Aus dem Institut für Epidemiologie

Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH)

Direktorin: Prof. Dr. Annette Peters

Metabolomics to identify biomarkers and influencing factors of Type 2 Diabetes

Metabolomics zur Identifizierung von Biomarkern und Einflussfaktoren auf Diabetes Typ 2

Kumulative Habilitationsleistung

Zur Erlangung der Venia Legendi
der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München
für das Fach Molekulare Epidemiologie

vorgelegt von
Dr. sc. hum. Rui Wang-Sattler
2021

Table of contents

1 Introduction

The present cumulative thesis provides a summary of my research activities in metabolomics studies of type 2 diabetes (T2D). The focus of my metabolomics research has been to identify novel metabolite biomarkers of (pre-) diabetes (Wang-Sattler *et al,* 2012) and factors that influence T2D such as age (Yu *et al*, 2012; Chat *et al*, 2019), smoking (Wang-Sattler *et al*, 2008; Xu *et al,* 2013) and alcohol drinking (Jaremek *et al,* 2013) as well as to characterize the effects of metformin intake (Xu *et al*, 2015; Brandmaier *et al,* 2015; Adam *et al,* 2016; Adam *et al,* 2017).

1.1 Type 2 diabetes

T2D is a lifelong, incapacitating disease affecting multiple organs (WHO, 2016). The International Diabetes Federation (IDF) estimates that 463 million adults (20-79 years) are living with diabetes worldwide (IDF Diabetes Atlas 2019). The prevalence of diabetes is rapidly increasing and it is projected that 693 million people will suffer from diabetes by 2045 (Roden *et al*, 2019). T2D is defined as increased blood glucose level, i.e. hyperglycaemia, due to pancreatic β - cell dysfunction, impaired insulin secretion and insulin resistance (Stumvoll *et al*, 2005; Buse *et al*, 2009).

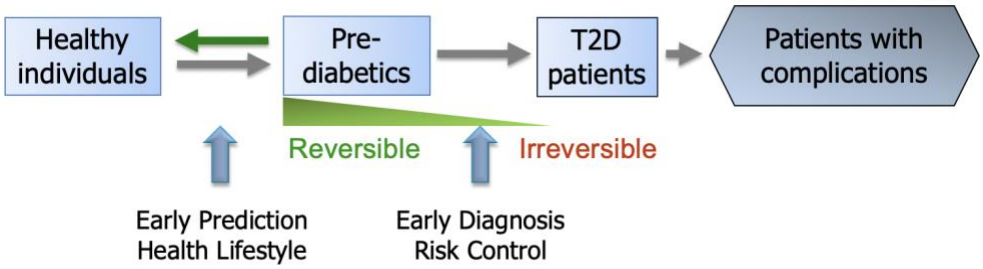

Figure 1 A simplified development of diabetes from healthy state toward complications of T2D

A state of pre-diabetes (i.e., impaired fasting glucose and/or impaired glucose tolerance (IGT)) with only slightly elevated blood glucose levels may precede T2D for years (Figure 1). The development of T2D in pre-diabetic individuals can be delayed or prevented by dietary changes and increased physical activity (Knowler *et al,* 2002; Tuomilehto *et al,* 2001). Pre-diabetic individuals can be divided in two subgroups: those that progress to developing diabetes or those that return to normoglycaemia. Both subgroups have a similar annualized conversion rate of 5%–10% (Tabak *et al,* 2012). Therefore, the early detection of pre-diabetes is thus of crucial importance for the development of personalized strategies to prevent T2D.

Both T2D and pre-diabetes are associated with devastating chronic complications including macrovascular disease (e.g. cardiovascular disease, which is

the major cause of death in western countries) and microvascular disorders leading to damage of the small blood vessels of the kidney (nephropathy), eye (retinopathy) and peripheral nerves (neuropathy). These complications impose an immense burden on the quality of life of the patients (ADA, 2020). Novel means to prevent these devastating diabetic complications are urgently needed. Therefore, the aims of my research focused on the identification of candidate biomarkers of pre-diabetes and T2D using metabolomics approach.

1.2 Metabolomics

Metabolomics is still a relatively new approach for studying metabolic changes connected to disease development and progression, as well as for finding predictive biomarkers to enable early interventions (Nicholson *et al,* 2003; Gieger *et al,* 2008; Illig *et al,* 2010). The central dogma of molecular biology states that DNA makes RNA and RNA makes protein (Crick 1970). The metabolome is downstream of the proteome and is dependent of the genetic background, physiological conditions, lifestyle and environmental factors (Figure 2). While the genome informs about 'what can happen', the metabolome reflects 'what has happened and is happening'.

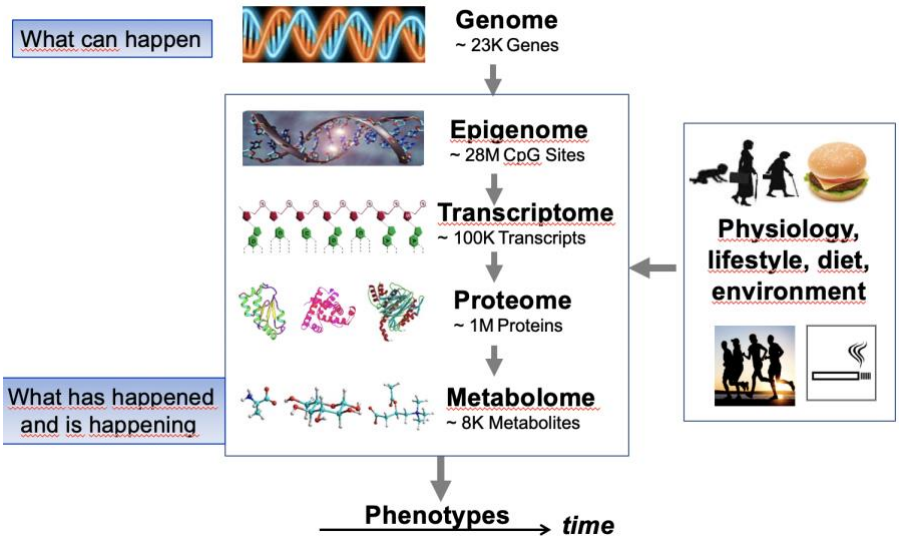

Figure 2 The metabolome is a major determinant of the visible phenotypes

The metabolome is comprised of all small molecules that can found within a biological sample such as blood. These molecules include endogenous metabolites that are the intermediary products of metabolism associated with energy storage and utilization (e.g. building blocks of proteins, carbohydrates and lipids, regulators of gene expression, signaling molecules) as well as exogenous metabolites (e.g. drugs, food additives, toxins) (Suhre *et al*, 2011). The metabolome, as the entirety of metabolites,

is a real-time functional portrait of the cell's or the organism's states and interaction with environment (Newgard *et al,* 2009; Walker *et al*, 2019).

Linking metabolic profiles to a given phenotypic outcome is therefore an efficient and highly promising tool for capturing the complexity of the metabolic process and identifying healthy and disease-linked states (He *et al*, 2012; Wischart *et al,* 2019). As metabolic changes occur already before the appearance of a clinical phenotype, they are very useful in identifying the early indicators and assessing the pre-clinical stages of disease. As detection of pre-clinical phenotypes is the key for preventive therapy, metabolomics studies are of enormous value in current medical research with a great impact on clinical applications.

2 Stringent quality control of metabolite profiles

The power of metabolomics lies in its ability to detect and measure numerous small molecules in a single approach. However, as the complicated analytical conditions are not optimized for each quantified metabolite with the current technology, stringent quality control of metabolite profiles is necessary for a reliable identification of candidate biomarkers.

2.1 Quantification of serum samples

During my metabolomics research, we have primarily analyzed metabolite profiles of blood samples of the population-based cohort, KORA (Kooperative Gesundheitsforschung in der Region Augsburg) (Wichmann *et al*, 2005). The metabolomics measurements have been performed in the Metabolomics Platform of the Genome analysis center, Helmholtz Zentrum München (Römisch-Margl *et al,* 2011; Zukunft *et al*, 2018).

Table 1 Measurement of serum samples of KORA S4 and F4 participants using targeted approaches

	KORA S4	**KORA F4**		
No. of samples	1608	1068	952	1060
Time of measurements	Mar. - Apr.2011	Aug.-Sep.2008	Nov.-Dec.2008	Mar. 2009
Batch no. of measurement	-	1	2	3
Used Absolute*IDQ*™ kits	p180	p150		
Targeted (used) no. of metabolites	188 (140)	163 (131)		

The serum samples from participants obtained at baseline (KORA S4) and follow-up (KORA F4) were measured with targeted approaches (Mittelstrass *et al*, 2011; Wang-Sattler *et al*, 2012). In KORA, 188 and 163 metabolites were

simultaneously measured with the Biocrates AbsoluteIDQ™ p180 Kit (KORA S4) and p150 Kit (F4) (Table 1).

2.2 Quality control of metabolite profiles in KORA S4 and F4

Stringent quality control procedures are briefly outlined below and details can be found in our publications (Wang-Sattler *et al*, 2012; Yu *et al*, 2012). Since the measurements of KORA S4 and F4 samples were conducted with two different kits at different time points corresponding to different batches, we had to apply separate quality control procedures.

For KORA S4, at each kit plate, five reference (human plasma pooled material, Seralab) and three zero (PBS) samples were measured in addition to the serum samples. To ensure data quality, each metabolite had to meet two criteria: 1) its coefficient of variance in the reference samples had to be smaller than 25%; 2) 50% of all measured sample concentrations for the metabolite had to be above the limit of detection, defined as 3 × median of the zero samples. In total, 140 metabolites passed the quality control (Wang-Sattler *et al*, 2012).

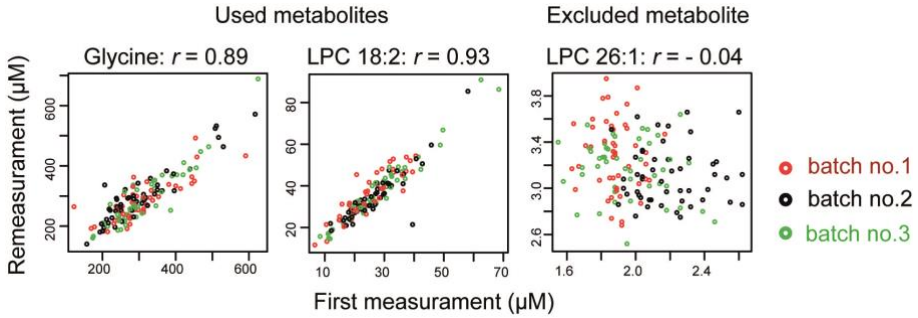

Figure 3 Metabolite concentrations of two duplicated measurement for 144 KORA F4 samples

For KORA F4, as we did not quantify the reference samples for each kit plate, we have re-measured 144 samples randomly selected from 3 batches (Figure 3). Each metabolite had to meet three criteria: 1) the average value of its coefficient of variance in the 3 quality controls (provided by the manufacturer on each kit plate) had to be smaller than 25%; 2) 90% of all measured sample concentrations for the metabolite had to be above the limit of detection; and 3) the correlation coefficient between two duplicate measurements of the metabolite in 144 re-measured samples had to be above 0.5. In total, 131 metabolites passed the three quality controls (Yu *et al*, 2012; Figure 3 and Table 1).

A total of 123 metabolites was measured in both KORA S4 and F4 studies, including sum of hexoses, 20 acylcarnitines, 14 amino acids, 13 sphinomyelins, 33 diacyl phosphatidycholines, 34 acyl-alkyl phosphatidycholines and 8 lysophosphatidylcholines (LPCs) (Xu *et al,* 2013).

3 Metabolomics characterization of factors influencing T2D

T2D is a chronic and complex disease that is caused not only by genetics but also environmental and physiological factors including nutrition (Knowler *et al*, 2002; Roden *et al,* 2019). Using the targeted and non-targeted metabolomics data from the KORA studies, we have investigated various factors that influence development of T2D including age, smoking, alcohol consumption and usage of metformin.

3.1 Human serum metabolite profiles are age dependent

Increased age is one of the largest risk factors for many chronic diseases. The American Diabetes Association (ADA) recommended that the testing for pre-diabetes and / or T2D should begin at the age of 45 years (ADA 2020). For diabetes, it has shown that the two most frequent T2D subgroups are those related with moderate age and moderate obesity (average prevalence of 39.6% and 21.0%, respectively) (Ahlqvist *et al*, 2018).

We have identified chronological age-related metabolites that were independent of BMI using the KORA F4 study which has a wide age range from 32 to 81 years (Yu *et al,* 2012, Figure 4).

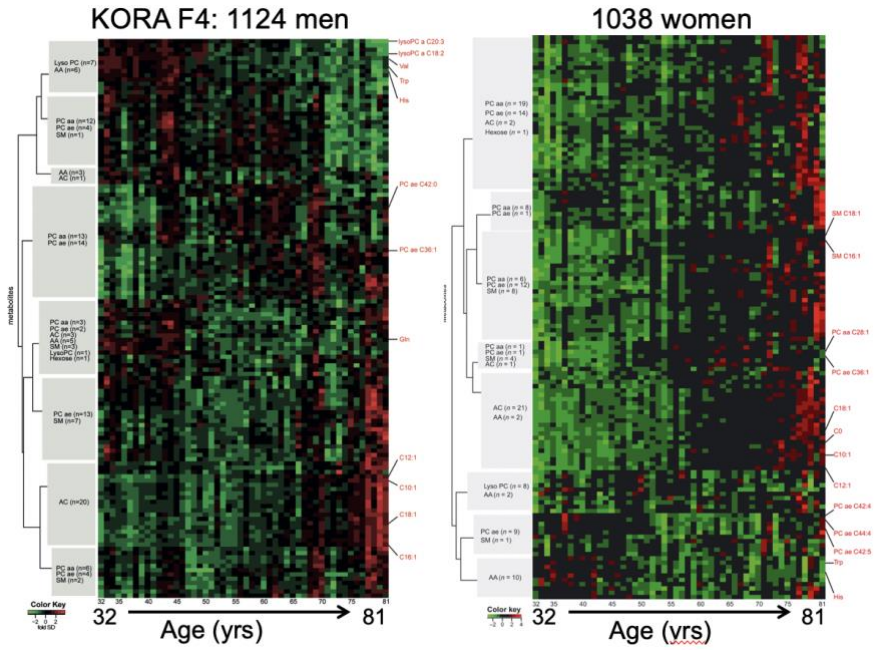

Figure 4 Heat maps of mean residuals for each year of age in men and women, respectively.

To investigate the “true” age-related metabolites, we have specifically excluded participants with major metabolic diseases (e.g. hypertension, T2D, obesity), adjusted for BMI and conducted sex-separated analyses. Our results indicate the age dependency of metabolic profiles that could reflect different aging processes due to

oxidative stress, alterations in cell morphology, beta oxidation capacity, and vascular function.

This study was however limited by: 1) its cross-sectional study design; 2) metabolite profiles that were related to chronological age but not to biological / physiological age; 3) limited replication conducted only in women from the TwinsUK study (Yu *et al,* 2012).

Recently, we had the possibility to analyze the longitudinal data of healthy individuals from two independent human cohorts (Chat *et al*, 2019). We identified aging-associated metabolites that were independent of chronological age and deepened our understanding of the long-term metabolite changes during aging. In this study, we attempted to overcome the limitations of our previous age study and conducted sex-stratified longitudinal analyses with two-time-point metabolomics data of the same person. We also adjusted for baseline chronological age besides other confounders using generalized estimation equation models. Moreover, we replicated our discoveries in both women and men of the CARLA-Study (Cardiovascular Disease, Living and Ageing in Halle) (Chat *et al*, 2019). Finally, we applied stringent exclusion criteria and used only 'healthy participants' in both discovery and replication studies (Figure 5).

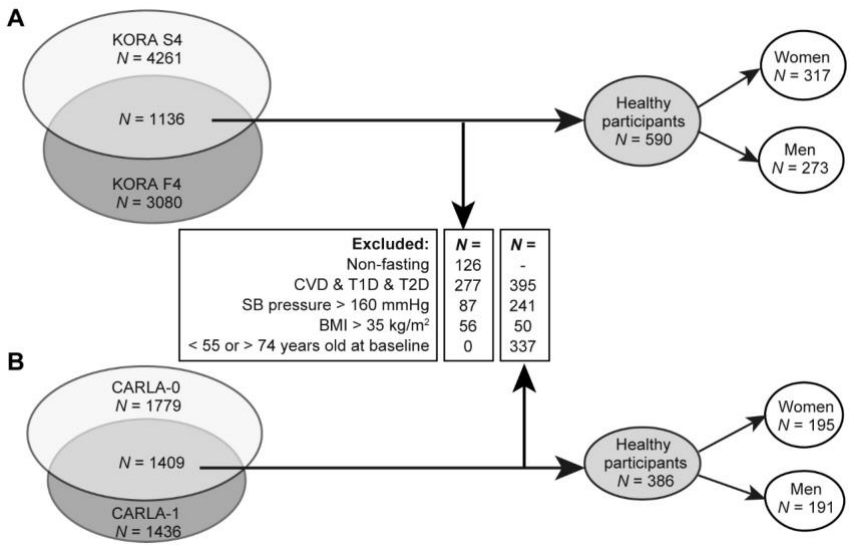

Figure 5 Flow diagram showing the inclusion and exclusion procedures of the study population in KORA (A) and in CARLA (B).

In addition to aging, lifestyle plays a prominent role in the development of T2D. We therefore characterized the influence of smoking and alcohol consumption on the metabolite concentration.

3.2 Smoking cessation reversed smoking-related changes in metabolites

Tobacco use increases the risk of death from many chronic diseases (e.g. ischemic heart disease, cancer and T2D). In 2016, cigarette smoking caused over 7.1 million deaths (5.1 million in men, 2.0 million in women) (The Tobacco Atlas, 2018). Based on the targeted metabolomics approaches, we identified smoking-associated metabolites in a pilot study with KORA F3 (Wang-Sattler *et al*, 2008) and a longitudinal investigation using two-time-point data of KORA S4 to F4 (Xu *et al*, 2013).

In the KORA F3 study, we analyzed 198 metabolites in 283 serum samples of male participants (aged 55 - 79 years). We found 23 smoking-associated lipid metabolites (e.g. diacyl-phosphatidylcholines, PC aa C34:1). Except for 3 acyl-alkyl-phosphatidylcholines (e.g. PC ae C40:6), their levels were higher in smokers compared to former and non-smokers (see box plots in Figure 6).

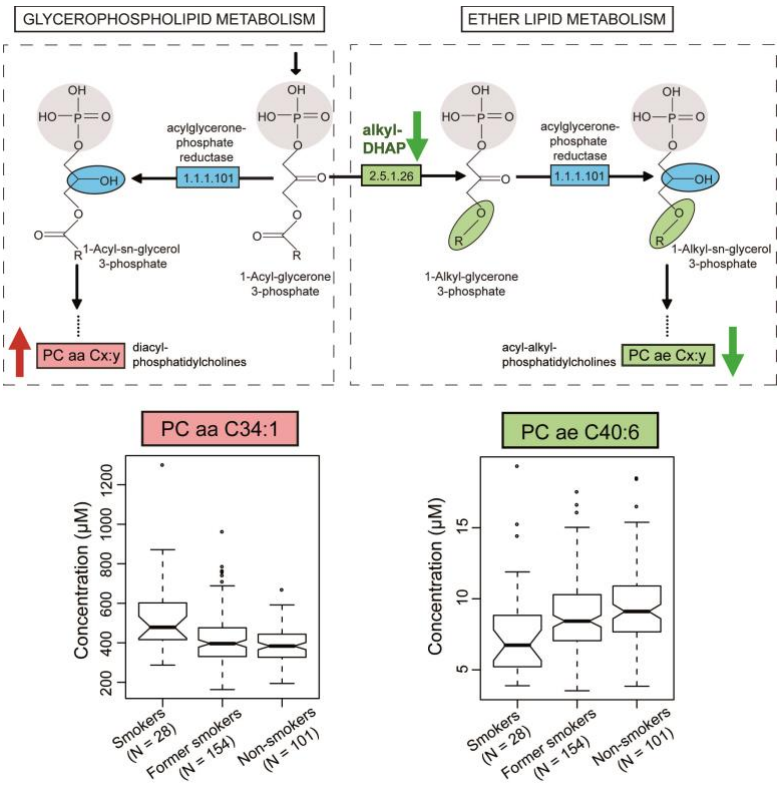

Figure 6 Smoking associated pathways of the glycerophospholipid- and ether lipid metabolism.

