Metabolomics analyses to better understand complex phenotypes

ZHONGHAO YU

München 2013

Aus dem HelmholtzZentrum München
Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH)
Institut für Epidemiologie II
Leitung: Dr. Annette Peters

Metabolomics analyses to better understand complex phenotypes

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

vorgelegt von
Zhonghao Yu
aus
Shanghai

Jahr

2013

**Gedruckt mit Genehmigung der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München**

Betreuer: Prof. Dr. Thomas Illig

Zweitgutachter: Priv. Doz. Dr. Alexander Faußner

Dekan: Prof. Dr. med. Dr. h. c. M. Reiser, FACR, FRCR

Tag der mündlichen Prüfung: 09.12.2013

Eidesstattliche Versicherung

Yu, Zhonghao

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Thema

Metabolomics analyses to better understand complex phenotypes

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 05.08.2014

Ort, Datum

Unterschrift Doktorandin/Doktorand

Acknowledgements

First of all, I want to thank for Prof. Dr. Thomas Illig, CEO and scientific head of Hannover Unified Biobank (HUB), Hannover Medical School, former head of Research Unit of Molecular Epidemiology, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health for his continuous support since the start of my PhD study. His guidance helps me not only in my research, my writing of this thesis but also in my career path.

I would also like to dedicate my gratefulness to my direct supervisor Dr. Rui Wang-Sattler, head of the Group leader of Molecular Epidemiology - Metabolism Research Unit of Molecular Epidemiology, who spend great amount of time and energy in supervising my research work as well as this thesis. She not only made a lot of effort in making the data available for the studies described in the thesis but also gave valuable insights and organized numerous fruitful discussions with experts in the fields.

Moreover, the work of this thesis would not be possible had there not be the diligent work from both the Genome Analysis Center (GAC) Metabolomic Platform (led by Prof. Dr. Jurek Adamsik) and the MONICA/KORA studies (led by Prof. Dr. Dr. H.-Erich Wichmann and Prof. Dr. Annette Peters). Thanks to the people from GAC for establishing the platform and making metabolite profiling possible and all the staffs who involved in the planning and conducting the MONIC/KORA study for the founding of such an excellent cohort.

I would not forget all the help and support I had from my current and former colleagues. Your kindness helps me to finish my PhD work and to adapt the life in Germany.

Special thanks to my wife, my parents and my friends who were always being unconditionally supportive and unbelievably patient.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Metabolomics

1.1.1 Metabolites and metabolism

Metabolites are the intermediates or end products produced by the cellular processes of a certain organism. Their levels can be regarded as the ultimate responses of the biological systems to genetic and/or environmental challenges (Fiehn, 2002). Metabolism is constituted by a set of chemical reactions and transformations which are needed to maintain life. It comprises two parts, the catabolism which is the breakdown of molecules to obtain energy and the anabolism which is the synthesis of all compounds needed by the cells.

The metabolites play critical roles in biological systems due to their involvement in cellular and physiological energetics, structure, and signaling (Vinayavekhin et al., 2010). Moreover, unlike RNA and proteins, metabolites are not directly coded into the genome. Therefore, one of the major goals in human biology is to understand the biochemical pathways which comprise the human metabolism as well as to study their relations to different human diseases.

1.1.2 Metabolomics

The ‘omics’ technologies, which provide information regarding detailed content of the cells, tissues, organs or biofluids in large scales with a high throughput manner, are becoming more popular in biomedical studies (Rochfort, 2005). Metabolome, coined less than two decades ago (Oliver et al., 1998), is similar to other '-ome' terminologies, and is defined as the total complement of small-molecule metabolites found in or produced by an organism (Mayr, 2008). Metabolomics is regarded as the studies of metabolome, with a view to understanding complex biological systems on a large scale using high-throughput identification and quantification techniques with statistical methods to cope with the huge datasets produced. (Brown et al., 2005; Kaddurah-Daouk et al., 2008; Psychogios et al., 2011).

Over the past few years, the scientific community has witnessed the advent of this so-called 'omics' era. Studies of single genes, single mRNA transcripts, single proteins and single metabolites have been moved to those encompassed the entire genomes, transcriptomes, proteomes and metabolomes (Kaddurah-Daouk et al., 2008). More investigators are now seeking to understand the complex biological systems on a larger scale other than by simply using the traditional reductionistic approach (Brown et al., 2005; Mayr, 2008). Along with the other three 'omics' –genomics, transcriptomics, and proteomics-, metabolomics has added a new piece of building block to the fast emerging field of systems biology. Together, they provide powerful tools with which to analyses physiological and disease-induced biological states at the molecular level, taking into account both the organism's intrinsic properties, i.e. genetic factors, and the effects of lifestyle, diet, and environment. Many attempts have been made to discover the link between genetics and metabolite concentrations (Gieger et al., 2008; Illig et al., 2010; Suhre et al., 2011), whilst other scholars have sought to unveil the association between metabolite profiles and general phenotypes (Mittelstrass et al., 2011; Wang-Sattler et al., 2008; Yu et al., 2011), In addition to these investigations, various other studies have attempted to both predict the behavior of diseases (Floegel et al., 2012; Wang-Sattler et al., 2012) and use metabolite concentrations to ascertain the disease etiology hidden behind the metabolomics data (German et al., 2005a).

However, the scale and coverage of metabolomics is in no comparison to the other 'omics'. The exact number of metabolites in human metabolome is still a matter of debate and numbers ranging from a few thousand to tens of thousands of have been proposed (Kaddurah-Daouk et al., 2008). Up till now, it has remained impossible to measure the whole metabolome using one single analytic method. Researchers have had to carefully choose appropriate technologies based on their desired results from the metabolome. New fields, such as lipidomics, have come into existence to study the subgroup instead of the whole metabolome (Shevchenko and Simons, 2010; Wenk, 2005). One of the major reasons behind this limitation is the chemical complexity and the concentration range in the whole metabolome. In contrast, the building blocks for genome, transcriptome and proteome are relatively limited. There are four to five nucleotides and approximately 20 primary amino acids and several of their derivatives (e.g. methylated nucleotides, phosphorylated proteins), which do not exist in metabolome. Moreover, the range of the metabolite concentrations varied dramatically (e.g. from pM to mM) and there is no available instrument that can cover such a range without differential dilution (Brown et al., 2005).

1.1.3 Techniques used in metabolite concentration measurements

Two analytic methods, namely nuclear magnetic resonance (NMR) and mass spectrometry (MS) are most widely used in metabolomics studies for different analytical approaches including profiling-, non-targeted-, and targeted- metabolomics. These approaches have been developed to meet the distinct requirements for different study aims (Psychogios et al., 2011).

NMR can detect a wide range of biochemical metabolites and is considered to be robust and reproducible (Mayr, 2008). However, the NMR technology suffers from low sensitivity (on the order of 10 μmol/L) and high initial instrument investments (Spratlin et al., 2009). MS-based methods were used in the metabolomics measurements represented in this thesis as such platform is available in the Helmholtz Centre Munich.

The mass to charge ratio (m/z) is a dimensionless value used in mass spectrometric experiments, and is formed by dividing the mass number of an ion by its charge number. The quantity measured by MS is the mass-to-charge ratio of ions formed from molecules, usually separated by chromatography because the power of this technology depends on separation along with detection. The MS technology is highly sensitive, typically at the pictogram level, which makes the detection of metabolites with low concentration possible (Spratlin et al., 2009). The current applications of metabolomics have two major platforms: gas chromatography MS (GC-MS) and liquid chromatography MS (LC-MS). GC-MS is more suitable when it comes to measuring the non-polar metabolites with lower molecular weight whilst LC-MS is preferred to measure those polar ones with higher molecular weight (Artati et al., 2012). In the metabolomic analysis we presented in this thesis (Mittelstrass et al., 2011; Wang-Sattler et al., 2012; Yu et al., 2011), targeted metabolite profiling using electrospray ionization (ESI) tandem mass spectrometry (MS/MS) was also performed. The details of the platform will be provided in the third section of Chapter 2.

1.2 Epidemiology studies

Epidemiology is the study of the distributions and determinants of health-related states or events (including diseases), and the application of this study to the control of diseases and to help improve other health-related problems (Susser, 1973).

1.2.1 Study type in epidemiology

To investigate the questions of disease development and other health-related problems, it is crucial to choose the appropriate study design. Epidemiological studies can be classified as either observational or experimental based on whether the investigator intervenes. In this thesis, analytical observational studies were used.

The three most common types of observational study are, the cross-sectional study, the case-control study, and the cohort study. In a cross-sectional study, the measurement of the exposure and effect are conducted at the same time. It is relatively easy and inexpensive to conduct, although it is difficult to assess the reasons, if any, for the associations. In a case-control study, people with a disease (or other outcome variable) of interest are recruited, along with a suitable control group. The aim is to investigate the causes behind the diseases, and particularly rare diseases. Cohort studies begin with a group of people who are free of disease or who are classified into subgroups according to certain exposures. Cohort studies provide the best information about the causation of disease and the most direct measurement of the risk of developing disease (Beaglehole et al., 2006). As a variation of the case-control study, the nested case-control study uses only a subset of controls which are selected for each case from that case's risk set from the cohort and compared to cases.

The studies involved in this thesis are one cohort study, one nested case-control study and population based cross-sectional studies.

1.2.2 Confounders

The disease status and health parameters investigated in epidemiology studies are generally referred to as phenotypes. Risk factors (i.e. factors which can potentially change the phenotype status) are referred to as either environmental / genetic / physiological (age, sex) factors, or as covariates. It should be noted that all non-genetic factors, including e.g. environmental exposures such as fine dust particles, but also life-style parameters like smoking and age, are generally termed environmental or physiological factors. Association analysis quantifies the relation between phenotype and environmental and/or genetic factors through statistical analysis (e.g. regression). Estimated effect sizes describe the relative change in the phenotype due to different covariate values. In association analysis, it is common that a third parameter (i.e. risk factor) correlates with both the phenotype and the environmental factor. Such a parameter is referred to as a confounding factor or confounding variable and must be accounted for in the association analysis to evaluate the real effect of the factor of interest.

1.3 Statistical and bioinformatical analysis

During the development of 'omics' studies, statistics as well as bioinformatics, have become an important tool both in finding effective signals among huge amount of data and in collecting and integrating information from different sources either for public use or for the purpose of a specific study. These techniques essentially refer to the science of managing and analyzing biological data using advanced computing techniques (German et al., 2005b).

The nature of the data acquired in the metabolomics studies is similar to those in other 'omics' studies: high in dimension with a relatively small number of observations. The major goal in metabolomics studies related to life science research is to identify biomarkers and to understand the mechanistic basis for biological difference (e.g. healthy vs. diseased). The machine learning methods which have been applied for years are suitable for this purpose with such data property. Both unsupervised (e.g. principle component analysis (PCA), clustering) and supervised methods (e.g. random forest, partial least square (PLS)) can be used to find the features, which are crucial to the phenotypes (e.g. the development of the disease) but which have been buried under the huge amount of data.

1.4 Metabolomic variations in complex phenotypes

Although the measurements of the metabolome are not as mature as in the other 'omics', valuable information is generated from metabolomics. Many studies have investigated the associations between metabolic variations and different disease such as metabolic diseases, cancer, and infectious diseases (Spratlin et al., 2009; Vinayavekhin et al., 2010). However, many studies have also shown that complex phenotypes, including environmental factors such as cigarette smoking (Wang-Sattler et al., 2008), fasting status (Rubio-Aliaga et al., 2011), age (Yu et al., 2012), sex (Mittelstrass et al., 2011), body mass index (BMI) (Jourdan et al., 2012), and physical activity/challenges (Krug et al., 2012) could all produce influential metabolite concentration levels in the human body. Moreover, different sample matrices could also affect the final readout of the metabolite concentrations (Yu et al., 2011). In order to find the real metabolic perturbations related with disease etiology, specific consideration must be given to those features that can also contribute to the metabolic variations. In the following chapters we will present our studies on two sources of these variations, namely the sample matrix and the sex effect on the metabolite concentration variations.

1.4.1 Metabolomic variations in plasma and serum

One source of the metabolic variations is rooted in the different collection procedurals of human blood. Human plasma and serum are most commonly used in biomedical experiments and clinical tests. However, different matrices usually produce different results in tests (Beheshti et al., 1994) and thus are preferred under different circumstances. For example, heparin confounds some cardiac troponin I assay and thus serum is preferred for the measurement of cardiac troponins I and T (Gerhardt et al., 2000; Jaffe et al., 2000), whereas plasma is favored in oral glucose tolerance tests for type 2 diabetes proposed in the diagnosis guideline (Sacks et al., 2002). As reviewed by

Mannello (Mannello, 2008), the use of an incorrect matrix can lead to an improper diagnosis.

Blood is composed of two parts: a cellular component consisting of red and white blood cells and platelets, and a liquid carrier, known as plasma or serum. The major difference between plasma and serum depends on whether an anti-coagulate agent is introduced during the blood collection procedure. The coagulation cascade is blocked in plasma and only centrifugation is required to remove or decant the most buoyant (non-cellular) portion. In contrast, with regards to serum, the coagulation is started through a series of interconnected self-amplifying, zymogen-enzyme conversions that penultimately produce thrombin. In the final step of the coagulation cascade, FIIa hydrolyses fibrinogen into fibrin units which oligomerize into a fine mesh, which in turn, cases blood to gel or clot (Vogler and Siedlecki, 2009). During the clotting process, platelets can release proteins (e.g. pro-inflammatory cytokines (Schnabel et al., 2009)) as well as metabolites (e.g. sphingosine-1-phosphate (Yatomi et al., 1997)) into the serum. Both plasma and serum are aqueous solutions (approximately 95% water) and contain a variety of substances including proteins and peptides (such as albumins, globulins, lipoproteins, enzymes and hormones), nutrients (such as carbohydrates, lipids and amino acids), electrolytes, organic wastes and a variety of other small organic molecules suspended or dissolved in them (Psychogios et al., 2011). Several studies have already examined the potential proteomic differences caused by different blood collecting procedures (Barelli et al., 2007; Tammen et al., 2005). Since metabolomics is a newly developed discipline compared to the other 'omics', there are only a few recent studies related to this subject (e.g. comparing different biofluids (Bando et al., 2010) as is also the case for studies comparing plasma and serum from animal blood (Ayache et al., 2006)). Moreover, two studies using small samples of around 15 human participants have addressed this issue with conflicting results. Teahan et al. reported minimal differences between the two matrices while Liu et al. observed changes ranging from 0.03 to 18-fold (Liu et al., 2010; Teahan et al., 2006).

In the third chapter of this thesis, I will present our study (Yu et al., 2011) which was performed using a targeted metabolomics study of 163 metabolites to compare plasma and serum samples from 377 individuals. The results showed a good reproducibility of metabolite concentrations in both plasma and serum, although somewhat better in plasma. There was also a clear discrimination between the metabolite profiles of plasma and serum. Metabolite concentrations were generally higher in serum, yet still highly correlated between the two matrices. Furthermore, serum revealed more potential biomarkers than plasma when comparisons were made between different phenotypes.

1.4.2 Metabolomic variations in sex

I will also explore a second source of metabolic variation in this thesis, namely the effect brought about by sexual dimorphisms. Sex refers to the classification of males and females according to their reproductive organs. Historically, the scientific community assumed that apart from the reproductive system, differences in cellular or molecular levels did not exist or were not relevant (Wizemann and Pardue, 2001). In a survey of studies published in 2004 and spanning nine different medical journals found that only 37% of participants were women (24% when it comes to drug trials) whilst only 13% of studies analyzed data by sex (Kim et al., 2010). Over the past decades, new discoveries in basic human biology have made it increasingly apparent that many normal physiological functions—and, in many cases, pathological functions—are influenced either directly or indirectly by sex-based differences in biology. Gender inequalities have been increasingly recognized and different studies showed that there is a strong correlation between sex and the incidence, prevalence, age at onset, symptoms and severity of a disease, as well as the reaction to drugs (Fairweather and Rose, 2004; Mostertz W, 2010).

With this in mind, it is important to determine for which aspects and to what extent gender influences metabolomics. To study the gender effect on metabolomics, I report the results (Mittelstrass et al., 2011) in the third chapter with a systematical assessment of the effect from sex on serum metabolites in a large population-based cohort (Holle et al., 2005) and with the replication of most of the findings.

1.5 Identification of type 2 diabetes candidate biomarker

Metabolic disorders such as type 2 diabetes (T2D) are an obvious choice for this application of metabolomics. Indeed, this is because many of the underlying causes of these disorders are thought to result from dys-regulation in small molecule metabolism.

T2D is defined by increased blood glucose levels due to pancreatic beta-cell dysfunction and insulin resistance without evidence for specific causes, such as autoimmune destruction of pancreatic beta-cells (Krebs et al., 2002; Muoio and Newgard, 2008; Stumvoll et al., 2005). Diabetes has reached epidemic proportions and as of 2011 had affects more than 360 million individuals worldwide. Moreover, the number of people with type 2 diabetes is expected to reach more than 550 million by the year 2030.

A state of pre-diabetes (i.e., impaired fasting glucose (IFG) and/or impaired glucose tolerance (IGT)) with only slightly elevated blood glucose levels can accompany an individual for years before the onset of T2D (McGarry, 2002; Tabák et al., 2012) . The development of diabetes in pre-diabetic individuals can be prevented or delayed by dietary changes and increased physical activity (Knowler et al., 2002; Tuomilehto et al., 2001). However, no specific biomarkers that result in an effective prevention have been reported. Metabolomics studies allow metabolites involved in disease mechanisms to be discovered by monitoring metabolite level changes in predisposed individuals compared with healthy ones (Newgard et al., 2009; Pietiläinen et al., 2011; Rhee et al., 2011; Shaham et al., 2008; Zhao et al., 2010). Altered metabolite levels may serve as diagnostic biomarkers and enable preventive actions. Previous cross-sectional metabolomics studies of T2D were either based on small sample sizes (Pietiläinen et al., 2011; Shaham et al., 2008; Wopereis et al., 2009; Zhao et al., 2010) or did not place sufficient emphasis on the influence of common risk factors of T2D (Newgard et al., 2009). Recent work based on prospective nested case–control studies with relatively large samples (Rhee et al., 2011; Wang et al., 2011), five branched-chain and aromatic amino acids were identified as predictors of T2D (Wang et al., 2011). Here, in the third section of Chapter 3, I will present our attempt to (i) reliably identify candidate biomarkers of pre-diabetes and (ii) build metabolite–protein networks to understand diabetes-related metabolic pathways using various comprehensive large-scale approaches with measured metabolite concentration profiles.

Chapter 2 Materials and Methods

2.1 Population based KORA cohort

KORA (Cooperative Health Research in the Region Augsburg) was used in the analysis of this thesis. Written informed consent was obtained from each KORA participant. The study was approved by the ethics committee of the Bavarian Medical Association.

KORA is a regional research platform for population-based surveys and subsequent follow-up studies in the fields of epidemiology, health economics, and health care research. In 1996, KORA was established to continue and expand the MONICA (Monitoring of Trends and Determinants of Cardiovascular Disease) project in Augsburg. The available pool of study participants allows for cohort, case-control and family studies (Holle et al., 2005).

The individuals of KORA were sampled in a two-stage procedure. In the first step, Augsburg and the 16 communities were selected using cluster sampling. In a second step, stratified random sampling was performed in each community (*MONICA-Projekt, Region Augsburg*, 1986). Four cross-sectional studies, KORA survey 1 (S1) to survey 4 (S4) were performed at five-year intervals. Follow-up studies of S3 and S4 were conducted in around seven to ten years after each survey.

The KORA survey 3 (S3) was conducted in 1994/1995 with a 10 years later (2004/2005) follow up (F3) while the KORA survey 4 (S4) was conducted in 1999/2001 with a 7 years later (2006/2008) follow-up survey (F4).

In all surveys, baseline information on socio-demographic variables, risk factors (smoking, alcohol consumption, physical activity, etc.), medical history and family history of chronic diseases, medication use, and more was gathered by trained medical staff during an extensive standardized face-to-face interview. In addition, a standardized medical examination including blood pressure measurements and anthropometric measurements were performed on all the participants (Holle et al., 2005).

Three studies in KORA (F3, S4 and F4) were used in the analyses (Mittelstrass et al., 2011; Wang-Sattler et al., 2012; Yu et al., 2011) presented in this thesis. Plasma and serum samples collected from 377 participants in the KORA F3 were used to elaborate the metabolic variation between two different blood matrices. In the study of sex dimorphism of metabolomics, serum samples from 3080 KORA F4 individuals were used as discovery population and KORA F3 were served as the replication population. To find the biomarkers for (pre-) diabetes, 4261 KORA S4 and 3080 KORA F4 individuals were used as discovery population in both cross-sectional and longitudinal manners.

2.2 Blood Sample collections

To measure the metabolite concentrations in human blood, plasma and/or serum samples were collected from the KORA participants. The blood was drawn into S-Monovettes tubes (SARSTEDT AG & Co., Nümbrecht, Germany) in the morning between 08:00 and 10:30 after a period of overnight fasting for at least eight hours. EDTA plasma were shaken gently and thoroughly for 15 minutes followed by centrifugation at 2750 g for 15 minutes at 15°C. Serum tubes were gently inverted twice, followed by 30 min resting at room temperature, to obtain complete coagulation. They were then centrifuged at 2750 g at 15°C for 10 min. Plasma and serum was filled into synthetic straws, which were stored in liquid nitrogen until the metabolic analyses were conducted. Plasma and serum samples from KORA F3 participants and serum samples from KORA S4 and F4 were used in the analysis. (Jourdan et al., 2012; Mittelstrass et al., 2011; Wang-Sattler et al., 2012; Yu et al., 2011)

2.3 Quantification of metabolite concentration profiles

Two commercially available kits from Biocrates (Biocrates Life Sciences AG, Innsbruck, Austria) were used in the metabolomics measurements including the Absolute*IDQ*™ kit *p*150 and the Absolute*IDQ*™ kit *p*180.

2.3.1 Absolute*IDQ*^{TM} kit *p*150

The Absolute*IDQ*™ kit *p*150 used a targeted metabolite profiling named electrospray ionization (ESI) tandem mass spectrometry (MS/MS). This technique has been described in detail elsewhere (Weinberger and Graber, 2005; Weinberger, 2008). Briefly, the assay preparation was done by an automated robotics system (Hamilton Robotics GmbH) on special double-filter plates with 96 wells. These plates also contain the isotope labeled non-radioactive internal standards, blank samples (PBS) and quality controls. Assays used 10µl serum or plasma samples and include phenylisothiocyanate (PITC)-derivatisation of amino acids, extraction with organic solvent and several liquid handling steps. Flow injection analysis (FIA) coupled with multiple reaction monitoring scans (FIA MS/MS) on an API 4000 QTrap instrument (Applied Biosystems) was used for quantification of amino acids, acylcarnitines, sphingomyelins, phosphatidylcholines, and hexose. Concentrations were calculated and evaluated in the Met*IQ* software provided by the manufacturer. It compared measured analytes in a defined extracted ion count section to those of specific labeled internal standards or nonlabeled, nonphysiological standards (semiquantitative) provided by the kit plate. This method has been proven to be in conformance with the "Guidance for Industry — Bioanalytical Method Validation" published by the FDA (Food and Drug Administration), which implies the proof of reproducibility within a given error range (Altmaier et al., 2011; Römisch-Margl et al., 2011).

Plasma and serum samples from KORA F3, serum samples from KORA F4 were measured using this kit for metabolite concentration profiles.

2.3.2 Absolute*IDQ*™ kit p180

The Absolute*IDQ*™ kit *p*180 is an upgrade of the Absolute*IDQ*™ kit *p*150. It used the combination of FIA-MS and LC-MS to detect the metabolite concentrations. Metabolite concentrations measured using the Absolute*IDQ*™ kit *p*180 were preceded according to the manufacturer’s instructions on an API4000™ LC/MS/MS System equipped with an electrospray ionization source. Samples (10 µl) were pipetted onto the spots of the kit plate. The plate was centrifuged at 100 g for 2 min, receiving about 250 µl sample in plate 1 (FIA plate). The upper plate was removed, and 150 µl of each sample was transferred into a second plate (LC-MS plate). HPLC water (150 µl) was added to the LC-MS plate, and 500 µl of MS running solvent (Biocrates solvent diluted in methanol) was added to the FIA plate. The LC-MS plate was measured first by scheduled multiple reaction monitoring, and the FIA plate was stored at 4°C. Concentrations were calculated and evaluated in the Analyst/MetIQ software by comparing measured analytes in a defined extracted ion count section to those of specific labeled internal standards or nonlabeled, nonphysiological standards (semiquantitative) provided by the kit plate. (Schmerler et al., 2012)

The serum samples from KORA S4 were measured using this kit for metabolite concentration profiles.

2.3.3 Metabolites measured

In total, up to 190 different metabolites were quantified by these two kits. Absolute*IDQ*™ kit *p*150 can measure 163 metabolites, including 14 amino acids (13 proteinogenic and ornithine), hexose (sum of hexoses, around 90 – 95% glucose), free carnitine (C0) and 40 other acylcarnitines (Cx:y), 15 sphingomyelins (SMx:y), 77 phosphatidylcholines (PCs, diacyl (aa) and acyl-alkyl (ae)) and 15 lyso-phosphatidylcholines (LPCs). The lipid side chain composition is abbreviated as Cx:y, with x denoting the number of carbons in the side chain and y denoting the number of double-bonds. The Absolute*IDQ*™ kit *p*180 can measure 186 metabolites, including 21 amino acids (19 proteinogenic, citrulline and ornithine), hexose, free carnitine, 39 acylcarnitines, 15 sphingomyelins, 90 phosphatidylcholines (14 LPCs and 76 PCs) as well as 19 biogenic amines. The overlap of these two kits is 159 metabolites. Full biochemical names and abbreviations are provided in Table 1.

