Gesundheitswesen der Technischen Universität Berlin eingereichte Dissertation mit dem Titel *„Serum metabolites and their association with risk of type 2 diabetes and cardiovascular diseases: a targeted metabolomics approach in EPIC-Potsdam“* selbstständig angefertigt und verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt wurden. Die geltende Promotionsordnung der Technischen Universität Berlin (Promotionsordnung Dr. P.H. vom 16.03. 1999) ist mir bekannt. Teile der Arbeit sind im Rahmen des Promotionsvorhabens bereits veröffentlicht worden und als solche gekennzeichnet.* Weiterhin versichere ich, die Arbeit an keiner anderen Hochschule oder Fachhochschule eingereicht zu haben.

Anna Flögel **Berlin, den 19.12.2012**

*Relevante Publikationen (Originalarbeiten):

1.) Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, Joost HG, Boeing H, Pischon T: *Reliability of Serum Metabolite Concentrations over a 4-Month Period Using a Targeted Metabolomic Approach.* PLoS One 2011;6:e21103

2.) Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, Fritsche A, Haring HU, Hrabe de Angelis M, Peters A, Roden M, Prehn C, Wang-Sattler R, Illig T, Schulze MB, Adamski J, Boeing H, Pischon T: *Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach*. Diabetes 2012; Oct.4th (Epub ahead, print scheduled for February 2013)