We further found consistently reduced ratios of PC ae Cx:y to PC aa Cx:y in smokers. The synthesis of these phospholipids is regulated by the enzyme alkylglycerone phosphate (alkyl-DHAP) synthase (AGPS) involved in the ether lipid and glycerophospholipid metabolism (Figure 6). Notably, our results were consistent with the strong down-regulation of AGPS expression in smokers in the study that analyzed gene expression in human lung tissues (Gruber *et al*, 2006). Our data

suggested that smoking is associated with plasmalogen-deficiency disorders, caused by reduced or abolished activity of the peroxisomal AGPS. Our findings provide new insight into the pathophysiology of smoking addiction. Activation of the enzyme alkyl-DHAP synthase by small molecules could provide novel routes for smoking therapy. The pilot study was limited by its cross-sectional study design with a relatively small number of smokers (28 men) and metabolite quantification performed solely in men (Wang-Sattler *et al*, 2008).

We followed this up by an extended analysis in KORA S4 with a larger sample of smokers (125 men, 70 women) and confirmed a part of the findings from the pilot study and suggested additional metabolites (Xu *et al,* 2013). Among the 23 smoking-related metabolites identified in the pilot study, 11 metabolites were measured using the Absolute*IDQ*™ p180 kit in KORA S4, of which 5 were validated in men. Additionally, 6 metabolites were identified in female smokers and 3 metabolites were found both in female and male smokers. Furthermore, 4 metabolites in smokers (3 acyl-alkyl-PCs in men, one in women) showed a significant association with pack years.

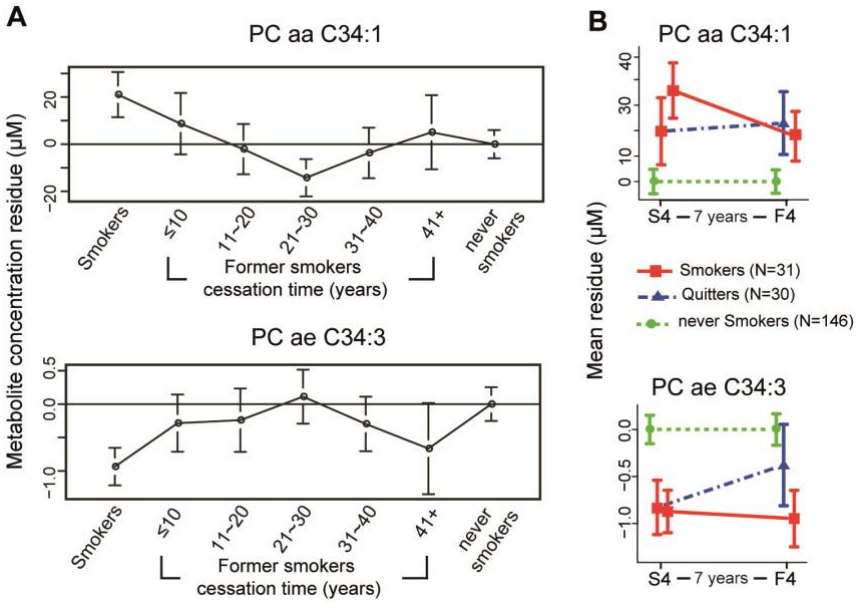

Figure 7 Residues of 2 metabolites concentration in relation to smoking cessation time (A) and changes of smokers, former smokers and never smokers from baseline S4 to 7-year follow-up F4 (B).

More specifically, our longitudinal analysis using two-time-point data (31 smokers, 146 never smokers and 30 quitters, people who were smokers in baseline KORA S4 but former smokers in seven-years follow up F4) showed that smoking-associated metabolite changes are reversible after smoking cessation (Xu *et al,* 2013, see Figure 7).

3.3 Alcohol-induced metabolomic differences

Harmful use of alcohol causes 3 million deaths every year. Alcohol consumption is the world's third largest risk factor for disease burden and is a causal factor for more than 200 diseases including T2D (WHO 2018). In the KORA F4 study, daily alcohol consumption was estimated by a detailed questionnaire filled by each study participant according to their past week alcohol use, separated to weekdays and weekend days. We have categorized the study participants according to their daily alcohol intake into two categories, light drinkers (LD, women < 20g/day and men < 40g/day) and moderate-to-heavy drinkers (MHD, women ≥ 20g/day and men ≥ 40g/day). We have also included alcohol abstainers (alcohol intake of 0 g/day) and separated them into those with or without lipid-lowering medications (Jaremek *et al,* 2013, see plots A & B in Figure 8).

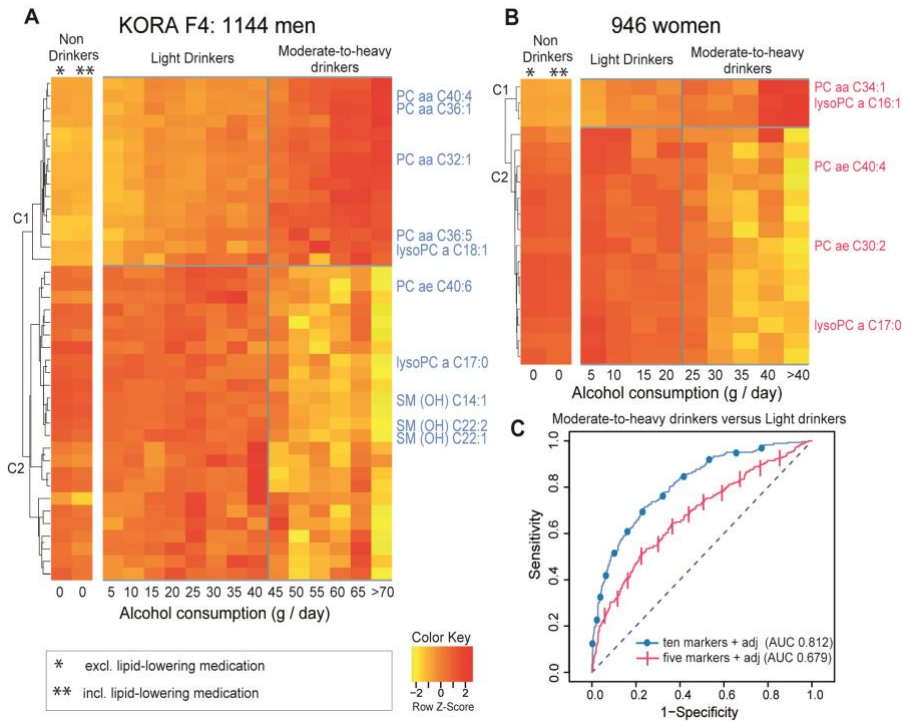

Figure 8 Heatmap of alcohol consumption in men (A) and women (B) and AUC values (C)

In addition to the regression analysis (adjusted for age, BMI, smoking, HDL and triglyceride), we applied machine learning (e.g. random forest) and selected 10 / 5 independent metabolites in men / women that significantly differed in concentration between MHD and LD (see plots A & B in Figure 8). For these metabolites, the respective area under the Receiver Operating Characteristic curves (AUC) were 81.2% / 67.9% in men / women, respectively (see plot C in Figure 8). These AUC values provide a good to moderate sensitivity and specificity for the discrimination of MHD to LD (Jaremek *et al,* 2013).

3.4 Effects of metformin usage

Metformin is used as a first-line oral medication to treat patients with pre-diabetes and T2D for more than 50 years. However, the underlying mechanism is not fully understood. Here, we used both targeted and non-targeted metabolomics approaches to investigate the effects of metformin usage (Xu *et al*, 2015; Brandmaier *et al,* 2015; Adam *et al,* 2016; Adam *et al,* 2017).

3.4.1 Metformin lowers the levels of 3 metabolites and LDL cholesterol

To study the effect of metformin intake on the metabolite profiles of KORA S4 and F4 participants, we first excluded the T2D patients treated with both metformin and insulin. We found that metformin treatment was associated with significantly decreased levels of 3 acyl-alkyl PCs in T2D patients and replicated the findings in two independent Dutch cohorts (the Erasmus Rucphen Family study, ERF, and the Netherlands Twin Register, NTR) (Xu *et al*, 2015). The longitudinal examination of T2D patients that started metformin treatment during the follow-up of 7 years further confirmed that the decrease in these 3 metabolites was related to metformin therapy.

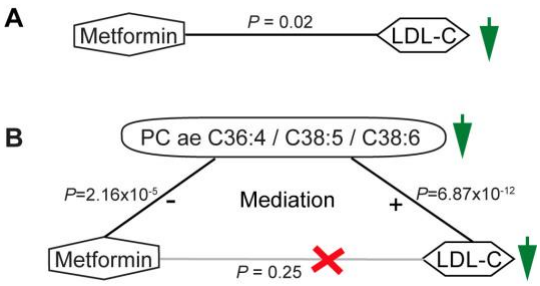

Figure 9 Mediation analyses. A: The association between metformin and LDL-C without consideration of the 3 metformin-associated metabolites. B: The results of the mediation analysis; the red cross indicates that the direct association between metformin and LDL-C is not significant anymore.

By mediation analysis using the prospective data of 912 KORA participants, we found that the reduction in LDL-C levels in metformin-treated T2D patients was partially mediated by the three acyl-alkyl PCs (Figure 9).

To identify the genes associated with the three metabolites, we applied phenotype set enrichment analysis in a subset of KORA F4 individuals (n= 1,809) with available genotyping and metabolite profiles. Blood concentrations of these three acyl-alkyl PCs were significantly associated with several single nucleotide polymorphisms (SNPs) that were enriched in the transcribed or flanking regions of 17 genes. By building a gene-metabolite interaction network, we identified the organ-specific AMPK pathway that could be one of the underlying mechanisms of metformin activity. The activity of the AMPK complex in the hypothalamus is inhibited by leptin and metformin, whereas its activity in the liver is activated by metformin and leptin (see the right part of Figure 10). Our network results suggested that metformin intake could activate AMPK and consequently suppress the enzyme activities of fatty acid desaturases

(FADS), which would reduce the circulating levels of the 3 acyl-alkyl PCs and LDL-C. Our findings that metformin intake lowers the levels of the harmful LDL-C furthermore suggest its potential benefit in the prevention of cardiovascular disease (Xu *et al*, 2015; Brandmaier *et al,* 2015).

3.4.2 Metformin effect on non-targeted metabolite profiles in T2D patients and in multiple murine tissues

The non-targeted metabolomics data of KORA S4 and F4 were measured with the Metabolon analytical technologies (Metabolon, Inc., Durham, NC). In this study, we used 363 metabolites in the baseline KORA S4, and 353 metabolites in the 7-year follow-up F4 study. The number of overlapping metabolites in KORA S4 and F4 was 312 (Adam *et al*, 2016). Based on the cross-sectional and longitudinal investigations, and adjustments for multiple covariates such as age, BMI, lifestyle, clinical measurements and medication usage, we found that lower value of citrulline was significantly associated with metformin treatment (Adam *et al*, 2016).

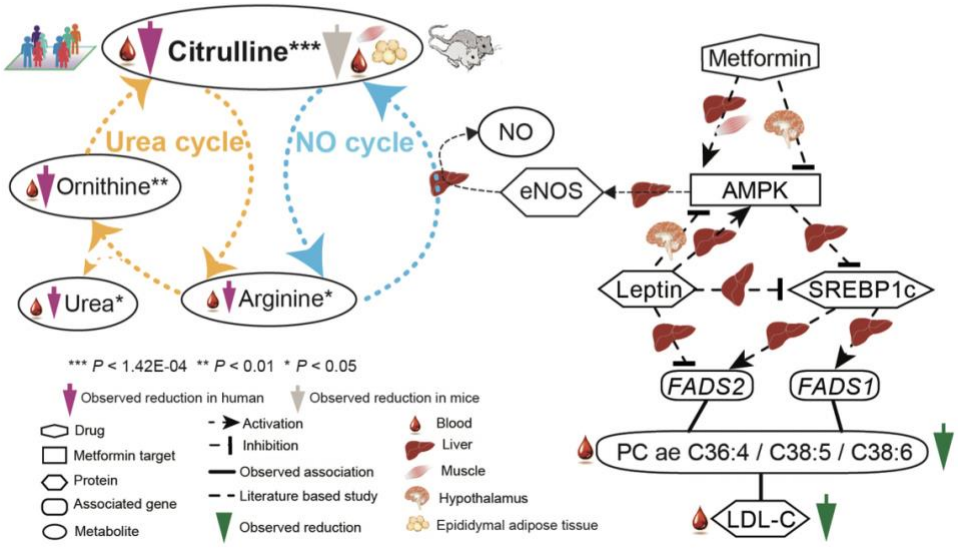

Figure 10 Effects of metformin - Lowered LDL-C and enhancing NO production

To corroborate our findings, we further analyzed four murine tissues from the Mouse200 intervention study including metformin treatment. In mice, we could validate the significantly lower citrulline values due to metformin treatment in plasma, skeletal muscle and adipose tissue but not in the liver. The reduced values of citrulline are most likely the result of metformin's pleiotropic effects on the interlocked urea and nitric oxide cycles (see the left part of Figure 10). The translational data derived from multiple murine tissues corroborated and complemented the findings from the human cohort (Adam *et al,* 2016; Adam *et al,* 2017).

4 Identification of novel metabolite biomarkers of (pre-) diabetes

We used both cross-sectional and prospective KORA studies to identify reliable candidate metabolite biomarkers of pre-diabetes and T2D, and to understand diabetes-related metabolic pathways by metabolite-protein networks (Wang-Sattler *et al,* 2012). Additionally, we have replicated our results in the European Prospective Investigation into Cancer and Nutrition (EPIC) – Potsdam cohort (Floegel *et al*, 2013).

4.1 Identification of independent candidate biomarkers

To avoid the potential influence of anti-diabetic medication (e.g. metformin), participants with known T2D diagnosis were excluded from our analysis. In our logistic regression analysis, we have adjusted for many known diabetes risk indicators (e.g. age, sex, BMI, smoking, alcohol consumption, systolic blood pressure, HDL cholesterol, HbA_{1C}, fasting glucose and fasting insulin). Furthermore, to identify unique and independent biomarker candidates, we employed two additional approaches (the non – parametric random forest and the parametric stepwise selection).

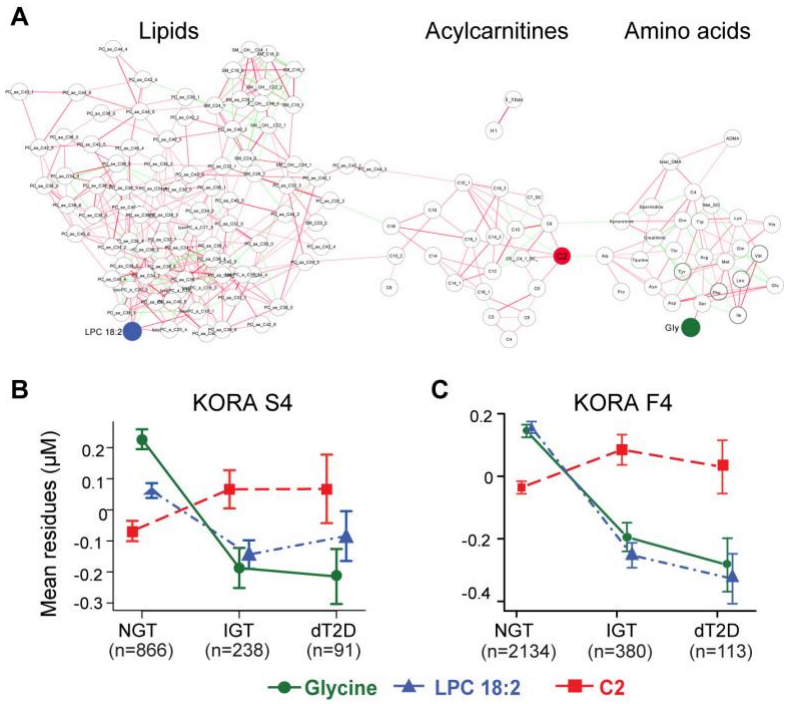

Figure 11 Identified three metabolites in three groups: participants with normal glucose tolerance (NGT), impaired glucose tolerance (IGT) and newly diagnosed T2D (dT2D).

We identified three metabolites (glycine, LPC (18:2) and acetylcarnitine C2) that had significantly altered levels in IGT individuals as compared to those with normal glucose tolerance (NGT) in both cross-sectional KORA S4 and F4 studies (Figure 11).

4.2 Glycine and LPC 18:2 predict risk of IGT and T2D

Using the prospective KORA S4 → F4 cohort, we further investigated the predictive value for IGT and T2D of the identified metabolites. We calculated the AUC values for each metabolite, each T2D risk indicator and their combinations (Table 2). By comparing the baseline metabolites concentrations of 118 incident IGT individuals with 471 NGT controls, the highest AUC value was obtained with age (0.580) among 11 tested risk indicators. Our results further confirmed that age was the leading risk factor in predicting the future onset of pre-diabetes. For the same comparison, the AUC value of LPC (18:2) was 0.610, which even bigger than age (Table 2).

Table 2 AUC values for each metabolite and each diabetes risk indicator and their combinations.

	118 incident IGT vs. 471 NGT	**91 incident T2D vs. 885 non-T2D**
Metabolite		
Glycine	0.546	0.604
LPC 18:2	0.610	0.606
C2	0.521	0.530
Glycine + LPC (18:2) + C2	0.622	0.635
Single T2D risk indicator		
Age	0.580	0.629
Sex	0.519	0.584
BMI	0.576	0.685
Physical activity	0.550	0.530
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		
age, sex, BMI, physical activity, alcohol intake, smoking, Systolic BP, HDL cholesterol, HbA_{1C}, Fasting glucose and fasting insulin	0.656	0.818
Metabolites combined with T2D risk indicators		
Glycine + LPC (18:2) + C2 + combined T2D risk indicators	0.682	0.828

When 91 incident T2D participants were compared with 785 people who remained diabetes-free during the 7-year follow-up, fasting glucose had the highest AUC value (0.735) for T2D. The AUC values of combined T2D risk indicators were 0.656 and 0.818 for IGT and T2D, respectively. Addition of the 3 identified metabolites to the combined T2D risk indicators, the AUC values increased 2.6% (= 0.682 - 0.656) and 1.0% (= 0.828 - 0.818) for IGT and T2D, respectively. Our identified three metabolites improved prediction of IGT and/or T2D (see Table 2).

4.3 Replicated 4 out of the 5 amino acids in nested case-control setting

Based on prospective nested case-control study design in the Framingham Offspring Study, five branched-chain and aromatic amino acids were identified as predictors of T2D (Wang *et al,* 2011). In our population-based prospective investigation of 843 KORA individuals, we conducted linear regression analysis between baseline metabolite concentrations (KORA S4) and follow-up 2-h glucose values after oral glucose tolerance test (KORA F4). We found that none of the previously reported five amino acids were significantly associated with T2D, indicating that they cannot predict risk of glucose tolerance in the population-based settings (Wang-Sattler *et al*, 2012).