Table 1: Full biochemical names, abbreviation, all metabolites measured by Biocrates Absolute*IDQ*™ kits *p*150 and *p*180

Abbrevation	Full biochemical name	Abbrevation	Full biochemical name
C0	Carnitine	PC aa C36:0	Phosphatidylcholine diacyl C36:0

C2	Acetylcarnitine	PC aa C36:1	Phosphatidylcholine diacyl C36:1
C3	Propionylcarnitine	PC aa C36:2	Phosphatidylcholine diacyl C36:2
C3-OH	Hydroxypropionylcarnitine	PC aa C36:3	Phosphatidylcholine diacyl C36:3
C3:1	Propenonylcarnitine	PC aa C36:4	Phosphatidylcholine diacyl C36:4
C4	Butyrylcarnitine	PC aa C36:5	Phosphatidylcholine diacyl C36:5
C4-OH	Hydroxybutyrylcarnitine	PC aa C36:6	Phosphatidylcholine diacyl C36:6
C4:1	Butenylcarnitine	PC aa C38:0	Phosphatidylcholine diacyl C38:0
C5	Valerylcarnitine	PC aa C38:1	Phosphatidylcholine diacyl C38:1
C5-DC	Glutarylcarnitine	PC aa C38:3	Phosphatidylcholine diacyl C38:3
C5-M-DC	Methylglutarylcarnitine	PC aa C38:4	Phosphatidylcholine diacyl C38:4
C5-OH	Hydroxyvalerylcarnitine	PC aa C38:5	Phosphatidylcholine diacyl C38:5
C5:1	Tiglylcarnitine	PC aa C38:6	Phosphatidylcholine diacyl C38:6
C5:1-DC	Glutaconylcarnitine	PC aa C40:1	Phosphatidylcholine diacyl C40:1
C6	Hexanoylcarnitine	PC aa C40:2	Phosphatidylcholine diacyl C40:2
C6:1	Hexenoylcarnitine	PC aa C40:3	Phosphatidylcholine diacyl C40:3
C7-DC	Pimelylcarnitine	PC aa C40:4	Phosphatidylcholine diacyl C40:4
C8	Octanoylcarnitine	PC aa C40:5	Phosphatidylcholine diacyl C40:5
C8:1	Octenoylcarnitine	PC aa C40:6	Phosphatidylcholine diacyl C40:6
C9	Nonaylcarnitine	PC aa C42:0	Phosphatidylcholine diacyl C42:0
C10	Decanoylcarnitine	PC aa C42:1	Phosphatidylcholine diacyl C42:1
C10:1	Decenoylcarnitine	PC aa C42:2	Phosphatidylcholine diacyl C42:2
C10:2	Decadienylcarnitine	PC aa C42:4	Phosphatidylcholine diacyl C42:4
C12	Dodecanoylcarnitine	PC aa C42:5	Phosphatidylcholine diacyl C42:5
C12-DC	Dodecanedioylcarnitine	PC aa C42:6	Phosphatidylcholine diacyl C42:6
C12:1	Dodecenoylcarnitine	PC ae C30:0	Phosphatidylcholine acyl-akyl C30:0
C14	Tetradecanoylcarnitine	PC ae C30:1	Phosphatidylcholine acyl-akyl C30:1
C14:1	Tetradecenoylcarnitine	PC ae C30:2	Phosphatidylcholine acyl-akyl C30:2
C14:1-OH	Hydroxytetradecenoylcarnitine	PC ae C32:1	Phosphatidylcholine acyl-akyl C32:1
C14:2	Tetradecadienylcarnitine	PC ae C32:2	Phosphatidylcholine acyl-akyl C32:2
C14:2-OH	Hydroxytetradecadienylcarnitine	PC ae C34:0	Phosphatidylcholine acyl-akyl C34:0
C16	Hexadecanoylcarnitine	PC ae C34:1	Phosphatidylcholine acyl-akyl C34:1
C16-OH	Hydroxyhexadecanoylcarnitine	PC ae C34:2	Phosphatidylcholine acyl-akyl C34:2
C16:1	Hexadecenoylcarnitine	PC ae C34:3	Phosphatidylcholine acyl-akyl C34:3
C16:1-OH	Hydroxyhexadecenoylcarnitine	PC ae C36:0	Phosphatidylcholine acyl-akyl C36:0
C16:2	Hexadecadienylcarnitine	PC ae C36:1	Phosphatidylcholine acyl-akyl C36:1
C16:2-OH	Hydroxyhexadecadienylcarnitine	PC ae C36:2	Phosphatidylcholine acyl-akyl C36:2
C18	Octadecanoylcarnitine	PC ae C36:3	Phosphatidylcholine acyl-akyl C36:3
C18:1	Octadecenoylcarnitine	PC ae C36:4	Phosphatidylcholine acyl-akyl C36:4
C18:1-OH	Hydroxyoctadecenoylcarnitine	PC ae C36:5	Phosphatidylcholine acyl-akyl C36:5
C18:2	Octadecadienylcarnitine	PC ae C38:0	Phosphatidylcholine acyl-akyl C38:0
Ala	Alanine	PC ae C38:1	Phosphatidylcholine acyl-akyl C38:1
Arg	Arginine	PC ae C38:2	Phosphatidylcholine acyl-akyl C38:2
Asn	Asparagine	PC ae C38:3	Phosphatidylcholine acyl-akyl C38:3
Asp	Aspartate	PC ae C38:4	Phosphatidylcholine acyl-akyl C38:4
Cit	Citrulline	PC ae C38:5	Phosphatidylcholine acyl-akyl C38:5
Gln	Glutamine	PC ae C38:6	Phosphatidylcholine acyl-akyl C38:6
Glu	Glutamate	PC ae C40:0	Phosphatidylcholine acyl-akyl C40:0

Gly	Glycine	PC ae C40:1	Phosphatidylcholine acyl-akyl C40:1
His	Histidine	PC ae C40:2	Phosphatidylcholine acyl-akyl C40:2
Ile	Isoleucine	PC ae C40:3	Phosphatidylcholine acyl-akyl C40:3
Leu	Leucine	PC ae C40:4	Phosphatidylcholine acyl-akyl C40:4
Lys	Lysine	PC ae C40:5	Phosphatidylcholine acyl-akyl C40:5
Met	Methionine	PC ae C40:6	Phosphatidylcholine acyl-akyl C40:6
Orn	Ornithine	PC ae C42:0	Phosphatidylcholine acyl-akyl C42:0
Phe	Phenylalanine	PC ae C42:1	Phosphatidylcholine acyl-akyl C42:1
Pro	Proline	PC ae C42:2	Phosphatidylcholine acyl-akyl C42:2
Ser	Serine	PC ae C42:3	Phosphatidylcholine acyl-akyl C42:3
Thr	Threonine	PC ae C42:4	Phosphatidylcholine acyl-akyl C42:4
Trp	Tryptophan	PC ae C42:5	Phosphatidylcholine acyl-akyl C42:5
Tyr	Tyrosine	PC ae C44:3	Phosphatidylcholine acyl-akyl C44:3
Val	Valine	PC ae C44:4	Phosphatidylcholine acyl-akyl C44:4
xLeu	Leucine/Isoleucine	PC ae C44:5	Phosphatidylcholine acyl-akyl C44:5
Ac Orn	Acetylornithine	PC ae C44:6	Phosphatidylcholine acyl-akyl C44:6
ADMA	Asymmetric dimethylarginine	LPC a C14:0	lysoPhosphatidylcholine acyl C14:0
SDMA	Symmetric Dimethylarginine	LPC a C16:0	lysoPhosphatidylcholine acyl C16:0
total DMA	Sum of ADMA and SDMA	LPC a C16:1	lysoPhosphatidylcholine acyl C16:1
alpha AAA	alpha-Aminoadipic acid	LPC a C17:0	lysoPhosphatidylcholine acyl C17:0
Carnosine	Carnosine	LPC a C18:0	lysoPhosphatidylcholine acyl C18:0
Creatinine	Creatinine	LPC a C18:1	lysoPhosphatidylcholine acyl C18:1
Histamine	Histamine	LPC a C18:2	lysoPhosphatidylcholine acyl C18:2
Kynurenine	Kynurenine	LPC a C6:0	lysoPhosphatidylcholine acyl C6:0
Met SO	Methioninesulfoxide	LPC a C20:3	lysoPhosphatidylcholine acyl C20:3
Nitro-Tyr	Nitrotyrosine	LPC a C20:4	lysoPhosphatidylcholine acyl C20:4
OH-Pro	Hydroxyproline	LPC a C24:0	lysoPhosphatidylcholine acyl C24:0
PEA	Phenylethylamine	LPC a C26:0	lysoPhosphatidylcholine acyl C26:0
Putrescine	Putrescine	LPC a C26:1	lysoPhosphatidylcholine acyl C26:1
Sarcosine	Sarcosine	LPC a C28:0	lysoPhosphatidylcholine acyl C28:0
Serotonin	Serotonin	LPC a C28:1	lysoPhosphatidylcholine acyl C28:1
Spermidine	Spermidine	SM C16:0	Sphingomyeline C16:0
Spermine	Spermine	SM C16:1	Sphingomyeline C16:1
Taurine	Taurine	SM C18:0	Sphingomyeline C18:0
PC aa C24:0	Phosphatidylcholine diacyl C24:0	SM C18:1	Sphingomyeline C18:1
PC aa C26:0	Phosphatidylcholine diacyl C26:0	SM C20:2	Sphingomyeline C20:2
PC aa C28:1	Phosphatidylcholine diacyl C28:1	SM C22:3	Sphingomyeline C22:3
PC aa C30:0	Phosphatidylcholine diacyl C30:0	SM C24:0	Sphingomyeline C24:0
PC aa C30:2	Phosphatidylcholine diacyl C30:2	SM C24:1	Sphingomyeline C24:1
PC aa C32:0	Phosphatidylcholine diacyl C32:0	SM C26:0	Sphingomyeline C26:0 [#]
PC aa C32:1	Phosphatidylcholine diacyl C32:1	SM C26:1	Sphingomyeline C26:1
PC aa C32:2	Phosphatidylcholine diacyl C32:2	SM (OH) C14:1	Hydroxysphingomyeline C14:1
PC aa C32:3	Phosphatidylcholine diacyl C32:3	SM (OH) C16:1	Hydroxysphingomyeline C16:1
PC aa C34:1	Phosphatidylcholine diacyl C34:1	SM (OH) C22:1	Hydroxysphingomyeline C22:1
PC aa C34:2	Phosphatidylcholine diacyl C34:2	SM (OH) C22:2	Hydroxysphingomyeline C22:2
PC aa C34:3	Phosphatidylcholine diacyl C34:3	SM (OH) C24:1	Hydroxysphingomyeline C24:1
PC aa C34:4	Phosphatidylcholine diacyl C34:4	H1	Hexose

2.3.4 Quality controls for metabolomic measurements

2.3.4.1 KORA F3

The plasma and serum samples measured using Biocrates *p*150 kit had 83 individuals with duplicated measurements (for both plasma and serum). We therefore used the following criteria for data quality control: a metabolite is used in further analysis only if (I) the average value of the coefficient of variance (CV) of the three quality control samples (representing human plasma samples provided by the manufacturer in each kit plate) was smaller than 0.25; (II) the mean concentration of the metabolite over all samples was above 0.1 μM or over 90% of the samples have their metabolite concentration above the limit of detection (LOD). The LODs were set to three times the values of zero samples; (III) the Pearson's correlation coefficient (*r*) between the two repeated measurements of the 83 samples in either specimen exceeded 0.5. Altogether, 25 quantified and 97 semi-quantified metabolites passed all three criteria (Table 2).

Table 2: Summary of metabolites in plasma and serum samples of KORA F3

The abbreviations of 163 metabolite name are shown in the first column. The next three columns list the values of coefficient of variance (CV) of quality controls, percentage of individuals above limit of detection (LOD), and Person's correlation coefficient (*r*) of repeated measurements, respectively, for each metabolite. The following two columns exhibit the mean concentration (μM) and standard deviation (SD) of each metabolite in plasma and serum. The last three columns show the mean concentration difference, the correlation coefficient (r) and the *p*-value of paired Wilcoxon test of each metabolite between plasma and serum, respectively.

Metabolite abbreviations	CV of quality controls	Concentrations above LOD(%)	*r* of repated measurements	Mean ±SD (μM) in plasma	Mean ±SD (μM) in serum	Relative mean difference (%)	*r* between plasma and serum	*p*-value of Wilcoxon test
C0	0.10	99.87	0.85	42.64 ± 9.67	47.15 ± 11.11	9.94	0.88	5.66E-10
C10	0.12	95.36	0.93	9.88 ± 3.95	10.94 ± 4.72	9.63	0.96	1.24E-08
C10:1	0.09	55.17	0.92	0.45 ± 0.16	0.48 ± 0.18	5.42	0.93	2.76E-04
C12	0.12	94.03	0.93	0.28 ± 0.14	0.31 ± 0.16	9.97	0.97	1.09E-08
C12:1	0.14	17.24	0.90	0.14 ± 0.04	0.15 ± 0.05	5.92	0.88	6.13E-04
C14:1	0.15	99.87	0.93	0.24 ± 0.08	0.27 ± 0.1	14.42	0.92	2.31E-12
C14:2	0.12	97.21	0.94	0.1 ± 0.05	0.11 ± 0.06	11.44	0.93	1.33E-07
C16	0.09	99.87	0.93	0.36 ± 0.14	0.4 ± 0.16	11.00	0.90	1.20E-07
C18	0.16	99.60	0.92	0.19 ± 0.06	0.21 ± 0.07	9.86	0.87	4.29E-06
C18:1	0.10	99.87	0.93	0.13 ± 0.04	0.16 ± 0.06	13.76	0.88	8.44E-10
C18:2	0.08	99.87	0.89	0.16 ± 0.05	0.18 ± 0.06	11.17	0.85	3.53E-07
C2	0.10	99.87	0.79	0.16 ± 0.04	0.18 ± 0.05	10.36	0.81	3.30E-07
C3	0.11	99.87	0.89	0.04 ± 0.01	0.04 ± 0.02	7.00	0.86	2.00E-03
C4	0.12	95.49	0.79	0.13 ± 0.03	0.15 ± 0.04	14.26	0.73	7.69E-10
C5	0.12	96.55	0.82	0.05 ± 0.01	0.06 ± 0.02	10.17	0.80	2.00E-06
C8	0.08	59.15	0.84	0.15 ± 0.04	0.17 ± 0.05	10.98	0.72	2.73E-06

C8:1	0.09	91.64	0.82	0.06 ± 0.02	0.06 ± 0.02	2.95	0.66	2.90E-01
Arg	0.08	99.87	0.53	88.51 ± 17.48	146.67 ± 20.35	49.99	0.50	2.55E-15
Gln	0.11	100.00	0.52	589.99 ± 83.97	646.12 ± 89.1	9.06	0.47	7.55E-07
Gly	0.10	100.00	0.86	252.17 ± 75.49	316.6 ± 73.93	24.04	0.82	1.26E-14
His	0.13	99.87	0.46	91.58 ± 14.11	103.68 ± 16.54	12.21	0.59	3.82E-10
Met	0.11	99.87	0.74	32.23 ± 6.4	37.62 ± 7.28	15.34	0.67	1.94E-11
Orn	0.10	99.87	0.75	79.28 ± 20.7	81.09 ± 18.26	2.85	0.70	2.38E-01
Phe	0.16	99.87	0.62	52.77 ± 10.03	70.68 ± 14.52	28.46	0.58	6.97E-15
Pro	0.10	100.00	0.91	208.4 ± 62.39	232.93 ± 63.91	11.32	0.89	1.52E-09
Ser	0.15	99.87	0.72	89.93 ± 22.13	128.48 ± 26.92	35.88	0.77	2.55E-15
Thr	0.11	99.87	0.78	98.35 ± 27.23	108.66 ± 27.05	10.48	0.84	1.50E-07
Trp	0.10	99.87	0.65	76.47 ± 10.12	88.96 ± 12.78	14.83	0.56	7.68E-13
Tyr	0.14	99.87	0.71	78.98 ± 17.57	88.12 ± 18.4	10.97	0.71	1.65E-07
Val	0.11	100.00	0.74	286.24 ± 52.04	309.77 ± 60.07	7.54	0.66	3.87E-05
xLeu	0.10	100.00	0.78	233.81 ± 50	264.22 ± 56.52	11.94	0.68	2.88E-08
PC aa C24:0	0.23	62.73	0.30	0.65 ± 0.09	0.67 ± 0.1	3.72	0.61	6.86E-03
PC aa C28:1	0.09	99.87	0.88	3.28 ± 0.95	3.66 ± 1.03	10.82	0.86	9.10E-08
PC aa C30:0	0.11	99.87	0.95	5.26 ± 1.98	5.75 ± 2.23	8.18	0.92	8.61E-06
PC aa C32:0	0.06	99.87	0.85	15.23 ± 4.32	17.19 ± 4.51	12.05	0.85	2.20E-09
PC aa C32:1	0.11	99.87	0.97	20.17 ± 12.61	22.36 ± 13.43	9.87	0.95	1.18E-07
PC aa C32:2	0.21	99.87	0.93	4.65 ± 1.96	5.1 ± 2.15	8.70	0.91	5.17E-05
PC aa C32:3	0.09	99.87	0.85	0.5 ± 0.15	0.56 ± 0.16	12.58	0.82	1.48E-08
PC aa C34:1	0.05	100.00	0.84	234.94 ± 79.75	265.48 ± 83.75	12.06	0.90	5.18E-10
PC aa C34:2	0.13	100.00	0.64	395.28 ± 112.48	440.52 ± 105.24	11.17	0.85	1.10E-08
PC aa C34:3	0.06	100.00	0.91	17.03 ± 5.78	19.1 ± 6.61	10.97	0.89	5.08E-08
PC aa C34:4	0.10	99.87	0.94	2.31 ± 0.91	2.59 ± 1.02	11.10	0.92	2.64E-08
PC aa C36:0	0.13	99.87	0.77	3.02 ± 0.84	3.39 ± 0.96	11.31	0.86	1.81E-08

PC aa C36:1	0.12	99.87	0.89	51.78 ± 16.77	58.86 ± 18.19	12.46	0.87	2.20E-09
PC aa C36:2	0.06	100.00	0.84	247.4 ± 70.35	276.48 ± 68.83	11.31	0.85	1.02E-08
PC aa C36:3	0.06	100.00	0.89	143.8 ± 50.91	162.57 ± 49.61	12.46	0.89	2.10E-09
PC aa C36:4	0.07	100.00	0.82	206.61 ± 57.88	232.18 ± 57.54	11.83	0.85	2.72E-09
PC aa C36:5	0.08	100.00	0.97	29.73 ± 16.77	33.85 ± 19.19	11.68	0.96	2.30E-09
PC aa C36:6	0.16	99.87	0.95	1.15 ± 0.49	1.27 ± 0.56	9.30	0.93	3.05E-06
PC aa C38:0	0.11	99.87	0.84	3.09 ± 0.84	3.52 ± 0.98	12.44	0.88	2.04E-10
PC aa C38:1	0.22	99.73	0.52	0.61 ± 0.61	0.79 ± 0.78	25.38	0.89	3.96E-05
PC aa C38:3	0.07	100.00	0.92	52.57 ± 18.23	58.69 ± 18.15	11.25	0.88	2.93E-08
PC aa C38:4	0.06	100.00	0.89	111.2 ± 31.75	123.59 ± 32.51	10.70	0.86	2.53E-08
PC aa C38:5	0.07	100.00	0.91	58.75 ± 17.61	66.37 ± 20.34	11.62	0.88	1.24E-09
PC aa C38:6	0.07	100.00	0.90	82.47 ± 26.46	92.59 ± 28.9	11.43	0.90	3.82E-09
PC aa C40:1	0.18	15.38	0.74	0.41 ± 0.09	0.45 ± 0.1	7.70	0.79	3.08E-06
PC aa C40:4	0.07	99.87	0.91	3.68 ± 1.31	4.04 ± 1.29	9.44	0.88	1.42E-06
PC aa C40:5	0.06	99.87	0.94	11.18 ± 3.75	12.41 ± 4.16	10.12	0.89	1.64E-07
PC aa C40:6	0.07	100.00	0.93	27.39 ± 9.51	30.27 ± 10.13	10.00	0.90	2.25E-07
PC aa C42:0	0.17	99.87	0.91	0.53 ± 0.17	0.58 ± 0.2	9.26	0.90	1.89E-06
PC aa C42:1	0.19	99.87	0.83	0.26 ± 0.07	0.29 ± 0.09	10.05	0.84	8.80E-07
PC aa C42:2	0.17	99.87	0.84	0.19 ± 0.06	0.21 ± 0.08	8.76	0.84	1.90E-04
PC aa C42:5	0.21	100.00	0.87	0.37 ± 0.12	0.41 ± 0.14	9.03	0.87	8.64E-06
PC aa C42:6	0.14	62.07	0.85	0.56 ± 0.13	0.58 ± 0.13	3.29	0.80	4.13E-02
PC ae C30:0	0.23	99.73	0.87	0.41 ± 0.14	0.46 ± 0.16	9.35	0.88	4.24E-06
PC ae C30:2	0.20	87.40	0.66	0.11 ± 0.03	0.12 ± 0.04	8.46	0.65	3.48E-03
PC ae C32:1	0.07	100.00	0.86	2.97 ± 0.79	3.37 ± 0.93	12.27	0.82	4.34E-09
PC ae C32:2	0.13	99.87	0.86	0.71 ± 0.19	0.8 ± 0.22	11.52	0.82	3.94E-08
PC ae C34:0	0.08	99.87	0.91	1.72 ± 0.53	1.99 ± 0.62	13.76	0.88	3.36E-10
PC ae C34:1	0.06	99.87	0.87	10.53 ± 2.54	11.96 ± 3.08	12.04	0.82	4.27E-09

PC ae C34:2	0.05	100.00	0.89	12.23 ± 3.42	14.02 ± 4.01	13.24	0.84	1.37E-09
PC ae C34:3	0.05	99.87	0.92	8.23 ± 2.73	9.4 ± 3.08	13.05	0.89	5.58E-10
PC ae C36:1	0.10	99.87	0.87	8.24 ± 1.98	9.36 ± 2.59	11.63	0.83	1.73E-08
PC ae C36:2	0.07	99.87	0.91	14.62 ± 4.17	16.52 ± 4.65	12.05	0.86	7.63E-09
PC ae C36:3	0.06	99.87	0.88	8.08 ± 2.05	9.19 ± 2.39	12.47	0.82	1.78E-09
PC ae C36:4	0.06	100.00	0.85	20.1 ± 4.78	22.98 ± 5.56	13.10	0.79	1.19E-09
PC ae C36:5	0.06	99.87	0.88	13.92 ± 3.5	15.99 ± 4.21	13.47	0.84	2.50E-10
PC ae C38:0	0.22	99.87	0.85	1.99 ± 0.7	2.27 ± 0.8	12.81	0.92	8.54E-10
PC ae C38:1	0.25	99.60	0.27	0.61 ± 0.22	0.68 ± 0.29	6.28	0.49	4.04E-02
PC ae C38:2	0.14	99.87	0.84	1.86 ± 0.51	2.16 ± 0.63	13.94	0.78	6.62E-09
PC ae C38:3	0.10	100.00	0.87	3.88 ± 0.96	4.37 ± 1.13	11.47	0.82	9.66E-09
PC ae C38:4	0.06	99.87	0.81	14.81 ± 3.07	16.72 ± 3.41	11.99	0.75	3.16E-09
PC ae C38:5	0.06	100.00	0.81	18.63 ± 3.68	21.24 ± 4.37	12.78	0.75	4.48E-10
PC ae C38:6	0.07	100.00	0.86	7.95 ± 1.94	9.07 ± 2.35	12.80	0.84	6.63E-10
PC ae C40:0	0.10	2.25	0.76	9.33 ± 1.74	10.02 ± 1.93	6.86	0.89	1.02E-08
PC ae C40:1	0.25	99.87	0.66	1.41 ± 0.31	1.56 ± 0.41	8.80	0.80	1.63E-06
PC ae C40:2	0.18	99.87	0.88	1.87 ± 0.49	2.1 ± 0.59	10.70	0.81	6.16E-07
PC ae C40:3	0.22	100.00	0.81	1.03 ± 0.25	1.19 ± 0.28	13.90	0.81	4.15E-11
PC ae C40:4	0.15	99.87	0.82	2.35 ± 0.48	2.66 ± 0.58	12.06	0.75	2.77E-09
PC ae C40:5	0.08	99.87	0.84	3.28 ± 0.62	3.66 ± 0.79	10.27	0.74	9.93E-08
PC ae C40:6	0.07	100.00	0.89	4.89 ± 1.28	5.51 ± 1.52	11.26	0.86	7.80E-09
PC ae C42:1	0.14	99.87	0.66	0.34 ± 0.08	0.36 ± 0.09	5.01	0.73	6.33E-03
PC ae C42:2	0.19	99.87	0.82	0.59 ± 0.15	0.65 ± 0.19	9.18	0.85	1.12E-06
PC ae C42:3	0.24	99.87	0.75	0.75 ± 0.16	0.84 ± 0.23	9.32	0.78	4.52E-06
PC ae C42:4	0.15	100.00	0.87	0.82 ± 0.21	0.91 ± 0.26	9.14	0.80	7.35E-06
PC ae C42:5	0.16	99.73	0.85	1.99 ± 0.46	2.21 ± 0.54	10.16	0.85	5.65E-09
PC ae C44:3	0.15	99.87	0.62	0.11 ± 0.03	0.12 ± 0.03	5.81	0.64	1.37E-02

PC ae C44:4	0.18	99.87	0.85	0.39 ± 0.11	0.43 ± 0.14	8.43	0.82	6.49E-05
PC ae C44:5	0.12	99.87	0.90	1.74 ± 0.48	1.93 ± 0.62	8.88	0.88	3.54E-07
PC ae C44:6	0.08	99.87	0.91	1.14 ± 0.31	1.28 ± 0.37	10.42	0.86	7.34E-08
LPC a C14:0	0.24	7.56	0.71	3.87 ± 1.24	4.28 ± 1.34	10.63	0.93	1.00E-09
LPC a C16:0	0.07	100.00	0.77	101.23 ± 19.61	130.59 ± 25.12	25.23	0.63	8.09E-15
LPC a C16:1	0.06	99.87	0.92	3.38 ± 1.41	4.13 ± 1.7	19.96	0.93	3.28E-13
LPC a C17:0	0.13	99.87	0.87	1.86 ± 0.54	2.38 ± 0.73	23.83	0.83	7.18E-14
LPC a C18:0	0.15	100.00	0.77	29.81 ± 6.43	39.2 ± 8.01	27.30	0.71	3.41E-15
LPC a C18:1	0.11	99.87	0.89	21.09 ± 6.29	26.35 ± 7.59	22.08	0.87	3.64E-14
LPC a C18:2	0.07	99.87	0.92	26.13 ± 8.45	30.01 ± 9.82	13.68	0.89	2.78E-10
LPC a C20:3	0.10	99.87	0.89	2.5 ± 0.78	2.9 ± 0.98	13.62	0.91	9.81E-11
LPC a C20:4	0.10	100.00	0.87	6.64 ± 1.9	8.04 ± 2.51	18.31	0.89	1.14E-12
LPC a C28:0	0.25	24.14	0.21	0.31 ± 0.08	0.34 ± 0.09	10.34	0.58	9.40E-05
SM (OH) C14:1	0.09	99.87	0.87	6.51 ± 1.84	7.03 ± 2.09	7.23	0.84	5.32E-04
SM (OH) C16:1	0.10	99.87	0.87	3.65 ± 1.02	4 ± 1.19	8.51	0.81	1.26E-04
SM (OH) C22:1	0.14	100.00	0.86	12.93 ± 3.82	14.1 ± 4.93	7.15	0.84	1.04E-03
SM (OH) C22:2	0.16	100.00	0.86	10.61 ± 3.1	11.59 ± 3.85	7.69	0.82	5.68E-04
SM (OH) C24:1	0.20	100.00	0.86	1.25 ± 0.44	1.3 ± 0.48	3.76	0.81	1.05E-01
SM C16:0	0.09	100.00	0.77	105.82 ± 19.62	115.47 ± 25.6	7.91	0.65	1.45E-04
SM C16:1	0.08	100.00	0.80	16.32 ± 3.36	17.74 ± 4.24	7.69	0.72	2.06E-04
SM C18:0	0.12	100.00	0.78	23.11 ± 5.86	25.57 ± 7.21	9.27	0.76	5.57E-05
SM C18:1	0.09	100.00	0.85	11.92 ± 3.13	13.1 ± 3.64	9.15	0.77	7.95E-05
SM C24:0	0.11	99.87	0.82	19.9 ± 4.51	21.37 ± 6.03	5.85	0.75	4.84E-03
SM C24:1	0.12	100.00	0.78	47.75 ± 12.04	52.22 ± 14.54	8.12	0.75	4.02E-04
H1	0.10	99.87	0.86	5586.66 ± 1412.33	6009.7 ± 1567.69	6.99	0.88	1.77E-06
The metabolites below are excluded for analysis								
C10:2	0.20	72.55	0.34	0.04 ± 0.01	0.05 ± 0.01	11.1	0.04	9.20E-08

C12-DC	0.14	0.00	0.03	0.06 ± 0.02	0.06 ± 0.02	-1.9	0.06	2.29E-04
C14	0.11	71.22	0.37	0.06 ± 0.01	0.07 ± 0.02	-1.8	0.06	2.12E-07
C14:1-OH	0.12	72.55	0.48	0.02 ± 0	0.02 ± 0.01	-11.8	0.02	1.60E-05
C14:2-OH	0.14	32.89	0.19	0.01 ± 0	0.01 ± 0	8	0.01	9.21E-06
C16-OH	0.23	2.65	0.45	0.01 ± 0	0.01 ± 0	10.5	0.01	3.73E-02
C16:1	0.14	2.52	0.82	0.05 ± 0.01	0.05 ± 0.01	9	0.05	1.72E-02
C16:1-OH	0.13	5.17	0.31	0.01 ± 0.01	0.01 ± 0.01	0	0.01	1.36E-04
C16:2	0.17	74.14	0.39	0.01 ± 0	0.01 ± 0	-13.3	0.01	3.14E-02
C16:2-OH	0.18	9.15	0.23	0.01 ± 0	0.01 ± 0	-7.4	0.01	3.55E-01
C18:1-OH	0.20	0.00	0.27	0.01 ± 0	0.01 ± 0	-30.8	0.01	1.13E-02
C3-DC (C4-OH)	0.21	27.98	0.57	0.06 ± 0.03	0.06 ± 0.03	-19.3	0.06	1.58E-07
C5-OH (C3-DC-M)	0.25	48.01	0.06	0.02 ± 0	0.02 ± 0	-21.7	0.02	3.54E-04
C3:OH	0.42	0.13	-0.08	0.01 ± 0	0.01 ± 0	19	0.01	3.83E-02
C3:1	0.36	0.00	0.32	0.02 ± 0.01	0.02 ± 0.01	-10.5	0.02	2.78E-05
C4:1	0.20	19.63	0.52	0.02 ± 0.01	0.02 ± 0.01	0	0.02	6.10E-06
C6 (C4:1-DC)	0.20	80.24	0.41	0.03 ± 0.01	0.03 ± 0.01	0	0.03	1.30E-01
C5-DC (C6-OH)	0.19	64.19	0.21	0.03 ± 0.01	0.04 ± 0.01	-5.1	0.03	1.58E-03
C5-M-DC	0.33	1.72	0.36	0.04 ± 0.01	0.04 ± 0.02	4.1	0.04	6.44E-06
C5:1	0.21	0.40	0.31	0.01 ± 0	0.01 ± 0	7.1	0.01	2.65E-03
C5:1-DC	0.32	24.93	0.90	0.08 ± 0.03	0.09 ± 0.04	-8.7	0.08	4.92E-05
C6:1	0.22	0.00	0.34	0.02 ± 0.01	0.02 ± 0.01	0	0.02	9.84E-01
C7-DC	0.12	62.07	0.69	0.04 ± 0.01	0.05 ± 0.02	4.9	0.04	7.08E-10
C9	0.16	86.60	0.63	0.06 ± 0.02	0.06 ± 0.03	1.7	0.06	4.34E-01
PC aa C26:0	0.21	6.63	0.23	0.09 ± 0.02	0.1 ± 0.03	15.4	0.09	1.98E-01
PC aa C30:2	0.55	73.87	-0.07	0.01 ± 0.04	0.03 ± 0.07	200	0.01	1.78E-01
PC aa C40:2	0.26	99.87	0.50	0.33 ± 0.09	0.37 ± 0.1	-26.2	0.33	3.84E-05

PC aa C40:3	0.27	99.87	0.67	0.56 ± 0.15	0.63 ± 0.18	-10.9	0.56	1.35E-08
PC aa C42:4	0.25	99.87	0.52	0.19 ± 0.04	0.21 ± 0.04	-24.6	0.19	3.02E-06
PC ae C30:1	0.41	69.89	0.33	0.06 ± 0.08	0.07 ± 0.11	-191.1	0.06	8.02E-01
PC ae C36:0	0.34	99.87	0.76	0.88 ± 0.24	1.03 ± 0.32	-23.7	0.88	6.47E-10
PC ae C42:0	0.32	18.04	0.50	0.49 ± 0.09	0.53 ± 0.11	-7.8	0.49	2.98E-06
LPC a C24:0	0.21	23.61	0.13	0.21 ± 0.05	0.22 ± 0.05	13.7	0.21	1.20E-01
LPC a C26:0	0.27	30.90	-0.06	0.26 ± 0.08	0.29 ± 0.08	-2.3	0.26	1.54E-03
LPC a C26:1	0.10	0.00	0.07	1.98 ± 0.57	1.99 ± 0.54	2.1	0.29	8.43E-01
LPC a C28:1	0.26	98.41	0.64	0.44 ± 0.15	0.49 ± 0.17	7.5	0.44	5.43E-05
LPC a C6:0	0.32	29.97	-0.08	0.01 ± 0.01	0.01 ± 0.01	6.5	0.01	7.92E-01
SM C20:2	0.31	99.87	0.44	0.27 ± 0.13	0.31 ± 0.17	-2.4	0.27	2.08E-03
SM C22:3	0.77	88.20	0.09	0.01 ± 0.03	0.03 ± 0.08	-3.85	0.01	1.71E-01
SM C26:0	0.52	100.00	0.63	0.19 ± 0.07	0.2 ± 0.08	80.9	0.19	1.12E-01
SM C26:1	0.40	99.87	0.80	0.41 ± 0.14	0.43 ± 0.17	15.2	0.41	3.21E-01

2.3.4.2 KORA S4

For each kit plate, five references (human plasma pooled material, Seralab) and three zero samples (PBS) were measured in addition to the KORA samples. To ensure data quality, each metabolite had to meet the following two criteria: (1) the coefficient of variance (CV) for the metabolite in the total 110 reference samples should be smaller than 25%. In total, seven outliers were removed because their concentrations were larger than the mean plus 5s.d.; (2) 50% of all measured sample concentrations for the metabolite should be above the limit of detection (LOD), which is defined as 3 times median of the three zero samples. In total, 140 metabolites passed the quality controls (Table 3): one hexose (H1), 21 acylcarnitines, 21 amino acids, 8 biogenic amines, 13 SMs, 33 diacyl (aa) PCs , 35 acyl-alkyl (ae) PCs and 8 LPCs. Concentrations of all analyzed metabolites are reported in mM.