To replicate the identified five amino acids (Wang *et al,* 2011), we specifically matched our baseline samples to the incident cases of T2D in the KORA study using the same method described in Wang *et al,* 2011 (i.e. matched for age, BMI and fasting glucose). We replicated four out of the five amino acids (isoleucine, leucine, valine and tyrosine) excepted for phenylalanine. In contrast, phenylalanine, but none of these four amino acids, was identified to be significantly associated with risk of T2D in the population-based EPIC-Potsdam study (Floegel *et al*, 2013).

As expected, the three 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 included about 50% pre-diabetic individuals, which is significantly higher than in the general population (about 15%). Most likely, concentration changes in these amino acids occur at a later stage of T2D progression (IGT to T2D) whereas the metabolites identified in our studies may be more suitable for early disease stage (NGT to IGT) (Wang-Sattler *et al*, 2012).

4.4 Metabolite-protein interaction networks

To investigate the underlying molecular mechanisms of the three identified metabolites in relation to T2D, we explored their associations with 46 T2D-related genes, which were known at that time, by protein-metabolite interaction networks (Szklarczyk *et al,* 2011; Wishart *et al,* 2009). Specifically, we have manually checked for the biochemical links between each metabolite with pathway-related proteins and T2D-related genes and identified a network of seven T2D-related genes and four enzymes (see Figure 12).

Our network results suggest that the observed decrease in glycine in pre-diabetics and diabetics may result from insulin resistance. The enzyme 5-aminolevulinate synthase 1 (ALAS-H) catalyzes the condensation of glycine and

succinyl-CoA into 5-aminolevuinate acid (ALA). Lower circulating values of glycine in individuals with IGT and T2D could potentially reflect an altered activity of ALAS-H. Production of ALA is the first step in heme biosynthesis and ALA deficiency has been recently reported to cause IGT and insulin resistance (Saitoh *et al*, 2018). Moreover, two double-blind and randomized placebo-controlled trials conducted in Hawaii and Japan suggested that the oral supplementation with ALA can protect from mild hyperglycemia and help prevent T2D (Rodriguez *et al*, 2012; Higashikawa *et al*, 2013).

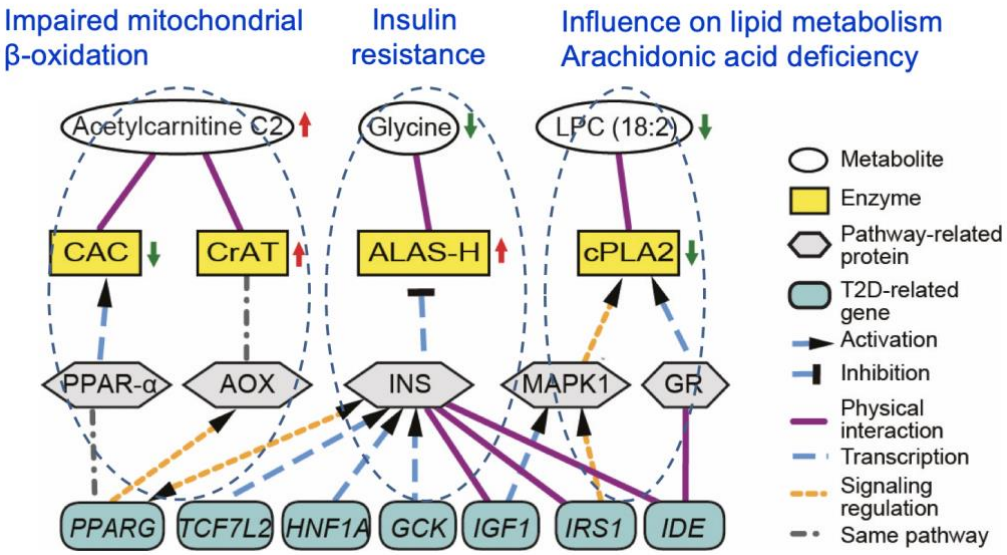

Figure 12 The identified 3 metabolites linked with 7 T2D-related genes

Acetylcarnitine is produced by the mitochondrial matrix enzyme, CrAT from carnitine and acetyl-CoA, a molecule that is both a product of fatty acid β-oxidation and glucose oxidation and can be used by citric acid cycle for energy generation. The positive association between C2 metabolite and pre-diabetes found in our studies may be a consequence of impaired mitochondrial β-oxidation. Mitochondrial dysfunction has been proposed as a central cause of insulin resistance (Kim *et al*, 2008; Fazakerley *et al*, 2018).

Decreased concentration of LPC (18:2) was also found to be biomarkers of myocardial infarction in the KORA S4 study, and this metabolite was replicated in KORA S2 and AGES- REFINE (Age, Gene/Environment Susceptibility and the Risk Evaluation For Infarct Estimates) - Reykjavik studies (Ward-Caviness *et al*, 2017). Currently, we are investigating this metabolite using multi-levels of OMICs data to deeper understand LPC (18:2) related pathophysiological pathways.

5 Summary and future perspective

5.1 Summary of 14 pleiotropic metabolite biomarkers

We have separately identified a number of candidate metabolite biomarkers of pre-diabetes, T2D, metformin intake, aging, smoking and alcohol drinking. Taken together, 14 metabolites were associated with more than two of these investigated influencing factors of diabetes and / or with (pre-) diabetes (Table 3).

Table 3 Overview of 14 pleiotropic metabolite biomarkers

Metabolites	Pre-diabetes	T2D	Metformin intake	Aging	Smoking	Alcohol drinking
C2	↑			↑		
Glycine	↓	↓				
LPC (18:2)	↓	↓		↓	↓	
Isoleucine		↑		↓		
Leucine		↑		↓		
Valine		↑		↓		
Phenylalanine		↑		↓		
Tryosine		↑		↑		
PC ae C38:6			↓	↓	↓	
Arginine			↓	↓	↑	
Ornithine			↓	↑	↑	
PC aa C34:1				↑	↑	↑
PC ae C40:6				↓	↓	↓
PC ae C34:3				↑	↓	↓

Lower blood concentration of LPC (18:2) was detected in individuals with pre-diabetes and T2D, and was significantly associated with aging and smoking (Wang-Sattler *et al*, 2012; Chat *et al*, 2019; Xu *et al*, 2013). Furthermore, higher PC aa C34:1 but lower PC ae C40:6 levels were consistently associated with aging, smoking and alcohol drinking (Chat *et al*, 2019; Xu *et al*, 2013; Jaremek *et al*, 2013). As we reported in our pilot study of smoking, the synthesis of these phospholipids is regulated by the enzyme AGPS (Wang-Sattler *et al*, 2008; Figure 6). The reported down-regulation of AGPS expression in lung tissues of smokers led us to conclude that smoking could be related with plasmalogen-deficiency disorders, caused by reduced or abolished activity of the peroxisomal AGPS (Gruber *et al*, 2006; Wang-Sattler *et al*, 2008). Additionally, the Agps's gene expression was found significantly decreased in the aged murine cerebellum, which consequently influences several downstream cellular processes, including synthesis of plasmalogens (Popesco *et al*, 2008).

Phenylalanine and three branched chain amino acids (isoleucine, leucine and valine) exhibited opposite associations with T2D and aging (Wang *et al*, 2011; Wang-Sattler *et al*, 2012; Floegel *et al*, 2013; Chat *et al*, 2019). This suggests that lower

concentration of those amino acids may be protective for T2D, but harmful for longevity. The contradictory effects of these amino acids should be further elucidated using causality approaches before considering their supplementation to patients with diabetes. Tryosine was positively associated with T2D and aging, and these results may suggest lower concentration of tryosine is protective for both T2D and aging (Wang *et al*, 2011; Chat *et al*, 2019).

Increased level of ornithine was associated with aging and smoking (Chat *et al*, 2019; Xu *et al*, 2013, Table 3). We additionally observed the ornithine level decreased by 14.8% in quitters compared with smokers during a 7-years follow-up, suggesting a normalization of its level after smoking cessation (Xu *et al*, 2013). Furthermore, metformin treatment lowered ornithine levels in T2D patients (Adam *et al*, 2016). Therefore, our results suggest that metformin exert pleiotropic effects on human metabolism. Indeed, apart from being a safe, effective and globally affordable glucose-lowering medication for the treatment of (pre-) diabetes, metformin therapy appears to protect from cancer, inflammation, and age-related pathologies and was proposed as a potential geroprotector in translational aging research (Piskovatska *et al*, 2019).

5.2 Summary

We conducted stringent quality control of metabolite profiles to identify reliably and robust candidate biomarkers of pre-diabetes and T2D. This enabled us to systematically characterize numerous factors that influence T2D onset and progression such as age, smoking, alcohol consumption and metformin usage. Our identified metabolite biomarkers (glycine, LPC (18:2) and C2) might precede other metabolites (branched-chain and aromatic amino acids) in the ability to detect pre-diabetes and monitor T2D progression.

5.3 Future perspective

My future research will focus on the investigation of complications of diabetes (e.g. chronic kidney diseases) using newly available multi-time-point and multi-levels of OMICs data (e.g. genomics, epigenomics, transcriptomics, proteomics and metabolomics). This will enable us to conduct causality analysis (e.g. mendelian randomization) and deepen the understanding of our identified candidate biomarkers including the 14 pleiotropic metabolites. Furthermore, additional types of “big data” include personalized clinical phenotypes further increasing the dimensions and complexity of ‘personalized OMICs data’. Novel methods for integration of personalized OMICs data will be a challenging for me. My final goal to translate our discoveries into personalized healthcare, a future without diabetes.

6 References

ADA, American Diabetes Association, Standards of Medical Care in Diabetes – 2020. *Diabetes care* 2020 Jan; 43 (Suppl 1): S1-S2. doi: 10.2337/dc20-Sint.

Adam J, Brandmaier S, Leonhardt J, Scheerer MF, Mohney RP, Xu T, Bi J, Rotter M, Troll M, Chi S, Heier M, Herder C, Rathmann W, Giani G, Adamski J, Illig T, Strauch K, Li Y, Gieger C, Peters A, Suhre K, Ankerst D, Meitinger T, Hrabe de Angelis M, Roden M, Neschen S, Kastenmüller G, **Wang-Sattler R***. (2016) Metformin Effect on Non-Targeted Metabolite Profiles in Patients with Type 2 Diabetes and Multiple Murine Tissues. *Diabetes*. 65(12):3776-3785. doi: 10.2337/db16-0512. *Correspondence

Adam J, Brandmaier S, Troll M, Rotter M, Mohney RP, Heier M, Adamski J, Li Y, Neschen S, Kastenmüller G, Suhre K, Ankerst D, Meitinger T, **Wang-Sattler R***. (2017) Response to Comment on Adam et al. *Diabetes* 66(5):e3-e4. doi: 10.2337/dbi17-0010.

Ahlqvist E, Storm P, Käräjämäki A, Martinell M, Dorkhan M, Carlsson AC, Vikman P, Prasad RB, Mansour Aly D, Almgren P, Wessman M, Shaat N, Spégel P, Mulder H, Lindholm E, Melander O, Hansson O, Malmqvist U, Lernmark A, Lahti K, Forsén T, Tuomi T, Rosengren AH, Groop L. (2018) Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet. Diabetes Endocrinol*. 6(5):361-369. doi: 10.1016/S2213-8587(18)30051-2.

Buse JB, Caprio S, Cefalu WT, Ceriello A, Del Prato S, Inzucchi SE, McLaughlin S, Phillips GL 2nd, Robertson RP, Rubino F, Kahn R, Kirkman MS. (2009) How do we define cure of diabetes? *Diabetes Care*. 32(11):2133-5. doi: 10.2337/dc09-9036.

Brandmaier S, Xu T, Illig T, Suhre K, Adamski J, **Wang-Sattler R***. (2015) Response to comment on Xu et al. Effects of Metformin on Metabolite Profiles and LDL Cholesterol in Patients With Type 2 Diabetes. *Diabetes Care* 38(12):e216-7. doi:10.2337/dci15-0022.

Chak CM, Lacruz ME, Adam J, Brandmaier S, Covic M, Huang J, Meisinger C, Tiller D, Prehn C, Adamski J, Berger U, Gieger C, Peters A, Kluttig A, **Wang-Sattler R***. (2019) Ageing Investigation Using Two-Time-Point Metabolomics Data from KORA and CARLA Studies. *Metabolites* 9(3). pii: E44. doi: 10.3390/metabo9030044.

Crick F (1970). Central dogma of molecular biology. *Nature* 227 (5258): 561–3.

Fazakerley DJ, Minard AY, Krycer JR, Thomas KC, Stöckli J, Harney DJ, Burchfield JG, Maghzal GJ, Caldwell ST, Hartley RC, Stocker R, Murphy MP, James DE (2018) Mitochondrial oxidative stress causes insulin resistance without disrupting oxidative phosphorylation. *J Biol Chem*. 293(19):7315-7328. doi: 10.1074/jbc.RA117.001254.

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. (2013). Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* 62(2):639-48. doi: 10.2337/db12-0495.

Gieger C, Geistlinger L, Altmaier E, Hrabé de Angelis M, Kronenberg F, Meitinger T, Mewes HW, Wichmann HE, Weinberger KM, Adamski J, Illig T, Suhre K. (2008) Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* 4(11):e1000282. doi: 10.1371/journal.pgen.1000282.

Gruber MP, Coldren CD, Woolum MD, Cosgrove GP, Zeng C, Barón AE, Moore MD, Cool CD, Worthen GS, Brown KK, Geraci MW. (2006) Human lung project: evaluating variance of gene expression in the human lung. *Am J Respir Cell Mol Bio*l 35: 65-71. doi: 10.1165/rcmb.2004-0261OC.

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[#]. (2012) Schizophrenia shows a unique metabolomics signature in plasma. *Transl Psychiatry* 2:e149. doi: 10.1038/tp.2012.76. doi: 10.1038/tp.2012.76. [#]Equal contribution

Higashikawa F1, Noda M, Awaya T, Tanaka T, Sugiyama M. (2013) 5-aminolevulinic acid, a precursor of heme, reduces both fasting and postprandial glucose levels in mildly hyperglycemic subjects. *Nutrition.* 29(7-8):1030-6. doi: 10.1016/j.nut.2013.02.008.

Illig T, Gieger C, Zhai G, Römisch-Margl W, **Wang-Sattler R**, Prehn C, Altmaier E, Kastenmüller G, Kato BS, Mewes HW, Meitinger T, de Angelis MH, Kronenberg F, Soranzo N, Wichmann HE, Spector TD, Adamski J, Suhre K. (2010) A genome-wide perspective of genetic variation in human metabolism. *Nat Genet* 42: 137-141. doi: 10.1038/ng.507.

Jaremek M, Yu Z, Mangino M, Mangino M, Mittelstrass K, Prehn C, Singmann P, Xu T, Dahmen N, Weinberger K, Suhre K, Peters A, Döring A, Hauner H, Adamski J, Illig T, Spector T, **Wang-Sattler R***. (2013) Alcohol-induced metabolomic differences in humans. *Transl. Psychiatry* 3:e276. doi: 10.1038/tp.2013.55.

Kim JA, Wei Y, Sowers JR. (2008) Role of mitochondrial dysfunction in insulin resistance. *Circ Res.*102(4):401-14. doi: 10.1161/CIRCRESAHA.107.165472.

Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM; Diabetes Prevention Program Research Group. (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 346: 393-403. doi: 10.1056/NEJMoa012512.

Mittelstrass K, Ried JS, Yu Z, 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. (2011) Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PloS Genet* 7(8):e1002215. doi: 10.1371/journal.pgen.1002215.

Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, Haqq AM, Shah SH, Arlotto M, Slentz CA, Rochon J, Gallup D, Ilkayeva O, Wenner BR, Yancy WS Jr, Eisenson H, Musante G, Surwit RS, Millington DS, Butler MD, Svetkey LP. (2009) A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab* 9: 311-326. doi: 10.1016/j.cmet.2009.02.002.

Nicholson JK, Wilson ID. (2003) Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nat Rev Drug Discov* Aug;2(8):668-76. doi: 10.1038/nrd1157.

Piskovatska V, Stefanyshyn N, Storey KB, Vaiserman AM, Lushchak O. (2019) Metformin as a geroprotector: experimental and clinical evidence. *Biogerontology.* 20(1):33-48. doi: 10.1007/s10522-018-9773-5.

Popesco MC, Lin S, Wang Z, Ma ZJ, Friedman L, Frostholm A, Rotter A. (2008) Serial analysis of gene expression profiles of adult and aged mouse cerebellum. *Neurobiol Aging*. 5:774-88. doi: 10.1016/j.neurobiolaging.2006.12.006.

Roden M, Shulman GI. (2019) The integrative biology of type 2 diabetes. *Nature*. Dec;576(7785):51-60. doi: 10.1038/s41586-019-1797-8.

Römisch-Margl W, Prehn C, Bogumil R, Röhring C, Suhre K, Adamski J. (2012) Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics* volume 8, Issue 1, pp 133–142.

Rodriguez BL, Curb JD, Davis J, Shintani T, Perez MH, Apau-Ludlum N, Johnson C, Harrigan RC. (2012) Use of the dietary supplement 5-aminiolevulinic acid (5-ALA) and its relationship with glucose levels and hemoglobin A1C among individuals with prediabetes. *Clin Transl Sci.* 5(4):314-20. doi: 10.1111/j.1752-8062.2012.00421.x.

Saitoh S, Okano S, Nohara H, Nakano H, Shirasawa N, Naito A, Yamamoto M, Kelly VP, Takahashi K, Tanaka T, Nakajima M, Nakajima O. (2018) 5-aminolevulinic acid (ALA) deficiency causes impaired glucose tolerance and insulin resistance coincident with an attenuation of mitochondrial function in aged mice. *PLoS One*.13(1):e0189593. doi: 10.1371/journal.pone.0189593.

Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wägele B, Altmaier E; CARDIoGRAM, Deloukas P, Erdmann J, Grundberg E, Hammond CJ, de Angelis MH, Kastenmüller G, Köttgen A, Kronenberg F, Mangino M, Meisinger C, Meitinger

T, Mewes HW, Milburn MV, Prehn C, Raffler J, Ried JS, Römisch-Margl W, Samani NJ, Small KS, Wichmann HE, Zhai G, Illig T, Spector TD, Adamski J, Soranzo N, Gieger C. (2011) Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 477(7362):54-60. doi: 10.1038/nature10354.

Stumvoll M, Goldstein BJ, van Haeften TW. (2005) Type 2 diabetes: principles of pathogenesis and therapy. *Lancet.* 365(9467):1333-46. doi: 10.1016/S0140-6736(05)61032-X.

Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C. (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 39: D561-568. doi: 10.1093/nar/gkq973.

Tabak AG, Herder C, Rathmann W, Brunner EJ, Kivimaki M. (2012) Prediabetes: a high-risk state for diabetes development. *Lancet;* 379(9833):2279-90. doi: 10.1016/S0140-6736(12)60283-9.

Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, Keinänen-Kiukaanniemi S, Laakso M, Louheranta A, Rastas M, Salminen V, Uusitupa M; Finnish Diabetes Prevention Study Group. (2001) Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med* 344**:** 1343-1350. doi: 10.1056/NEJM200105033441801.

Walker DI, Valvi D, Rothman N, Lan Q, Miller GW, Jones DP. (2019) The metabolome: A key measure for exposome research in epidemiology. *Curr Epidemiol Rep*. 6:93-103.

Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, O'Donnell CJ, Carr SA, Mootha VK, Florez JC, Souza A, Melander O, Clish CB, Gerszten RE. (2011) Metabolite profiles and the risk of developing diabetes. *Nat Med* **17:** 448-453. doi: 10.1038/nm.2307.

Wang-Sattler R*, Yu Z, Herder C, Messias AC, Floegel A, He Y, Heim K, Campillos M, Holzapfel C, Thorand B, Grallert H, Xu T, Bader E, Huth C, Mittelstrass K, 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, Hrabé de Angelis M, Rathmann W, Suhre K, Prokisch H, Peters A, Meitinger T, Roden M, Wichmann HE, Pischon T, Adamski J, Illig T. (2012) Novel biomarkers for pre-diabetes identified by metabolomics. *Mol. Syst. Biol.* Sep 25;8:615. doi: 10.1038/msb.2012.43.

Wang-Sattler R*, Yu Y, Mittelstrass K, Lattka E, Altmaier E, Gieger C, Ladwig KH, Dahmen N, Weinberger KM, Hao P, Liu L, Li Y, Wichmann HE, 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*: 2008;3(12):e3863. doi: 10.1371/journal.pone.0003863.

Ward-Caviness CK, Xu T, Aspelund T, Thorand B, Montrone C, Meisinger C, Dunger-Kaltenbach I, Zierer A, Yu Z, Helgadottir IR, Harris TB, Launer LJ, Ganna A, Lind L, Eiriksdottir G, Waldenberger M, Prehn C, Suhre K, Illig T, Adamski J, Ruepp A, Koenig W, Gudnason V, Emilsson V, **Wang-Sattler R***, Peters A. (2017) Improvement of myocardial infarction risk prediction via inflammation-associated metabolite biomarkers. *Heart*. 103(16):1278-1285. doi: 10.1136/heartjnl-2016-310789.

WHO, Global report on diabetes. 2016

WHO, Global status report on alcohol and health. 2018

Wichmann HE, Gieger C, Illig T. (2005) KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* 67 Suppl 1: S26-30. doi: 10.1055/s-2005-858226.

Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, Mandal R, Sinelnikov I, Xia J, Jia L, Cruz JA, Lim E, Sobsey CA, Shrivastava S, Huang P, Liu P, Fang L, Peng J, Fradette R, Cheng D, Tzur D, Clements M, Lewis A, De Souza A, Zuniga A, Dawe M, Xiong Y, Clive D, Greiner R, Nazyrova A, Shaykhutdinov R, Li L, Vogel HJ, Forsythe I. (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* 37**:** D603-10. doi: 10.1093/nar/gkn810.

Xu T, Holzapfel C, Dong X, Bader E, Yu Z, Prehn C, Perstorfer K, Jaremek M, Roemisch-Margl W, Rathmann W, Li Y, Wichmann HE, Wallaschofski H, Ladwig KH, Theis F, Suhre K, Adamski J, Illig T, Peters A, **Wang-Sattler R***. (2013) Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *BMC Med*. Mar 4;11:60. doi: 10.1186/1741-7015-11-60.

Xu T, Brandmaier S, Messias AC, Herder C, Draisma HH, Demirkan A, Yu Z, Ried JS, Haller T, Heier M, Campillos M, Fobo G, Stark R, Holzapfel C, Adam J, Chi S, Rotter M, Panni T, Quante AS, He Y, Prehn C, Roemisch-Margl W, Kastenmüller G, Willemsen G, Pool R, Kasa K, van Dijk KW, Hankemeier T, Meisinger C, Thorand B, Ruepp A, Hrabé de Angelis M, Li Y, Wichmann HE, Stratmann B, Strauch K, Metspalu A, Gieger C, Suhre K, Adamski J, Illig T, Rathmann W, Roden M, Peters A, van Duijn CM, Boomsma DI, Meitinger T, **Wang-Sattler R***. (2015) Effects of Metformin on Metabolite Profiles and LDL Cholesterol in Patients With Type 2 Diabetes. *Diabetes Care* 2015 Oct;38(10):1858-67. doi: 10.2337/dc15-0658.

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***. (2012) Human serum metabolic profiles are age dependent. *Aging cell*, Dec;11(6):960-7. doi: 10.1111/j.1474-9726.2012.00865.x.

Zukunft S, Prehn C, Röhring C, Möller G, Hrabě de Angelis M, Adamski J, Tokarz J. (2018) High-throughput extraction and quantification method for targeted metabolomics in murine tissues. *Metabolomics*. 14(1):18. doi: 10.1007/s11306-017-1312-x.

7 Acknowledgement

I am very grateful to:

Prof. Dr. Annette Peters for being my mentor. Though having many obligations and duties as director of the EPI and speaker of KORA cohort amongst others, she was always available for discussion and support. I am grateful for supporting me and my group with constructive comments and being the official promoter of my doctoral students.

Prof. Dr. Konstantin Strauch, as being my mentor and providing me the opportunity to teach the 'Genetic Epidemiology' course since 2011. Although he is currently working at the University of Mainz, he continues to support me as member of my '*Fachmentorat*'.

PD Dr. Andreas Lechner, as member of my '*Fachmentora*t'. Especially, being a medical doctor he provided very valuable input on the relevance and importance of translational research. He also strongly supported me during the EU EIT Health project DeTecT2D.

I thank the current and past members of my group:

Marcela Covic, Stefan Brandmaier, Tao Xu, Jonathan Adam, Zhonghao Yu, Ying He, Markus Rotter, Erik Bader, Choiwai Maggie Chat, Marta Jaremek, Jialing Huang and Li Wang for their hard work, collaboration, joint achievement and celebration (e.g. Chinese food & German cakes), as well as many other interactions and support, such as help with translation of complicated German documents, and exchange about different cultures and backgrounds.

I am grateful to the board members of KORA and data-owners of the OMICs data:

Christian Gieger, Harald Grallert, Melanie Waldenberger, Gabi Anton, Cornelia Huth, Barbara Thorand, Margit Heier, Christa Mesinger, Wolfgang Rathmann, Erich Wichmann, Birgit Linkohr, Jerzy Adamski, Karsten Suhre, Thomas Illig, Thomas Meitinger for design and conducting the KORA cohort and / or generating multi-time-point and multi-levels OMICs data, which enabled my research.

I am grateful to my colleagues:

Brigitte Fröhlich, Astrid Glaser, Susanne Vogt, Antonia Flaquer, Anne Quante, Susanne Breitner, Nadine Lindemann, Hans Demski, Pamela Matias Garcia, Mohamed Elhadad, Rory Wilson, Maike Groenewold, Natascha Strauss, Humberto Chavez Moruno, Conny Prehn, Julia Scarpa, Werner Römisch-Margl, Gabi Karstenmüller, Monica Campillos, Anja Ludolph, Heilmaier Herbert, Thomas Hendel, Anni Paulini, Liane Thaller, Gabriele Kölbl, Angelika Andreas for their great support, collaboration and joint teaching, as well as many small, but very helpful business matters (ordering print paper and get signatures).

I thank

Yixue Li, Wei He, Hong He, Yashu Wang, Shen Chi, Lifang Mu, Peijian Zou, Yang Wang, Wensi Vennekate, Siwei Zhang, Xu Dong for their support and encouragement.

Finally, I thank

My parents and my five sisters for their love and encouragement, and last but not least, I thank **Katharina** and **Michael** for encouragement and much more…

8 Enumerate papers

8.1 Yu Z, …, Wang-Sattler R*, Human serum metabolic profiles are age dependent. *Aging Cell* 11(6):960-7 (2012)

https://doi.org/10.1111/j.1474-9726.2012.00865.x

8.2 Xu T, …, Wang-Sattler R*, Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *BMC Med*. 11:60 (2013)

https://doi.org/10.1186/1741-7015-11-60

8.3 Jaremek M, …, Wang-Sattler R*, Alcohol-induced metabolomic differences in humans. *Transl. Psychiatry* 3:e276 (2013)

https://doi.org/10.1038/tp.2013.55

8.4 Xu T, …, Wang-Sattler R*, Effects of Metformin on Metabolite Profiles and LDL Cholesterol in Patients With Type 2 Diabetes. *Diabetes Care* 38(10):1858-67 (2015)

https://doi.org/10.2337/dc15-0658

8.5 Adam J, …, Wang-Sattler R*, Metformin Effect on Non-Targeted Metabolite Profiles in Patients with Type 2 Diabetes and Multiple Murine Tissues. *Diabetes* 65(12): 3776-85 (2016)

https://doi.org/10.2337/db16-0512

8.6 Wang-Sattler R*, …, Illig T, Novel biomarkers for pre-diabetes identified by metabolomics. *Mol. Syst. Biol*. 8:615 (2012)

https://doi.org/10.1038/msb.2012.43

Aging Cell (2012) **11,** pp960–967 Doi: 10.1111/j.1474-9726.2012.00865.x

Human serum metabolic profiles are age dependent

Zhonghao Yu,[1]* Guangju Zhai,[2,3]* Paula Singmann,[1]* Ying He,[4,5] Tao Xu,[1] Cornelia Prehn,[6] Werner Römisch-Margl,[7] Eva Lattka,[1] Christian Gieger,[8] Nicole Soranzo,[2,9] Joachim Heinrich,[10] Marie Standl,[10] Elisabeth Thiering,[10] Kirstin Mittelstraß,[1] Heinz-Erich Wichmann,[10,11,12] Annette Peters,[1,13,14] Karsten Suhre,[7,15,16] Yixue Li,[4,5] Jerzy Adamski,[6,17] Tim D. Spector,[2†] Thomas Illig,[1,18†] and Rui Wang-Sattler[1†]

[1]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, 85764 Neuherberg, Germany
[2]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK
[3]Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St John's, NL, Canada
[4]Shanghai Center for Bioinformation Technology, 200235 Shanghai, China
[5]Key Lab of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 200031 Shanghai, China
[6]Genome Analysis Center, Institute of Experimental Genetics, Helmholtz Zentrum München, 85764 Neuherberg, Germany
[7]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany
[8]Institute of Genetic Epidemiology, Helmholtz Zentrum München, 85764 Neuherberg, Germany
[9]Wellcome Trust Sanger Institute Genome Campus, Hinxton, UK
[10]Institute of Epidemiology I, Helmholtz Zentrum München, 85764 Neuherberg, Germany
[11]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany
[12]Klinikum Grosshadern, Munich, Germany
[13]Institute of Epidemiology II, Helmholtz Zentrum München, 85764 Neuherberg, Germany
[14]Department of Environmental Health, Harvard School of Public Health Adjunct Associate Professor of Environmental Epidemiology, Boston, MA, USA
[15]Faculty of Biology, Ludwig-Maximilians-Universität, 82152 Planegg-Martinsried, Germany
[16]Department of Physiology and Biophysics, Weill Cornell Medical College in Qatar, 24144 Education City–Qatar Foundation, Doha, Qatar
[17]Institute of Experimental Genetics, Life and Food Science Center Weihenstephan, Technische Universität München, 85354 Freising-Weihenstephan, Germany
[18]Hannover Unified Biobank, Hannover Medical School, 30625 Hannover, Germany

Correspondence
Rui Wang-Sattler, Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, 85764 Neuherberg, Germany. Tel.: +49 89 3187 3978; fax: +49 89 3187 2428; e-mail: rui.wang-sattler@helmholtz-muenchen.de

Carried out at the Research Unit of Molecular Epidemiology, Helmholtz ZentrumMünchen, 85764 Neuherberg, Germany, and the Department of Twin Research and Genetic Epidemiology, King's College London, London, UK.
*Zhonghao Yu, Guangju Zhai and Paula Singmann contributed equally.
†Tim D. Spector, Thomas Illig and Rui Wang-Sattler contributed equally.
Co-corresponding authors. Tim D. Spector, Department of Twin Research and Genetic Epidemiology, King's College London, London, UK and Thomas Illig, Hannover Unified Biobank, Hannover Medical School, 30625 Hannover, Germany. Tel.: +49 89 3187 3978; fax: +49 89 3187 2428; e-mails: tim.spector@kcl.ac.uk; illig@helmholtz-muenchen.de

Accepted for publication *16 July 2012*

Summary

Understanding the complexity of aging is of utmost importance. This can now be addressed by the novel and powerful approach of metabolomics. However, to date, only a few metabolic studies based on large samples are available. Here, we provide novel and specific information on age-related metabolite concentration changes in human homeostasis. We report results from two population-based studies: the KORA F4 study from Germany as a discovery cohort, with 1038 female and 1124 male participants (32–81 years), and the TwinsUK study as replication, with 724 female participants. Targeted metabolomics of fasting serum samples quantified 131 metabolites by FIA-MS/MS. Among these, 71/34 metabolites were significantly associated with age in women/men (BMI adjusted). We further identified a set of 13 independent metabolites in women (with *P* values ranging from 4.6×10^{-04} to 7.8×10^{-42}, $\alpha_{corr} = 0.004$). Eleven of these 13 metabolites were replicated in the TwinsUK study, including seven metabolite concentrations that increased with age (C0, C10:1, C12:1, C18:1, SM C16:1, SM C18:1, and PC aa C28:1), while histidine decreased. These results indicate that metabolic profiles are age dependent and might reflect different aging processes, such as incomplete mitochondrial fatty acid oxidation. The use of metabolomics will increase our understanding of aging networks and may lead to discoveries that help enhance healthy aging.

Key words: age; aging; epidemiology; metabolomics; population-based study.

Introduction

Life expectancy in humans has dramatically increased throughout the world (Ferrucci *et al.*, 2008). This process entails great challenges for world population in terms of health and economics, and complex social issues emerge, in particular for the future of healthcare systems, as aging is often accompanied by disabilities and diseases such as cardiovascular diseases, chronic lower respiratory tract disease, Alzheimer's disease, chronic joint symptoms, arthritis, and diabetes (Butler *et al.*, 2008; Wijsman *et al.*, 2011). Thus, understanding the physiology of aging is of tremendous importance to allow populations to grow old, disease-free and with a good quality of life.

Aging is a very complex process because many transformations happen to the human organisms that affect all levels, from organ systems to cell organelles, and lead to a wide variety of altered functions. However, this process is incompletely understood. Genetic and environmental influences seem to be involved, but other approaches and insights are needed (Karasik *et al.*, 2005; Piper & Bartke, 2008; Kerber *et al.*, 2009; Deelen *et al.*, 2011). Popular theories of aging include those implicating free radicals, accumulation of glycated proteins (AGEs), involvement of chronic low-grade inflammation, altered action of several hormones or chromosome telomere shortening (Szibor & Holtz, 2003; Franceschi *et al.*, 2007; Yan *et al.*, 2007; Jiang *et al.*, 2008; Simm *et al.*, 2008; Perheentupa & Huhtaniemi, 2009). Nonetheless, the largest volume of knowledge stems from non-human studies. Based on genetic studies in animal models, there are several complex pathways known to be involved in aging mechanisms that also are clearly linked to metabolism (Toth & Tchernof, 2000; Dennis *et al.*, 2009; Feltes *et al.*, 2011; Partridge *et al.*, 2011).

Metabolomics is a key technology of modern systems biology that focuses on obtaining an integral depiction of the current metabolic status of an organism, associated with physiological and pathophysiological processes (Psychogios *et al.*, 2011). Numerous small molecules are measured that can be both endogenous and exogenous. These ideally represent the whole range of intermediate metabolic pathways and may serve as biomarkers, indicating distinct physiological and/or pathophysiological states of an organism (He *et al.*, 2012). Metabolomics is therefore a valuable tool for investigating in a single approach all the various ways in which metabolism is influenced, and then linking these influences to the phenotypic outcome of interest, thus a very promising tool for capturing the complexity of the aging process.

Only a few aging studies with metabolomics analyses have been conducted in animal models (in rats, mice and dogs) (Williams *et al.*, 2005; Wang *et al.*, 2007; Nevedomskaya *et al.*, 2010). The very few published results on aging in adult humans have been small. A study of 269 individuals (both men and women) analyzed the human plasma metabolome with age using NMR and thus obtaining a very different set of metabolites to the current study (Lawton *et al.*, 2008; Orešič, 2009). Nikkilä and colleagues performed a metabolomics study on early childhood, following 59 children from birth to an age of 4 years, and identified previously unknown metabolic changes with age (Nikkila *et al.*, 2008).

The objectives of the current study were to characterize the metabolic profile of a large group of subjects with a wide age range (32–81 years) and identify metabolites relates to chronological age independent of BMI. Such metabolites may subsequently be investigated in future studies to establish whether they associate with health conditions and aging mechanisms.

Results

Study population

Individuals without metabolic diseases (e.g. hypertension, type 2 diabetes and obesity) were used in both studies of the Cooperative Health Research in the Region of Augsburg () (Holle *et al.*, 2005; Wichmann *et al.*, 2005) and the UK Adult Twin Registry (TwinsUK) (Spector & Williams, 2006; Moayyeri *et al.*, 2012). KORA F4 sample involved 1038 women and 1124 men, aged 32–81 years (Table 1). Women and men of the KORA participants were about the same average age (54 years old) and had comparable mean BMIs (around 26 kg m^{-2}). The TwinsUK contained 742 women, aged 19–82 years, with a mean age of 58 years and BMI of about 26 kg m^{-2}. Comparing the discovery KORA women with the replication TwinsUK individuals, KORA women are about 4 years younger and have a slightly higher BMI (Table 1).

Serum concentrations of 163 metabolites were measured in all fasting participants; the 163 metabolites and their characteristics are summarized in Table S1 (Supporting information).

Table 1 Population characteristics of KORA F4 and Twins UK

	KORA F4		TwinsUK
	Males	Females	Females
N	1124	1038	742
Age (years)†	53.6 ± 12.5	54.1 ± 13.1	57.7 ± 10.6
BMI (kg m^{-2})†	25.9 ± 3.9	27.1 ± 3.2	25.6 ± 3.7

†Values of age and BMI are shown mean ± standard deviation (SD).

Discovery of age-associated metabolites: general analyses procedure

Owing to prior results from KORA F4, which showed strong metabolic differences between women and men (Mittelstrass *et al.*, 2011), we conducted strictly sex-separated analyses. Furthermore, we found that the BMI was significantly correlated with both age (Pearson's $r = 0.26$, $P = 2.2 \times 10^{-16}$) and metabolite concentrations (Table S1) in KORA F4. To investigate the 'true' age-related metabolites, BMI was adjusted in subsequent analyses, and residuals of metabolite concentrations from linear regression against BMI were used in the graphs. We first plotted heat maps of mean residuals for 131 metabolites for each year of age as an explorative tool. Second, we used linear regression models for each metabolite at individual level to identify age-related metabolites. Moreover, as some of these metabolites correlated with each other, especially for the same classes (Fig. S1), we employed two additional statistical methods, the nonparametric random forest and the parametric stepwise selection methods, to identify unique and independent biomarker candidates. Finally, the set of age-related metabolites were replicated in the TwinsUK study. Smoother plots of this subset of metabolites were displayed to characterize age-associated changes regarding metabolite concentrations (Figs S3–S5).

Heat map of KORA F4 revealed association between age and metabolite

We display heat maps of normalized mean metabolite residuals for each year of age in both women and men. The resulting heat map for women displayed a clear increase in metabolite concentrations with respect to age in KORA F4, in particular with most acylcarnitines (ACs) and diacyl phosphatidylcholines (PC aa), while others showed gradual decreases with age, for example, for most amino acids (AAs) (Fig. 1). A similar trend for most ACs and AAs were also observed in the heat map for men; however, most of the Lyso PCs decreased with age in men. In general, the results of heat maps of women and men showed different patterns of changes and clusters (the heat map result of men is shown in Fig. S2).