Table 3: Characteristics of the 188 targeted metabolites in KORA S4 measured by Absolute*IDQ*™ kit p180 and the 163 metabolites in KORA F4 measured by Absolute*IDQ*™ kit p150

	KORA S4			KORA F4			
Abbreviation	**CV (%)**	**% > LOD**	**Application**	***r***	**% > LOD**	**CV**	**Application**
C0	5.8	99.63	Used	0.88	100.00	6.7%	Used
C2	6.3	99.63	Used	0.94	100.00	9.4%	Used
C3	10.0	99.63	Used	0.86	100.00	8.0%	Used
C3:1	32.8	3.72	Excluded	0.05	0.36	76.6%	Excluded
C3-OH	44.7	2.85	Excluded	-0.11	0.10	37.5%	Excluded
C4	9.7	99.63	Used	0.89	100.00	8.8%	Used
C4:1	22.2	46.25	Excluded	0.04	5.65	34.7%	Excluded
C4-OH (C3-DC)	21.1	18.95	Excluded	0.47	8.40	35.5%	Excluded
C5	10.8	98.70	Used	0.81	95.56	14.2%	Used
C5:1	22.9	1.80	Excluded	0.37	0.75	26.1%	Excluded

C5:1-DC	40.0	24.83	Excluded	0.13	12.48	42.4%	Excluded
C5-DC (C6-OH)	29.4	61.36	Excluded	0.15	27.06	21.0%	Excluded
C5-M-DC	28.0	2.48	Excluded	0.18	0.95	42.9%	Excluded
C5-OH (C3-DC-M)	26.9	19.69	Excluded	0.25	55.10	28.7%	Excluded
C6(C4:1-DC)	21.8	65.33	Used	0.85	76.67	13.6%	Used
C6:1	30.7	5.20	Excluded	0.07	0.33	32.4%	Excluded
C7-DC	18.4	70.53	Used	0.79	61.34	34.4%	Excluded
C8	13.2	60.62	Used	0.89	51.54	16.3%	Used
C8:1				0.92	96.01	8.4%	Used
C9	23.6	97.28	Used	0.84	83.73	20.8%	Used
C10	11.7	99.07	Used	0.93	94.08	11.4%	Used
C10:1	11.2	74.80	Used	0.83	48.66	10.4%	Used
C10:2	16.0	94.86	Used	0.51	50.49	14.5%	Used
C12	12.2	96.41	Used	0.86	87.35	10.4%	Used
C12:1	15.2	26.75	Excluded	0.73	13.69	13.0%	Used
C12-DC	12.3	0.00	Excluded	0.05	0.00	12.2%	Excluded
C14	15.8	96.66	Used	0.54	51.67	12.6%	Used
C14:1	11.4	99.63	Used	0.81	100.00	16.9%	Used
C14:1-OH	28.9	74.92	Excluded	0.70	67.35	16.4%	Used
C14:2	18.3	98.33	Used	0.87	98.82	11.6%	Used
C14:2-OH	35.1	47.00	Excluded	0.27	38.04	17.4%	Excluded
C16	11.3	99.63	Used	0.84	100.00	8.9%	Used
C16:1	18.1	77.83	Used	0.71	2.78	10.2%	Used
C16:1-OH	26.5	26.01	Excluded	0.38	2.25	17.5%	Excluded
C16:2	34.0	87.49	Excluded	0.57	70.69	19.4%	Used
C16:2-OH	30.1	5.76	Excluded	0.32	4.67	16.6%	Excluded
C16-OH	33.0	16.28	Excluded	0.20	3.33	24.1%	Excluded
C18	15.7	99.63	Used	0.69	99.80	13.7%	Used
C18:1	9.7	99.57	Used	0.87	98.33	10.2%	Used
C18:1-OH	44.6	7.37	Excluded	0.06	0.95	33.4%	Excluded
C18:2	10.5	99.57	Used	0.81	100.00	9.4%	Used
Ala	13.7	99.50	Used				
Arg	13.2	99.26	Used	0.59	100.00	8.2%	Used
Asn	11.1	99.57	Used				
Asp	12.2	99.44	Used				
Cit	12.7	99.44	Used				
Gln	12.8	99.57	Used	0.62	100.00	9.9%	Used
Glu	15.8	99.57	Used				
Gly	13.2	99.50	Used	0.89	100.00	7.9%	Used
His	12.9	99.38	Used	0.69	100.00	8.3%	Used
Ile	13.9	99.63	Used				
Leu	12.9	98.58	Used				
xLeu				0.74	100.00	8.2%	Used
Lys	15.5	99.69	Used				
Met	13.5	99.69	Used	0.53	100.00	9.7%	Used
Orn	14.9	99.63	Used	0.75	100.00	9.4%	Used
Phe	12.2	99.57	Used	0.62	100.00	8.4%	Used
Pro	11.8	99.63	Used	0.89	100.00	7.4%	Used

Ser	13.6	99.44	Used	0.62	100.00	9.6%	Used
Thr	18.3	99.13	Used	0.71	100.00	12.1%	Used
Trp	12.9	99.63	Used	0.51	100.00	7.5%	Used
Tyr	14.7	99.57	Used	0.66	100.00	8.6%	Used
Val	13.5	99.63	Used	0.69	100.00	19.6%	Used
Ac-Orn	20.8	79.07	Used				
ADMA	17.4	66.50	Used				
SDMA	32.4	97.34	Excluded				
total-DMA	20.3	99.20	Used				
alpha-AAA	32.0	97.34	Excluded				
Carnosine	89.8	4.02	Excluded				
Creatinine	14.7	99.38	Used				
Histamine	43.5	89.97	Excluded				
Kynurenine	11.3	97.28	Used				
Met-SO	20.9	96.66	Used				
Nitro-Tyr	58.4	7.55	Excluded				
OH-Pro	NA	2.11	Excluded				
PEA	NA	0.56	Excluded				
Putrescine	53.2	93.75	Excluded				
Sarcosine	28.7	4.40	Excluded				
Serotonin	38.0	99.32	Excluded				
Spermidine	24.1	98.51	Used				
Spermine	8.5	9.29	Excluded				
Taurine	13.7	96.90	Used				
DOPA	19.5	44.58	Excluded				
Dopamine	NA	0.06	Excluded				
LPC a C14:0	6.8	0.00	Excluded	0.45	21.24	23.8%	Excluded
LPC a C16:0	6.9	99.81	Used	0.75	100.00	8.8%	Used
LPC a C16:1	7.0	99.69	Used	0.84	100.00	8.6%	Used
LPC a C17:0	7.3	99.63	Used	0.84	100.00	12.7%	Used
LPC a C18:0	7.2	99.81	Used	0.80	100.00	9.7%	Used
LPC a C18:1	6.8	99.75	Used	0.84	100.00	9.2%	Used
LPC a C18:2	6.9	99.75	Used	0.93	100.00	8.8%	Used
LPC a C20:3	8.8	99.63	Used	0.77	100.00	9.0%	Used
LPC a C20:4	7.3	99.69	Used	0.87	100.00	9.0%	Used
LPC a C24:0	32.0	23.22	Excluded	0.09	12.45	21.1%	Excluded
LPC a C26:0	44.4	43.72	Excluded	0.09	59.58	31.0%	Excluded
LPC a C26:1	9.5	0.00	Excluded	-0.04	0.00	7.9%	Excluded
LPC a C28:0	37.0	23.47	Excluded	0.17	49.61	29.1%	Excluded
LPC a C28:1	35.5	98.64	Excluded	0.29	99.84	22.6%	Excluded
LPC a C6:0				-0.14	33.33	62.5%	Excluded
PC aa C24:0	45.9	69.35	Excluded	0.11	72.55	26.5%	Excluded
PC aa C26:0	27.2	5.63	Excluded	0.09	11.54	32.9%	Excluded
PC aa C28:1	9.5	99.63	Used	0.87	100.00	9.8%	Used
PC aa C30:0	9.4	99.63	Used	0.89	100.00	7.8%	Used
PC aa C30:2	89.9	31.33	Excluded	0.12	4.22	81.6%	Excluded
PC aa C32:0	8.4	99.81	Used	0.83	100.00	7.1%	Used
PC aa C32:1	9.2	99.81	Used	0.96	100.00	7.4%	Used
PC aa C32:2	12.3	99.81	Used	0.91	99.93	11.1%	Used

PC aa C32:3	9.2	99.75	Used	0.79	100.00	8.9%	Used
PC aa C34:1	7.1	99.88	Used	0.83	100.00	7.2%	Used
PC aa C34:2	7.0	99.88	Used	0.75	100.00	7.7%	Used
PC aa C34:3	6.3	99.88	Used	0.91	100.00	8.6%	Used
PC aa C34:4	6.8	99.81	Used	0.92	100.00	8.0%	Used
PC aa C36:0	11.6	99.63	Used	0.74	100.00	17.4%	Used
PC aa C36:1	6.9	99.88	Used	0.84	100.00	8.5%	Used
PC aa C36:2	6.6	99.88	Used	0.80	100.00	6.7%	Used
PC aa C36:3	6.5	99.88	Used	0.86	100.00	7.5%	Used
PC aa C36:4	6.3	99.94	Used	0.87	100.00	7.8%	Used
PC aa C36:5	6.7	99.81	Used	0.82	100.00	8.6%	Used
PC aa C36:6	9.5	99.75	Used	0.89	100.00	11.1%	Used
PC aa C38:0	8.8	99.63	Used	0.86	100.00	13.8%	Used
PC aa C38:1	27.0	99.75	Excluded	0.34	99.84	18.1%	Excluded
PC aa C38:3	6.9	99.88	Used	0.86	100.00	7.6%	Used
PC aa C38:4	5.7	99.88	Used	0.88	100.00	7.3%	Used
PC aa C38:5	5.6	99.88	Used	0.83	100.00	7.9%	Used
PC aa C38:6	6.9	100.00	Used	0.93	100.00	8.1%	Used
PC ae C40:0				0.87	1.05	4.8%	Used
PC aa C40:1	11.7	14.24	Excluded	0.51	8.66	13.5%	Used
PC aa C40:2	14.9	99.63	Used	0.51	100.00	11.7%	Used
PC aa C40:3	13.9	99.75	Used	0.60	100.00	11.2%	Used
PC aa C40:4	6.8	99.81	Used	0.86	100.00	7.6%	Used
PC aa C40:5	6.5	99.75	Used	0.89	100.00	7.0%	Used
PC aa C40:6	6.1	99.63	Used	0.93	100.00	7.1%	Used
PC aa C42:0	9.2	99.88	Used	0.85	99.97	12.3%	Used
PC aa C42:1	12.0	99.69	Used	0.72	100.00	14.8%	Used
PC aa C42:2	13.5	99.69	Used	0.56	100.00	14.6%	Used
PC aa C42:4	11.0	99.81	Used	0.51	100.00	11.7%	Used
PC aa C42:5	11.3	99.69	Used	0.75	100.00	10.6%	Used
PC aa C42:6	10.7	95.42	Used	0.62	60.16	12.5%	Used
PC ae C30:0	19.7	99.57	Used	0.76	98.86	18.1%	Used
PC ae C30:1	77.9	82.35	Excluded	0.18	94.12	41.7%	Excluded
PC ae C30:2	25.2	99.57	Excluded	0.65	86.34	17.5%	Used
PC ae C32:1	9.3	99.81	Used	0.83	100.00	8.0%	Used
PC ae C32:2	12.2	99.63	Used	0.77	100.00	11.6%	Used
PC ae C34:0	9.6	99.81	Used	0.82	100.00	7.9%	Used
PC ae C34:1	7.4	99.81	Used	0.87	100.00	7.6%	Used
PC ae C34:2	7.2	99.88	Used	0.90	100.00	7.6%	Used
PC ae C34:3	6.9	99.88	Used	0.91	100.00	7.9%	Used
PC ae C36:0	22.7	99.63	Used	0.35	100.00	35.6%	Excluded
PC ae C36:1	7.9	99.75	Used	0.85	100.00	9.8%	Used
PC ae C36:2	7.0	99.88	Used	0.92	100.00	8.3%	Used
PC ae C36:3	7.1	99.88	Used	0.86	100.00	8.1%	Used
PC ae C36:4	6.3	99.88	Used	0.87	100.00	7.9%	Used
PC ae C36:5	6.1	99.81	Used	0.89	100.00	8.0%	Used
PC ae C38:0	8.1	99.63	Used	0.81	100.00	10.8%	Used
PC ae C38:1	14.7	99.50	Used	0.48	100.00	12.4%	Used
PC ae C38:2	11.7	99.75	Used	0.73	100.00	10.3%	Used

PC ae C38:3	7.0	99.94	Used	0.85	100.00	9.2%	Used
PC ae C38:4	6.1	100.00	Used	0.82	100.00	8.6%	Used
PC ae C38:5	5.9	100.00	Used	0.82	100.00	8.3%	Used
PC ae C38:6	6.5	99.88	Used	0.85	100.00	8.1%	Used
PC ae C40:1	11.1	99.63	Used	0.68	100.00	10.5%	Used
PC ae C40:2	8.3	99.88	Used	0.85	100.00	9.5%	Used
PC ae C40:3	9.0	99.94	Used	0.73	100.00	9.5%	Used
PC ae C40:4	8.7	99.63	Used	0.82	100.00	9.6%	Used
PC ae C40:5	6.5	99.88	Used	0.78	100.00	8.3%	Used
PC ae C40:6	6.9	99.94	Used	0.88	100.00	8.6%	Used
PC ae C42:0	13.8	36.35	Excluded	0.60	14.87	15.7%	Used
PC ae C42:1	16.0	99.57	Used	0.51	100.00	11.5%	Used
PC ae C42:2	11.5	99.69	Used	0.69	100.00	12.8%	Used
PC ae C42:3	9.8	99.88	Used	0.80	100.00	10.8%	Used
PC ae C42:4	7.8	99.63	Used	0.78	100.00	9.2%	Used
PC ae C42:5	7.4	99.57	Used	0.86	99.97	7.4%	Used
PC ae C44:3	24.3	99.69	Used	0.50	100.00	12.5%	Used
PC ae C44:4	12.1	99.69	Used	0.71	100.00	11.4%	Used
PC ae C44:5	7.4	99.69	Used	0.86	100.00	8.0%	Used
PC ae C44:6	7.8	99.63	Used	0.89	100.00	7.7%	Used
SM (OH) C14:1	11.0	99.63	Used	0.91	100.00	7.7%	Used
SM (OH) C16:1	11.0	100.00	Used	0.86	100.00	8.8%	Used
SM (OH) C22:1	11.2	99.88	Used	0.82	100.00	11.2%	Used
SM (OH) C22:2	11.2	99.88	Used	0.87	100.00	10.3%	Used
SM (OH) C24:1	15.1	99.75	Used	0.75	100.00	15.1%	Used
SM C16:0	10.6	99.88	Used	0.73	100.00	8.0%	Used
SM C16:1	9.9	99.88	Used	0.84	100.00	7.5%	Used
SM C18:0	9.8	99.81	Used	0.79	100.00	9.0%	Used
SM C18:1	9.4	99.88	Used	0.84	100.00	8.2%	Used
SM C20:2	16.2	99.81	Used	0.61	99.93	12.6%	Used
SM C22:3	NA	0.37	Excluded	-0.04	55.85	57.6%	Excluded
SM C24:0	11.9	99.75	Used	0.78	100.00	10.7%	Used
SM C24:1	12.1	99.88	Used	0.75	100.00	10.0%	Used
SM C26:0	31.8	99.81	Excluded	0.46	100.00	67.8%	Excluded
SM C26:1	21.2	99.75	Used	0.69	100.00	20.8%	Used
H1	5.2	99.81	Used	0.69	100.00	6.3%	Used

2.3.4.3 KORA F4

To ensure data quality, metabolites had to meet three criteria: (1) average value of coefficient of variance (CV) of the three QCs should be smaller than 25%. (2) 90% of all measured sample concentrations should be above the limit of detection (LOD). (3)

Correlation coefficients between two duplicated measurements of 144 re-measured samples should be above 0.5 (Table 3). In total, 131 metabolites passed the three quality controls.

2.4 Gene expression profiling

Peripheral blood was drawn under fasting conditions from 599 KORA S4 individuals at the same time as the serum samples used for metabolic profiling were prepared. Blood samples were collected directly in PAXgene (TM) Blood RNA tubes (PreAnalytiX). The RNA extraction was performed using the PAXgene Blood miRNA kit (PreAnalytiX). Purity and integrity of RNA was assessed on the Bioanalyzer (Agilent) with the 6000 Nano LabChip reagent set (Agilent). In all, 500 ng of RNA was reverse-transcribed into cRNA and biotin-UTP labeled, using the Illumina TotalPrep-96 RNA Amplification Kit (Ambion). In all, 3000 ng of cRNA was hybridized to the Illumina HumanHT-12 v3 Expression BeadChip. Chips were washed, detected and scanned according to manufacturer's instructions. Raw data were exported from the Illumina 'GenomeStudio' Software to R. The data were converted into logarithmic scores and normalized using the quantile method (Bolstad et al., 2003). The sample sets comprised 383 individuals with NGT, 104 with IGT and 26 with dT2D. The known T2D individuals were removed as had been done for the metabolomics analysis.

2.5 Statistical analysis

All statistical calculations were performed under the R statistical environment (http://www.r-project.org/).

2.5.1 Delta (difference in metabolite concentration means for males and females).

For comparison of metabolite concentrations between males and females we used the delta (Δ), as it describes the difference in concentration means for males and females for a specific metabolite relative to the mean metabolite concentration in males. Therefore the difference of mean metabolite concentration in males and mean metabolite concentration in females is calculated and then divided by the mean metabolite concentration in males. For example, a value of Δ = 50% means that the mean metabolite concentration in females is 50% lower than that in males.

2.5.2 Correlations

A correlation exists between two variables when one of them is related to the other. Pearson's (product moment) correlation coefficient (*r*) measures the strength of the linear relationship between the paired x- and y-quantitative values in a sample (Triola et al., 2006). Its value is computed as:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}}$$

Where n is the number of pairs of data present.

In order to investigate how strong the different metabolites correlate with each other and the sex-specific effects propagate through the underlying metabolic network, we calculated full-order partial correlation coefficients between all pairs of metabolites. The resulting partial correlation networks are commonly referred to as Gaussian graphical models (GGMs), which we have previously demonstrated to be useful for the analysis of direct metabolite-metabolite effects in the same population cohort (Krumsiek et al., 2011).

2.5.3 Regression

2.5.3.1 Linear regression

Metabolite concentration differences between males and females were investigated by linear regression analysis. The basic model contains the log-transformed metabolite as dependent variable and sex as explanatory variable with both age and BMI as covariates. To correct for multiple testing, the Bonferroni-correction was applied. The *P*-value cutoff for significance was set at 0.05/131 = 3.84E-4. In the replication, we also applied Bonferroni-correction.

Moreover, we also analyzed the influence of anthropometric phenotypes, diseases and environmental factors by including different covariates to the linear regression and comparison of the structure of the results. Four models which differed in the use of one or more additional covariates were performed. The covariates in each model beside age are waist hip ratio (WHR), lipid parameters (HDL and LDL cholesterol, triglycerides), T2D, alcohol consumption and smoking. Furthermore, a meta-analysis of the discovery and the replication sample with a fixed effect model was analyzed to reveal the sex-specific effects of metabolite concentrations.

Associations between metabolite concentrations and 2-h glucose value were also explored by linear regression. β estimates were calculated from the regression analyses. The concentration of each metabolite was log-transformed and normalized to have a mean of zero and a standard deviation (s.d.) of one. Various risk factors in the linear regression were added as covariates, and the same significance level 3.6E-4 was adopted.

2.5.3.2 Logistic regression

Logistic regression was used to identify metabolites showed significantly different concentrations between groups when we look for early biomarkers of T2D. Odds ratios (ORs) for single metabolites were calculated between two groups. The concentration of each metabolite was scaled to have a mean of zero and an s.d. of one; thus, all reported OR values correspond to the change per s.d. of metabolite concentration. Various T2D risk factors were added to the logistic regression analysis as covariates. To handle false discovery from multiple comparisons, the cutoff point for significance was calculated according to the Bonferroni correction, at a level of 3.6E-4 (for a total use of 140 metabolites at the 5% level). Because the metabolites were correlated within well-defined biological groups (e.g., 8 LPCs, 33 diacyl PCs, 35 acyl-alkyl PCs and 13 SMs), this correction was considered conservative. Additionally, the categorized metabolite concentrations and combined scores (see below) were analyzed, and the ORs were calculated across quartiles. To test the trend across quartiles, we assigned all individuals either the median value of the concentrations or the combined scores, and obtained the *P*-values using the same regression model.

2.5.3.3 Combination of metabolites

After identified early biomarkers for T2D, we obtain the combined scores of these metabolites, the scaled metabolite concentrations (mean = 0, s.d. = 1) were first modeled with multivariate logistic regression containing all confounding variables. The coefficients of these metabolites from the model were then used to calculate a weighted sum for each individual. In accordance with the decreasing trend of glycine and LPC (18:2), we inverted these values as the combined scores.

2.5.3.4 Residuals of metabolite concentrations

To avoid the influence of other confounding factors when plotting the concentration of metabolites, we used the residuals from a linear regression model. Metabolite concentrations were log-transformed and scaled (mean = 0, s.d. = 1), and the residuals were then deduced from the linear regression that included the corresponding confounding factors.

2.5.4 Machine learning methods

2.5.4.1 Random forest stepwise selection methods and candidate biomarker selection

To select candidate biomarkers, we applied two more methods, the random forest (Breiman, 2001) and stepwise selection, which assess the metabolites as a group while logistic regression evaluates one metabolite at a time.

Between the NGT and the IGT groups, supervised classification method random forest was first used to select the metabolites among the 30 highest ranking variables of importance score, meaning they can best separate the individuals between the two groups. These metabolites showed most impact on whether or not individuals can be assigned correctly to their diabetes status in the internal permutation test of random forest. T2D risk indicators (i.e. age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP, HDL cholesterol, HbA$_{1c}$, fasting glucose, fasting insulin) were also included in this method with all the metabolites.

We further select the metabolites using stepwise selection on the logistic regression model. Metabolites which showed significantly different concentrations between the compared groups in logistic regression and also being selected using random forest were used in this model along with all the risk indicators. They were added and dropped from the model one by one. Akaike's Information Criterion (AIC) was used to evaluate the performance of these subsets of metabolites used in the models. The model with minimal AIC was chosen and metabolites left in this model are the potential independent markers to best distinguish IGT from NGT individuals and the correlated metabolites with less separation power were dropped. The area under the receiver-operating-characteristic curves (AUC) was used to evaluate the models and a likelihood ratio test was used to compare the models.

2.5.4.2 Partial least square analysis

Partial least square (PLS) (Lorber et al., 1987), or projection to latent structures by means of partial least squares is a supervised machine learning method. It relates a matrix X to a vector y (or to a matrix Y). The x-data are transformed into a set of a few intermediate linear latent variables (components) using linear combination. The purpose is to maximize the covariance between the components and the vector y (or matrix Y).

The PLS analysis was carried out using the R package pls to investigate the metabolic profiles serum and plasma as well as of males and females. The concentrations of each metabolite were transformed into a mean of zero and an s.d. of one before the analysis. Data was visualized by plotting the scores of the first two components against each other, where each point represented an individual (serum/plasma or male/female) sample.

2.5.5 Network analysis

Metabolite–protein interactions from the Human Metabolome Database (HMDB) (Wishart et al., 2009) and protein–protein interactions in the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (Szklarczyk et al., 2011) were used to construct a network containing relationships between metabolites, enzymes, other proteins and

T2D-related genes. The candidate metabolites were assigned to HMDB IDs using the metaP-Server (Kastenmüller et al., 2011), and their associated enzymes were derived according to the annotations provided by HMDB. These enzymes were connected to the 46 T2D related genes (considered at that point), allowing for one intermediate protein (proteins other than the T2D related genes or the integrating enzymes) through STRING protein functional interactions and optimized by eliminating edges with a STRING score of 0.7 and undirected paths. The sub-networks were connected by the shortest path from metabolites to T2D-related genes.

Chapter 3 Results

This chapter is divided into three parts. The first two parts clarify the potential influence of blood matrices and sex on metabolic variation. The third part presents the results on finding early biomarker for T2D as well as the attempts to find the potential underlying mechanism.

3.1 Metabolomics differences between EDTA plasma and serum

We analyzed the concentrations of 122 metabolites after quality control in both EDTA plasma and serum collected from 377 German participants of the KORA F3 study (Holle et al., 2005; Wang-Sattler et al., 2008). These plasma and serum samples were measured separately in 10 plates. In order to reduce potential bias and authenticate our findings, we randomly chose 83 participants from these 377 individuals and measured the metabolite concentration profiles in two further plates with the same technology, this time, including both plasma samples and their corresponding serum samples from each person within the same plates. All these relatedly measured samples were randomly distributed on the plates.

3.1.1 Good reproducibility in serum and better in plasma

Both plasma and serum samples which displayed good stability in the metabolites were measured. The metabolite concentrations from the repeated measurements on the 83 samples showed a high correlation between the first and the second measurements (Figure 1) with mean Person's correlation coefficients (*r*) of all the

122 metabolites being 0.83 and 0.80 for plasma and serum, respectively. Most of the metabolites showed an *r* value higher than 0.6 except for a few outliers. The reproducibility was significantly better for plasma than for serum (*P* = 0.01, paired t-test), despite that the absolute mean differences in *r* values were rather small.

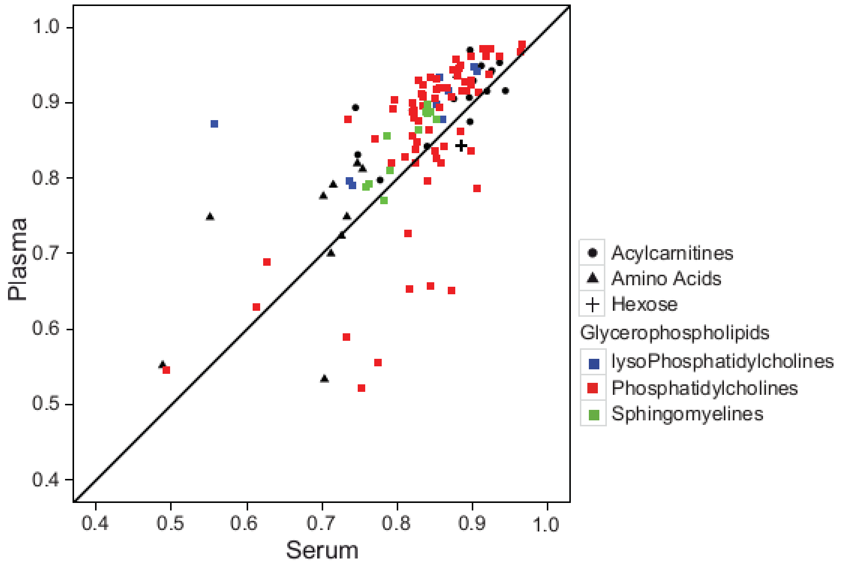

Figure 1: Correlation between repeated measurements of plasma and serum metabolites.

Pearson's correlation coefficients (*r*) between repeated measurements of metabolite concentrations were plotted. *r* values in serum are plotted against *r* values in plasma. Different shapes represent different groups of metabolites: solid circle for acylcarnitines, triangle for amino acids, cross for hexose, and square for glycerophospholipid. Different colors of squares represent different subgroups of glycerophospholipids: blue for lyso-phosphatidylcholine, red for phosphatidylcholine, and green for sphingomyeline.

3.1.2 High correlation between plasma and serum metabolite concentrations and higher concentrations in serum.

Results showed that metabolite concentrations were generally higher in serum than in plasma (Figure 2). Out of the 122 metabolites we analyzed, 104 (85%) have significantly higher concentrations (t-test) in serum and the average value of the relative difference over all metabolites was around 11.7% higher in serum.

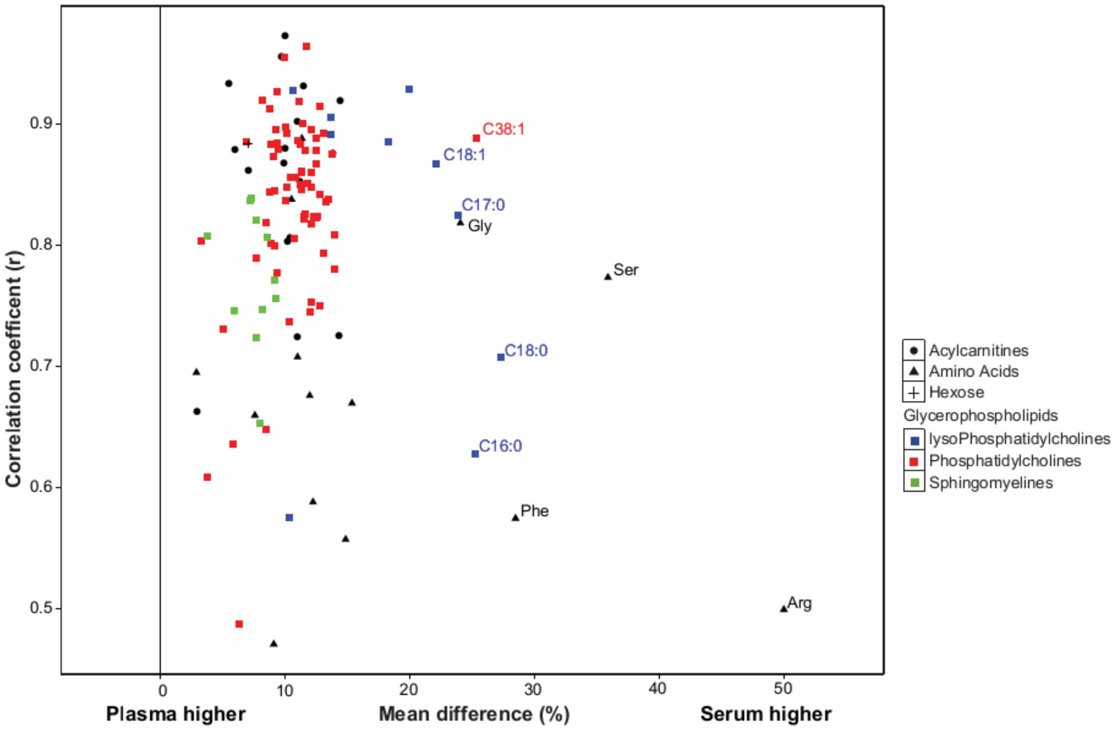

Figure 2: Relative concentration differences and correlation coefficients between plasma and serum for individual metabolites

The X-axis indicates the mean value of the relative concentration difference. Shapes represent different groups of metabolites: Acylcarnitines (•), Amino acids (▴), Hexose (+), and Glycerophospholipid (▪). Colors represent different subgroups of glycerophospholipids: lysoPhosphatidylcholine (blue), Phosphatidylcholine (red), and Sphingomyeline (green). Metabolite names are indicated for metabolites with a mean relative concentration difference larger than 20%.

We also performed a PLS analysis on 377 KORA individuals. The result demonstrated that plasma samples were clearly separated from serum samples (Figure 3). In addition, we observed an overall high correlation (mean r = 0.816 ± 0.1) between the values in these two matrices, indicating that the differences of metabolite concentrations between both matrices are due to systematic changes across all individuals. This is especially true for most acylcarnitines (mean r = 0.866 ± 0.09) and glycerophospholipids (mean r = 0.826 ± 0.09). However, for amino acids, the correlation between the two matrices was significantly lower (mean r = 0.676 ± 0.13) compared to all the metabolites (p = 0.004, t-test) (Figure 2). Among the metabolites with significantly higher concentrations in serum, nine metabolites had relative concentration differences greater than 20% (Figure 2). Arginine had the highest concentration difference, displaying a nearly 50% higher concentration in serum with a lower correlation (r = 0.50) between the two matrices, while diacyl PC C38:1, which was 26% higher in serum than in plasma, still kept a good correlation (r = 0.88). Four LPC (C16:0, C17:0, C18:0, C18:1) and three other amino acids (serine, phenylalanine, glycine) were also found to have more than 20% higher concentrations in serum. Moreover, from the PLS result (Figure 3) we observed similar shapes of both the serum samples and the plasma samples, even though they were clustered into two groups. The size of the group of serum samples was larger than that of the plasma group. These observations were consistent with the high correlation between metabolite concentrations in plasma and serum and a higher absolute concentration in serum.

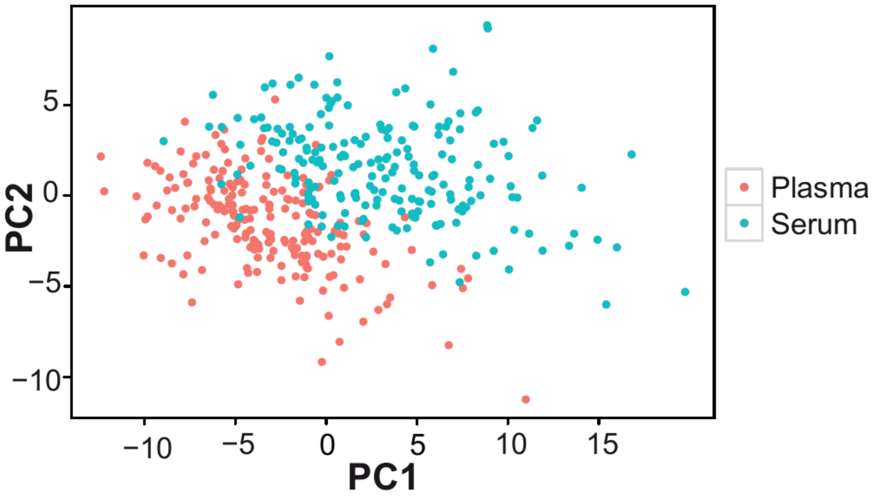

Figure 3: Separation of plasma and serum metabolite profiles

The dot plot presented the results from the partial least squares (PLS) analysis. Scores of the first two PLS components were plotted against each other. Each point indicates either a plasma (red) or serum (blue) sample.

3.1.3 Higher sensitivity in serum

We also noticed that serum provided higher sensitivity than plasma, when metabolite concentrations were compared between subjects with different phenotypes. For example, 40 metabolites in serum were identified to have a significantly different mean concentration in T2D patients vs. non-diseased individuals, while plasma only revealed 25. Similar results were also observed when comparing male against female individuals, as well as when comparing smokers against nonsmokers, serum always containing larger number of significantly different metabolites (Table 4). Furthermore, for each of the three phenotypes, all significantly different metabolites that were identified using plasma were among those identified using serum. The metabolites that failed to be identified in plasma were, nevertheless, close to the borderline of significance.

Table 4: Numbers of significant different metabolite in plasma and serum

	Plasma (n=377)	Serum (n=377)
T2D (n = 51)vs. non-T2D (n =. 326)	25	40
Males (n = 197) vs. Female (n = 180)	62	69
Smoker (n = 45) vs. non-smoker (n = 332)	4	6

3.2 Sexual dimorphisms in metabolomics

3.2.1 Phenotypic metabotype differences between males and females

All phenotypic analysis steps were performed on population based cohort data of KORA F4 (1452 males and 1552 females) and KORA F3 (197 males and 180 females) with fasting serum concentrations of 131 metabolites after quality control. The metabolites covered a biologically relevant panel that could be divided into five subgroups such as amino acids, hexose, acylcarnitines and phospholipids. A PLS analysis of all metabolites showed that there were major differences in serum metabolite concentrations between males and females, as the first two components from the PLS analysis showed clearly clustered pattern for different sexes (Figure 4). This is true for both the KORA F4 population and the replication samples in KORA F3.

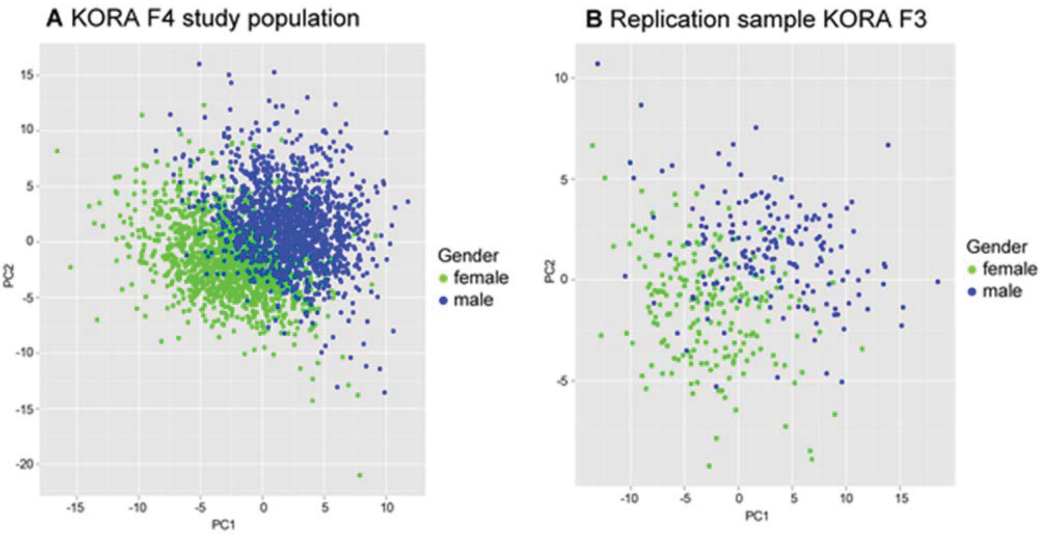

Figure 4: Separation of males and females metabolite profiles

Partial least square analyses show that males and females are clustered into two different groups using the 131 metabolite concentrations in males and females. (A) in KORA F4. (B) in KORA F3. Each point represents an individual and different color stands for different gender: green for female and blue for male.

Motivated by the global gender differences in metabolite concentrations shown by PLS analysis, we further investigate the effect of sex on each metabolite. We performed linear regression with the log-transformed concentration as dependent and sex as the explanatory variable for each metabolite. In the regression model, age and BMI were also used as covariates. The regression results revealed in 102 of the total 131 metabolites (*p*-value below the Bonferroni-corrected significance level of 3.86×10^{-4}) significant effects of gender. Moreover, at least one metabolite in each subgroup including amino acids, acylcarnitines, PCs, LPCs and SMs showed significant sex-specific differences in metabolite concentrations.

The linear regression analysis showed that the concentrations of most amino acids were significantly higher in males except for glycine (effect of sex: β = -0.13, *P*-value = 2.36 x 10^{-46}) and serine (effect of sex: β = -0.13, *P*-value = 1.0 x 10^{-12}). Both of them exhibited higher concentrations in females. The relative sex-specific difference for glycine was Δ = 214%, which means that the mean concentration in men was 114% lower than that in women. The levels of most serum acylcarnitines were significantly higher in males compared to females. The concentrations of PC (both PC ae Cx:y and PC aa Cx:y) tended to be significantly lower in males compared to females. The most significant difference between the two sexes could be seen for the PC aa C32:3 (Δ = 217.9%, *P*-value = 4.4 x 10^{-108}), whereas LPC concentrations were higher in males compared to females. In contrast, the concentrations of most sphingomyelins were significantly lower in males than in females. The concentration of hexose, which is the sum of C6-sugars, was significantly higher in males compared to females (Δ = 7.3%, *p*-value = 6.2 x 10^{-27}).

The adjustment for different covariates (e.g. waist-hip ratio (WHR), HDL (high density lipoprotein), LDL (low density lipoprotein), triglycerides, T2D, smoking, and high alcohol consumption) did not affect the sex-specific differences in the metabolite concentrations extensively. The majority of the high significant sex-effects remained significant. In particular, the adjustments for lipid parameter (HDL, LDL and triglycerides), T2D, smoking, and high alcohol consumption did not influence our main findings. If WHR was included into the linear regression model as covariate instead of

BMI or as an additional covariate in addition to BMI, the *P*-values of the sex-effect on metabolites changed, but for most metabolites the gender differences remained significant. Interestingly, seven PC aa Cx:ys and LPC a C17:0 showed significant differences between sexes while adjusting for age and WHR but not for age and BMI adjustment. We refer the interested reader to Table 6. As replication the same linear regression approach (covariates: age, BMI) was applied to the KORA F3 cohort which included 377 individuals. Despite this smaller sample size for 63 of 102 metabolites with a significant effect of sex in KORA F4, the effect of sex in KORA F3 had the same direction and a significant *P*-value lower than the Bonferroni-corrected replication significance level corrected for the 102 metabolites taken forward to replication (0.05/102 = 4.9×10^{-4}). That means 61.8% of the sex-specific differences could be replicated. A combined meta-analysis of KORA F4 and KORA F3 revealed 113 metabolites with a significant effect of sex (Bonferroni-corrected meta-analysis significance level: *P*-value < 3.86×10^{-4}).

3.2.2 Sex-Specific Effects in the Metabolic Network

We further investigated how groups of metabolites share pairwise correlations, that mean similar effects, and how the sex specific effects propagate through the metabolic network. Therefore we calculated a partial correlation matrix between all metabolites, corrected against age, sex and BMI (Krumsiek et al., 2011). The resulting network, which is also referred to as a Gaussian graphical model, was annotated with the results from the linear regression analysis to get a comprehensive picture of sex-effects in this data-driven metabolic network (Figure 5). We applied a cut-off of $r = 0.3$ (r represents the partial correlation coefficient) in order to emphasize strong inter-metabolite effects. We observed a general structuring of the network into members from similar metabolic classes, e.g. the amino acids, the PC, SM and acylcarnitines (Figure 6). Direct correlations between metabolites, as represented by partial correlation coefficients, are rare in this metabolite panel with only around 1% of all partial correlations showing a strong effect above $r = 0.3$. For this specific cut-off we obtained 14 non-singleton groups, which can be regarded as independently regulated phenotypes within the measured metabolite panel. Detailed description of the distribution of partial correlations and the group structure in the network can be found in Figure 6 and Figure 7. The low connectedness of the network is in line with previous findings (Krumsiek et al., 2011) which demonstrated that Gaussian graphical models are sparsely connected on the one hand, but specifically exclude indirect metabolic interactions on the other hand.

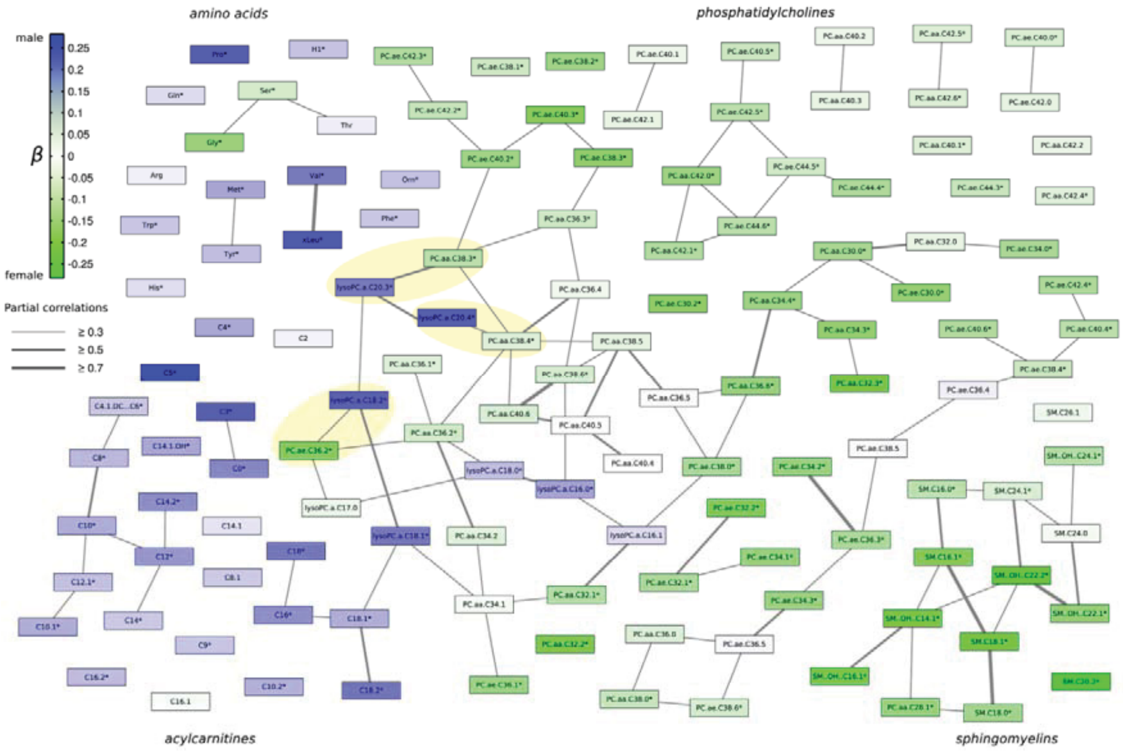

Figure 5: Gaussian graphical model of all measured metabolites illustrating the correlation strength and the propagation of gender-specific effects through the underlying metabolic network

Each node represents one metabolite whereas edge weights correspond to the strength of partial correlation. Only edges with a partial correlation above r = 0.3 are shown. Node coloring represents the strength of association (measured using β from linear regression analysis) towards either males or females. Metabolite names marked with a star * represent significantly different metabolites between genders. Yellow highlighted pairs of metabolites differ by a C18:0 fatty acid residue.

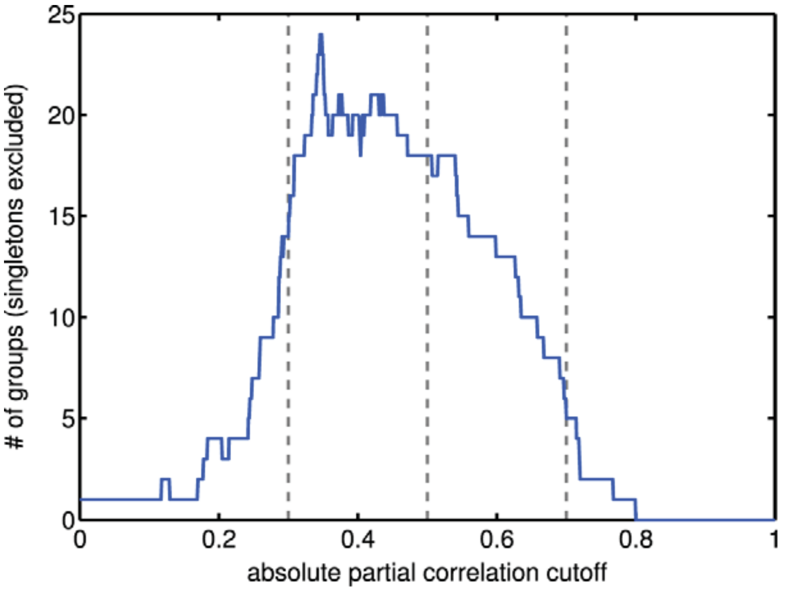

Figure 6: Distribution of partial correlation coefficients

Partial correlations centered around zero with a shift towards positive high values. When applying a correlation cutoff of $r = 0.3$, we are left with 109 out of 8515 correlation values (1.28%)

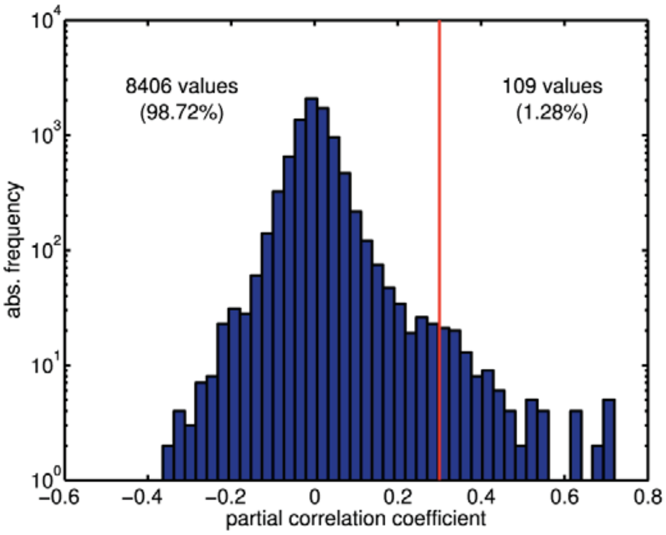

Figure 7: Numbers of clustered groups in the GGM as a function of the absolute partial correlation cutoff

Note that we did not count singleton metabolites without any partial correlation above threshold here. Most non-singleton groups emerge in the cutoff range between 0.3 and 0.7, which corresponds to the Figure in the main manuscript. For our lower cutoff of 0.3, we obtain 14 groups, which can here be regarded as *independent phenotypes* in the metabolite pool

Strikingly, sex-specific effects appear to be localized with respect to metabolic classes and connections in the partial correlation matrix. For instance, while most sphingomyelin concentrations have been shown to be higher in females, we also observe them to be a connected component in the GGM. Similarly, acylcarnitines are higher in males and also share partial correlation edges, mostly with other acylcarnitines (Figure 5). Interestingly, we observed three metabolite pairs from the PC aa and LPC classes, respectively, which constitute a side chain length difference of 18 carbon atoms (yellow shaded metabolite pairs, Figure 5).

3.3 Detecting novel pre-diabetic markers using metabolomics approach

3.3.1 Study participants

Individuals with known T2D were identified by physician validated self-reporting (Rathmann et al., 2010) and excluded from our analysis, to avoid potential influence from anti-diabetic medication with non-fasting participants and individuals with missing values (Figure 8A).

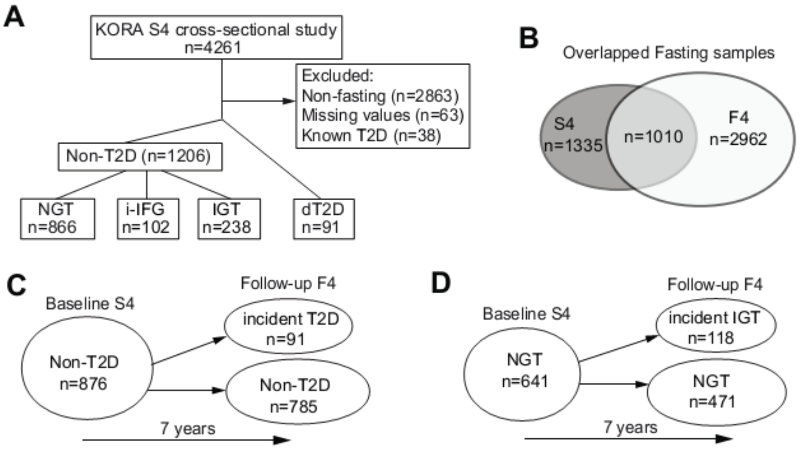

Figure 8: Population description

Metabolomics screens in the KORA cohort, at baseline S4 (A), overlapped between S4 and F4 (B) and prospective (C, D). Participant numbers are shown. Normal glucose tolerance (NGT), isolated impaired fasting glucose (i-IFG), impaired glucose tolerance (IGT), type 2 diabetes mellitus (T2D) and newly diagnosed T2D (dT2D). Non-T2D individuals include NGT, i-IFG and IGT participants.

Based on both fasting and 2-h glucose values (i.e., 2 h post oral 75 g glucose load), individuals were defined according to the WHO diagnostic criteria to have normal glucose tolerance (NGT), isolated IFG (i-IFG), IGT or newly diagnosed T2D (dT2D) (Meisinger et al., 2010; Rathmann et al., 2009) (Table 5). The sample sets include 91 dT2D patients and 1206 individuals with non-T2D, including 866 participants with NGT, 102 with i-IFG and 238 with IGT, in the cross-sectional KORA S4 (Figure 8A; study characteristics are shown in Table 6). Of the 1010 individuals in a fasting state who participated at baseline and follow-up surveys (Figure 8B, study characteristics of the KORA F4 survey are shown in Table 7), 876 of them were non-diabetic at baseline. Out of these, about 10% developed T2D (i.e., 91 incident T2D) (Figure 8C). From the 641 individuals with NGT at baseline, 18% developed IGT (i.e., 118 incident IGT) 7 years later (Figure 8D). The study characteristics of the prospective KORA S4-F4 are shown in Table 8.

Table 5: Classification based on fasting and 2-h glucose values according to the WHO diagnostic criteria

Abbreviations: NGT, normal glucose tolerance; i-IFG, isolated impaired fasting glucose, IGT, impaired glucose tolerance; dT2D, newly-diagnosed type 2 diabetes.

	Fasting glucose values (mg/dl)		2-h glucose values (mg/dl)
NGT	<110	and	<140
i-IFG	110 ≤ and < 126	and	<140
IGT	<126	and	140 ≤ and < 200
dT2D	≥ 126	and / or	≥ 200

Table 6: Characteristics of the KORA S4 cross-sectional study sample

Abbreviations: NGT, normal glucose tolerance; i-IFG, isolated impaired fasting glucose; IGT, impaired glucose tolerance; dT2D, newly-diagnosed type 2 diabetes; BP, blood pressure; HDL, high-density lipoprotein; LDL, low-density lipoprotein. Percentages of individuals or means ± SD are given for each variable and each group (NGT, i-IFG, IGT and dT2D).

Clinical and laboratory parameters	NGT	i-IFG	IGT	dT2D
N	866	102	238	91
Age (years)	63.5 ± 5.5	64.1 ± 5.2	65.2 ± 5.2	65.9 ± 5.4
Sex (female) (%)	52.2	30.4	44.9	41.8
BMI (kg/m^2)	27.7 ± 4.1	29.2 ± 4	29.6 ± 4.1	30.2 ± 3.9
Physical activity (% >1h per week)	46.7	35.3	39.9	36.3
Alcohol intake* (%):	20.2	20.5	25.2	24.2
Current smoker (%)	14.8	10.8	10.9	23.1
Systolic BP (mm-Hg)	131.7 ± 18.9	138.9 ± 17.9	140.7 ± 19.8	146.8 ± 21.5
HDL cholesterol (mg/dl)	60.5 ± 16.4	55.7 ± 15.9	55.7 ± 15.1	50.0 ± 15.8
LDL cholesterol (mg/dl)	154.5 ± 39.8	152.1 ± 37.7	155.2 ± 38.6	146.1 ± 44.6
Triglycerides (mg/dl)	120.7 ± 68.3	145.0 ± 96.0	146.6 ± 80.0	170.6 ± 107.1
HbA$_{1c}$ (%)	5.56 ± 0.33	5.62 ± 0.33	5.66 ± 0.39	6.21 ± 0.83
Fasting glucose (mg/dl)	95.6 ± 7.1	114.2 ± 3.7	104.5 ± 9.7	133.2 ± 31.7
2-h glucose (mg/dl)	102.1 ± 21.0	109.3 ± 18.7	163.4 ± 16.4	232.1 ± 63.7
Fasting insulin (µU/ml)	10.48 ± 7.28	16.26 ± 9.67	13.92 ± 9.53	17.70 ± 12.61

* ≥ 20g/day for women; ≥ 40g/day for men.

Table 7: Cross-sectional analysis: Characteristics of the KORA F4 follow-up study sample

Abbreviations: NGT, normal glucose tolerance; i-IFG, isolated impaired fasting glucose, IGT, impaired glucose tolerance; dT2D, newly-diagnosed type 2 diabetes; BP, blood pressure; HDL, high-density lipoprotein; LDL, low-density lipoprotein. Percentages of individuals or means ± SD are given for each variable and each group (NGT, i-IFG, IGT and dT2D).