Identification of a metabolite set for age

We performed multiple linear regression analysis between metabolite concentrations and age with BMI as covariate. In the discovery cohort, we found 71 metabolite concentrations in women and 34 metabolites in men that were significantly associated with age (P values $< 3.8 \times 10^{-4}$). As some of these metabolites are expected to correlate with each other, we employed two additional statistical methods (the nonparametric random forest and the parametric stepwise selection, see below) to identify independent biomarker candidates. Of the 71/34 metabolites, 13/12 were found to contain independent information in women and men, respectively (mean concentrations, beta and P values for women are shown in Table 2, and for men in Table S2) (Supporting information). Five metabolites (C12:1, C18:1, His, Trp, and PC ae C36:1) were found both in women and men.

Smoother plots were drawn to characterize the trends and courses of metabolite concentration changes with age in women (Fig. S3) and men (Fig. S4). The smoothing method was 'loess', which is a locally weighted regression robust against a small fraction of outliers. As already stated for the heat map, some metabolite concentrations showed increasing, while some showed decreasing, trends with age, but smoother plots provided deeper insight. For example, the larger proportion of the identified metabolites showed linear associations with age for both women and

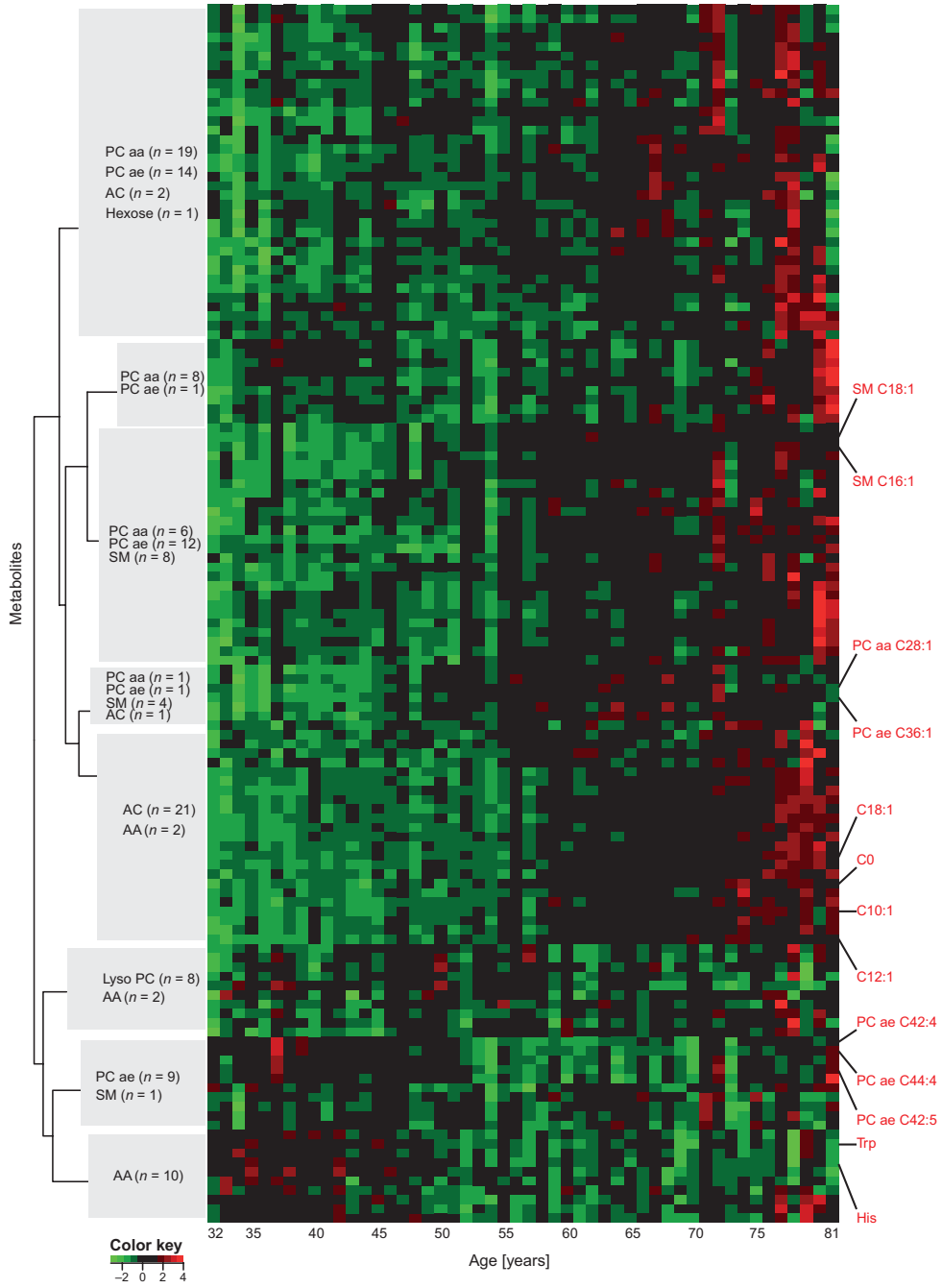

Fig. 1 Heat map of the fold standard deviation changes between ages, and clustering of these changes, over all ages in 1038 women from KORA F4. The heat map shows changes of x-fold standard deviation from the overall mean concentration for each age year in a color-coded way. Green squares represent a decrease, and red squares an increase. Gray boxes represent groups of metabolites with similar changes with number of metabolites in parentheses. Metabolite names in red indicate our set of 13 metabolites. AA, amino acid; AC, acylcarnitines; PC aa, phosphatidylcholinediacyl; PC ae, phosphatidylcholine acyl-alkyl; and lyso PC a, lysophosphatidylcholine acyl.

men, with the exception of three metabolites in women (PC ae C42:4, PC ae 42:5, and PC ae C44:4) and one in men (Gln), which exhibited a decrease around age of 51.

Replication of the metabolite set in the TwinsUK

For replication purpose, we used an independent sample of 742 women subjects derived from the TwinsUK cohort, for whom fasting serum metabolomics data were available. The metabolites showed comparable mean concentrations in both studies. Among the 13 age-related metabolites found in the KORA study, 11 metabolites were well replicated in the TwinsUK sample, with $P < 0.004$ for the significance level after adjustment for multiple testing with the Bonferroni correction method (Table 2), and the effect direction was the same as in the discovery sample (except for one metabolite, PC ae C36:1), with borderline significance in the TwinsUK sample ($P = 2.00 \times 10^{-03}$). We also drew smoother plots for the TwinsUK sample and obtained similar curve shapes (with the exception of PC ae C36:1) (Fig. S5).

Table 2 Potential biomarkers for aging in women from KORA F4 and TwinsUK

	Discovery sample			Replication sample				
	KORA F4 females			TwinsUK females			Meta-analysis	
Marker	Mean† ± SD	β‡ (SE)	*P* value§	Mean ± SD	β (SE)	*P* value§	β (SE)	*P* value§
C0	32.93 ± 6.79	0.45 (0.05)	2.77E−19	38.29 ± 9.38	0.26 (0.04)	5.30E−10¶	0.33 (0.03)	1.04E−26
C10:1	0.15 ± 0.05	79.67 (6.69)	7.96E−31	0.19 ± 0.06	24.15 (6.38)	7.53E−05¶	50.59 (4.62)	6.08E−28
C12:1	0.14 ± 0.04	105.20 (8.06)	2.93E−36	0.17 ± 0.05	28.64 (8.60)	1.00E−03¶	69.4 (5.88)	3.87E−32
C18:1	0.12 ± 0.03	118.02 (10.86)	3.65E−26	0.19 ± 0.05	33.09 (7.51)	5.40E−06¶	60.57 (6.18)	1.07E−22
His	97.41 ± 13.67	−0.18 (0.02)	3.15E−13	98.22 ± 28.93	−0.10 (0.02)	2.90E−06¶	−0.14 (0.01)	4.18E−23
Trp	80.2 ± 8.85	−0.24 (0.04)	1.29E−10	86.66 ± 16.63	−0.035 (0.03)	0.18	−0.11 (0.02)	5.81E−06
PC aa C28:1	3.56 ± 0.89	4.65 (0.36)	2.10E−35	4.17 ± 1.27	1.54 (0.33)	2.10E−06¶	2.96 (0.24)	4.59E−34
PC ae C36:1	8.94 ± 2.05	2.15 (0.15)	7.75E−42	12.14 ± 5.23	−0.31 (0.16)	2.00E−03	1 (0.11)	6.75E−20
PC ae C42:4	1.08 ± 0.25	−6.44 (1.39)	4.25E−06	1.18 ± 0.46	−5.84 (1.13)	1.20E−07¶	−6.08 (0.88)	4.13E−12
PC ae C42:5	2.49 ± 0.50	−2.42 (0.69)	4.57E−04	2.71 ± 0.98	−2.57 (0.52)	3.20E−07¶	−2.52 (0.42)	1.38E−09
PC ae C44:4	0.46 ± 0.11	−15.23 (3.05)	6.76E−07	0.51 ± 0.18	−10.56 (2.52)	1.00E−05¶	−12.45 (1.94)	1.45E−10
SM C16:1	16.80 ± 2.97	1.27 (0.11)	4.89E−28	19.58 ± 4.79	0.38 (0.09)	1.00E−05¶	0.74 (0.07)	3.74E−26
SM C18:1	11.96 ± 2.56	1.32 (0.13)	3.36E−22	12.38 ± 3.50	0.42 (0.12)	1.00E−03¶	0.33 (0.03)	1.04E−26

†Mean concentration in μM from serum.
‡β estimate represents changes per year of age, adjusted for BMI.
§Corrected significance level of $\alpha_{corr} = 0.004$ (correction for 13 tests according to Bonferroni method).
¶Replication succeeded for these markers.

Furthermore, we analyzed both study samples in a meta-analysis, which virtually gave the same results as in the KORA female sample alone. All 13 metabolites of the marker set were significant, with *P* values ranging from 5.18×10^{-6} to 3.87×10^{-32}.

Discussion

We found strong associations between age and the human metabolome and identified 11 metabolites that associated with age in women in both discovery and replication cohort. Our finding of 12 metabolites associated with age in men was not yet tested for confirmation.

The extent of sexual dimorphism in the metabolome has recently been shown (Mittelstrass *et al.*, 2011), stressing the need for sex stratification. Even accounting for the lack of a replication sample, the results in men are less impressive for unclear reasons. Lifestyle and different life experiences probably play prominent roles in this. Nonetheless, five metabolites could be identified in both men and women in the discovery cohort, so that these may be considered to be metabolites associated with aging in the population as a whole.

We explored biological alterations that might have led to the observed changes in metabolite concentrations and could link our observations to common aging theories. As we had excluded participants with major metabolic diseases, we assume that our findings may be representative for the general aging population.

Phospholipids are major components of cell membranes. Changes in the cellular membrane that include a G protein coupled receptor may relate to aging, which affects at least G protein activity, cell morphology, and cell homeostasis (Naru *et al.*, 2008).A common feature of cellular senescence is an increased cell surface. Naru *et al.* (2008) found that senescent cells had a higher uptake of PC species with long-chain fatty acid residues. Additionally, the number of special lipid rafts, termed caveolae, increased with senescence. Presumably, PC species were integrated into caveolae, which contain caveolin-1 as a crucial component capable of cell cycle suppression at the G0/G1 phase. Thus, altered consumption of phospholipids because of specific membranaceous demands could be associated with senescence, possibly involving the scavenger receptor SR-BI as the major mediator of selective phospholipid (PC, SM, and PE) uptake from particles as HDL and LDL (Engelmann & Wiedmann, 2010),

Sphingomyelins are further important components of cell membranes, especially neuronal cells, as they influence membrane fluidity and can promote signal transduction. While both Ichi *et al.* (2009) and Corre *et al.* (2010) reported the connection between oxidative stress and sphingomyelin metabolism, in their studies, oxidative stress was shown to accelerate degradation of sphingomyelins to ceramide, which would be inconsistent with our observation that elderly subjects had elevated SM levels. However, another study identified cells that had adapted to chronic oxidative stress by altering sphingomyelin metabolism and making major changes in membrane composition, leading to stabilization (Clement *et al.*, 2009). The observed SM level elevations might indicate that aging humans have effective mechanisms to protect cells from oxidative stress, accompanied by changes in SM metabolism and incorporation of SM into cell membranes.

Elevated serum levels of acylcarnitines could be due to different underlying causes. AC species are found as a consequence of incompletely oxidized fatty acids because of an excess of beta oxidation capacity and related pathways (i.e. higher rate of substrate use than energy demand, with accumulated acyl-CoA converted to AC that then exits cells and tissues), but also as a consequence of oxidative stress (Noland *et al.*, 2009). Acylcarnitines in turn show up in the blood. A recent rodent study suggested increased AC levels in blood could be healthy (Noland *et al.*, 2009). In the study, Wistar rats maintained on a high-fat diet exhibited diminished carnitine and increased AC levels in skeletal muscle cells because of perturbations in mitochondrial fuel utilization, for example, they had incomplete fatty acid oxidation. Supplemented carnitine led to AC efflux, which in turn showed up in blood accompanied by improved metabolism and glucose tolerance. Thus, the carnitine shuttle system is considered to be a prominent factor in maintaining mitochondrial performance and glucose homeostasis.

The observed higher levels of AC with advanced age might indicate that the aging processes counteract oxidative damage from the mitochondria via the carnitine-acylcarnitine shuttle.

The metabolic fate of the amino acid histidine has two possible routes. Its presence in blood at higher ages could indicate an advanced tissue demand. The first histidine-consuming pathway is its metabolism to the biogenic amine histamine by a decarboxylation step. Histamine is involved in local immune responses and can act as neurotransmitter. However, to our knowledge, an association between histidine or histamine and immunosenescence has not been reported so far. The second metabolic pathway that consumes histidine produces carnosine. This dipeptide from beta-alanine and L-histidine is found in virtually all tissues, in particular in skeletal muscle cells and different brain cells (Derave *et al.*, 2010). Owing to its antioxidant characteristics, carnosine is considered to be a natural anti-aging substance capable of suppressing oxidative damage, glycation of proteins, and scavenging toxic age-related molecules (Hipkiss, 2010). For instance, carnosine was shown to capture lipoxidation products and prevent protein cross-linking (Zhu *et al.*, 2009). Assuming that histidine levels are lower owing to its consumption by carnosine biosynthesis with advancing age, decreased histidine in our study might reflect a response to oxidative stress.

Aging is understood as a continuous and dynamic remodeling process of the human organism accompanied by numerous losses and gains on different levels, including intermediate metabolism and cell function (Barbieri *et al.*, 2009). Our results of a large number of age-associated metabolites and the replicated set of 11 most representative ones might reflect these processes and allow us to draw a more integral picture of the aging organism. Overall, we observed age-dependent differences in PC, SM, AC, and AA levels that might be linked to altered cell membrane composition, mitochondrial metabolism, and counteracting oxidative stress.

While the present study was large in participant number and metabolically well characterized, it should be noted that metabolite profiles were related to chronological not physiological age and that only a cross-sectional study design was used. It is also not yet known whether the changes in the 11 confirmed metabolites represent neutral changes with age, changes causal to physiological aspects of aging or beneficial responses to damaging agents. Further research on the metabolic surrounding of the metabolite set could eventually lead to early determination of a person's potential for healthy aging at beginning of the remodeling process.

In summary, we identified a set of 11 significantly associated and replicated markers for age in women using the German KORA and TwinsUK studies. Literature gives indications that these markers might be linked to aging processes such as oxidative stress, alterations in cell morphology, beta oxidation capacity, and vascular function. This study shows the power of metabolomics to better understand the phenotype of aging in the human population and to link this knowledge in functional studies to aging pathways.

Experimental procedures

Sample Source

KORA is a population-based research platform with subsequent follow-up studies in the fields of epidemiology, health economics, and health-care research (Holle *et al.*, 2005; Wichmann *et al.*, 2005). It is based on interviews and medical and laboratory examinations, as well as biological samples. Four surveys were conducted with 18 079 participants who live in the city of Augsburg (Southern Germany) and 16 surrounding towns and villages. KORA S4 consists of representative samples from 4261 individuals who live in Augsburg, who were examined during 1999–2001. During the years 2006–2008, 3080 participants took part in a follow-up (KORA F4) survey of the one conducted 7 years ago. For all studies, we obtained written consent from participants and approval from the ethics committee of the Bavarian medical association.

To avoid the potential influences from type 2 diabetes, hypertension and obesity, we excluded a total of 918 subjects from KORA F4 for subsequent analyses, resulting in 2162 subjects aged 32–81 years. Among the excluded were 20 experimental failures, 18 nonfasting subjects, 332 type 2 diabetics, 80 subjects without fasting glucose or 2 h glucose measurement, 77 subjects with systolic blood pressure > 160 mmHg, and 153 subjects with BMI > 35 kg m^{-2}. Further removals followed during statistical analyses (n = 239, see statistical section below).

Sampling

Blood was drawn into serum gel tubes in the morning between 8:00 and 10:30 am after a fasting period of at least 8 h. Tubes were gently inverted twice, followed by 30 min resting at room temperature to obtain complete coagulation. For serum collection, blood was centrifuged at 2750 ***g*** at 15 °C for 10 min. Serum was frozen at –80 °C until execution of metabolic analyses.

Metabolite measurements

The targeted metabolomics approach was based on measurements with the Absolute*IDQ*^{TM} p150 kit (BIOCRATES Life Sciences AG, Innsbruck, Austria), allowing simultaneous quantification of 163 metabolites. The method conforms with FDA-Guidelines 'Guidance for Industry – Bioanalytical Method Validation (May 2001)', which implies proof of reproducibility within a given error range. The assay procedures and nomenclature have been described previously in detail (Zhai *et al.*, 2010; Mittelstrass *et al.*, 2011; Römisch-Margl *et al.*, 2011; Yu *et al.*, 2011). Metabolite measurements were adjusted for batch effect as we have described previously (Mittelstrass *et al.*, 2011).

To ensure data quality, each metabolite had to meet the three same criteria we used before (Mittelstrass *et al.*, 2011): (i) average value of the coefficient of variance (CV) for the metabolite in the three QCs should be smaller than 25%; (ii) 90% of all measured sample concentrations for the metabolite should be above the limit of detection (LOD); and (iii) the correlation coefficient between two duplicate measurements of the metabolite in 144 re-measured samples should be above 0.5. In total, 131 metabolites passed the three quality controls, and the final metabolomics dataset contained the sum of hexoses (H1), 14 amino acids (AA), 24 acylcarnitines (AC), 13 sphingomyelins (SMs), 34 diacyl phosphatidylcholines (PC aa), 37 acyl-alkyl PCs (PC ae), and eight lyso PCs. Table S1 (Supporting information) summarizes the characteristics of all 163 metabolites measured in KORA F4.

Statistics

Removal of outliers

To detect outliers, concentrations obtained for the remaining 131 metabolites were first scaled to have a mean of zero and a standard deviation (SD) of one and projected onto the unit sphere, and Mahalanobis distances for each individual were then calculated using the Robust principal components algorithm (Filzmoser *et al.*, 2008) and were calculated separately for men and women. For each group, the mean Mahalanobis dis-

tance plus three times variance were defined as the cut-off. About 239 individuals whose distances were greater than these cut offs were identified as outliers.

Residuals of metabolite concentrations

To avoid the influence of BMI when plotting the concentration of metabolites, we used the residuals from the linear regression model rather than use the absolute concentration. Log-transformed metabolite concentrations were treated as dependent, and the BMI as an independent, variable in the linear regression model, and the residuals from the model were used as the residuals for each metabolite concentrations. Regressions were performed for women and men separately.