Clinical and laboratory parameters	**NGT**	**i-IFG**	**IGT**	**dT2D**
N	2134	112	380	113
Age (years)	52.8 ± 12.6	61.2 ± 10.9	63.8 ± 10.9	65.4 ± 10.3
Sex (female) (%)	54.4	33.9	51.3	40.7
BMI (kg/m^2)	26.6 ± 4.3	29.9 ± 4.6	29.7 ± 4.9	30.9 ± 4.4
Physical activity (% >1h per week)	58.1	45.5	50.3	47.8
Alcohol intake* (%)	17.4	20.5	17.4	21.2
Smoker (%)	20.6	9.6	8.7	13.3
Systolic BP (mm-Hg)	119.2 ± 17.4	130.8 ± 19.5	127.6 ± 18.6	131.8 ± 17.6
HDL cholesterol (mg/dl)	57.6 ± 14.4	50.7 ± 13.5	54.3 ± 14.4	48.2 ± 12.5
LDL cholesterol (mg/dl)	134.9 ± 34.2	145.2 ± 36.1	144.2 ± 35.7	138.2 ± 34.6
Triglycerides (mg/dl)	110.9 ± 74.5	154.5 ± 87.7	145.9 ± 85.9	129.2 ± 162.3
HbA_{1c} (%)	5.36 ± 0.30	5.69 ± 0.32	5.64 ± 0.35	6.24 ± 0.98
Fasting glucose (mg/dl)	91.7 ± 7.6	113.8 ± 3.5	100.1 ± 10.6	123.7 ± 28.6
2-h glucose (mg/dl)	97.7 ± 20.8	109.9 ± 17.1	161.7 ± 17.1	219.9 ± 60.9

* ≥ 20 g/day for women; ≥ 40 g/day for men

Table 8: Characteristics of the KORA S4 → F4 prospective study samples

Abbreviations: BP, blood pressure; HDL, high-density lipoprotein; LDL, low-density lipoprotein. Percentages of individuals or means ± SD are given for each variable and each group.

	NGT at baseline (n=589)		**Non-T2D at baseline (n=876)**	
	Remained NGT at follow-up	**Developed IGT at follow-up**	**Remained Non-T2D at follow-up**	**Developed T2D at follow-up**
N	471	118	785	91
Age (years)	62.4 ± 5.4	63.9 ± 5.5	62.9 ± 5.4	65.5 ± 5.2
Sex (female) (%)	52.2	55.9	50.8	34.1
BMI (kg/m^2)	27.2 ± 3.8	28.2 ± 3.9	27.9 ± 4	30.2 ± 3.6
Physical activity (% >1h per week)	52.9	43.2	52.2	58.2
Alcohol intake* (%)	19.9	20.3	20.6	19.8
Smoker (%)	14.6	9.3	12.0	14.3
Systolic BP (mm-Hg)	129.6 ± 18.2	134.2 ± 18.7	132.4 ± 18.6	137.8 ± 19
HDL cholesterol (mg/dl)	61.3 ± 16.8	58.9 ± 16.2	60.0 ± 16.5	51.9 ± 12.4
LDL cholesterol (mg/dl)	153.9 ± 38.4	156.9 ± 42.7	154.5 ± 39.5	157.7 ± 41.6

Triglycerides (mg/dl)	118.1 ± 63.9	129.5 ± 79.0	125.0 ± 70.0	151.2 ± 74.2
HbA$_{1c}$ (%)	5.54 ± 0.33	5.59 ± 0.34	5.6 ± 0.3	5.8 ± 0.4
Fasting glucose (mg/dl)	94.7 ± 6.9	96.6 ± 7.1	97.7 ± 8.8	106.1 ± 10.1
2-h glucose (mg/dl)	98.2 ± 20.5	109.9 ± 16.8	109.3 ± 28	145.9 ± 32.3
Fasting insulin (μU/ml)	9.91 ± 6.48	11.79 ± 8.83	11.0 ± 7.6	16.2 ± 9.6

* ≥ 20g/day for women; ≥ 40g/day for men

3.3.2 Analyses strategies

We first screened for significantly differed metabolites concentrations among four groups (dT2D, IGT, i-IFG and NGT) for 140 metabolites with cross-sectional studies in KORA S4, and for 131 metabolites in KORA F4. Three IGT-specific metabolites were identified and further investigated in the prospective KORA S4-F4 cohort, to examine whether the baseline metabolite concentrations can predict incident IGT and T2D, and whether they are associated with glucose tolerance 7 years later. Our results are based on a prospective population-based cohort, which differed from previous nested case–control study (Wang et al., 2011). We also performed analysis with same study design using our data. The obtained results provided clues to explain the differences between the two sets of biomarkers. The three metabolites were also replicated in an independent European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam cohort (Wang-Sattler et al., 2012). Finally, the relevance of the identified metabolites was further investigated using bioinformatical analysis to construct the protein-metabolite interaction networks which also combined with the gene expression data.

3.3.3 Identification of novel pre-diabetes metabolites distinct from known T2D risk indicators

To identify metabolites with altered concentrations between the individuals with NGT, i-IFG, IGT and dT2D, we first examined five pairwise comparisons (i-IFG, IGT and dT2D versus NGT, as well as dT2D versus either i-IFG or IGT) in the cross-sectional KORA S4. Based on multivariate logistic regression analysis, 26 metabolite concentrations differed significantly (*P*-values $< 3.6 \times 10^{-4}$) between two groups in at least one of the five comparisons (Figure 9A; odds ratios (ORs) and P-values are shown in Table 9). These associations were independent of age, sex, body mass index (BMI), physical activity, alcohol intake, smoking, systolic blood pressure (BP) and HDL cholesterol (model 1). As expected, the level of total hexose H1, which is mainly represented by glucose (Pearson's correlation coefficient value *r* between H1 and fasting glucose reached 0.85; Table 10), was significantly different in all five comparisons. The significantly changed metabolite panel differed from NGT to i-IFG or to IGT. Most of the significantly altered metabolite concentrations were found between individuals with dT2D and IGT as compared with NGT (Table 11A).

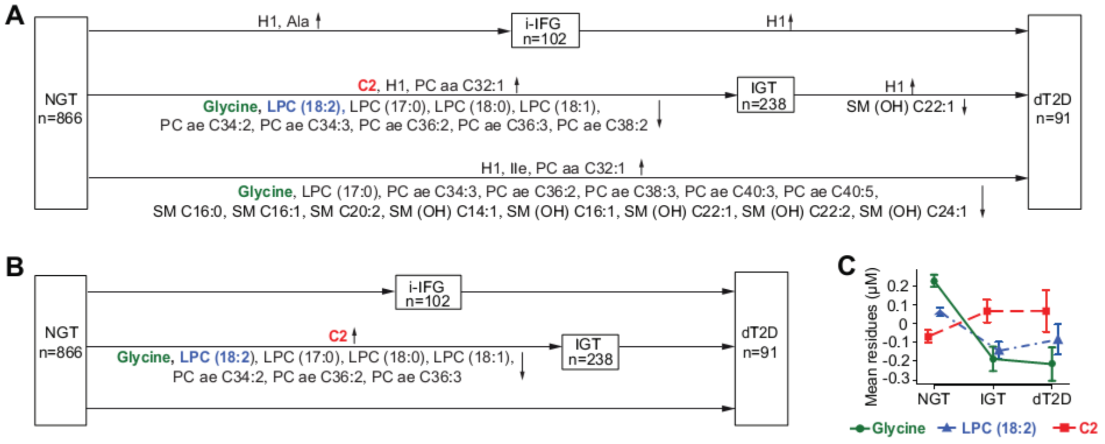

Figure 9: Differences in metabolite concentrations from cross-sectional analysis of KORA S4

Plots (A, B) show the names of metabolites with significantly different concentrations in multivariate logistic regression analyses (after the Bonferroni correction for multiple testing withPo3.6104) in the five pairwise comparisons of model 1 and model 2. Plot (C) shows the average residues of the concentrations with standard errors of the three metabolites (glycine, LPC (18:2) and acetylcarnitine C2) for the NGT, IGT and dT2D groups. Plot (A) shows the results with adjustment for model 1 (age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP and HDL cholesterol), whereas plots (B, C) have additional adjustments for HbA1c , fasting glucose and fasting insulin (model 2). Residuals were calculated from linear regression model (formula: metabolite concentration ~ model 2). For further information, see Supplementary Table 13.

Table 9: Odds ratios (ORs) and P-values in five pairwise comparisons with two adjusted models in the KORA S4

ORs were calculated with multivariate logistic regression analysis with adjustment for age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP and HDL cholesterol in model 1; model 2 includes those variable in model 1 plus HbA$_{1c}$, fasting glucose and fasting insulin. CI denotes confidence interval.

Metabolite	**Model 1**		**Model 2**	
	OR (95% CI), per SD	***P*-value**	**OR (95% CI), per SD**	***P*-value**
		238 IGT vs. 866 NGT		
Glycine	0.65(0.53-0.78)	5.6E-06	0.67(0.54-0.81)	8.6E-05
LPC (18:2)	0.58(0.47-0.7)	1.3E-07	0.58(0.46-0.72)	2.1E-06
C2	1.37(1.18-1.59)	3.8E-05	1.38(1.16-1.64)	2.4E-04
		91 dT2D vs. 866 NGT		

Glycine	0.47(0.33-0.65)	1.1E-05	0.44(0.22-0.83)	1.6E-02
LPC (18:2)	0.62(0.44-0.85)	4.1E-03	0.61(0.32-1.07)	1.1E-01
C2	1.17(0.94-1.45)	1.5E-01	1.71(1.14-2.52)	6.8E-03
		91 dT2D vs. 234 IGT		
Glycine	0.81(0.61-1.07)	1.5E-01	0.76(0.51-1.1)	1.6E-01
LPC (18:2)	0.91(0.69-1.19)	4.8E-01	0.84(0.57-1.22)	3.7E-01
C2	0.93(0.71-1.2)	5.9E-01	1.27(0.87-1.86)	2.2E-01
		102 i-IFG vs. 866 NGT		
Glycine	0.75(0.57-0.98)	3.9E-02	0.62 *	1.0E+00
LPC (18:2)	0.99(0.77-1.26)	9.6E-01	0.79 *	1.0E+00
C2	1.2(0.99-1.46)	5.9E-02	0.18 *	1.0E+00
		91 dT2D vs. 102 i-IFG		
Glycine	0.62(0.43-0.87)	7.8E-03	0.62(0.4-0.93)	2.5E-02
LPC (18:2)	0.62(0.43-0.89)	1.1E-02	0.54(0.33-0.84)	8.9E-03
C2	0.92(0.66-1.27)	6.2E-01	1.23(0.82-1.85)	3.1E-01

* Fasting glucose values were added as co-variants to the model 2, resulting in a perfect separation between i-IFG and NGT.

Table 10: Cross-sectional analysis: Pearson's correlation coefficients (r) between metabolite concentrations and clinical/laboratory parameters in the KORA S4 survey

30 metabolites were chosen to be included in this correlation table. Among them, 26 presented significance difference in at least one of the five pairwise comparisons shown in Figure 8 and the four additional amino acids described by Wang *et al* (Wang et al, 2011). Abbreviations: BP, blood pressure; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

Metabolite	Age	BMI	Systolic BP	HDL	LDL	Tri-glyceride	HbA$_{1c}$	Fasting Glucose	2-h glucose	Fasting insulin	HOMA-B	HOMA-IR
Glycine	0.01	-0.10	-0.06	0.17	0.06	-0.14	-0.09	-0.16	-0.19	-0.14	-0.06	-0.16
LPC (18:2)	-0.12	**-0.34**	-0.07	0.12	0.03	-0.07	-0.13	-0.15	**-0.28**	**-0.24**	**-0.24**	-0.19
C2	0.15	0.08	0.08	0.02	-0.01	0.04	-0.02	0.07	0.11	0.05	0.01	0.04
Alanine	0.07	0.19	0.13	-0.15	0.07	0.26	0.12	0.20	0.16	**0.24**	**0.23**	0.13
Isoleucine	0.02	0.19	0.17	**-0.31**	-0.01	**0.26**	0.09	**0.23**	0.18	0.19	0.11	0.17
PC aa C32:1	-0.04	0.05	0.10	0.16	0.06	**0.35**	0.01	0.07	0.17	0.08	0.07	0.05
PC ae C34:2	-0.04	-0.2	-0.11	**0.46**	**0.26**	-0.17	-0.05	**-0.21**	**-0.22**	-0.17	**-0.21**	-0.08
PC ae C34:3	-0.06	**-0.27**	-0.1	**0.54**	**0.21**	**-0.25**	-0.11	**-0.27**	**-0.28**	**-0.24**	**-0.28**	-0.14
PC ae C36:2	0.00	**-0.26**	-0.14	**0.34**	**0.32**	-0.03	-0.04	**-0.22**	**-0.22**	-0.19	**-0.21**	-0.08
PC ae C36:3	-0.05	-0.16	-0.07	**0.47**	**0.27**	-0.1	-0.05	-0.19	-0.2	-0.13	-0.17	-0.05
PC ae C38:2	-0.02	-0.19	-0.09	**0.33**	**0.34**	0.04	-0.09	**-0.21**	-0.18	-0.15	-0.17	-0.03
PC ae C38:3	0.02	-0.07	-0.09	**0.28**	**0.39**	0.09	0.00	-0.16	-0.11	-0.07	-0.09	0.03
PC ae C40:3	0.02	-0.12	-0.11	**0.38**	**0.33**	-0.12	-0.11	**-0.25**	-0.16	-0.11	-0.14	0.03
PC ae C40:5	-0.01	-0.2	-0.06	**0.4**	**0.26**	-0.09	-0.1	-0.18	-0.16	-0.14	-0.17	-0.09
LPC (17:0)	-0.01	**-0.28**	-0.14	0.08	**0.22**	-0.04	-0.08	-0.2	**-0.26**	**-0.22**	**-0.22**	-0.12
LPC (18:0)	-0.04	-0.13	-0.02	0.01	**0.32**	0.16	-0.03	-0.09	-0.15	-0.09	-0.1	-0.05
LPC (18:1)	-0.1	**-0.29**	-0.03	0.17	0.05	0.07	-0.12	-0.12	-0.2	**-0.21**	**-0.21**	-0.19
SM (OH) C14:1	0.08	-0.09	-0.11	**0.26**	**0.35**	-0.2	-0.07	-0.2	-0.16	-0.1	-0.12	0.02
SM (OH) C16:1	0.08	-0.08	-0.13	**0.22**	**0.38**	**-0.21**	-0.06	-0.18	-0.13	-0.08	-0.11	0.03
SM (OH) C22:1	-0.05	-0.05	-0.11	**0.25**	**0.5**	-0.12	-0.06	-0.15	-0.11	-0.03	-0.07	0.07
SM (OH) C22:2	-0.01	-0.08	-0.16	**0.39**	**0.37**	**-0.23**	-0.08	**-0.24**	-0.19	-0.1	-0.14	0.03
SM (OH) C24:1	-0.02	-0.07	-0.1	0.18	**0.42**	-0.15	-0.11	-0.16	-0.14	-0.05	-0.07	0.05
SM C16:0	0.07	-0.11	-0.03	**0.31**	**0.52**	**-0.21**	-0.13	**-0.21**	-0.17	-0.1	-0.14	-0.01
SM C16:1	0.08	0.11	-0.07	**0.38**	**0.43**	-0.09	-0.07	-0.18	-0.12	-0.03	-0.08	0.06
SM C20:2	0.06	0.01	-0.08	**0.26**	**0.22**	**-0.21**	-0.06	-0.17	-0.11	-0.05	-0.08	0.02

H1	0.06	**0.28**	0.19	**-0.22**	-0.04	**0.21**	**0.54**	**0.85**	**0.62**	**0.36**	-0.02	**0.50**
Leucine	-0.01	0.19	0.15	**-0.26**	0.06	**0.21**	0.09	0.19	0.14	0.14	0.06	0.13
Valine	0.00	**0.27**	0.16	**-0.26**	0.05	**0.22**	0.08	**0.22**	0.18	**0.23**	0.16	**0.22**
Tryosine	0.06	**0.27**	0.16	-0.09	-0.02	0.13	0.06	0.17	0.16	**0.23**	0.16	**0.22**
Phenylalanine	0.04	**0.23**	0.17	-0.13	0.05	0.15	0.07	0.14	0.12	**0.21**	0.15	0.18

Table 11: Cross-sectional analysis: Odds ratios (ORs) and P-values in five pairwise comparisons in the KORA S4

(A) In addition to Table 13, we list here the 23 additional metabolites that show significant concentration differences in at least one pairwise comparison using multivariate logistic regression analysis with adjustment for age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP and HDL cholesterol, as for model 1.

A.	**i-IFG vs. NGT**		**IGT vs. NGT**		**dT2D vs. NGT**		**dT2D vs. i-IFG**		**dT2D vs. IGT**	
Metabolites	**OR (95% CI), per SD**	***P*-value**	**OR (95% CI), per SD**	***P*-value**	**OR (95% CI), per SD**	***P*-value**	**OR (95% CI), per SD**	***P*-value**	**OR (95% CI), per SD**	***P*-value**
PC aa C32:1	1.07 (0.85, 1.31)	5.59E-01	1.43 (1.23, 1.68)	8.01E-06	1.62 (1.30, 2.03)	2.34E-05	1.64 (1.11, 2.53)	1.76E-02	1.34 (1.01, 1.79)	4.16E-02
PC ae C34:2	0.72 (0.55, 0.93)	1.43E-02	0.67 (0.55, 0.80)	2.19E-05	0.67 (0.50, 0.90)	8.32E-03	0.96 (0.67, 1.39)	8.39E-01	0.95 (0.70, 1.28)	7.29E-01
PC ae C34:3	0.67 (0.50, 0.88)	5.47E-03	0.64 (0.52, 0.77)	9.43E-06	0.44 (0.31, 0.63)	9.41E-06	0.68 (0.44, 1.02)	6.88E-02	0.72 (0.52, 0.99)	4.45E-02
PC ae C36:2	0.74 (0.56, 0.96)	2.39E-02	0.64 (0.53, 0.77)	4.25E-06	0.57 (0.42, 0.76)	2.54E-04	0.80 (0.56, 1.13)	2.07E-01	0.78 (0.57, 1.05)	1.06E-01
PC ae C36:3	0.76 (0.59, 0.97)	3.10E-02	0.70 (0.59, 0.84)	1.53E-04	0.67 (0.50, 0.89)	7.28E-03	0.90 (0.63, 1.28)	5.48E-01	0.90 (0.66, 1.21)	4.74E-01
PC ae C38:2	0.79 (0.61, 1.01)	6.21E-02	0.72 (0.60, 0.86)	2.95E-04	0.67 (0.50, 0.88)	5.45E-03	0.86 (0.60, 1.21)	3.85E-01	0.80 (0.61, 1.06)	1.27E-01
PC ae C38:3	0.82 (0.63, 1.04)	1.10E-01	0.82 (0.69, 0.98)	2.76E-02	0.57 (0.42, 0.75)	1.51E-04	0.72 (0.51, 1.02)	6.57E-02	0.65 (0.48, 0.88)	5.77E-03
PC ae C40:3	0.71 (0.55, 0.92)	1.04E-02	0.79 (0.66, 0.94)	9.06E-03	0.57 (0.42, 0.76)	2.28E-04	0.84 (0.59, 1.20)	3.42E-01	0.68 (0.49, 0.93)	1.67E-02
PC ae C40:5	0.80 (0.63, 1.02)	7.88E-02	0.76 (0.64, 0.90)	2.00E-03	0.59 (0.44, 0.78)	2.32E-04	0.77 (0.54, 1.10)	1.59E-01	0.81 (0.60, 1.07)	1.47E-01
LPC a C17:0	0.80 (0.62, 1.02)	8.29E-02	0.56 (0.46, 0.68)	4.96E-09	0.46 (0.33, 0.62)	1.59E-06	0.65 (0.45, 0.92)	1.69E-02	0.76 (0.57, 1.00)	5.77E-02
LPC a C18:0	0.97 (0.78, 1.21)	8.09E-01	0.71 (0.60, 0.83)	4.34E-05	0.70 (0.54, 0.90)	7.50E-03	0.71 (0.51, 0.98)	4.10E-02	0.90 (0.70, 1.16)	4.26E-01
LPC a C18:1	0.96 (0.75, 1.21)	7.48E-01	0.66 (0.54, 0.80)	2.23E-05	0.78 (0.58, 1.04)	9.76E-02	0.83 (0.58, 1.18)	3.07E-01	1.11 (0.85, 1.46)	4.39E-01
SM (OH) C14:1	0.79 (0.61, 1.02)	7.33E-02	0.72 (0.60, 0.87)	5.99E-04	0.46 (0.33, 0.63)	2.34E-06	0.65 (0.44, 0.95)	2.65E-02	0.63 (0.45, 0.86)	4.26E-03
SM (OH) C16:1	0.79 (0.61, 1.01)	6.30E-02	0.82 (0.69, 0.98)	3.00E-02	0.51 (0.37, 0.69)	1.58E-05	0.71 (0.48, 1.04)	8.60E-02	0.60 (0.42, 0.83)	2.70E-03

SM (OH) C22:1	0.81 (0.64, 1.03)	9.51E-02	0.96 (0.81, 1.14)	6.43E-01	0.57 (0.42, 0.75)	1.05E-04	0.69 (0.48, 0.99)	4.46E-02	0.55 (0.40, 0.75)	2.54E-04
SM (OH) C22:2	0.71 (0.54, 0.94)	1.86E-02	0.71 (0.58, 0.87)	9.41E-04	0.40 (0.28, 0.56)	2.94E-07	0.60 (0.40, 0.89)	1.25E-02	0.55 (0.38, 0.77)	8.29E-04
SM (OH) C24:1	0.88 (0.70, 1.10)	2.66E-01	0.88 (0.75, 1.03)	1.18E-01	0.54 (0.40, 0.72)	4.15E-05	0.62 (0.44, 0.88)	8.12E-03	0.59 (0.43, 0.80)	8.14E-04
SM C16:0	0.75 (0.59, 0.95)	1.83E-02	0.80 (0.67, 0.94)	7.45E-03	0.49 (0.36, 0.65)	1.16E-06	0.64 (0.44, 0.92)	1.89E-02	0.63 (0.46, 0.84)	2.38E-03
SM C16:1	0.78 (0.6, 1.02)	7.39E-02	0.78 (0.64, 0.94)	8.69E-03	0.49 (0.36, 0.67)	1.16E-05	0.63 (0.42, 0.93)	2.20E-02	0.64 (0.46, 0.89)	1.01E-02
SM C20:2	0.69 (0.52, 0.91)	8.86E-03	0.76 (0.63, 0.91)	3.39E-03	0.54 (0.39, 0.74)	1.95E-04	0.84 (0.58, 1.21)	3.50E-01	0.80 (0.58, 1.09)	1.69E-01
H1	5.40 (3.98, 7.52)	1.89E-25	2.38 (1.96, 2.92)	1.68E-17	10.18 (6.61, 16.61)	4.23E-23	7.83 (3.63, 19.31)	1.30E-06	3.56 (2.44, 5.43)	4.80E-10
Ala	1.45 (1.18, 1.78)	3.39E-04	1.21 (1.04, 1.40)	1.29E-02	1.50 (1.2, 1.88)	4.36E-04	0.98 (0.71, 1.36)	9.19E-01	1.25 (0.96, 1.64)	9.71E-02
Ile	1.12 (0.89, 1.40)	3.41E-01	1.12 (0.95, 1.32)	1.84E-01	1.64 (1.29, 2.1)	5.39E-05	1.57 (1.12, 2.27)	1.09E-02	1.47 (1.11, 1.96)	7.72E-03

(B) Six additional metabolites that show significant concentration differences in at least one pairwise comparison, using multivariate logistic regression analysis with adjustment for model 1 plus HbA$_{1c}$, fasting glucose and fasting insulin, are given.

B.	**IGT vs. NGT**		**dT2D vs. NGT**		**dT2D vs. i-IFG**		**dT2D vs. IGT**	
	OR (95% CI), per SD	***P*-value**	**OR (95% CI), per SD**	***P*-value**	**OR (95% CI), per SD**	***P*-value**	**OR (95% CI), per SD**	***P*-value**
PC ae C34:2	0.66 (0.53, 0.81)	9.95E-05	0.62 (0.34, 1.07)	9.79E-02	0.86 (0.54, 1.37)	5.29E-01	0.91 (0.6, 1.38)	6.58E-01
PC ae C36:2	0.65 (0.52, 0.80)	8.06E-05	0.45 (0.23, 0.84)	1.52E-02	0.65 (0.39, 1.04)	7.93E-02	0.70 (0.45, 1.06)	9.98E-02
PC ae C36:3	0.67 (0.55, 0.83)	1.83E-04	0.54 (0.29, 0.97)	4.28E-02	0.76 (0.49, 1.18)	2.34E-01	0.84 (0.55, 1.26)	3.94E-01
LPC a C17:0	0.60(0.49, 0.74)	3.07E-06	0.49 (0.25, 0.86)	2.04E-02	0.57 (0.36, 0.88)	1.25E-02	0.68 (0.46, 0.99)	5.03E-02
LPC a C18:1	0.62 (0.50, 0.77)	1.49E-05	0.88 (0.51, 1.46)	6.38E-01	0.74 (0.47, 1.16)	1.91E-01	1.02 (0.71, 1.46)	9.24E-01

To investigate whether HbA1c, fasting glucose and fasting insulin levels mediate the shown associations, these were added as covariates to the regression analysis (model 2) in addition to model 1 (Figure 9B).We observed that, under these conditions, no metabolite differed significantly when comparing individuals with dT2D to those with NGT, suggesting that these metabolites are associated with HbA1c, fasting glucose and fasting insulin levels (*r* values are shown in Table 10). Only nine metabolite concentrations significantly differed between IGT and NGT individuals (Table 9; Table 11B). These metabolites therefore represent novel biomarker candidates, and are independent from the known risk indicators for T2D. The logistic regression analysis was based on each single metabolite, and some of these metabolites are expected to correlate with each other. To further assess the metabolites as a group, we employed two additional statistical methods (the non-parametric random forest and the parametric stepwise selection) to identify unique and independent biomarker candidates. Out of the nine metabolites, five molecules (i.e., glycine, LPC (18:2), LPC (17:0), LPC (18:1) and C2) were select after random forest, and LPC (17:0) and LPC (18:1) were then removed after the stepwise selection. Thus, three molecules were found to contain independent information: glycine (adjusted OR = 0.67 (0.54 - 0.81), P = 8.6 × 10-5), LPC (18:2) (OR = 0.58 (0.46 - 0.72), P = 2.1 × 10-6) and acetylcarnitine C2 (OR = 1.38 (1.16 - 1.64), P = 2.4 × 10-4) (Figure 9C). Similar results were observed in the follow-up KORA F4 study (Figure 10). For instance, when 380 IGT individuals were compared with 2134 NGT participants, these three metabolites were also found to be highly significantly different (glycine, OR = 0.64 (0.55 - 0.75), P = 9.3 x 10-8; LPC (18:2), OR = 0.47 (0.38 - 0.57), P = 2.1 x 10-13; and C2, OR = 1.33 (1.17 – 1.49), P = 4.9 x 10 -6) (Table 12).

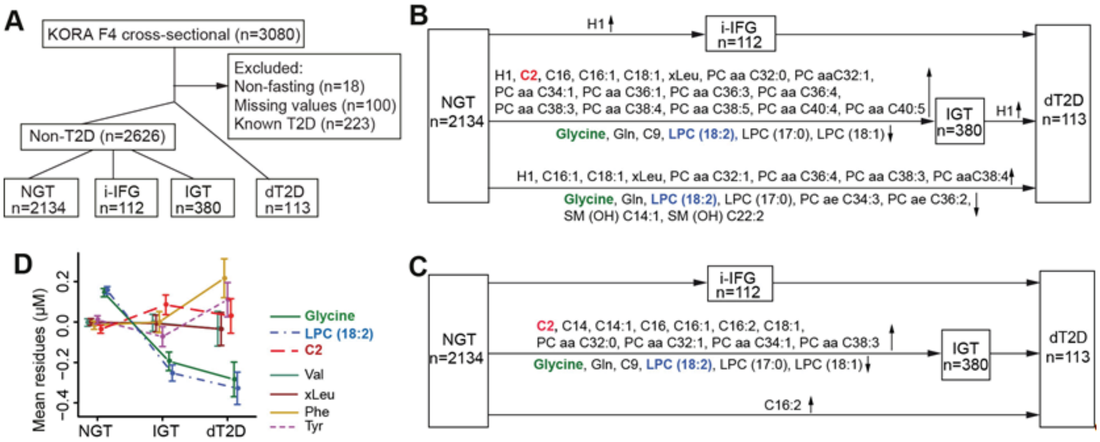
A
KORA F4 cross-sectional (n=3080)
Excluded:
Non-fasting (n=18)
Missing values (n=100)
Known T2D (n=223)
Non-T2D (n=2626)
NGT n=2134
i-IFG n=112
IGT n=380
dT2D n=113
B
NGT n=2134
H1
i-IFG n=112
H1, C2, C16, C16:1, C18:1, xLeu, PC aa C32:0, PC aaC32:1, PC aa C34:1, PC aa C36:1, PC aa C36:3, PC aa C36:4, PC aa C38:3, PC aa C38:4, PC aa C38:5, PC aa C40:4, PC aa C40:5
Glycine, Gln, C9, LPC (18:2), LPC (17:0), LPC (18:1)
IGT n=380
H1
dT2D n=113
H1, C16:1, C18:1, xLeu, PC aa C32:1, PC aa C36:4, PC aa C38:3, PC aaC38:4
Glycine, Gln, LPC (18:2), LPC (17:0), PC ae C34:3, PC ae C36:2, SM (OH) C14:1, SM (OH) C22:2
D
Mean residues (µM)
−0.4
−0.2
0.0
0.2
NGT
IGT
dT2D
Glycine
LPC (18:2)
C2
Val
xLeu
Phe
Tyr
C
NGT n=2134
i-IFG n=112
C2, C14, C14:1, C16, C16:1, C16:2, C18:1, PC aa C32:0, PC aa C32:1, PC aa C34:1, PC aa C38:3
Glycine, Gln, C9, LPC (18:2), LPC (17:0), LPC (18:1)
IGT n=380
C16:2
dT2D n=113

Figure 10: Differences in metabolite concentrations from cross-sectional analysis in KORA F4

Plot A demonstrates the study population in the KORA F4. Plots B and C show the names of metabolites with significantly different concentrations in multivariate logistic regression analyses (after the Bonferroni correction for multiple testing with *P* < 3.6 x 10^{-4}) in the five pairwise comparisons. The plot shows the results with adjustment for model 1 (age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP and HDL cholesterol), whereas plots B and D additionally show the adjustment for HbA1c and fasting glucose (model 2). Plot D shows the average residues of the concentrations with standard errors of glycine, LPC (18:2) and acetylcarnitine C2, as well as xLeu (isoleucine and leucine), valine, phenylalanine and tyrosine, for the NGT, IGT and dT2D groups.