Heat map

Correlations were calculated for each metabolite pairs, and values were displayed using heat map. For participants with the same age, mean values were calculated for residuals from linear regression against BMI for each metabolite. These values were then scaled to a mean of 0 and a standard deviation of 1, to display a heat map that showed the changes of the metabolite concentrations with the increase of age; the color change in the heat map represents the concentration deviation from the mean value.

Linear regression analysis

Linear regression was applied to model the relationship between age and the concentration of each metabolite, with BMI used as covariate. Metabolite concentrations were log-transformed to achieve normality. Regressions were done for men and women separately. To handle false discovery rates from multiple comparisons, the cut point for significance was calculated according to the Bonferroni correction, at a level of 3.8×10^{-4} (for a total use of 131 metabolites at 5% level).

Smoother plots

Smoother plots were drawn for each metabolite of the set of metabolites with the R function 'qplot' (package 'ggplot2') using the options geom = smooth, method = loess, and span = 0.5, producing smoother plots with locally weighted regression (*loess*) applying a smoothing span of 0.5, which results in medium smoothing. The method computes outlier robust locally weighted regression fitted values by fitting local polynomials, using weights and results in the (loess) curve as shown in our smoother plots. Further information about the method has been previously published (Cleveland & Devlin 1988). For better visualization, plots were truncated to observations between the first and 99th percentiles.

Criteria for metabolite selection

Multivariate linear regression, random forest, and a stepwise selection of linear regression methods were applied: metabolites were chosen if they both were significant in linear regression for every single metabolite, adjusting for BMI and also for the top 30 most important variables in random forest method, in which both the 131 metabolites and BMI were severed as variables. The chosen metabolites with BMI as co-variable were further selected based on a stepwise selection of multi-variables linear regression according to the Akaike information criterion (AIC) value.

All calculations were done with R statistical platform, version 2.12 (http://www.r-project.org/).

Replication

TwinsUK is a UK-wide twin registry sample of 11 000 adults, founded in 1993 with the aim to explore the genetic epidemiology of common adult diseases (Spector & Williams, 2006). The cohort has been tested to be generalizable to UK population singletons, with no population stratification for a wide variety of musculoskeletal, CVD, or metabolic traits (Andrew *et al.*, 2001). Over 7000 twins have attended detailed clinical examinations, with a wide range of phenotypes, over the last 18 years. Blood samples were taken after at least 6 h fasting at each visit. Samples were immediately inverted three times, followed by 40 min resting at 4 °C, to obtain complete coagulation. Samples were then centrifuged for 10 min at 1439 ***g***. Serum was removed from the centrifuged tubes as the top, yellow, clear layer of liquid. Aliquot in 4×1.5 mL skirted microcentrifuge tubes was then stored in a –45 °C freezer until sampling. About 1237 twins were selected for the targeted metabolomic profiling for either osteoarthritis or genetic studies. Metabolites were measured using the same metabolomics platform (Biocrates metabolomic assay kit; BIOCRATES Life Sciences AG, Innsbruck, Austria) and following an identical protocol as for the KORA study, at the Genome Analysis Centre of the Helmholtz Centre Munich (Zhai *et al.*, 2010). To replicate the KORA F4 study, data on the 13 age-related metabolites identified in the KORA F4 aging study were retrieved, and the association between age and these 13 serum metabolites were analyzed by robust regression modeling that takes into account twin relatedness.

Of the data on a total of 1237 individuals with metabolomic data available, 44 men were excluded. Following the KORA F4 study's exclusion criteria, we further excluded 64 individuals with systolic blood pressure > 160 mmHg, 14 individuals with type 2 diabetes, 45 individuals with BMI > 35 kg m^{-2}, and 328 individuals for whom no data for blood pressure, fasting serum glucose levels, or diabetes diagnosis data were available. A total of 742 female individuals were included in the final analysis.

Meta-analysis

For the meta-analyses of KORA and TwinsUK women, a fixed effects model was used.

Acknowledgments

We express our appreciation to all KORA and TwinsUK study participants for donating their blood and time. We thank the field staff in Augsburg conducting the KORA studies. The KORA group consisted of H.E. Wichmann (speaker), A. Peters, C. Meisinger, T. Illig, R. Holle, and J. John as well as their co-workers, who were responsible for the design and conduction of the studies. We thank the staff from the Institute of Epidemiology and the Helmholtz ZentrumMünchen, and the Genome Analysis Center Metabolomics Platform who helped in the sample logistics, data and straw collection, and metabolomic measurements, and especially J. Scarpa, K. Sckell, B. Hochstrat, F. Scharl, N. Lindemann, H. Chavez, and A. Sabunchi. We thank M. Sattler and V. A. Raker for comments and suggestions.

Authors' contribution

GZ, KS, JA, TDS, TI, and RWS designed the studies; CP, WRM, and RWS performed the metabolomic measurements; ZY, GZ, PS, YH, TX, JA, and RWS analyzed the data; PS, MS, ET, CG, JH, AP, JA, TI, and RWS interpreted the data; PS, GZ, ZY, TDS, and RWS wrote the manuscript; and all authors commented on and corrected the manuscript.

Financial competing interests

Part of this project was supported by the German Federal Ministry of Education and Research (BMBF) by funding the Project 'SysMBo: Systems Biology of Metabotypes' (FKZ: 0315494A), and the project 'Metabolomics of ageing' (FKZ: 01DO12030), the German Center for Diabetes (DZD) and the NGFN. The TwinsUK study was funded by the Wellcome Trust, European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F2-2008-201865-GEFOS and (FP7/2007-2013), and the FP-5 GenomEUtwin Project (QLG2-CT-2002-01254). The study also receives support from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. TDS is an NIHR senior Investigator. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Nonfinancial

None.

References

Andrew T, Hart DJ, Snieder H, de Lange M, Spector TD, MacGregor AJ (2001) Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res.* **4**, 464–477.

Barbieri M, Boccardi V, Papa M, Paolisso G (2009) Metabolic journey to healthy longevity. *Horm. Res.* **71**(Suppl 1), 24–27.

Butler RN, Miller RA, Perry D, Carnes BA, Williams TF, Cassel C, Brody J, Bernard MA, Partridge L, Kirkwood T, Martin GM, Olshansky SJ (2008) New model of health promotion and disease prevention for the 21st century. *BMJ* **337**, a399.

Clement AB, Gamerdinger M, Tamboli IY, Lutjohann D, Walter J, Greeve I, Gimpl G, Behl C (2009) Adaptation of neuronal cells to chronic oxidative stress is associated with altered cholesterol and sphingolipid homeostasis and lysosomal function. *J. Neurochem.* **111**, 669–682.

Cleveland WS, Devlin SJ (1988) Locally-weighted regression: an approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* **83**, 596–610.

Corre I, Niaudet C, Paris F (2010) Plasma membrane signaling induced by ionizing radiation. *Mutat. Res.* **704**, 61–67.

Deelen J, Beekman M, Uh HW, Helmer Q, Kuningas M, Christiansen L, Kremer D, van der Breggen R, Suchiman HE, Lakenberg N, van den Akker EB,Passtoors WM, Tiemeier H, van Heemst D, de Craen AJ, Rivadeneira F, de Geus EJ, Perola M, van der Ouderaa FJ, Gunn DA, Boomsma DI, Uitterlinden AG, Christensen K, van Duijn CM, Heijmans BT, Houwing-Duistermaat JJ, Westendorp RG, Slagboom PE (2011) Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* **10**, 686–698.

Dennis JW, Nabi IR, Demetriou M (2009) Metabolism, cell surface organization, and disease. *Cell* **139**, 1229–1241.

Derave W, Everaert I, Beeckman S, Baguet A (2010) Muscle carnosine metabolism and beta-alanine supplementation in relation to exercise and training. *Sports Med.* **40**, 247–263.

Engelmann B, Wiedmann MK (2010) Cellular phospholipid uptake: flexible paths to coregulate the functions of intracellular lipids. *Biochim. Biophys. Acta* **1801**, 609–616.

Feltes BC, de Faria Poloni J, Bonatto D (2011) The developmental aging and origins of health and disease hypotheses explained by different protein networks. *Biogerontology* **12**, 293–308.

Ferrucci L, Giallauria F, Guralnik JM (2008) Epidemiology of aging. *Radiol. Clin. North Am.* **46**, 643–652, v.

Filzmoser P, Maronna R, Werner M (2008). Outlier identification in high dimension. *Comput. Stat. Data Anal.*, **52**, 1694ß1711.

Franceschi C, Capri M, Monti D, Giunta S, Olivieri F, Sevini F, Panourgia MP, Invidia L, Celani L, Scurti M, Cevenini E, Castellani GC, Salvioli S (2007) Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans. *Mech. Ageing Dev.* **128**, 92–105.

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 (2012) Schizophrenia shows a unique metabolomics signature in plasma. *Transl. Psychiatry* (in press).

Hipkiss AR (2010) Aging, proteotoxicity, mitochondria, glycation, NAD and carnosine: possible inter-relationships and resolution of the oxygen paradox. *Front. Aging Neurosci.* **2**, 10.

Holle R, Happich M, Lowel H, Wichmann HE (2005) KORA – a research platform for population based health research. *Gesundheitswesen* **67**, S19–S25.

Ichi I, Kamikawa C, Nakagawa T, Kobayashi K, Kataoka R, Nagata E, Kitamura Y, Nakazaki C, Matsura T, Kojo S (2009) Neutral sphingomyelinase-induced ceramide accumulation by oxidative stress during carbon tetrachloride intoxication. *Toxicology* **261**, 33–40.

Jiang H, Schiffer E, Song Z, Wang J, Zurbig P, Thedieck K, Moes S, Bantel H, Saal N, Jantos J, Brecht M, Jeno P, Hall MN, Hager K, Manns MP, Hecker H, Ganser A, Dohner K, Bartke A, Meissner C, Mischak H, Ju Z, Rudolph KL (2008) Proteins induced by telomere dysfunction and DNA damage represent biomarkers of human aging and disease. *Proc. Natl. Acad. Sci. USA* **105**, 11299–11304.

Karasik D, Demissie S, Cupples LA, Kiel DP (2005) Disentangling the genetic determinants of human aging: biological age as an alternative to the use of survival measures. *J. Gerontol. A Biol. Sci. Med. Sci.* **60**, 574–587.

Kerber RA, O'Brien E, Cawthon RM (2009) Gene expression profiles associated with aging and mortality in humans. *Aging Cell* **8**, 239–250.

Lawton KA, Berger A, Mitchell M, Milgram KE, Evans AM, Guo L, Hanson RW, Kalhan SC, Ryals JA, Milburn MV (2008) Analysis of the adult human plasma metabolome. *Pharmacogenomics* **9**, 383–397.

Mittelstrass K, Ried JS, Yu Z, 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 (2011) Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet.* **7**, e1002215.

Moayyeri A, Hammond CJ, Valdes AM, Spector TD (2012) Cohort profile: TwinsUK and Healthy Ageing Twin Study. *Int. J. Epidemiol.*

Naru E, Takanezawa Y, Kobayashi M, Misaki Y, Kaji K, Arakane K (2008) Increased levels of a particular phosphatidylcholine species in senescent human dermal fibroblasts in vitro. *Hum. Cell* **21**, 70–78.

Nevedomskaya E, Meissner A, Goraler S, de Waard M, Ridwan Y, Zondag G, van der Pluijm I, Deelder AM, Mayboroda OA (2010) Metabolic profiling of accelerated aging ERCC1 d/- mice. *J. Proteome Res.* **9**, 3680–3687.

Nikkila J, Sysi-Aho M, Ermolov A, Seppanen-Laakso T, Simell O, Kaski S, Oresic M (2008) Gender-dependent progression of systemic metabolic states in early childhood. *Mol. Syst. Biol.* **4**, 197.

Noland RC, Koves TR, Seiler SE, Lum H, Lust RM, Ilkayeva O, Stevens RD, Hegardt FG, Muoio DM (2009) Carnitine insufficiency caused by aging and overnutrition compromises mitochondrial performance and metabolic control. *J. Biol. Chem.* **284**, 22840–22852.

Orešič M (2009) Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction. *Nutr. Metab. Cardiovasc. Dis.* **19**, 816–824.

Partridge L, Thornton J, Bates G (2011) The new science of ageing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 6–8.

Perheentupa A, Huhtaniemi I (2009) Aging of the human ovary and testis. *Mol. Cell. Endocrinol.* **299**, 2–13.

Piper MD, Bartke A (2008) Diet and aging. *Cell Metab.* **8**, 99–104.

Psychogios N, Hau DD, Peng J, Guo AC, 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 TL, Smith SR, Bamforth F, Greiner R, McManus B, Newman JW, Goodfriend T, Wishart DS (2011) The human serum metabolome. *PLoS ONE* **6**, e16957.

Römisch-Margl W, Prehn C, Bogumil R, Roehring C, Suhre K, Adamski J (2011) Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics* **8**, 133–142.

Simm A, Nass N, Bartling B, Hofmann B, Silber RE, Navarrete Santos A (2008) Potential biomarkers of ageing. *Biol. Chem.* **389**, 257–265.

Spector TD, Williams FM (2006) The UK Adult Twin Registry (TwinsUK). *Twin Res. Hum. Genet.* **9**, 899–906.

Szibor M, Holtz J (2003) Mitochondrial ageing. *Basic Res. Cardiol.* **98**, 210–218.

Toth MJ, Tchernof A (2000) Lipid metabolism in the elderly. *Eur. J. Clin. Nutr.* **54**(Suppl 3), S121–S125.

Wang Y, Lawler D, Larson B, Ramadan Z, Kochhar S, Holmes E, Nicholson JK (2007) Metabonomic investigations of aging and caloric restriction in a life-long dog study. *J. Proteome Res.* **6**, 1846–1854.

Wichmann HE, Gieger C, Illig T (2005) KORA-gen – resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67**(Suppl 1), S26–S30.

Wijsman CA, Rozing MP, Streefland TC, le Cessie S, Mooijaart SP, Slagboom PE, Westendorp RG, Pijl H, van Heemst D (2011) Familial longevity is marked by enhanced insulin sensitivity. *Aging Cell* **10**, 114–121.

Williams RE, Lenz EM, Lowden JS, Rantalainen M, Wilson ID (2005) The metabonomics of aging and development in the rat: an investigation into the effect of age on the profile of endogenous metabolites in the urine of male rats using 1H NMR and HPLC-TOF MS. *Mol. BioSyst.* **1**, 166–175.

Yan SF, D'Agati V, Schmidt AM, Ramasamy R (2007) Receptor for advanced glycation endproducts (RAGE): a formidable force in the pathogenesis of the cardiovascular complications of diabetes & aging. *Curr. Mol. Med.* **7**, 699–710.

Yu Z, Kastenmuller G, He Y, Belcredi P, Moller 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 (2011) Differences between human plasma and serum metabolite profiles. *PLoS ONE* **6**, e21230.

Zhai G, Wang-Sattler R, Hart DJ, Arden NK, Hakim AJ, Illig T, Spector TD (2010) Serum branched-chain amino acid to histidine ratio: a novel metabolomic biomarker of knee osteoarthritis. *Ann. Rheum. Dis.* **69**, 1227–1231.

Zhu X, Gallogly MM, Mieyal JJ, Anderson VE, Sayre LM (2009) Covalent cross-linking of glutathione and carnosine to proteins by 4-oxo-2-nonenal. *Chem. Res. Toxicol.* **22**, 1050–1059.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site:

Fig. S1 Correlation matrix for women in KORA F4.

Fig. S2 Heat map of the fold standard deviation changes between ages, and clustering of these changes over all ages, in 1124 males from KORA F4.

Fig. S3 Concentration trends of 13 potential biomarkers in women from KORA F4.

Fig. S4 Concentration trends of 12 selected metabolites in males from KORA F4.

Fig. S5 Concentration trends of the set of 13 metabolites in women from the TwinsUK study.

Table S1 Characteristics of 163 metabolites in KORA F4.

Table S2 Potential biomarkers for age in men from KORA F4.

Xu et al. BMC Medicine 2013, 11:60
http://www.biomedcentral.com/1741-7015/11/60

BMC Medicine

RESEARCH ARTICLE Open Access

Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study

Tao Xu[1], Christina Holzapfel[1,2], Xiao Dong[3,4], Erik Bader[1], Zhonghao Yu[1], Cornelia Prehn[5], Katrin Perstorfer[1], Marta Jaremek[1], Werner Roemisch-Margl[6], Wolfgang Rathmann[7], Yixue Li[3], H -Erich Wichmann[8,9], Henri Wallaschofski[10], Karl H Ladwig[11,12], Fabian Theis[6,13], Karsten Suhre[6,14,15], Jerzy Adamski[5,16], Thomas Illig[1,17], Annette Peters[1,11,18] and Rui Wang-Sattler[1*]

Abstract

Background: Metabolomics helps to identify links between environmental exposures and intermediate biomarkers of disturbed pathways. We previously reported variations in phosphatidylcholines in male smokers compared with non-smokers in a cross-sectional pilot study with a small sample size, but knowledge of the reversibility of smoking effects on metabolite profiles is limited. Here, we extend our metabolomics study with a large prospective study including female smokers and quitters.

Methods: Using targeted metabolomics approach, we quantified 140 metabolite concentrations for 1,241 fasting serum samples in the population-based Cooperative Health Research in the Region of Augsburg (KORA) human cohort at two time points: baseline survey conducted between 1999 and 2001 and follow-up after seven years. Metabolite profiles were compared among groups of current smokers, former smokers and never smokers, and were further assessed for their reversibility after smoking cessation. Changes in metabolite concentrations from baseline to the follow-up were investigated in a longitudinal analysis comparing current smokers, never smokers and smoking quitters, who were current smokers at baseline but former smokers by the time of follow-up. In addition, we constructed protein-metabolite networks with smoking-related genes and metabolites.

Results: We identified 21 smoking-related metabolites in the baseline investigation (18 in men and six in women, with three overlaps) enriched in amino acid and lipid pathways, which were significantly different between current smokers and never smokers. Moreover, 19 out of the 21 metabolites were found to be reversible in former smokers. In the follow-up study, 13 reversible metabolites in men were measured, of which 10 were confirmed to be reversible in male quitters. Protein-metabolite networks are proposed to explain the consistent reversibility of smoking effects on metabolites.

Conclusions: We showed that smoking-related changes in human serum metabolites are reversible after smoking cessation, consistent with the known cardiovascular risk reduction. The metabolites identified may serve as potential biomarkers to evaluate the status of smoking cessation and characterize smoking-related diseases.

Keywords: metabolic network, metabolomics, molecular epidemiology, smoking, smoking cessation

* Correspondence: rui.wang-sattler@helmholtz-muenchen.de
[1]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Ingolstädter Landstraße 1, Neuherberg, 85764, Germany
Full list of author information is available at the end of the article

Background

Smoking is responsible for 90% of all lung cancers, accounts for 25% of cancer deaths worldwide [1-3] and is a significant risk factor for cardiovascular disease (CVD) [4-7]. The benefits of smoking cessation are remarkable. Risk of CVD is reduced in former smokers (FS) compared with current smokers (CS) [8-10]; mortality and future cardiac events both decline in FS [11,12]. Nevertheless, for cancers, especially for adenocarcinoma, the risk remains high in FS compared with never smokers (NS) [13,14]. Studies have made attempts to find the molecular basis for the influence of smoking and smoking cessation on cardiovascular risks. For instance, smoking is associated with the increase of several CVD-related inflammatory markers, for example, c-reactive protein and fibrinogen [15-17], and smoking cessation could largely reduce the level of these markers [18]. However, there is also evidence that other molecular changes associated with smoking are permanent, for example, loss of heterozygosity and hypermethylation in the promoter regions of cancer-related genes [19-23].

The metabolomics approach provides a functional readout of activities located downstream of the gene expression level that are more closely related to the physiological status [24] and, thus, may be particularly useful for the study of environmental influences, namely the 'exposome' [25]. Studying a strong environmental factor, for example a lifestyle-related exposure to smoking, may be considered a very powerful approach for understanding the links between environmental exposure and the metabolome. In human lung epithelial cells, it has been shown that metabolite concentration changes in various pathways, for example, the urea cycle and polyamine metabolism and lipid metabolism under smoke exposure [26]. In a pilot study with 283 male participants from the Cooperative Research in the Region of Augsburg (KORA) F3 in Germany, we have shown that levels of diacyl-phosphatidylcholines (PCs) were higher in 28 CS compared with 101 NS, except for acyl-alkyl-PCs [1]. The reduced ratios of acyl-alkyl-to diacyl-PCs in CS may be regulated by the enzyme alkyl-dihydroxyacetone phosphate in both ether lipid and glycerophospholipid pathways [1]. However, little has been reported about the reversibility of the metabolite profile upon smoking cessation, which is important for comprehensive understanding of smoking effects. It is also known that metabolite profile is different between men and women [25], but whether lifestyle factors such as smoking may induce different metabolite patterns in men and women is still unknown.

In this study, we analyzed the association between smoking and the concentration of metabolites in 1,241 serum samples from the KORA baseline survey 4 (S4) and follow-up (F4) study, aiming to extend the knowledge of smoking-associated metabolites beyond our pilot study by including female CS at two time points over seven years, to investigate whether smoking-associated changes in metabolite profile are reversible after smoking cessation, and to provide insights into the pathophysiological consequences of smoking in protein-metabolite networks.

Methods

Ethics statement

Written informed consent was obtained from KORA S4 and F4 participants. The KORA study was approved by the ethics committee of the Bavarian Medical Association in Munich, Germany.

Study population

The KORA surveys are population-based studies conducted in the Region of Augsburg in Germany [27,28]. Four surveys were conducted with 18,079 participants recruited from 1984 to 2001. The S4 consists of 4,261 individuals (25 to 74 years old) examined from 1999 to 2001. From 2006 to 2008, 3,080 participants (with an age range of 32 to 81 years) took part in the F4 survey. Each participant completed a lifestyle questionnaire providing information on a number of parameters including smoking status (current, former, never). Serum samples for metabolomics analysis were collected in parallel in the KORA S4 and F4 survey as described elsewhere [29-31].

For metabolite profiles, serum samples from 1,614 people aged 55 to 74 years old were available [29]. Participants with non-fasting status (N = 216) or missing values (N = 22) were excluded from the analysis. We further excluded 145 people in KORA S4 and 116 people in the longitudinal data of KORA S4 → F4, whose spouses were CS, to rule out passive smoking effects. Furthermore, metabolite concentrations of serum samples from 1,036 participants were measured in both KORA S4 and F4.

Metabolite measurements

Liquid handling of serum samples (10 μl) was performed with Hamilton star robot (Hamilton Bonaduz AG, Bonaduz, Switzerland) and prepared for quantification using the Absolute*IDQ* P180 and P150 kits (BIOCRATES Life Science AG, Innsbruck, Austria) for the KORA S4 and F4 surveys, respectively. This allowed simultaneous quantification of 188 or 163 metabolites using liquid chromatography and flow injection analysis mass spectrometry as described previously [32,33]. The complete analytical process was monitored by quality control steps, reference samples and the MetIQ software package, which is an integral part of the Absolute *IDQ* kit.

Because the two datasets were generated by different platforms, different quality control processes were introduced. The metabolite data quality control procedure for the KORA S4 samples was described in our recently

published work [29]. There were 140 metabolites that passed the two quality controls: one hexose, 21 amino acids, eight biogenic amines, 21 acylcarnitines, 13 sphingomyelins (SMs), eight lysoPCs, 33 diacyl-PCs (PC aa Cx:y) and 35 acyl-alkyl-PCs (PC ae Cx:y). Lipid side chain composition is abbreviated as Cx:y, where × denotes the number of carbons in the side chain and y the number of double bonds. The precise position of the double bonds and the distribution of the carbon atoms in different fatty acid side chains cannot be determined with this technology. Concentrations of all analyzed metabolites are reported in μmol/L (μM). The data cleaning procedure for the KORA F4 samples has previously been described in detail [24,30]. In total, 121 metabolites were measured in both S4 and F4, and used for the prospective study.

Statistical analysis

Differences in population characteristics (CS, FS and NS) were tested by a two-tailed student's *t*-test. The metabolite concentrations were log transformed for normalization. We tested cross-sectional association of each metabolite with smoking using logistic regression models adjusted for age, body mass index (BMI) and alcohol consumption (see Figure 1). To correct for multiple testing, false discovery rate (FDR) was calculated using the Benjamini-Hochberg method [34] and the cut-off for statistical significance was set at FDR <0.05.

Linear regression models were used to investigate whether smoking intensities measured in pack years and cessation time are associated with metabolite concentrations. In the case of CS, the years of smoking were calculated as the time period from starting smoking until the start of the survey. Pack year was calculated as the number of cigarettes per day multiplied by smoking duration and divided by 20 [35]. Cessation time (in years) was calculated according to the questionnaire. The models contained the log-transformed metabolite concentrations as the dependent variable and the smoking intensities as the explanatory variable, with age, BMI and alcohol consumption as covariates. Every unit change of one covariate corresponds to a relative change of the metabolite concentration by Δ (%):

$$\Delta = (\exp(\beta_i) - 1) \times 100\%$$

where β_i indicates the estimate of *ith* covariate in the model.

To assess the role of smoking cessation for the quitters, who were CS at S4 but FS at F4, we fitted the linear mixed models to the longitudinal data of KORA S4 → F4. The models contained the fixed effect of smoking status (CS, FS and NS), age, BMI and alcohol consumption with a random effect assigned to each participant. All calculations were performed in R (version 2.14.1).

Network and pathway analysis

We retrieved protein-protein interactions from the databases of the Search Tool for the Retrieval of Interacting Genes/Proteins [36] and the relationships between enzymes and metabolites from the Human Metabolome Database [37] to construct protein-metabolite networks containing links between metabolites, enzymes and smoking-related genes. Genes and metabolites were connected allowing for at most one intermediate enzyme by Dijkstra's algorithm [38], and optimized by eliminating edges with Search Tool for the Retrieval of Interacting Genes/Proteins scores less than 0.7. Each edge in the networks was manually checked. We have implemented this method in our previous studies [29,39]. The analysis was performed using the R package igraph [40]. The network was visualized using Cytoscape [41]. Pathway analysis was performed by MetaboAnalyst [42].

Results

Characteristics of participants of the cross-sectional KORA S4

Participants were divided into three groups according to their self-reported smoking status. Population characteristics are shown in Table 1. On average, CS were two to three years younger and had a lower BMI than FS and NS. Male CS showed higher alcohol consumption than male NS, but there was no significant difference observed in women. Furthermore, the statistics showed differences in lifestyle factors between men and women. Alcohol consumption was higher in men than women ($P = 1.5e^{-11}$ (CS); $P = 2.2e^{-18}$ (FS); $P = 9.5e^{-17}$ (NS)), and smoking intensity (in pack years) was higher in male than in female CS ($P = 6.0e^{-6}$).

Metabolomic differences between current, former and never smokers

We identified 18 metabolites in men and six in women that were significantly different (FDR <0.05) between CS and NS. Three metabolites (PC ae C34:3, PC aa C36:1 and glutamate) were identified in both men and women showing the same pattern of variation (higher or lower) (Table 2). Compared with FS and NS, in male CS the concentrations of four unsaturated diacyl-PCs (PC aa C34:1, PC aa C36:1, PC aa C38:3 and PC aa C40:4) and five amino acids (arginine, aspartate, glutamate, ornithine and serine) were higher, whereas three saturated diacyl-PCs, one lysoPC and four acyl-alkyl-PCs, as well as kynurenine showed lower concentrations. In female CS, we found higher levels of carnitine and PC aa C32:1, and a lower level of hydroxysphingomyeline (SM (OH)) C22:2.

Among the 21 smoking-related metabolites (18 in men and six in women), 19 were found to be reversible (that is, significant difference between FS and CS but without

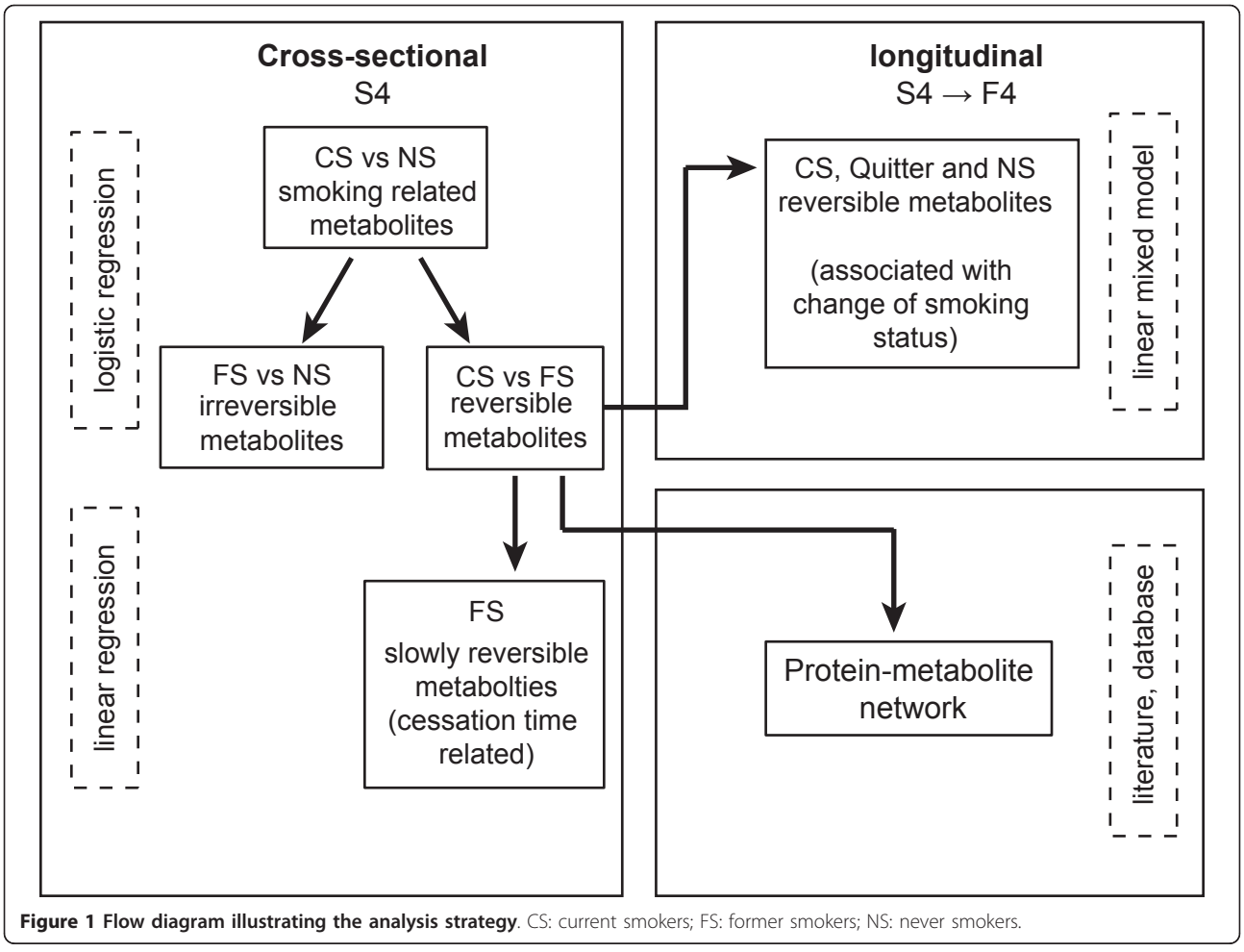

Figure 1 Flow diagram illustrating the analysis strategy. CS: current smokers; FS: former smokers; NS: never smokers.

Table 1 Characteristics of cross-sectional KORA S4.

	Current smoker	Former smoker	Never smoker	*P*[a]	
				Current versus former smoker	Current versus never smoker
Male (N = 646)					
N (%)	125 (19.3%)	321 (49.7%)	200 (31.0%)		
Age (years)	62.2 ±5.3	65.3 ± 5.3	64.1 ± 5.6	$7.9e^{-08}$	$3.0e^{-03}$
BMI (kg/m^2)	27.0 ±3.6	28.9 ±3.6	27.8 ±3.4	$1.5e^{-06}$	$6.5e^{-02}$
Alcohol consumption (g/day)	27.5 ±29.0	24.1 ±24.3	20.5 ±21.3	0.25	0.02
Pack years[b]	39.3 ±22.4				
Quit time[c] (years)		23.6 ±12.6			
Female (N = 595)					
N (%)	70 (11.8%)	130 (21.8%)	395 (66.4%)		
Age (years)	61.3 ±5.2	64.0 ±5.2	64.6 ±5.3	$7.5e^{-04}$	$5.9e^{-06}$
BMI (kg/m^2)	27.2 ±4.5	28.7 ±5.0	28.5 ±4.6	0.029	0.02
Alcohol consumption (g/day)	6.5 ±10.9	10.0 ±12.8	7.5 ±11.1	0.042	0.48
Pack years[b]	25.8 ±15.3				
Quit time[c] (years)		20.9 ±13.1			

The study characteristics of KORA S4 are shown separately for current, former and never smokers. Values are shown as mean ±SD when appropriate. [a]*P*-values are calculated by student's *t*-test; [b]calculated as the number of cigarettes consumed per day × years of smoking/20; [c] the time till the survey is conducted since the person has stopped smoking. BMI: body mass index.

Table 2 Smoking-related metabolites in KORA S4.

Metabolites	CS versus NS		CS versus FS		FS versus NS	
	Odds ratio (95% CI)	*P*	Odds ratio (95% CI)	*P*	Odds ratio (95% CI)	*P*
Men	**(125 versus 200)**		**(125 versus 321)**		**(321 versus 200)**	
Arginine	1.7 (1.3, 2.2)	2.6e^{-05a}	1.3 (1.0, 1.6)	0.03[a]	1.2 (1.0, 1.5)	0.03
Aspartate	1.6 (1.2, 2.0)	2.5e^{-04a}	1.4 (1.1, 1.7)	4.7e^{-03a}	1.1 (0.9, 1.3)	0.36
Glutamate	1.6 (1.2, 2.0)	6.2e^{-04a}	1.4 (1.1, 1.9)	0.02[a]	1.0 (0.8, 1.3)	0.88
Ornithine	1.4 (1.2, 1.9)	2.2e^{-03a}	1.3 (1.1, 1.7)	8.3e^{-03a}	1.0 (0.9, 1.2)	0.78
Serine	1.4 (1.1, 1.8)	3.5e^{-03a}	1.2 (1.0, 1.5)	0.12	1.1 (0.9, 1.4)	0.25
Kynurenine	0.6 (0.5, 0.9)	3.2e^{-03a}	0.7 (0.5, 0.9)	2.3e^{-03a}	1.0 (0.8, 1.2)	0.88
PC aa C32:3	0.7 (0.5, 0.9)	6.4e^{-03a}	0.8 (0.6, 1.0)	0.07	0.9 (0.7, 1.0)	0.12
PC aa C34:1	1.7 (1.3, 2.2)	2.0e^{-04a}	1.7 (1.3, 2.2)	2.5e^{-05a}	0.9 (0.8, 1.1)	0.49
PC aa C36:0	0.6 (0.5, 0.8)	3.5e^{-04a}	0.6 (0.5, 0.8)	2.7e^{-04a}	1.0 (0.8, 1.2)	0.72
PC aa C36:1	1.6 (1.2, 2.0)	9.4e^{-04a}	1.6 (1.3, 2.0)	8.2e^{-05a}	0.9 (0.8, 1.1)	0.33
PC aa C38:0	0.7 (0.5, 0.9)	2.1e^{-03a}	0.6 (0.5, 0.8)	1.2e^{-04a}	1.0 (0.9, 1.3)	0.64
PC aa C38:3	1.5 (1.1, 1.9)	3.4e^{-03a}	1.3 (1.1, 1.7)	0.01[a]	1.0 (0.8, 1.2)	0.85
PC aa C40:4	1.5 (1.2, 2.0)	3.4e^{-03a}	1.4 (1.1, 1.8)	3.6e^{-03a}	1.0 (0.8, 1.2)	0.86
PC ae C34:3	0.5 (0.4, 0.7)	3.3e^{-06a}	0.6 (0.5, 0.8)	6.0e^{-05a}	0.9 (0.7, 1.1)	0.23
PC ae C38:0	0.7 (0.5, 0.9)	2.1e^{-03a}	0.6 (0.5, 0.8)	6.7e^{-04a}	1.0 (0.8, 1.2)	0.94
PC ae C38:6	0.7 (0.5, 0.9)	4.8e^{-03a}	0.7 (0.5, 0.8)	6.6e^{-04a}	1.0 (0.8, 1.2)	0.97
PC ae C40:6	0.6 (0.5, 0.8)	8.8e^{-04a}	0.7 (0.5, 0.8)	8.9e^{-04a}	0.9 (0.8, 1.1)	0.33
lysoPC a C18:2	0.7 (0.5, 0.9)	3.3e^{-03a}	0.8 (0.6, 0.9)	0.046[a]	0.9 (0.7, 1.1)	0.23
Women	**(70 versus 395)**		**(70 versus 130)**		**(130 versus 395)**	
carnitine	1.8 (1.4, 2.4)	4.3e^{-05a}	1.5 (1.1, 2.1)	0.01[a]	1.1 (0.9, 1.4)	0.32
Glutamate	1.7 (1.3, 2.2)	1.2e^{-04a}	1.8 (1.3, 2.5)	1.1e^{-03a}	0.9 (0.7, 1.1)	0.17
PC aa C32:1	1.5 (1.1, 1.9)	2.1e^{-03a}	1.4 (1.0, 2.0)	0.03[a]	1.1 (0.9, 1.4)	0.24
PC aa C36:1	1.6 (1.2, 2.0)	1.1e^{-03a}	1.5 (1.1, 2.0)	0.02[a]	1.0 (0.8, 1.2)	0.87
PC ae C34:3	0.6 (0.4, 0.8)	7.7e^{-04a}	0.6 (0.4, 0.8)	2.5e^{-03a}	1.0 (0.8, 1.2)	0.94
SM (OH) C22:2	0.6 (0.5, 0.8)	2.1e^{-03a}	0.6 (0.4, 0.9)	4.9e^{-03a}	0.9 (0.7, 1.1)	0.35

Results of pair wise comparison by logistic regression of metabolites on smoking status adjusted for age, body mass index and alcohol consumption. Men and women were analyzed separately. We present all results with a false discovery rate (FDR) below 0.05 (in the comparison between CS and NS, the FDR was calculated by *P*-value adjusted for all 140 metabolites; for CS versus FS and FS versus NS, the FDR was calculated by *P*-value adjusted for the number of metabolites significantly different between CS and NS). Smoking-related metabolites found in both men and women are in bold. aa: diacyl-; ae: acyl-alkyl-; CI: confidence interval; CS: current smokers; FS: former smokers; NS: never smokers; PC: phosphatidylcholine; lysoPC: acyl-phosphatidylcholine; SM (OH): hydroxysphingomyeline. [a]FDR <0.05.

significant difference between FS and NS; FDR <0.05). No irreversible metabolite was observed (that is, significant difference between FS and NS). Serine and PC aa C32:3 in men were not classified because their concentrations were not significantly different between CS and FS or between FS and NS (Table 2). A heat map representing the concentration profiles of the 21 identified metabolites in CS, FS and NS is shown in Figure 2, demonstrating the reversibility of metabolites after smoking cessation.