Table 12: Cross-sectional analysis: ORs and P-values in five pairwise comparisons with two adjusted models in the KORA F4

ORs were calculated with multivariate logistic regression analysis with adjustment for age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP and HDL cholesterol in model 1; model 2 includes model 1 and additionally HbA_{1c} and fasting glucose. CI denotes confidence interval.

Metabolite	Model 1		Model 2	
	OR (95% CI), per SD	*P*-value	OR (95% CI), per SD	*P*-value

		380 IGT vs. 2134 NGT		
Glycine	0.64(0.55-0.74)	1.0E-08	0.64(0.55-0.75)	9.3E-08
LPC (18:2)	0.47(0.39-0.57)	3.0E-14	0.47(0.38-0.57)	2.1E-13
C2	1.29(1.15-1.44)	1.2E-05	1.33(1.17-1.49)	4.9E-06
		113 dT2D vs. 2134 NGT		
Glycine	0.45(0.33-0.61)	9.0E-07	0.42(0.23-0.70)	1.8E-03
LPC (18:2)	0.40(0.27-0.57)	1.6E-06	0.34(0.17-0.63)	1.0E-03
C2	1.24(1.12-1.61)	1.6E-03	1.36(0.99-1.85)	5.0E-02
		113 dT2D vs. 380 IGT		
Glycine	0.78(0.60-1.00)	5.6E-02	0.74(0.54-1.01)	6.4E-02
LPC (18:2)	0.90(0.69-1.15)	4.0E-01	0.68(0.48-0.95)	2.6E-02
C2	1.07(0.85-1.34)	5.4E-01	1.08(0.80-1.46)	6.0E-01
		112 i-IFG vs. 2134 NGT		
Glycine	0.85(0.65-1.08)	2.0E-01	3.97 *	1.0E+00
LPC (18:2)	0.76(0.57-1.01)	6.7E-02	1.29 *	1.0E+00
C2	1.05(0.86-1.26)	6.4E-01	0.91 *	1.0E+00
		113 dT2D vs. 112 i-IFG		
Glycine	0.71(0.51-0.95)	2.4E-02	0.78(0.56-1.08)	1.4E-01
LPC (18:2)	0.66(0.45-0.93)	2.0E-02	0.65(0.42-0.96)	3.5E-02
C2	1.34(1.00-1.85)	5.7E-02	1.35(0.97-1.90)	7.7E-02

Table 13 Prediction of IGT and T2D in the KORA cohort

Odds ratios (ORs, 95% confidence intervals) and P-values of multivariate logistic regression results are shown in (A) and (B) for IGT, and in (C) and (D) for T2D, respectively, whereas β estimates and P-values from linear regression analysis between metabolite concentration in baseline KORA S4 and 2-h glucose values in follow-up KORA F4 are shown in (E). All models were adjusted for age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP and HDL cholesterol.

Model	**Glycine**	**LPC (18:2)**	**C2**	**Glycine, LPC (18:2), C2**
A. Metabolite as continuous variable (n = 589)				
Per SD	0.75 (0.58-0.95)	0.72 (0.54-0.93)	0.92 (0.73-1.14)	0.36 (0.20-0.67)
P	0.02	0.02	0.50	0.001
B. Metabolite as categorical variable (n = 589)				
1st quartile	1.0 (reference)	1.0 (reference)	1.0 (reference)	1.0 (reference)
2nd quartile	1.0 (0.80-1.46)	0.96 (0.73-1.27)	0.89 (0.66-1.23)	0.54 (0.30-0.97)
3rd quartile	1.0 (0.74-1.34)	0.71 (0.51-0.99)	0.93 (0.69-1.26)	0.66 (0.37-1.18)
4th quartile	0.78 (0.55-1.06)	0.78 (0.54-1.12)	0.99 (0.73-1.35)	0.36 (0.19-0.69)
P for trend	0.06	0.05	0.79	0.0082
C. Metabolite as continuous variable (n = 876)				
Per SD	0.73 (0.55-0.97)	0.70 (0.51-0.94)	0.94 (0.74-1.18)	0.39 (0.21-0.71)

P	0.04	0.02	0.59	0.0002
D. Metabolite as categorical variable (n = 876)				
1st quartile	1.0 (reference)	1.0 (reference)	1.0 (reference)	1.0 (reference)
2nd quartile	0.87 (0.71-1.07)	0.95 (0.77-1.17)	1.05 (0.85-1.31)	0.50 (0.33-0.76)
3rd quartile	0.82 (0.67-1.01)	0.70 (0.56-0.88)	0.97 (0.78-1.19)	0.57 (0.38-0.88)
4th quartile	0.67 (0.54-0.84)	0.68 (0.54-0.88)	1.21 (0.98-1.50)	0.33 (0.21-0.52)
P for trend	0.00061	0.00021	0.19	1.8E-05
E. Linear regression (n = 843)				
β estimates*(95% CI)	-2.47 (-4.64,-0.29)	-4.57 (-6.90,-2.24)	1.02 (-1.11,3.15)	-4.23 (-6.52,-2.31)
P	0.026	0.00013	0.59	8.8E-05

*ß estimate indicates the future difference in the glucose tolerance corresponding to the one SD differences in the normalized baseline metabolite concentration.

3.3.4 Predicted risks of IGT and T2D

To investigate the predictive value for IGT and T2D of the three identified metabolites, we examined the associations between baseline metabolite concentrations and incident IGT and T2D using the prospective KORA S4 → F4 cohort (Table 8). We compared baseline metabolite concentrations in 118 incident IGT individuals with 471 NGT control individuals. We found that glycine and LPC (18:2), but not C2, were significantly different at the 5% level in both adjusted model 1 and model 2 (Table 13 and Table 14). Significant differences were additionally observed for glycine and LPC (18:2), but not for C2, at baseline concentrations between the 91 incident T2D individuals and 785 participants who remained diabetes-free (non-T2D). Each standard deviation (SD) increment of the combinations of the three metabolites was associated with a 33% decreased risk of future diabetes (OR = 0.39 (0.21-0.71), *P* = 0.0002). Individuals in the fourth quartile of the combined metabolite concentrations had a three-fold lower chance of developing diabetes (OR = 0.33 (0.21-0.52), *P* = 1.8 x 10^{-5}), compared to those whose serum levels were in the first quartile (i.e. combination of glycine, LPC (18:2) and C2), indicating a protective effect from higher concentrations of glycine and LPC (18:2) combined with a lower concentration of C2. With the full adjusted model 2, consistent results were obtained for LPC (18:2) but not for glycine (Table 18). When the three metabolites were added to the fully adjusted model 2, the area under the receiver-operating-characteristic curves (AUC) increased 2.6% (*P* = 0.015) and 1% (*P* = 0.058) for IGT and T2D, respectively (Figure 11, Table 19). Thus, this provides an improved prediction of IGT and T2D as compared to T2D risk indicators.

Table 14: Prospective analysis: prediction of IGT and T2D in the KORA cohort with full adjustment model

ORs were calculated with multivariate logistic regression analysis with adjustment for age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP, HDL cholesterol HbA_{1c}, fasting glucose and fasting insulin. CI denotes confidence interval.

	Incident IGT		**Incident T2D**	
	OR (95% CI), per SD,	***P*-value**	**OR (95% CI), per SD**	***P*-value**
Glycine	0.77 (0.60, 0.97)	0.031	0.85 (0.62, 1.14)	0.29
LPC (18:2)	0.70 (0.53, 0.92)	0.011	0.69 (0.49, 0.94)	0.022
C2	0.97 (0.77, 1.20)	0.79	0.90 (0.70, 1.14)	0.40

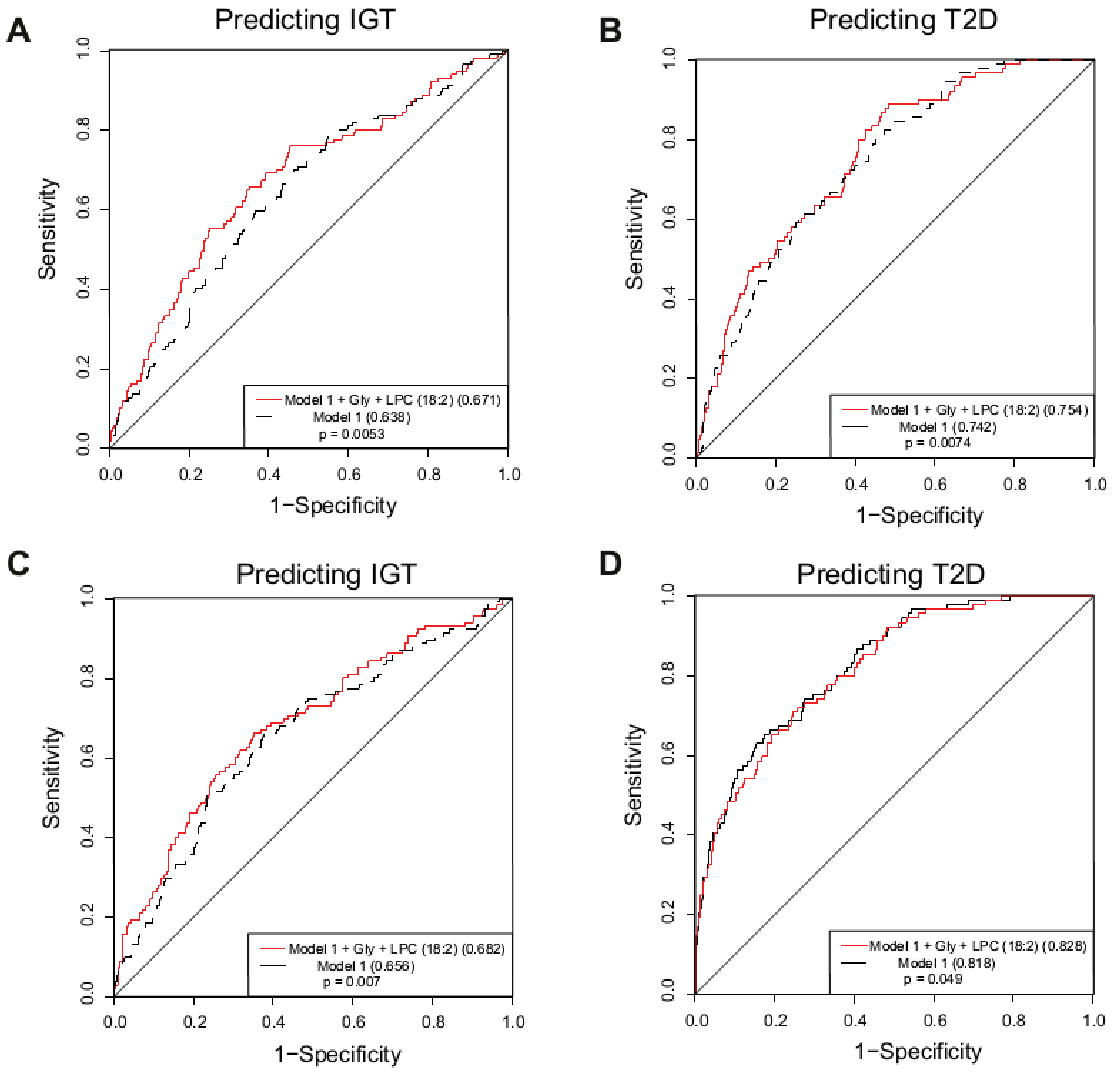

Figure 11: Prospective analysis: prediction of IGT and T2D using two adjustment models

Plots A-D show the AUC values predicting IGT or T2D using known T2D risk factors (model 1 or model 2) alone and in combination with three metabolites (glycine, LPC (18:2) and C2) and the *P*-values from likelihood ratio test comparing the two values.

Model 1 includes age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP and HDL. Model 2 includes the risk factors from model 1 plus HbA_{1c}, fasting glucose and fasting insulin.

Table 15: Prospective analysis: the area under the receiver-operating-characteristic curves (AUC) values for each metabolite and each diabetes risk indicator and their combinations

Model 1 includes age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP and HDL cholesterol, and model 2 includes model 1 plus HbA_{1c}, fasting glucose and insulin.

	IGT (118 incident IGT vs. 471 NGT)	**T2D (91 incident T2D vs. 885 non-T2D)**
Metabolite		
Glycine	0.546	0.604
LPC (18:2)	0.610	0.606
C2	0.521	0.53
Glycine + LPC (18:2) + C2	0.622	0.634
Single T2D risk indicator		
Age	0.580	0.629
Sex	0.519	0.584
BMI	0.576	0.685
Physical activity	0.550	0.53
Alcohol intake	0.501	0.505
Smoking	0.527	0.512
Systolic BP	0.569	0.583
HDL cholesterol	0.544	0.652
HbA_{1c}	0.538	0.688
Fasting glucose	0.575	0.735
Fasting insulin	0.562	0.707
Combined T2D risk indicators		
Model 1	0.638	0.742
Model 2	0.656	0.818
Metabolites combined with T2D risk indicators		
Glycine + LPC (18:2) + C2 + Model 1	0.671	0.754
Glycine + LPC (18:2) + C2 + Model 2	0.683	0.828

3.3.5 Baseline metabolite concentrations correlate with future glucose tolerance

We next investigated the associations between the baseline metabolite concentrations and the follow-up 2-h glucose values after an oral glucose tolerance test. Consistent results were observed for the three metabolites: glycine and LPC (18:2), but not acetylcarnitine C2 levels, were found to be significantly associated, indicating that glycine and LPC (18:2) predict glucose tolerance. Moreover, the three metabolites (glycine, LPC (18:2) and C2) revealed high significance even in the fully adjusted model 2 in the cross-sectional KORA S4 cohort (Table 16). As expected, a very significant association (P = 1.5 x 10^{-22}) was observed for hexose H1 in model 1, while no significance (P = 0.12) was observed for it in the fully adjusted model 2 (Table 16).

Table 16: Cross-sectional analysis: linear regression analysis between metabolite concentration and 2-h glucose values in the KORA S4 (n = 1297)

Beta estimates were calculated with multivariate linear regression analysis with adjustment for model 1 (age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP and HDL cholesterol), and model 2 includes model 1 plus HbA1c, fasting glucose and fasting insulin. CI denotes confidence interval.

	Model 1		Model 2	
	β estimates* (95% CI)	*P*-value	β estimates* (95% CI)	*P*-value
Glycine	-5.96 (-7.69, -4.24)	1.7E-11	-4.93 (-6.61, -3.26)	9.8E-09
LPC (18:2)	-6.98 (-8.82, -5.14)	1.9E-13	-6.47 (-8.24, -4.70)	1.4E-12
C2	3.93 (2.24, 5.63)	5.5E-06	3.81 (2.17, 5.45)	5.8E-06
H1	8.57 (6.88, 10.26)	1.5E-22	2.08 (-0.56, 4.72)	0.12
Isoleucine	0.017 (-1.89, 1.93)	0.99	-0.06 (-1.96, 1.85)	0.95
Leucine	-0.67 (-2.52, 1.20)	0.48	-0.71 (-2.56, 1.15)	0.45
Valine	0.68 (-1.15, 2.52)	0.46	0.03 (-1.75, 1.80)	0.98
Tryosine	-0.57 (-2.32, 1.18)	0.52	-1.09 (-2.81, 0.63)	0.21
Phenylalanine	-0.77 (-2.50, 0.97)	0.38	-0.90 (-2.59, 0.78)	0.29

*ß estimate indicates the future difference in the glucose intolerance corresponding to the one SD differences in the normalized baseline metabolite concentration.

3.3.6 Prospective population-based versus nested case-control designs

To investigate the predict value of the five branched-chain and aromatic amino acids (isoleucine, leucine, valine, tyrosine and phenylalanine) (Wang *et al*, 2011) in our study, we correlated the baseline metabolite concentrations with follow-up 2-h glucose values. We found none of them to be associated significantly, indicating that the five amino acids cannot predict risk of IGT (β estimates and *P*-values are shown in Table 17). Furthermore, none of these five amino acids showed associations with 2-h glucose values in the cross-sectional KORA S4 study (Table 16).

To replicate the identified five branched-chain and aromatic amino acids (Wang *et al*, 2011), we matched our baseline samples to the 91 incident T2D using the same method described previously (Wang *et al*, 2011). We replicated four out of the five branched-chain and aromatic amino acids (characteristics of the case-control and non-T2D samples are shown in Table 18; ORs and *P*-values are given in Table 19). As expected, the three identified IGT-specific metabolites did not significantly differ between the matched case control samples, because the selected controls were enriched with individuals accompanied by high-risk features such as obesity and elevated fasting glucose as described by Wang *et al* (Wang *et al*, 2011). In fact, the 91 matched controls include about 50% pre-diabetes individuals, which is significantly higher than the general population (about 15%).

Table 17: Prospective analysis: linear regression analysis between metabolite concentration in the KORA S4 and 2-h glucose values in the KORA F4 (n = 843)

Beta estimates were calculated with adjustment for age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP and HDL cholesterol.

	β estimates (95% CI)	***P*-value**
Isoleucine	1.10 (-1.38, 3.59)	0.38
Leucine	1.58 (-0.85, 4.02)	0.20
Valine	1.26 (-1.12, 3.64)	0.30
Tryosine	0.13 (-2.18, 2.44)	0.91
Phenylalanine	1.65 (-0.65, 3.94)	0.16

Table 18: Prospective analysis: characteristics of prospective nested case-control sample at baseline KORA S4

Abbreviations: BP, blood pressure; HDL, high-density lipoprotein; LDL, low-density lipoprotein. Percentages of individuals or means ± SD are given for each variable and each group (T2D at follow-up, matched controls and non-T2D).

Clinical and laboratory parameters	Case (T2D at follow-up)	Matched Controls	Non-T2D
N	91	91	1206
Age (years)	65.5 ± 5.2	65.3 ± 5.0	63.9 ± 5.5
Sex (female) (%)	33.1	33.1	0.49
BMI (kg/m^2)	30.2 ± 3.6	30.0 ± 3.4	28.1 ± 4.2
Physical activity (% >1h per week)	58.2	54.4	55.5
Alcohol intake* (%)	19.8	24.4	21.2
Smoker (%)	14.3	4.4	13.7
Systolic BP (mm-Hg)	137.8 ± 19.0	137.5 ± 15.9	134.1 ± 19.4
HDL cholesterol (mg/dl)	51.9 ± 12.7	55.7 ± 16.1	59.1 ± 16.3
LDL cholesterol (mg/dl)	157.7 ± 41.6	155.7 ± 37.3	154.4 ± 39.4
Triglycerides (mg/dl)	151.2 ± 74.3	130.0 ± 71.2	127.9 ± 74.3
HbA$_{1c}$ (%)	5.81 ± 0.39	5.64 ± 0.29	5.58 ± 0.35
Fasting glucose (mg/dl)	106.1 ± 10.0	105.4 ± 9.0	98.9 ± 9.5
2-h glucose (mg/dl)	145.9 ± 32.3	116.5 ± 28.7	114.8 ± 31.4
Fasting insulin (μU/ml)	16.21 ± 9.6	12.9 ± 7.2	11.6 ± 8.2

* ≥ 20 g/day for women; ≥ 40 g/day for men

Table 19: Prospective analysis: ORs and P-values in the comparison between prospective nested case-control samples

ORs were calculated with conditional multivariate logistic regression analysis with adjustment for age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP, HDL cholesterol in model 1; model 2 includes model 1 plus HbA1c and fasting glucose and fasting insulin. CI denotes confidence interval.

	Model 1		Model 2	
	ORs (95% CI), per SD	*P*-value	ORs (95% CI), per SD	*P*-value
Isoleucine	1.84 (1.25-2.71)	0.002	1.73 (1.15-2.60)	0.008
Leucine	1.51(1.06-2.14)	0.02	1.43(0.98-2.08)	0.06
Valine	1.52(1.08-2.13)	0.02	1.48(1.03-2.13)	0.03
Tryosine	1.50(1.06-2.14)	0.02	1.52(1.03-2.24)	0.03
Phenylalanine	1.21(0.88-1.67)	0.23	1.11(0.80-1.55)	0.53
Glycine	0.95(0.69-1.31)	0.76	1.03(0.74-1.44)	0.86
LPC (18:2)	0.77(0.55-1.10)	0.14	0.78(0.56-1.14)	0.21
C2	0.81(0.59-1.13)	0.22	0.80(0.57-1.14)	0.21

3.3.7 Metabolite-protein interaction networks confirmed by transcription levels

To investigate the underlying molecular mechanism for the three identified IGT metabolites, we studied their associations with T2D-related genes by analyzing protein-metabolite interaction networks (Szklarczyk *et al*, 2011; Wishart *et al*, 2009). Seven out of the 46 known T2D-related genes (*PPARG*, *TCF7L2*, *HNF1A*, *GCK*, *IGF1*, *IRS1* and *IDE*) were linked to these metabolites through related enzymes or proteins (Figure 12A; the list of 46 genes is shown in Table 20). To validate the networks, the links between metabolites, enzymes, pathway-related proteins and T2D-related genes were manually checked for biochemical relevance and classified into four groups: signaling regulation, transcription, physical interaction and the same pathway (Table 21).

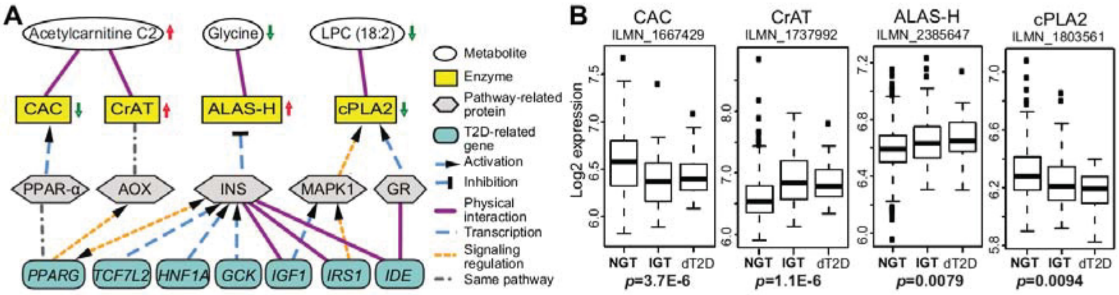

Figure 12: Three candidate metabolites for IGT associated with seven T2D-related genes

(A) Metabolites (white), enzymes (yellow), pathway-related proteins (grey) and T2D-related genes (blue) are represented with ellipses, rectangles, polygons, and rounded rectangles, respectively. Arrows next to the ellipses and rectangles indicate altered metabolite concentrations in persons with IGT as compared to NGT, and enzyme activities in individuals with IGT. The 21 connections between metabolites, enzymes, pathway-related proteins and T2D-related genes were divided after visual inspections into four categories: physical interaction (purple solid line), transcription (blue dash line), signaling regulation (orange dash line), and same pathway (grey dot and dash line). The activation or inhibition is indicated. For further information see Table 25. **(B)** Log-transformed gene expression results of the probes of CAC, CrAT, ALAS-H and cPLA2 in

383 individuals with NGT, 104 with IGT and 26 patients with dT2D are shown from cross-sectional analysis of the KORA S4 survey. The *P*-values were adjusted for sex, age, BMI, physical activity, alcohol intake, smoking, systolic BP, HDL cholesterol, HbA_{1c} and fasting glucose when IGT individuals were compared with NGT participants.

Table 20: The 46 T2D-related genes used in the network analysis

Abbreviations and full names of the 46 T2D-related genes are shown in the first and second column, respectively. The columns list approach, association and references.

Gene	Full name	Approach	Association	References
PPARG	Peroxisome proliferator-activated receptor gamma	*Candidate Gene Studies*	T2DM	(Scott *et al*, 2007)
TCF7L2	Transcription factor 7-like 2	*Large-scale association efforts*	T2DM/glucose	(Grant *et al*, 2006; Saxena *et al*, 2010; Scott *et al*, 2007; Sladek *et al*, 2007; WTCCC, 2007)
HNF1A	HNF1 homeobox A	*GWAS for T2D*	T2DM	(Voight *et al*, 2010)
GCK	Glucokinase (hexokinase 4)	*GWAS for Related Traits*	T2DM/Glucose/ HOMAB	(Dupuis *et al*, 2010)
IGF-1	Insulin like growth factor 1	*GWAS for Related Traits*	Insulin/HOMAIR	(Dupuis *et al*, 2010)
IRS1	Insulin receptor substrate 1	*GWAS for Related Traits*	T2DM	(Rung *et al*, 2009)
IDE/HHEX/KIF11	Insulin-degrading enzyme /Hematopoietically expressed homeobox / Kinesin family member 11	*GWAS for T2D*	T2DM	(Saxena *et al*, 2007; Scott *et al*, 2007; Sladek *et al*, 2007; Zeggini *et al*, 2007)
KCNJ11	Potassium inwardly-rectifying channel, subfamily J, member 11	*Candidate Gene Studies*	T2DM	(Scott *et al*, 2007)
CDKN2A/CDKN 2B	Cyclin-dependent kinase inhibitor 2A/2B	*GWAS for T2D*	T2DM	(Saxena *et al*, 2007; Scott *et al*, 2007; Zeggini *et al*, 2007)
NOTCH2	Notch 2	*GWAS for T2D*	T2DM	(Zeggini *et al*, 2008)
ZBED3	Zinc finger, BED-type containing 3	*GWAS for T2D*	T2DM	(Voight *et al*, 2010)
VPS13C	Vacuolar protein sorting 13	*GWAS for Related Traits*	Glucose	(Saxena *et al*, 2010)
FADS1	Fatty acid desaturase 1	*GWAS for Related Traits*	Glucose/HOMAB	(Dupuis *et al*, 2010)
GCKR	Glucokinase regulatory protein	*GWAS for Related Traits*	T2DM/Glucose/HOMAB/ INSULIN/HOMAIR	(Dupuis *et al*, 2010; Saxena *et al*, 2010)
MADD	Map kinase-activating death domain	*GWAS for Related Traits*	Glucose	(Dupuis *et al*, 2010)
PRC1	Protein regulating cytokinesis 1	*GWAS for T2D*	T2DM	(Voight *et al*, 2010)
GIPR	Gastric inhibitory polypeptide receptor	*GWAS for Related Traits*	Glucose/insulinogenic index	(Saxena *et al*, 2010)

ADCY5	Adenylate cyclase 5	*GWAS for Related Traits*	T2DM/Glucose/HOMAB	(Saxena *et al*, 2010)
CDKAL1	Cdk5 regulatory subunit-associated protein 1-like 1	*GWAS for T2D*	T2DM	(Saxena *et al*, 2007; Scott *et al*, 2007; Steinthorsdottir *et al*, 2007; WTCCC, 2007; Zeggini *et al*, 2007)
IGF2BP2	Insulin-like growth factor 2 mrna-binding protein 2	*GWAS for T2D*	T2DM	(Saxena *et al*, 2007; Scott *et al*, 2007; Zeggini *et al*, 2007)
WFS1	Wolfram syndrome 1	*Large-scale association efforts*	T2DM	(Sandhu *et al*, 2007)
HNF1B (=TCF2)	Hnf1 homeobox b	*Large-scale association efforts*	T2DM	(Gudmundsson *et al*, 2007)
TSPAN8-LGR5	Tetraspanin 8	*GWAS for T2D*	T2DM	(Zeggini *et al*, 2008)
ADAMTS9	A disintegrin-like and metalloproteinase with thrombospondin type 1 motif, 9	*GWAS for T2D*	T2DM	(Zeggini *et al*, 2008)
FTO	Fat mass- and obesity-associated gene	*GWAS for T2D*	T2DM/BMI	(Dina *et al*, 2007; Scott *et al*, 2007; WTCCC, 2007)
SLC30A8	Solute carrier family 30 (zinc transporter), member 8	*GWAS for T2D*	T2DM	(Saxena *et al*, 2007; Scott *et al*, 2007; Sladek *et al*, 2007; Zeggini *et al*, 2007)
CDC123-CAMK1D	Cell division cycle 123 homolog (S. Cerevisiae)	*GWAS for T2D*	T2DM	(Zeggini *et al*, 2008)
THADA	Thyroid adenoma-associated gene	*GWAS for T2D*	T2DM	(Zeggini *et al*, 2008)
JAZF1	Juxtaposed with another zinc finger gene 1	*GWAS for T2D*	T2DM	(Zeggini *et al*, 2008)
KCNQ1	Potassium channel, voltage-gated, kqt-like subfamily, member 1	*GWAS for T2D*	T2DM	(Unoki *et al*, 2008; Voight *et al*, 2010; Yasuda *et al*, 2008)
MTNR1B	Melatonin receptor 1b	*GWAS for Related Traits*	T2DM/Glucose/HOMAB	(Dupuis *et al*, 2010; Prokopenko *et al*, 2009)
DUSP9	Dual-specificity phosphatase 9	*GWAS for T2D*	T2DM	(Voight *et al*, 2010)
ZFAND6	Zinc finger, an1-type, domain-containing protein 6	*GWAS for T2D*	T2DM	(Voight *et al*, 2010)
CENTD2	Centaurin, delta-2	*GWAS for T2D*	T2DM	(Voight *et al*, 2010)
TP53INP1	Tumor protein p53-inducible nuclear protein 1	*GWAS for T2D*	T2DM	(Voight *et al*, 2010)
KLF14	Kruppel-like factor 14	*GWAS for T2D*	T2DM	(Voight *et al*, 2010)
BCL11A	B-cell cll/lymphoma 11a	*GWAS for T2D*	T2DM	(Voight *et al*, 2010)

CHCHD9	Coiled-coil-helix-coiled-coil-helix domain containing 9	*GWAS for T2D*	T2DM	(Voight *et al*, 2010)
HMGA2	High mobility group at-hook 2	*GWAS for T2D*	T2DM	(Voight *et al*, 2010)
DGKB-TMEM195	Diacylglycerol kinase, beta	*GWAS for Related Traits*	T2DM/Glucose/HOMAB	(Dupuis *et al*, 2010)
PROX1	Prospero-related homeobox 1	*GWAS for Related Traits*	T2DM/Glucose	(Dupuis *et al*, 2010)
SLC2A2	Solute carrier family 2	*GWAS for Related Traits*	Glucose	(Dupuis *et al*, 2010)
G6PC2	Glucose-6-phosphatase	*GWAS for Related Traits*	Glucose/HOMAB	(Dupuis *et al*, 2010)
GLIS3	Glis family zinc finger protein 3	*GWAS for Related Traits*	Glucose/HOMAB	(Dupuis *et al*, 2010)
ADRA2A	Alpha-2a-adrenergic receptor	*GWAS for Related Traits*	Glucose	(Dupuis *et al*, 2010)
CRY2	Cryptochrome 2	*GWAS for Related Traits*	Glucose	(Dupuis *et al*, 2010)

Table 21: The 21 links between metabolites, proteins and T2D-related genes

Names of metabolites, proteins, and T2D-related genes are shown in the first and second columns, respectively. The following columns list actions, interaction type, score and reference for each link, respectively. The four enzymes are: carnitine/acylcarnitine translocase (CAC), carnitine acetyltransferase (CrAT), 5-aminolevulinate synthase 1 (ALAS-H) and cytosolic phospholipase A2 (cPLA2). The five other proteins are: peroxisome proliferator activated receptor alpha (PPAR-α), acyl-CoA oxidase 1, palmitoyl (AOX), insulin precursor (INS), mitogen-activated protein kinase 1 (MAPK1) and glucocorticoid receptor (GR). The seven T2D-related genes are: peroxisome proliferator-activated receptor gamma (*PPARG*), Transcription factor 7-like 2 (*TCF7L2*), HNF1 homeobox A (*HNF1A*), Glucokinase (*GCK*), insulin-like growth factor 1 (*IGF1*), insulin receptor substrate 1 (*IRS1*) and insulin-degrading enzyme (*IDE*).

Metabolites/ proteins	**Proteins/ T2D-related genes**	**Actions**	**Interaction type**	**Score**	**References**
C2	CAC		Physical interaction	1	(Pande, 1975)
C2	CrAT		Physical interaction	1	(Bremer, 1983)
Gly	ALAS-H		Physical interaction	1	(Bishop, 1990)
LPC (18:2)	cPLA2		Physical interaction	1	(Sharp *et al*, 1991)
CAC	PPAR-α	PPARalpha induces the SLC25A20 expression	Transcription	0.923	(Tachibana *et al*, 2009)
PPAR-α	*PPARG*	PPAR signaling pathway	Same pathway	0.905	(Hihi *et al*, 2002)
CrAT	AOX	Peroxisome	Same pathway	0.754	(Lamers *et al*, 2011)
PPARG	AOX	PPARG -> AOX	Signaling regulation	0.772	(Waku *et al*, 2010)
ALAS-H	INS	INS -| ALAS-H	Transcription	0.899	(Scassa *et al*, 2004)

INS	*PPARG*	INS -> PPAR; PPAR -> INS	Signaling regulation	0.992	(Seto-Young *et al*, 2007)
INS	*TCF7L2*	TCF7L2 --> INS	Transcription	0.972	(Loder *et al*, 2008)
INS	*HNF1A*	HNF1A --> INS	Transcription	0.992	(Bartoov-Shifman *et al*, 2002)
INS	*GCK*	GCK regulate INS secretion	Transcription	0.993	(Hohmeier *et al*, 1997)
INS	*IGF1*	IGF-1 and growth hormone interact with insulin to modulate its control of carbohydrate metabolism	Physical interaction	0.994	(Yakar *et al*, 2004)
INS	*IRS1*		Physical interaction	0.999	(Giorgetti *et al*, 1993)
INS	*IDE*		Physical interaction	0.991	(Lee *et al*, 1996)
cPLA2	MAPK1	MAPK1 -> cPLA2	Signaling regulation	0.977	(Lin *et al*, 1993)
MAPK1	*IGF1*	IGF1 stimulate ERK2/MAPK1 activity	Transcription	0.98	(Kooijman *et al*, 2003)
MAPK1	*IRS1*	IRS1 -> MAPK1	Signaling regulation	0.99	(Yi *et al*, 2005)
cPLA2	GR	GR -> cPLA2	Transcription	0.818	(Guo *et al*, 2008)
GR	*IDE*		Physical interaction	0.927	(Kupfer *et al*, 1994)

Gene expression analysis in whole-blood samples of participants from the KORA S4 survey revealed significant variations (*P*-values ranging from 9.4×10^{-3} to 1.1×10^{-6}) of transcript levels of four enzymes, namely, carnitine/acylcarnitine translocase (CAC), carnitine acetyltransferase (CrAT), 5-aminolevulinate synthase 1 (ALAS-H) and cytosolic phospholipase A2 (cPLA2), which are known to be strongly associated with the levels of the three metabolites (Figure 12B). The clear relationship between changes in metabolites and transcription levels of associated enzymes strongly suggests that these metabolites are functionally associated with T2D genes in established pathways.

Chapter 4 Discussion

4.1 Plasma and serum

In the first part of the results chapter, we presented a robust analysis based on a large size of samples and highly reliable measurements of metabolites with stringent quality controls. The method, based on FIA MS/MS has been proven to be in conformance with the FDA-Guideline “Guidance for Industry - Bioanalytical Method Validation (May 2001)”, which implies proof of reproducibility within a given error range.

Our results give support to the good reproducibility of metabolite measurements in both plasma and serum. Moreover, plasma demonstrates to have a better reproducibility than serum, which may result from the less complicated collecting procedure for plasma, as it does not require time to coagulate and thus leads to less exposure time at the room temperature. The large sample size is not only powerful enough to detect metabolite concentration differences between the two matrices but also makes it possible to further characterize the relationship between them.

We observed that metabolite concentrations were generally higher in serum and this phenomenon may partly be explained by the so called volume displacement effect (Kronenberg et al., 1998) which means that deproteinization of serum eliminates the volume fraction of proteins and distributes the remaining small molecular weight constituents in a smaller volume, thus making them more concentrated and leading to a higher serum concentration. However, the volume displacement effect usually accounts for about 5% difference of the concentration, which means there are other reasons causing the differences we observed. Concentration differences in some metabolites were similar to those reported in previous studies and some differences were related to coagulation processes. The higher arginine concentration in serum has been observed before (Teerlink et al., 2002). The release of arginine from platelets during the coagulation process might account for this difference.

Our observations that concentrations of some LPCs were higher in serum are consistent with a former study (Aoki et al., 2002), who reported increased LPC concentrations, due to the release of phospholipases by platelets activated by thrombin, a process that also occurs upon coagulation. Glucose, which comprise the majority of hexose, was found in an earlier study (Ladenson et al., 1974) to be 5% lower in plasma than in serum. A similar difference was observed for hexose in our measurements. Although the exact reason for this observation is not clear, a shift in fluid from erythrocytes to plasma caused by anticoagulants might play a role (Sacks et al., 2002). Serum also demonstrated a higher sensitivity in biomarker detection in the three phenotypes (gender, diabetic status, smoking status) we chose. The generally higher metabolite concentrations in serum than in plasma could contribute to this advantage. Metabolite measurements in both matrices are subject to a certain level of background noise, which might affect measurement accuracy, especially for metabolites with low concentrations. Thus plasma is more prone to this effect than serum, where metabolite concentrations are generally higher. It was also proposed that the lower protein content in serum might benefit small molecule analyses and improve overall sensitivity (Denery et al., 2011). However, in our comparisons, the metabolites that differed significantly between two phenotypes in serum but not in plasma are, nevertheless, close to the significance level when plasma was used, an observation that is in agreement with the existence of high correlations between both matrices. The high correlations between plasma and serum measurements suggest that the shift in metabolite concentrations per se does not necessarily introduce a bias in epidemiological studies, although the higher concentrations in serum may provide some advantages. In general, our data indicate that metabolite profiles from either matrix can be analyzed, as long as the same blood sample is used. However, the better reproducibility in plasma and higher sensitivity in serum need to be taken into account, as they might influence the results for the identification of diagnostic biomarkers. Naturally, the metabolites we measured in our experiment represent only a small part of the human blood metabolome. Accordingly, it is yet to be determined in future studies whether similar observations can be made for other metabolites.

4.2 Sex dimorphism

There have been only a few studies addressing metabolic differences between males and females, and most of these studies were rather small in sample size and determined only a small number of metabolites (Döring et al., 2008; Geller et al., 2006). We investigated a number of 131 metabolites in a large population based study with sufficient statistical power to examine associations within subgroups. Our findings shed light on the sex-specific architectures of human metabolome and provided clues on biochemical mechanisms that might explain observed differences in susceptibility and time course of the development of common diseases in males and females. Our data provided new insights into sex-specific metabotype differences. Combining results from linear regression with partial correlation analysis (resulting in a Gaussian graphical model) yielded interesting insights into how sex-specific concentration differences spread over the metabolic network (Figure 3). The analysis suggests that sex-specific concentration differences affect whole metabolic pathways rather than being randomly spread over the different metabolites. In addition, we found three interesting inter-class associations between PCaa/PCae species and LPC species (highlighted in yellow in Figure 5). Those pairs shared a strong partial correlation but displayed differential concentration patterns with respect to gender effects. Furthermore, these pairs displayed a fatty acid residue difference of C18:0, indicating that this fatty acid species might be a key compound giving rise to opposing metabolic gender effects. Direct experimental evidence indicated a role for sphingolipids (SMs and ceramides) in several common complex chronic disease processes including atherosclerotic plaque formation, myocardial infarction, cardiomyopathy, pancreatic beta cell failure, insulin resistance, coronary heart disease and T2D (Holland and Summers, 2008; Yeboah et al., 2010). Evidences showed that in young children (between birth and 4 years old, with low levels of sex-hormones) there may already have been significant sex-specific differences in plasma sphingolipid concentrations (Nikkilä et al., 2008). Our observations described new sex-specific differences, while other lipid-derived molecules, like bile acids, were already demonstrated not to be sex-specific (Rodrigues et al., 1996). Therefore sphingomyelins represent important intermediate phenotypes. The concentration differences between males and females of acylcarnitines described in this study coincide with previous findings showing that carnitine (C0) and acetylcarnitine (C2) concentrations were higher in males than in females (Reuter et al., 2008; Slupsky et al., 2007). Phosphatidylcholines, as demonstrated in this study, are another gender-specific phenotype. Ghrelin (controlling energy homeostasis and pituitary hormone secretion in humans) levels have been shown to be similar in men and women and did not vary by menopausal status or in association with cortisol levels (Purnell et al., 2003). These findings of our and other studies urgently suggest when using metabolites for disease prediction sex has to be strictly taken into account. As global 'omics'-techniques are more and more refined to identify more compounds in single biological samples, the predictive power of these new technologies will greatly increase. Metabolite concentration profiles can be used as predictive biomarkers to indicate the presence or severity of a disease depending on sex. Our study provides new important insights into sex-specific differences of cell regulatory processes and underscores that studies should consider gender-specific effects in design and interpretation. Our findings also help to understand the biochemical mechanisms underlying sexual dimorphism, a phenomenon which may explain the differential susceptibility to common diseases in males and females.

4.3 Novel markers for pre-diabetes

Using a cross-sectional approach (KORA S4, F4), we analyzed 140 metabolites and identified three (glycine, LPC (18:2) and C2) that are IGT-specific metabolites with high statistical significance. Notably, these three metabolites are distinct from the currently known T2D risk indicators (e.g., age, BMI, systolic BP, HDL cholesterol, HbA1c, fasting glucose and fasting insulin). A prospective analysis (KORA S4-F4) shows that low levels of glycine and LPC at baseline predict the risks of developing IGT and/or T2D. Glycine and LPC especially were shown to be strong predictors of glucose tolerance, even 7 years before disease onset. Moreover, those two metabolites were independently replicated in the EPIC-Potsdam cross-sectional study. Finally, based on our analysis of interaction networks, and supported by gene expression profiles, we found that seven T2D-related genes are functionally associated with the three IGT candidate metabolites.

4.3.1 Different study designs reveal progression of IGT and T2D

From a methodological point of view, our study is unique with respect to the large sample sizes and the availability of metabolomics data from two time points. This allowed us to compare results generated with cross-sectional and prospective approaches directly, as well as with results from prospective population-based cohort and nested case–control designs. We found that individuals with IGT have elevated concentrations of the acetylcarnitine C2 as compared with NGT individuals only in the cross-sectional study, whereas C2 was unable to predict IGT and T2D seven years before the disease onset. We speculate that the acetylcarnitine C2 might be an event with a quick effect.

Our analysis could replicate four out of the five branched-chain and aromatic amino acids recently reported to be predictors of T2D using a nested case–control study design (Wang et al., 2011). However, the population-based prospective study employed in our study revealed that these five amino acids are in fact not associated with future 2-h glucose values. It should be taken into account, however, that more pre-diabetes individuals (~ 50%) were in the control group of that study design, and that these markers were unable to be extended to the general population (with only 0.4% improvement from the T2D risk indicators as reported in the Framingham Offspring Study) (Wang et al., 2011). Most likely, changes in these amino acids happen at a later stage in the development of T2D (e.g., from IGT to T2D); indeed, similar phenomenon was also observed in our study (Figure 10D). In contrast, we found that combined glycine, LPC (18:2) and C2 have 2.6 and 1% increment in predicting IGT and T2D in addition to the common risk indicators of T2D. This suggests they are better candidate for early biomarkers, and specifically from NGT to IGT, than the five amino acids.

4.3.1 IFG and IGT should be considered as two different phenotypes

By definition (WHO, 1999; ADA, 2010), individuals with IFG or IGT or both are considered as pre-diabetics. Yet we observed different behaviors regarding the change of the metabolite panel from NGT to i-IFG or to IGT, indicating that i-IFG and IGT are two different phenotypes. For future studies, we therefore suggest separating IFG from IGT.

4.3.2 Glycine

The observed decrease in the serum concentration of glycine in individuals with IGT and dT2D may result from insulin resistance (Pontiroli et al., 2004). It was already reported that insulin represses ALAS-H expression (Phillips and Kushner, 2005). As insulin sensitivity progressively decreases during diabetes development (Færch et al., 2009; McGarry, 2002; Stumvoll et al., 2005; Tabák et al., 16), it is expected that the expression levels of the enzyme increase in individuals with IGT and dT2D, since ALAS-H catalyzes the condensation of glycine and succinyl-CoA into 5-aminolevulinic acid (Bishop, 1990). This may explain our observation that glycine was lower in both individuals with IGT and those with dT2D. However, the level of fasting insulin in IGT and T2D individuals was higher than in NGT participants in the KORA S4 study, suggesting that yet undetected pathways may also play roles here.

4.3.3 Acetylcarnitine C2

Acetylcarnitine is produced by the mitochondrial matrix enzyme, CrAT, from carnitine and acetyl-CoA, a molecule that is a product of both fatty acid β-oxidation and glucose oxidation and can be used by the citric acid cycle for energy generation. We observed higher transcriptional level of CrAT in indivi duals with IGT and T2D, most probably due to an activation of the peroxisome proliferator activated receptor alpha (PPAR-a) pathway in peroxisomes (Horie et al., 1981). Higher expression of CrAT would explain the elevated levels of acetylcarnitine C2 in IGT individuals. Although it is not clear if mitochondrial CrAT is overexpressed when there is increased fatty acid b-oxidation (e.g., in diabetes; Noland et al, 2009), it is expected that additional acetylcarnitine will be formed by CrAT due to increased substrate availability (acetyl-CoA), thereby releasing pyruvate dehydrogenase inhibition by acetyl-CoA and stimulating glucose uptake and oxidation. An increase of acylcarnitines, and in particular of acetylcarnitine C2, is a hallmark in diabetic people (Adams et al., 2009). Cellular lipid levels are increased in humans with IGT or overt T2D who also may have altered mitochondrial function (Szendroedi et al., 2007). Together, these findings reflect an important role of increased cellular lipid metabolites and impaired mitochondrial b-oxidation in the development of insulin resistance (Koves et al., 2008; McGarry, 2002; Szendroedi et al., 2007).

4.3.4 LPC (18:2)

In our study, individuals with IGT and dT2D had lower cPLA2 transcription levels, suggesting reduced cPLA2 activity. As a result, a concomitant decrease in the concentration of arachidonic acid (AA), a product of cPLA2 activity, is expected. AA has been shown to inhibit glucose uptake by adipocytes (Malipa et al., 2008), in a mechanism that is probably insulin independent and that involves the GLUT-1 transporter. Therefore, our findings may point to regulatory effects in individuals with IGT, since the inhibition of AA production would result in an increased glucose uptake.

4.3.5 Limitations

While our metabolite profiles provide a snapshot of human metabolism, more detailed metabolic profile follow-ups, with longer time spans and more time points, are necessary to further evaluate the development of the novel biomarkers. Moreover, the influence from long-term dietary habits should not be ignored, even though we used only serum from fasting individuals (Altmaier et al., 2011; Primrose et al., 2011). Furthermore, additional tissue samples (e.g., muscle and adipocytes) and experimental approaches are needed to characterize the causal pathways in detail.

4.3.6 Conclusions

Three novel metabolites, glycine LPC (18:2) and C2, were identified as pre-diabetes-specific markers. Their changes might precede other branched-chain and aromatic amino acids markers in the progression of T2D. Combined levels of glycine, LPC (18:2) and C2 can predict risk not only for IGT but also for T2D. Targeting the pathways that involve these newly proposed potential biomarkers would help to take preventive steps against T2D at an earlier stage.

Summary

This thesis presented three metabolomics studies using the KORA cohort. The main aim of the thesis was to more thoroughly understand the role of the metabolome in complex phenotypes including differences in blood matrix, sex, and how the metabolite profiles change in a complex disease like type 2 diabetes (T2D).

All measured metabolites were filtered using strict quality controls to exclude artifacts. By collecting serum and plasma samples from the same 377 individuals, we found that the concentrations in plasma and serum were highly correlated, with both providing good reproducibility, although plasma was slightly better. On the contrary, serum showed higher concentrations and therefore is more likely to detect differences in the metabolite concentrations in serum.

With regards to the second topic of the thesis, we also demonstrated that 102 of 131 metabolites had significantly different metabolite concentrations by comparing males and females. Altogether, more than 3300 KORA individuals were analyzed and all analyses were Bonferroni corrected.

Furthermore, we quantified 140 metabolites in 4297 fasting serum samples from KORA with a view to identifying the candidate biomarkers of pre-diabetes. Three metabolites (glycine, LPC 18:2 and acetylcarnitine) were found to have significantly altered levels in impaired glucose tolerance (IGT) individuals. Lower levels of glycine and LPC were also proven to be predictive for IGT as well as for T2D. All these

identified metabolites were independent of previously identified diabetes risk factors. Further investigations including a systems biology approach were performed and we identified seven T2D-related genes which were linked to T2D through functional related enzymes; a theory which was confirmed by expression data.

Metabolomics, which studies the intermediates and end products of biological processes, is a useful tool in biomedical research, particularly for metabolic diseases. When proper quality controls are applied and the effects of the complex confounders (e.g. sex) are unveiled, the relationships between the metabolome and the diseases become even clearer. The findings in our T2D study proved that mining the metabolite profiles can help to detect novel disease markers as well as new pathways which can potentially be targeted to prevent the disease.

Zusammenfassung

In dieser Doktorarbeit werden drei Metabolomics-Studien der KORA Kohorte behandelt. Das Ziel dieser Doktorarbeit war es, ein besseres Verständnis der Rolle des Metabolismus von komplexen Phänotypen anhand von Unterschieden im Blutbild, des Geschlechts und anhand von Veränderungen des Metabolitenprofils bei multifaktoriellen Krankheiten wie Typ 2 Diabetes mellitus zu erhalten.

Um Artefakte auszuschließen wurden strikte Qualitätskontrollen aller gemessenen Metaboliten durchgeführt. Durch die Analyse von Blutplasma und -serum von 377 Personen konnten wir zeigen, dass die Konzentrationen der Metaboliten in Blutplasma und -serum stark korrelieren und darüber hinaus eine hohe Reproduzierbarkeit zeigen, bei der Blutplasma besser abschneidet. Im Gegensatz dazu zeigt das Blutserum höhere Metabolitenkonzentrationen und könnte deswegen besser für den Nachweis von Konzentrationsunterschieden geeignet sein.

Ein weiteres Ergebnis dieser Doktorarbeit war der Nachweis von signifikanten geschlechtsspezifischen Unterschieden der Konzentrationen von 102 der ausgewerteten 131 Metaboliten. Dabei wurden die Daten von mehr als 3300 Personen der KORA Kohorte verwendet und die Analysen einer konservativen Bonferroni-Korrektur unterzogen.

Darüber hinaus identifizierten wir potentielle Biomarker für Prä-Diabetes durch die Analyse von 140 Metaboliten in nüchtern abgegebenen Blutseren von 4297 Personen

der KORA Kohorte. Wir konnten zeigen, dass Personen mit gestörter Glukosetoleranz (IGT) signifikant unterschiedliche Konzentrationen von drei Metaboliten (Glycin, lysoPhosphatidylcholine (LPC) 18:2 und acetylcarnitine) im Vergleich zu gesunden Personen aufweisen. Darüber hinaus konnten wir nachweisen, dass geringere Konzentrationen der Metaboliten Glycin und LPC bei Probanden mit Typ 2 Diabetes oder IGT vorhanden sind. Die in dieser Studie identifizierten Metaboliten sind biologisch unabhängig von zuvor entdeckten Diabetes Risikofaktoren. Durch weitere Analysen und die Einbeziehung systembiologischer Ansätze entdeckten wir sieben Diabetesrisiko Susseptibilitätsgene, welche durch Expressionsdaten bestätigt wurden.

Metabolomics welches auf der Analyse von Stoffwechselzwischen- und Endprodukten basiert, ist eine wertvolle Methode besonders in der biomedizinischen Forschung, um Krankheitsmechanismen aufzuklären. Nachdem angemessene Qualitätskontrollen etabliert und der Einfluss von komplexen Störfaktoren (z.B. das Geschlecht) aufgeklärt wurden, konnte der Zusammenhang zwischen Krankheit und Metabolismus weiter an Klarheit gewinnen. Die Entdeckungen in unserer T2D Studie zeigen, dass die Analyse von Konzentrationsprofilen helfen kann neue Krankheitsrisikomarker genauso wie neue Wirkungspfade zu identifizieren, die möglicherweise das Ziel zur Heilung einer Krankheit sein könnten.

References

Adams, S.H., Hoppel, C.L., Lok, K.H., Zhao, L., Wong, S.W., Minkler, P.E., Hwang, D.H., Newman, J.W., Garvey, W.T., 2009. Plasma Acylcarnitine Profiles Suggest Incomplete Long-Chain Fatty Acid β-Oxidation and Altered Tricarboxylic Acid Cycle Activity in Type 2 Diabetic African-American Women. J. Nutr. 139, 1073–1081.

Altmaier, E., Kastenmüller, G., Römisch-Margl, W., Thorand, B., Weinberger, K., Illig, T., Adamski, J., Döring, A., Suhre, K., 2011. Questionnaire-based self-reported nutrition habits associate with serum metabolism as revealed by quantitative targeted metabolomics. European Journal of Epidemiology 26, 145–156.

Aoki, J., Taira, A., Takanezawa, Y., Kishi, Y., Hama, K., Kishimoto, T., Mizuno, K., Saku, K., Taguchi, R., Arai, H., 2002. Serum Lysophosphatidic Acid Is Produced through Diverse Phospholipase Pathways. Journal of Biological Chemistry 277, 48737 –48744.

Artati, A., Prehn, C., Möller, G., Adamski, J., 2012. Assay Tools for Metabolomics, in: Suhre, K. (Ed.), Genetics Meets Metabolomics. Springer New York, New York, NY, pp. 13–38.

Ayache, S., Panelli, M.C., Byrne, K.M., Slezak, S., Leitman, S.F., Marincola, F.M., Stroncek, D.F., 2006. Comparison of proteomic profiles of serum, plasma, and modified media supplements used for cell culture and expansion. J Transl Med. 4, 40.

Bando, K., Kawahara, R., Kunimatsu, T., Sakai, J., Kimura, J., Funabashi, H., Seki, T., Bamba, T., Fukusaki, E., 2010. Influences of biofluid sample collection and handling procedures on GC-MS based metabolomic studies. Journal of Bioscience and Bioengineering 110, 491–499.

Barelli, S., Crettaz, D., Thadikkaran, L., Rubin, O., Tissot, J.-D., 2007. Plasma/serum proteomics: pre-analytical issues. Expert Rev Proteomics 4, 363–370.

Beaglehole, R., Bonita, R., Kjellström, T., 2006. Basic Epidemiology, Second Edition, 2nd ed. World Health Organisation.

Beheshti, I., Wessels, L.M., Eckfeldt, J.H., 1994. EDTA-plasma vs serum differences in cholesterol, high-density-lipoprotein cholesterol, and triglyceride as measured by several methods. Clinical Chemistry 40, 2088–2092.

Bishop, D.F., 1990. Two different genes encode delta-aminolevulinate synthase in humans: nucleotide sequences of cDNAs for the housekeeping and erythroid genes. Nucleic Acids Res 18, 7187–7188.

Bolstad, B.M., Irizarry, R.A., Åstrand, M., Speed, T.P., 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19, 185–193.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

Brown, M., Dunn, W.B., Ellis, D.I., Goodacre, R., Handl, J., Knowles, J.D., O'Hagan, S., Spasić, I., Kell, D.B., 2005. A metabolome pipeline: from concept to data to knowledge. Metabolomics 1, 39–51.

Denery, J.R., Nunes, A.A.K., Dickerson, T.J., 2011. Characterization of Differences between Blood Sample Matrices in Untargeted Metabolomics. Analytical Chemistry 83, 1040–1047.

Döring, A., Gieger, C., Mehta, D., Gohlke, H., Prokisch, H., Coassin, S., Fischer, G., Henke, K., Klopp, N., Kronenberg, F., Paulweber, B., Pfeufer, A., Rosskopf, D., Völzke, H., Illig, T., Meitinger, T., Wichmann, H.-E., Meisinger, C., 2008. SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. Nat. Genet. 40, 430–436.

Færch, K., Vaag, A., Holst, J.J., Hansen, T., Jørgensen, T., Borch-Johnsen, K., 2009. Natural History of Insulin Sensitivity and Insulin Secretion in the Progression From Normal Glucose Tolerance to Impaired Fasting Glycemia and Impaired Glucose Tolerance: The Inter99 Study. Dia Care 32, 439–444.

Fairweather, D., Rose, N.R., 2004. Women and Autoimmune Diseases1. Emerg Infect Dis 10, 2005–2011.

Fiehn, O., 2002. Metabolomics – the link between genotypes and phenotypes. Plant Molecular Biology 48, 155–171.

Floegel, A., Stefan, N., Yu, Z., Mühlenbruch, K., Drogan, D., Joost, H.-G., Fritsche, A., Häring, H.-U., Angelis, M.H. de, Peters, A., Roden, M., Prehn, C., Wang-Sattler, R., Illig, T., Schulze, M.B., Adamski, J., Boeing, H., Pischon, T., 2012. Identification of Serum Metabolites Associated With Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach. Diabetes.

Geller, S.E., Adams, M.G., Carnes, M., 2006. Adherence to Federal Guidelines for Reporting of Sex and Race/Ethnicity in Clinical Trials. Journal of Women's Health 15, 1123–1131.

Gerhardt, W., Nordin, G., Herbert, A.-K., Burzell, B.L., Isaksson, A., Gustavsson, E., Haglund, S., Müller-Bardorff, M., Katus, H.A., 2000. Troponin T and I Assays Show Decreased Concentrations in Heparin Plasma Compared with Serum: Lower Recoveries in Early than in Late Phases of Myocardial Injury. Clinical Chemistry 46, 817–821.

German, J.B., Hammock, B.D., Watkins, S.M., 2005a. Metabolomics: building on a century of biochemistry to guide human health. Metabolomics 1, 3–9.

German, J.B., Watkins, S.M., Fay, L.-B., 2005b. Metabolomics in Practice: Emerging Knowledge to Guide Future Dietetic Advice toward Individualized Health. Journal of the American Dietetic Association 105, 1425–1432.

Gieger, C., Geistlinger, L., Altmaier, E., Hrabé de Angelis, M., Kronenberg, F., Meitinger, T., Mewes, H.-W., Wichmann, H.-E., Weinberger, K.M., Adamski, J., Illig, T., Suhre, K., 2008. Genetics Meets Metabolomics: A Genome-Wide Association Study of Metabolite Profiles in Human Serum. PLoS Genet 4, e1000282.

Holland, W.L., Summers, S.A., 2008. Sphingolipids, Insulin Resistance, and Metabolic Disease: New Insights from in Vivo Manipulation of Sphingolipid Metabolism. Endocrine Reviews 29, 381–402.

Holle, R., Happich, M., Löwel, H., Wichmann, H.E., 2005. KORA--a research platform for population based health research. Gesundheitswesen 67 Suppl 1, S19–25.

Horie, S., Ishii, H., Suga, T., 1981. Changes in Peroxisomal Fatty Acid Oxidation in the Diabetic Rat Liver. J Biochem 90, 1691–1696.

Illig, T., Gieger, C., Zhai, G., Romisch-Margl, W., Wang-Sattler, R., Prehn, C., Altmaier, E., Kastenmuller, G., Kato, B.S., Mewes, H.-W., Meitinger, T., De Angelis, M.H., Kronenberg, F., Soranzo, N., Wichmann, H.-E., Spector, T.D., Adamski, J., Suhre, K., 2010. A genome-wide perspective of genetic variation in human metabolism. Nat Genet 42, 137–141.

Jaffe, A.S., Ravkilde, J., Roberts, R., Naslund, U., Apple, F.S., Galvani, M., Katus, H., 2000. It's Time for a Change to a Troponin Standard. Circulation 102, 1216–1220.

Jourdan, C., Petersen, A.-K., Gieger, C., Döring, A., Illig, T., Wang-Sattler, R., Meisinger, C., Peters, A., Adamski, J., Prehn, C., Suhre, K., Altmaier, E., Kastenmüller, G., Römisch-Margl, W., Theis, F.J., Krumsiek, J., Wichmann, H.-E., Linseisen, J., 2012. Body Fat Free Mass Is Associated with the Serum Metabolite Profile in a Population-Based Study. PLoS ONE 7, e40009.

Kaddurah-Daouk, R., Kristal, B.S., Weinshilboum, R.M., 2008. Metabolomics: A Global Biochemical Approach to Drug Response and Disease. Annu. Rev. Pharmacol. Toxicol. 48, 653–683.

Kastenmüller, G., Römisch-Margl, W., Wägele, B., Altmaier, E., Suhre, K., 2011. metaP-Server: A Web-Based Metabolomics Data Analysis Tool 2011.

Kim, A.M., Tingen, C.M., Woodruff, T.K., 2010. Sex bias in trials and treatment must end. Nature 465, 688–689.

Knowler, W.C., Barrett-Connor, E., Fowler, S.E., Hamman, R.F., Lachin, J.M., Walker, E.A., Nathan, D.M., 2002. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N. Engl. J. Med. 346, 393–403.

Koves, T.R., Ussher, J.R., Noland, R.C., Slentz, D., Mosedale, M., Ilkayeva, O., Bain, J., Stevens, R., Dyck, J.R.B., Newgard, C.B., Lopaschuk, G.D., Muoio, D.M., 2008. Mitochondrial Overload and Incomplete Fatty Acid Oxidation Contribute to Skeletal Muscle Insulin Resistance. Cell Metabolism 7, 45–56.

Krebs, M., Krssak, M., Bernroider, E., Anderwald, C., Brehm, A., Meyerspeer, M., Nowotny, P., Roth, E., Waldhäusl, W., Roden, M., 2002. Mechanism of Amino Acid-Induced Skeletal Muscle Insulin Resistance in Humans. Diabetes 51, 599–605.

Kronenberg, F., Trenkwalder, E., Kronenberg, M.F., K\önig, P., Utermann, G., Dieplinger, H., 1998. Influence of hematocrit on the measurement of lipoproteins demonstrated by the example of lipoprotein (a). Kidney International 54, 1385–1389.

Krug, S., Kastenmüller, G., Stückler, F., Rist, M.J., Skurk, T., Sailer, M., Raffler, J., Römisch-Margl, W., Adamski, J., Prehn, C., Frank, T., Engel, K.-H., Hofmann, T., Luy, B., Zimmermann, R., Moritz, F., Schmitt-Kopplin, P., Krumsiek, J., Kremer, W., Huber, F., Oeh, U., Theis, F.J., Szymczak, W., Hauner, H., Suhre, K., Daniel, H., 2012. The dynamic range of the human metabolome revealed by challenges. FASEB J 26, 2607–2619.

Krumsiek, J., Suhre, K., Illig, T., Adamski, J., Theis, F., 2011. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. BMC Systems Biology 5, 21.

Ladenson, J.H., Tsai, L.M., Michael, J.M., Kessler, G., Joist, J.H., 1974. Serum versus heparinized plasma for eighteen common chemistry tests: is serum the appropriate specimen? Am. J. Clin. Pathol 62, 545–552.

Liu, L., Aa, J., Wang, G., Yan, B., Zhang, Y., Wang, X., Zhao, C., Cao, B., Shi, J., Li, M., Zheng, T., Zheng, Y., Hao, G., Zhou, F., Sun, J., Wu, Z., 2010. Differences in metabolite profile between blood plasma and serum. Analytical Biochemistry 406, 105–112.

Lorber, A., Wangen, L.E., Kowalski, B.R., 1987. A theoretical foundation for the PLS algorithm. Journal of Chemometrics 1, 19–31.

Malipa, A.C.A., Meintjes, R.A., Haag, M., 2008. Arachidonic acid and glucose uptake by freshly isolated human adipocytes. Cell Biochemistry and Function 26, 221–227.

Mannello, F., 2008. Serum or Plasma Samples?: The "Cinderella" Role of Blood Collection Procedures Preanalytical Methodological Issues Influence the Release and Activity of Circulating Matrix Metalloproteinases and Their Tissue Inhibitors, Hampering Diagnostic Trueness and Leading to Misinterpretation. Arterioscler Thromb Vasc Biol 28, 611–614.

Mayr, M., 2008. Metabolomics Ready for the Prime Time? Circ Cardiovasc Genet 1, 58–65.

McGarry, J.D., 2002. Banting Lecture 2001 Dysregulation of Fatty Acid Metabolism in the Etiology of Type 2 Diabetes. Diabetes 51, 7–18.

Meisinger, C., Strassburger, K., Heier, M., Thorand, B., Baumeister, S.E., Giani, G., Rathmann, W., 2010. Prevalence of undiagnosed diabetes and impaired glucose regulation in 35–59-year-old individuals in Southern Germany: the KORA F4 Study. Diabetic Medicine 27, 360–362.

Mittelstrass, K., Ried, J.S., Yu, Z., Krumsiek, J., Gieger, C., Prehn, C., Roemisch-Margl, W., Polonikov, A., Peters, A., Theis, F.J., others, 2011. Discovery of Sexual Dimorphisms in Metabolic and Genetic Biomarkers. PLoS Genetics 7, e1002215.

MONICA-Projekt, Region Augsburg: Survey sampling / by L. Chambless ..., 1986. . GSF-Forschungszentrum für Umwelt und Gesundheit.

Mostertz W, S.M., 2010. AGe- and sex-specific genomic profiles in non–small cell lung cancer. JAMA 303, 535–543.

Muoio, D.M., Newgard, C.B., 2008. Molecular and metabolic mechanisms of insulin resistance and [beta]-cell failure in type 2 diabetes. Nat Rev Mol Cell Biol 9, 193–205.

Newgard, C.B., An, J., Bain, J.R., Muehlbauer, M.J., Stevens, R.D., Lien, L.F., Haqq, A.M., Shah, S.H., Arlotto, M., Slentz, C.A., Rochon, J., Gallup, D., Ilkayeva, O., Wenner, B.R., Yancy Jr., W.S., Eisenson, H., Musante, G., Surwit, R.S., Millington, D.S., Butler, M.D., Svetkey, L.P., 2009. A Branched-Chain Amino Acid-Related Metabolic Signature that Differentiates Obese and Lean Humans and Contributes to Insulin Resistance. Cell Metabolism 9, 311–326.

Nikkilä, J., Sysi-Aho, M., Ermolov, A., Seppänen-Laakso, T., Simell, O., Kaski, S., Orešič, M., 2008. Gender-dependent progression of systemic metabolic states in early childhood. Mol Syst Biol 4, 197.

Oliver, S.G., Winson, M.K., Kell, D.B., Baganz, F., 1998. Systematic functional analysis of the yeast genome. Trends in Biotechnology 16, 373–378.

Phillips, J.D., Kushner, J.P., 2005. Fast track to the porphyrias. Nature Medicine 11, 1049–1050.

Pietiläinen, K.H., Róg, T., Seppänen-Laakso, T., Virtue, S., Gopalacharyulu, P., Tang, J., Rodriguez-Cuenca, S., Maciejewski, A., Naukkarinen, J., Ruskeepää, A.-L., Niemelä, P.S., Yetukuri, L., Tan, C.Y., Velagapudi, V., Castillo, S., Nygren, H., Hyötyläinen, T., Rissanen, A., Kaprio, J., Yki-Järvinen, H., Vattulainen, I., Vidal-Puig, A., Orešič, M., 2011. Association of Lipidome Remodeling in the Adipocyte Membrane with Acquired Obesity in Humans. PLoS Biol 9, e1000623.

Pontiroli, A.E., Pizzocri, P., Caumo, A., Perseghin, G., Luzi, L., 2004. Evaluation of insulin release and insulin sensitivity through oral glucose tolerance test: differences between NGT, IFG, IGT, and type 2 diabetes mellitus. A cross-sectional and follow-up study. Acta Diabetologica 41, 70–76.

Primrose, S., Draper, J., Elsom, R., Kirkpatrick, V., Mathers, J.C., Seal, C., Beckmann, M., Haldar, S., Beattie, J.H., Lodge, J.K., Jenab, M., Keun, H., Scalbert, A., 2011. Metabolomics and human nutrition. British Journal of Nutrition 105, 1277–1283.

Psychogios, N., Hau, D.D., Peng, J., Guo, A.C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T.L., Smith, S.R., Bamforth, F., Greiner, R., McManus, B., Newman, J.W., Goodfriend, T., Wishart, D.S., 2011. The Human Serum Metabolome. PLoS ONE 6, e16957.

Purnell, J.Q., Weigle, D.S., Breen, P., Cummings, D.E., 2003. Ghrelin Levels Correlate with Insulin Levels, Insulin Resistance, and High-Density Lipoprotein

Cholesterol, But Not with Gender, Menopausal Status, or Cortisol Levels in Humans. JCEM 88, 5747–5752.

Rathmann, W., Kowall, B., Heier, M., Herder, C., Holle, R., Thorand, B., Strassburger, K., Peters, A., Wichmann, H.-E., Giani, G., Meisinger, C., 2010. Prediction models for incident Type 2 diabetes mellitus in the older population: KORA S4/F4 cohort study. Diabetic Medicine 27, 1116–1123.

Rathmann, W., Strassburger, K., Heier, M., Holle, R., Thorand, B., Giani, G., Meisinger, C., 2009. 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. 26, 1212–1219.

Reuter, S.E., Evans, A.M., Chace, D.H., Fornasini, G., 2008. Determination of the reference range of endogenous plasma carnitines in healthy adults. Ann Clin Biochem 45, 585–592.

Rhee, E.P., Cheng, S., Larson, M.G., Walford, G.A., Lewis, G.D., McCabe, E., Yang, E., Farrell, L., Fox, C.S., O'Donnell, C.J., Carr, S.A., Vasan, R.S., Florez, J.C., Clish, C.B., Wang, T.J., Gerszten, R.E., 2011. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. J Clin Invest 121, 1402–1411.

Rochfort, S., 2005. Metabolomics Reviewed: A New "Omics" Platform Technology for Systems Biology and Implications for Natural Products Research. Journal of Natural Products 68, 1813–1820.

Rodrigues, C.M., Kren, B.T., Steer, C.J., Setchell, K.D., 1996. Formation of delta 22-bile acids in rats is not gender specific and occurs in the peroxisome. J. Lipid Res. 37, 540–550.

Römisch-Margl, W., Prehn, C., Bogumil, R., Röhring, C., Suhre, K., Adamski, J., 2011. Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. Metabolomics.

Rubio-Aliaga, I., De Roos, B., Duthie, S., Crosley, L., Mayer, C., Horgan, G., Colquhoun, I., Le Gall, G., Huber, F., Kremer, W., Rychlik, M., Wopereis, S., Van Ommen, B., Schmidt, G., Heim, C., Bouwman, F., Mariman, E., Mulholland, F., Johnson, I., Polley, A., Elliott, R., Daniel, H., 2011. Metabolomics of prolonged fasting in humans reveals new catabolic markers. Metabolomics 7, 375–387.

Sacks, D.B., Bruns, D.E., Goldstein, D.E., Maclaren, N.K., McDonald, J.M., Parrott, M., 2002. Guidelines and Recommendations for Laboratory Analysis in the Diagnosis and Management of Diabetes Mellitus. Clin Chem 48, 436–472.

Schmerler, D., Neugebauer, S., Ludewig, K., Bremer-Streck, S., Brunkhorst, F.M., Kiehntopf, M., 2012. Targeted metabolomics for discrimination of systemic inflammatory disorders in critically ill patients. J. Lipid Res. 53, 1369–1375.

Schnabel, R.B., Baumert, J., Barbalic, M., Dupuis, J., Ellinor, P.T., Durda, P., Dehghan, A., Bis, J.C., Illig, T., Morrison, A.C., Jenny, N.S., Keaney, J.F., Gieger, C., Tilley, C., Yamamoto, J.F., Khuseyinova, N., Heiss, G., Doyle, M., Blankenberg, S., Herder, C., Walston, J.D., Zhu, Y., Vasan, R.S., Klopp, N., Boerwinkle, E., Larson, M.G., Psaty, B.M., Peters, A., Ballantyne, C.M., Witteman, J.C.M., Hoogeveen, R.C., Benjamin, E.J., Koenig, W., Tracy, R.P., 2009. Duffy antigen receptor for chemokines (Darc) polymorphism regulates circulating concentrations of monocyte chemoattractant protein-1 and other inflammatory mediators. Blood blood–2009–05–221382.

Shaham, O., Wei, R., Wang, T.J., Ricciardi, C., Lewis, G.D., Vasan, R.S., Carr, S.A., Thadhani, R., Gerszten, R.E., Mootha, V.K., 2008. Metabolic profiling of the human response to a glucose challenge reveals distinct axes of insulin sensitivity. Molecular Systems Biology 4.

Shevchenko, A., Simons, K., 2010. Lipidomics: coming to grips with lipid diversity. Nature Reviews Molecular Cell Biology 11, 593–598.

Slupsky, C.M., Rankin, K.N., Wagner, J., Fu, H., Chang, D., Weljie, A.M., Saude, E.J., Lix, B., Adamko, D.J., Shah, S., Greiner, R., Sykes, B.D., Marrie, T.J., 2007. Investigations of the Effects of Gender, Diurnal Variation, and Age in Human Urinary Metabolomic Profiles. Anal. Chem. 79, 6995–7004.

Spratlin, J.L., Serkova, N.J., Eckhardt, S.G., 2009. Clinical Applications of Metabolomics in Oncology: A Review. Clin Cancer Res 15, 431–440.

Stumvoll, M., Goldstein, B.J., Van Haeften, T.W., 2005. Type 2 diabetes: principles of pathogenesis and therapy. The Lancet 365, 1333–1346.

Suhre, K., Shin, S.-Y., Petersen, A.-K., Mohney, R.P., Meredith, D., Wägele, B., Altmaier, E., CARDIoGRAM, Deloukas, P., Erdmann, J., Grundberg, E., Hammond, C.J., Angelis, M.H. de, Kastenmüller, G., Köttgen, A., Kronenberg, F., Mangino, M., Meisinger, C., Meitinger, T., Mewes, H.-W., Milburn, M.V., Prehn, C., Raffler, J., Ried, J.S., Römisch-Margl, W., Samani, N.J., Small, K.S., Wichmann, H.-E., Zhai, G., Illig, T., Spector, T.D., Adamski, J., Soranzo, N., Gieger, C., 2011. Human metabolic individuality in biomedical and pharmaceutical research. Nature 477, 54–60.

Susser, M., 1973. Causal thinking in the health sciences: concepts and strategies of epidemiology. Oxford University Press.

Szendroedi, J., Schmid, A.I., Chmelik, M., Toth, C., Brehm, A., Krssak, M., Nowotny, P., Wolzt, M., Waldhausl, W., Roden, M., 2007. Muscle Mitochondrial ATP Synthesis and Glucose Transport/Phosphorylation in Type 2 Diabetes. PLoS Med 4, e154.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., Jensen, L.J., Von Mering, C., 2011. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Research 39, D561–568.

Tabák, A.G., Herder, C., Rathmann, W., Brunner, E.J., Kivimäki, M., 16. Prediabetes: a high-risk state for diabetes development. The Lancet 379, 2279–2290.

Tabák, A.G., Jokela, M., Akbaraly, T.N., Brunner, E.J., Kivimäki, M., Witte, D.R., 2012. Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. The Lancet 373, 2215–2221.

Tammen, H., Schulte, I., Hess, R., Menzel, C., Kellmann, M., Mohring, T., Schulz-Knappe, P., 2005. Peptidomic analysis of human blood specimens: Comparison between plasma specimens and serum by differential peptide display. PROTEOMICS 5, 3414–3422.

Teahan, O., Gamble, S., Holmes, E., Waxman, J., Nicholson, J.K., Bevan, C., Keun, H.C., 2006. Impact of Analytical Bias in Metabonomic Studies of Human Blood Serum and Plasma. Analytical Chemistry 78, 4307–4318.

Teerlink, T., Nijveldt, R.J., De Jong, S., Van Leeuwen, P.A.M., 2002. Determination of Arginine, Asymmetric Dimethylarginine, and Symmetric Dimethylarginine in Human Plasma and Other Biological Samples by High-Performance Liquid Chromatography. Analytical Biochemistry 303, 131–137.

Triola, M.F., Goodman, W.M., LaBute, G., Law, R., MacKay, L., 2006. Elementary statistics. Pearson/Addison-Wesley.

Tuomilehto, J., Lindström, J., Eriksson, J.G., Valle, T.T., Hämäläinen, H., Ilanne-Parikka, P., Keinänen-Kiukaanniemi, S., Laakso, M., Louheranta, A., Rastas, M., Salminen, V., Aunola, S., Cepaitis, Z., Moltchanov, V., Hakumäki, M., Mannelin, M., Martikkala, V., Sundvall, J., Uusitupa, M., 2001. Prevention of Type 2 Diabetes Mellitus by Changes in Lifestyle among Subjects with Impaired Glucose Tolerance. New England Journal of Medicine 344, 1343–1350.

Vinayavekhin, N., Homan, E.A., Saghatelian, A., 2010. Exploring Disease through Metabolomics. ACS Chemical Biology 5, 91–103.

Vogler, E.A., Siedlecki, C.A., 2009. Contact activation of blood-plasma coagulation. Biomaterials 30, 1857–1869.

Wang, T.J., Larson, M.G., Vasan, R.S., Cheng, S., Rhee, E.P., McCabe, E., Lewis, G.D., Fox, C.S., Jacques, P.F., Fernandez, C., O'Donnell, C.J., Carr, S.A., Mootha, V.K., Florez, J.C., Souza, A., Melander, O., Clish, C.B., Gerszten, R.E., 2011. Metabolite profiles and the risk of developing diabetes. Nat Med advance online publication.

Wang-Sattler, R., Yu, Y., Mittelstrass, K., Lattka, E., Altmaier, E., Gieger, C., Ladwig, K.H., Dahmen, N., Weinberger, K.M., Hao, P., Liu, L., Li, Y., Wichmann, H.-E., Adamski, J., Suhre, K., Illig, T., 2008. Metabolic Profiling Reveals Distinct Variations Linked to Nicotine Consumption in Humans — First Results from the KORA Study. PLoS ONE 3.

Wang-Sattler, R., Yu, Z., Herder, C., Messias, A.C., Floegel, A., He, Y., Heim, K., Campillos, M., Holzapfel, C., Thorand, B., Grallert, H., Xu, T., Bader, E., Huth, C., Mittelstrass, K., Döring, 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, H.-G., Angelis, M.H. de, Rathmann, W., Suhre, K., Prokisch, H., Peters, A., Meitinger, T., Roden, M., Wichmann, H.-E., Pischon, T., Adamski, J., Illig, T., 2012. Novel biomarkers for pre-diabetes identified by metabolomics. Molecular Systems Biology 8.

Weinberger, K.M., 2008. Einsatz von Metabolomics zur Diagnose von Stoffwechselkrankheiten. Therapeutische Umschau 65, 0487–0491.

Weinberger, K.M., Graber, A., 2005. Using comprehensive metabolomics to identify novel biomarkers. Screening Trends in Drug Discovery 6, 42–45.

Wenk, M.R., 2005. The emerging field of lipidomics. Nature Reviews Drug Discovery 4, 594–610.

Wishart, D.S., Knox, C., Guo, A.C., Eisner, R., Young, N., Gautam, B., Hau, D.D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L., Cruz, J.A., Lim, E., Sobsey, C.A., 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, Y., Clive, D., Greiner, R., Nazyrova, A., Shaykhutdinov, R., Li, L., Vogel, H.J., Forsythe, I., 2009. HMDB: a knowledgebase for the human metabolome. Nucleic Acids Research 37, D603–610.

Wizemann, T.M., Pardue, M.L., 2001. Exploring the biological contributions to human health: does sex matter? National Academies Press.

Wopereis, S., Rubingh, C.M., Van Erk, M.J., Verheij, E.R., Van Vliet, T., Cnubben, N.H.P., Smilde, A.K., Van der Greef, J., Van Ommen, B., Hendriks, H.F.J., 2009. Metabolic Profiling of the Response to an Oral Glucose Tolerance Test Detects Subtle Metabolic Changes. PLoS ONE 4.

Yatomi, Y., Igarashi, Y., Yang, L., Hisano, N., Qi, R., Asazuma, N., Satoh, K., Ozaki, Y., Kume, S., 1997. Sphingosine 1-Phosphate, a Bioactive Sphingolipid Abundantly Stored in Platelets, Is a Normal Constituent of Human Plasma and Serum. J Biochem 121, 969–973.

Yeboah, J., McNamara, C., Jiang, X.-C., Tabas, I., Herrington, D.M., Burke, G.L., Shea, S., 2010. Association of Plasma Sphingomyelin Levels and Incident Coronary Heart Disease Events in an Adult Population Multi-Ethnic Study of Atherosclerosis. Arterioscler Thromb Vasc Biol 30, 628–633.

Yu, Z., Kastenmüller, G., He, Y., Belcredi, P., Möller, G., Prehn, C., Mendes, J., Wahl, S., Roemisch-Margl, W., Ceglarek, U., others, 2011. Differences between Human Plasma and Serum Metabolite Profiles. PLoS ONE 6, e21230.

Yu, Z., Zhai, G., Singmann, P., He, Y., Xu, T., Prehn, C., Römisch-Margl, W., Lattka, E., Gieger, C., Soranzo, N., Heinrich, J., Standl, M., Thiering, E., Mittelstraß, K., Wichmann, H.-E., Peters, A., Suhre, K., Li, Y., Adamski, J., Spector, T.D., Illig, T., Wang-Sattler, R., 2012. Human serum metabolic profiles are age dependent. Aging Cell 11, 960–967.

Zhao, X., Fritsche, J., Wang, J., Chen, J., Rittig, K., Schmitt-Kopplin, P., Fritsche, A., Häring, H.-U., Schleicher, E.D., Xu, G., Lehmann, R., 2010. Metabonomic fingerprints of fasting plasma and spot urine reveal human pre-diabetic metabolic traits. Metabolomics 6, 362–374.

Appendix

A1. List of publications

16. Siegert S, Yu Z, Wang-Sattler R, Illig T, Adamski J. Sex Dependency of Human Metabolic Profiles Revisited. Metabolomics 2(115), 2153-0769

15. Wahl S, Yu Z, Kleber M, Singmann P, Holzapfel C, He Y, Mittelstrass K, Polonikov A, Prehn C, Römisch-Margl W, Adamski J, Suhre K, Grallert H, Illig T, Wang-Sattler R*, Reinehr T. Childhood obesity is associated with changes in the serum metabolite profile. Obes Facts, 2012;5:660-670

14. Floegel A, Stefan N, Yu Z, Mühlenbruch K, Drogan D, Joost HG, Fritsche A, Häring 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. 2013; 62 (2), 639-648

13. 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, …, 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 Sep 25;8:615. doi: 10.1038/msb.2012.43.

12. He Y, Yu Z, Giegling I, Xie L, Hartmann AM, Prehn C, Adamski J, Kahn R , Li Y, Illig T, Wang-Sattler R*, Rujescu D. Schizophrenia shows a unique metabolomics signature in plasma. Transl Psychiatry. 2012 Aug 14;2:e149. doi: 10.1038/tp.2012.76.

11. Yu Z#, Zhai G#, Singmann P#, He Y, Xu T, Prehn C, Römisch-Margl W, Lattka E, Gieger C, Soranzo N, Heinrich J, Standl M, Thiering E, Mittelstraß K, Wichmann HE, Peters P, Suhre K, Li Y , Adamski J, Spector TD, Illig T, Wang-Sattler R*, Human serum metabolic profiles are age dependent. Aging cell, 2012 Jul 26. doi: 10.1111/j.1474-9726.2012.00865.x.

10. He Y#, Yu Z#, Ge D, Wang-Sattler R, Thiesen HJ, Xie L, Li Y. Cell type specificity of signaling: view from membrane receptors distribution and their downstream transduction networks. Protein & Cell, 2012 Sep;3(9):701-13. Epub 2012 Jul 16.

9. He Y, Zhang M, Ju Y, Yu Z, Lv D, Sun H, Yuan W, He F, Zhang J, Li H, Li J, Wang-Sattler R, Li Y, Zhang G, Xie L. dbDEPC 2.0: updated database of differentially expressed proteins in human cancers. Nucleic Acids Res. 2012 Jan;40(Database issue):D964-71. Epub 2011 Nov 16.

8. Kus V, Flachs P, Kuda O, …, Wang-Sattler R, Yu Z, Illig T, Kopecky J. Unmasking differential effects of rosiglitazone and pioglitazone in the combination treatment with n-3 fatty acids in mice fed a high-fat diet. PLoS One. 2011;6(11):e27126. Epub 2011 Nov 3.

7. Yu Z#, Mittelstrass K#, Ried JS#, Krumsiek J, Gieger C, Prehn C, Roemisch-Margl W, Polonikov A, Peters A, Theis FJ, Meitinger T, Kronenberg F, Weidinger S, Wichmann HE, Suhre K, Wang-Sattler R, Adamski J, Illig T. Discovery of sexual dimorphisms in metabolic and genetic biomarkers. PloS Genet, 2011 Aug;7(8):e1002215. Epub 2011 Aug 11.

6. Yu Z, Kastenmüller G, He Y, Belcredi P, Möller G, Prehn C, Mendes J, Wahl S, Roemisch-Margl W, Ceglarek U, Polonikov A, Dahmen N, Prokisch H, Xie L, Li Y, Wichmann HE, Peters A, Kronenberg F, Suhre K, Adamski J, Illig T, Wang-Sattler R*. Differences between human plasma and serum metabolite profiles. PLoS One: 2011;6(7):e21230. Epub 2011 Jul 8.

5. Hao P, …, Yu Z, …, Zhao G. Complete Sequencing and Pan-Genomic Analysis of Lactobacillus delbrueckii subsp. bulgaricus Reveal Its Genetic Basis for Industrial Yogurt Production. PLoS One: 2011; e15964.

4. Wang Z, Ding G, Yu Z, Liu L, Li Y. Modeling the age distribution of gene duplications in vertebrate genome using mixture density. Genomics. 2009;93(2):146-151.

3. Wang Z, Ding G, Yu Z, Liu L, Li Y.(2009) CHSMiner: a GUI tool to identify chromosomal homologous segments. Algorithms Mol Biol.; 4:2-2.

2. Yu Z#, Ding G#, Zhao J, Wang Z, Li Y, Tree of Life Based on Genome Context Networks. PLoS One, 2008; e3357

1. Zhao J, …, Yu Z, …, Li Y. Modular co-evolution of metabolic networks. BMC Bioinformatics. 2007;8(1):311.

# equal contribution

* Corresponding author