In women, SM (OH) C22:2 was significantly associated with cessation time (FDR <0.05); however, there was no such significant metabolite in men (Table S1 in Additional file 1), indicating a non-linear relationship between cessation time and the reversion of metabolite profile. In addition, we grouped the FS by stratified cessation years (0 to 10, 11 to 20, 21 to 30, 31 to 40, over 40 years). For some metabolites (for example, PC ae C38:0, PC aa C36:0 and ornithine), the greatest change of concentration occurred within the first 10 years of cessation compared with CS (Figure 3).

Within CS, we found kynurenine and PC ae C34:3, PC ae C38:0 and PC ae C38:6 in men, and PC aa C36:1 in women showing significant association with pack years. In the linear regression model, pack years showed a negative relation (parameter estimation $\beta < 0$) to these five metabolites (Table 3) (for example, one pack year increase will lead to a decrease of the kynurenine level in CS by 0.33%).

Prospective change of metabolite profiles (from KORA baseline S4 to follow-up F4)

The prospective dataset included 40 CS, 432 NS and 49 quitters (people who were CS in KORA S4 but FS in KORA F4) (Table 4). Among the 16 reversible metabolites in men, 13 (except kynurenine, glutamate and aspartate)

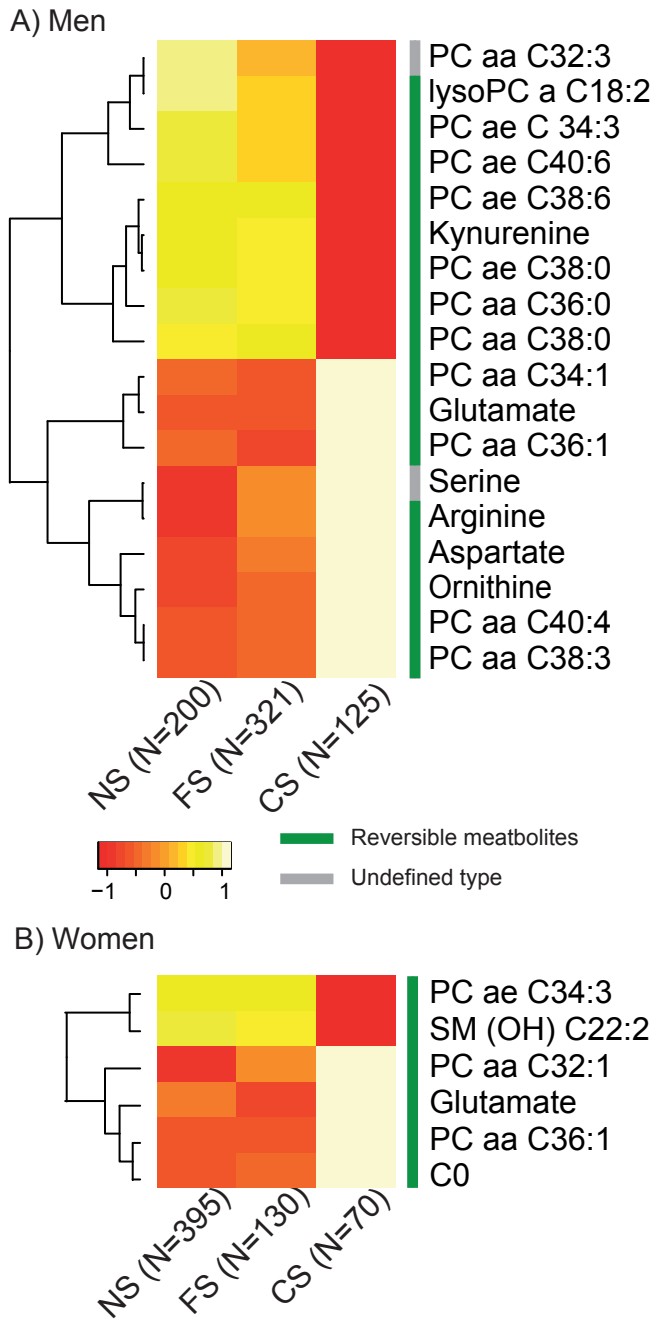

Figure 2 Heat maps of smoking-related metabolites in (A) men and (B) women. The heat map shows mean residues of smoking-related metabolites in CS, FS and NS and the reversibility after smoking cessation. The color of each cell in the heat map represents the relative mean concentration of each metabolite in NS, FS or CS. The number of samples in each group is provided. The bar besides the metabolite names indicates the reversibility of these metabolites after smoking cessation. aa: diacyl-; ae: acyl-alkyl-; C0: carnitine; CS: current smokers; FS: former smokers; lysoPC: acyl-phosphatidylcholine; NS: never smokers; PC: phosphatidylcholine; SM (OH): hydroxysphingomyeline.

were also measured in KORA F4 using a different kit (see Methods). We employed a linear mixed effect model to investigate the effects of smoking cessation on metabolite concentrations. Among these 13 metabolites, 10 metabolites showed a significant variation in quitters, with a period of smoking cessation from one to seven years, which indicated a reverting process. The arginine level decreased by 11.3% and ornithine by 14.8% in quitters compared with CS, whereas PC aa C36:0 increased by 18.5%. Figure 4 shows the prospective changes of the significant metabolites. For women, the same analysis was conducted. Because the number of female quitters was small (N = 10),

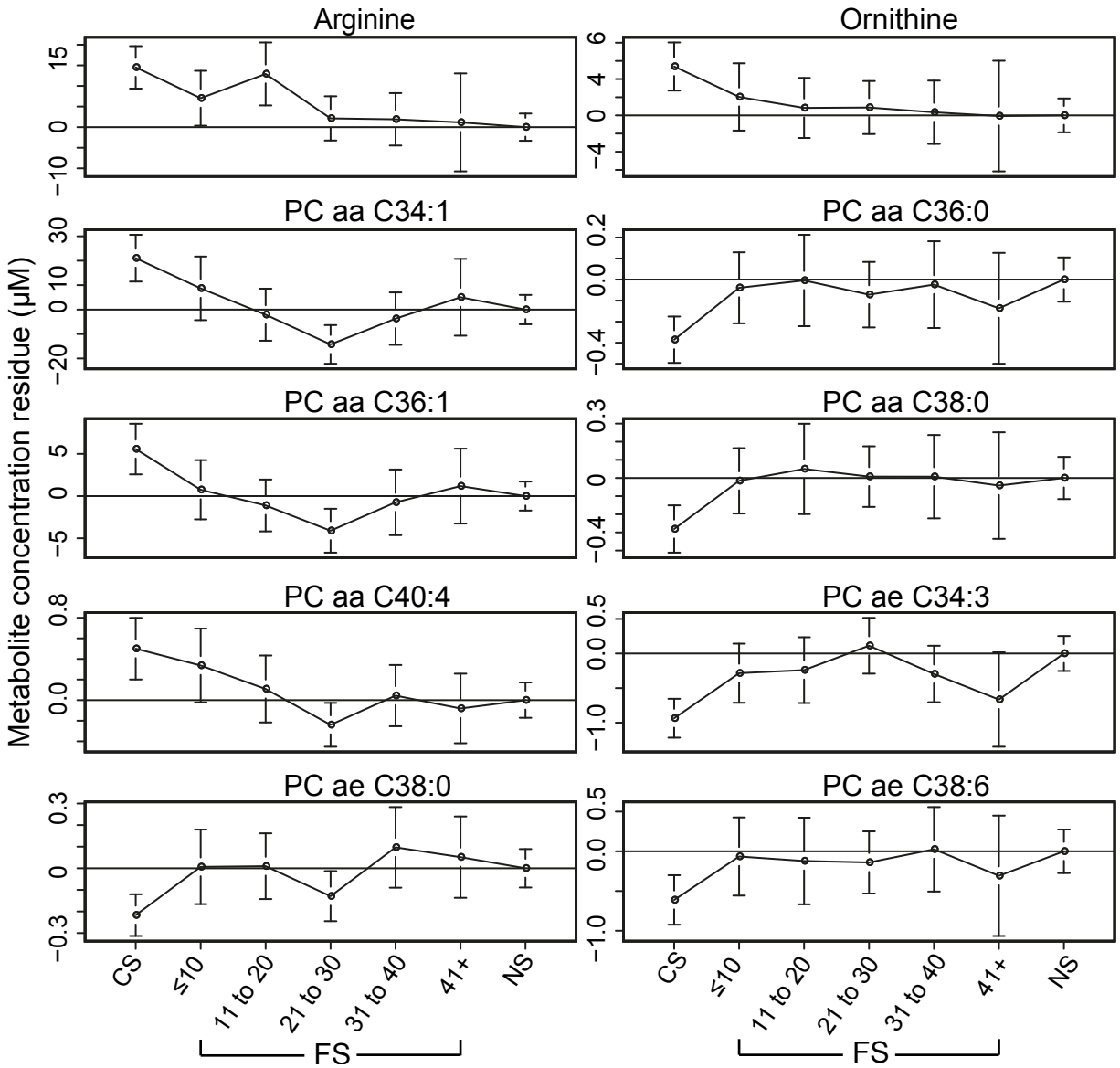

Figure 3 Metabolite concentration variations in relation to smoking cessation time. Taking NS as baseline, figures show the mean residuals of metabolites in different groups of CS and FS, giving the trend of metabolite variation with cessation time. FS were grouped by stratified cessation time (≤10, 11 to 20, 21 to 30, 31 to 40, 41+). Residuals were calculated by linear regression model (regression of metabolite concentration on age, body mass index and alcohol consumption). aa: diacyl-; ae: acyl-alkyl-; CS: current smokers; FS: former smokers; NS: never smokers; PC: phosphatidylcholine.

five metabolites that were measured in both KORA S4 and F4 showed borderline significance (P <0.05). However, none of these metabolites was found to be significant considering FDR <0.05 (see Table 5).

Smoking effects on metabolic network

Enrichment analysis of the 21 identified smoking-related metabolites on Kyoto Encyclopedia of Genes and Genomes pathways showed enrichment in a set of amino acid and lipid metabolism pathways (ether lipid, glycerophospholipid, arginine and proline metabolism). In addition, we analyzed the impact of the smoking-related metabolites in each pathway by measuring their structural importance (see Methods). These metabolites had high betweenness centrality and a strong impact on the enriched pathways (Figure 5 and Table S2 in Additional file 2).

To systematically investigate how the effects of smoking propagate over the metabolic networks, we evaluated the association between 175 smoking-related genes, previously reported [23], and the 21 smoking-related metabolites we

Table 3 Smoking intensity (pack years) related to metabolites

Metabolites	β estimate of pack year (95% confidence interval)×10^{-3}	Δ (%)	*P*
Men			
Arginine	-1.1 (-3.6, 1.4)	-0.11%	0.38
Aspartate	2.9 (-1.4, 7.1)	0.29%	0.20
Glutamate	2.9 (-1.2, 6.9)	0.29%	0.17
Ornithine	-2.4 (-5.2, 0.3)	-0.24%	0.09
Serine	1.1 (-1.3, 3.6)	0.11%	0.37
Kynurenine*	-3.3 (-6.1, -0.5)	-0.33%	0.02
PC aa C32:3	-1.4 (-4.3, 1.4)	-0.14%	0.33
PC aa C34:1	-0.9 (-3.5, 1.6)	-0.09%	0.48
PC aa C36:0	-2.3 (-4.9, 0.4)	-0.23%	0.09
PC aa C36:1	-1.4 (-4.6, 1.8)	-0.14%	0.39
PC aa C38:0	-2.1 (-4.9, 0.7)	-0.21%	0.15
PC aa C38:3	1.2 (-1.7, 4.1)	0.12%	0.43
PC aa C40:4	1.3 (-2.5, 5.1)	0.13%	0.51
PC ae C34:3*	-3.7 (-6.4, -0.9)	-0.37%	0.01
PC ae C38:0*	-3.6 (-6.6, -0.5)	-0.36%	0.02
PC ae C38:6*	-2.6 (-5.1, -0.1)	-0.26%	0.04
PC ae C40:6	-1.7 (-4.4, 1.0)	-0.17%	0.22
lysoPC a C18:2	-3.1 (-6.5, 0.3)	-0.31%	0.07
Women			
Carnitine	1.1 (-4.3, 6.5)	0.11%	0.70
PC aa C32:1	0.2 (-10.5, 10.9)	0.02%	0.97
PC aa C36:1*	6.9 (0.6, 13.2)	0.69%	0.04
PC ae C34:3	-2.7 (-7.7, 2.2)	-0.27%	0.54
SM (OH) C22:2	-2.8 (-7.7, 2.2)	-0.28%	0.28
Glutamate	2.2 (-7.8, 12.2)	0.22%	0.67

Results of linear regression of smoking intensity (pack years) on metabolite concentrations in men and women, adjusted for age, body mass index and alcohol consumption. All smoking-related metabolites presented in Table 2 are listed (*P <0.05). aa: diacyl-; ae: acyl-alkyl-; CS: current smokers; FS: former smokers; lysoPC: acyl-phosphatidylcholine; NS: never smokers; PC: phosphatidylcholine; SM (OH): hydroxysphingomyeline.

found in this study by analyzing protein-metabolite networks (see Methods). In men, 15 metabolites (lysoPC a C18:2, PC aa C32:3,PC aa C34:1, PC aa C36:0, PC aa C36:1, PC aa C38:0, PC aa C38:3, PC aa C40:4, PC ae C34:3, PC ae C38:0, PC ae C38:6, PC ae C40:6, arginine, glutamate and serine) were found to be linked with 11 genes (*ADH7, AKR1B1, DHRS3, FTL, GALE, GPC1, KRAS, S100A10, SLC7A11, SULF1, PLA2G10*) by related enzymes. In women, four metabolites (PC aa C36:1, PC ae C34:3, PC aa C32:1 and glutamate) were closely linked with nine genes (*ADH7, AKR1B1, DHRS3, FTL, GALE, GPC1, S100A10, SULF1, PLA2G10*) (Figure 6A and Table S3 in Additional file 3). Similar to enrichment analysis, the network in men and in women could be generally divided into glycerophospholipids and tightly associated proteins as well as amino acids and the associated genes and enzymes. A description of the protein-metabolite and protein-protein interactions was listed in Table S3 in Additional file 3.

The smoking effects on the networks were reversible. With regards to gene expressions, with the exception of *SULF1* and *PLA2G10*, all changes in the networks were reversible after smoking cessation [23]. All changes in metabolites in the network were also reversible, except serine.

Discussion

In this study, we have used an 'omics' approach to investigate the association of metabolite concentrations with smoking, delineated the reversion of metabolite variations after smoking cessation and demonstrated the results using protein-metabolite networks. We identified strong associations of various metabolites with smoking, and confirmed part of the findings of our pilot study [1]. Among the 23 smoking-related metabolites identified in the pilot study, 11 metabolites were measured in this study, five of which (four unsaturated diacyl-PCs and one acyl-alkyl-PC) were validated in men, based on about five-fold larger CS samples. Consistent patterns of smoking effects on metabolite profile were observed in the current study. Among all the smoking-related metabolites, in CS we found higher unsaturated diacyl-PCs, but lower acyl-alkyl-PCs and saturated diacyl-PCs, which may indicate generally increased levels of unsaturated fatty acids in CS. Unsaturated fatty acids are more

Table 4 Characteristics of the prospective dataset (KORA S4 → F4).

	Current smoker	Former smoker	Never smoker
Men (N = 207)			
N (%)	31 (15.0%)	30 (14.5%)	146 (70.5%)
Age at S4 (years)	60.2 ±5.3	63.0 ±5.0	63.0 ±5.5
Alcohol consumption (S4/F4)(g/day)	27.7 ±28.2/20.4 ±28.7	29.6 ±31.6/19.3 ±21.1	22.2 ±22.8/20.2 ±19.5
BMI (S4/F4) (kg/m^2)	26.8 ±2.9/26.9 ±3.3	28.5 ±3.8/28.9 ±3.9	27.6 ±3.3/27.8 ±3.4
Women (N = 314)			
N (%)	18 (5.7%)	10 (3.2%)	286 (91.1%)
Age at S4	61.0 ±5.1	59.5 ±3.1	63.6 ±5.1
Alcohol consumption (S4/F4)(g/day)	7.6 ±11.6/7.4 ±11.8	4.7 ±6.7/10.7 ±14.1	7.6 ±11.2/7.3 ±11.4
BMI (S4/F4) (kg/m^2)	27.9 ±5.1/27.7 ±5.3	26.9 ±3.9/27.4 ±5.1	28.6 ±4.5/28.9 ±4.7

Population characteristics were calculated based on 207 men and 314 women who participated in both the KORA S4 and F4 study. Values are provided as mean ± SD. BMI: body mass index.

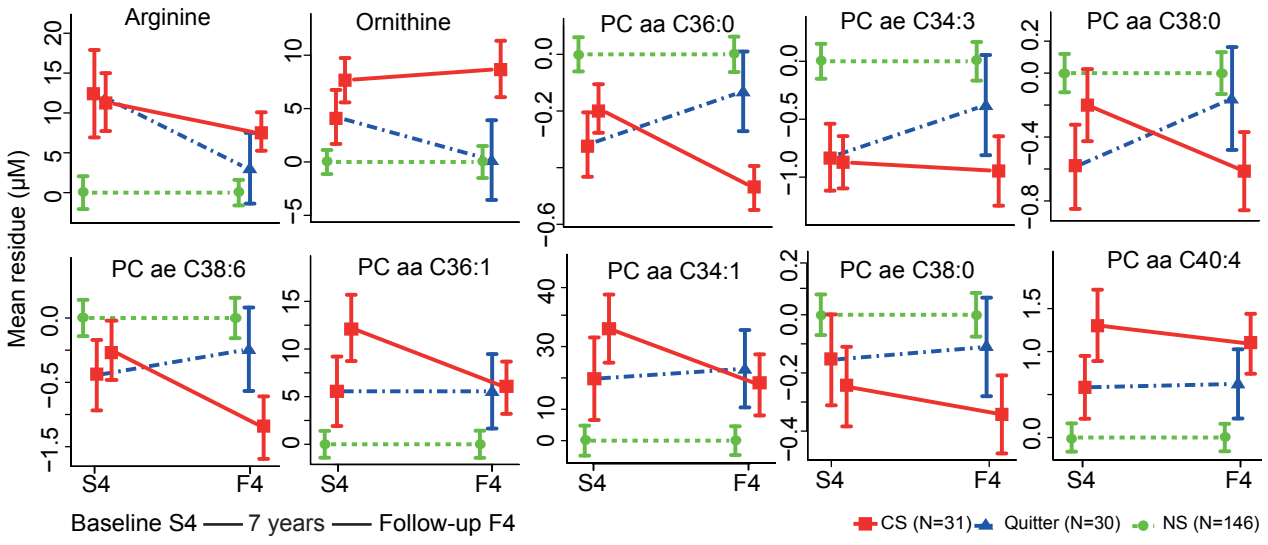

Figure 4 Changes of smoking-related metabolites in current, former and never smokers in KORA S4 → F4. Taking the NS as baseline, the concentration change of each metabolite is shown as the adjusted mean residue in KORA S4 and F4 in all three groups (CS, FS and NS). Only metabolites with significant prospective change in KORA S4 → F4 are shown in the figure. Residuals were calculated from a linear regression model (regression of metabolite concentration on age, body mass index and alcohol consumption). aa: diacyl-; ae: acyl-alkyl-; CS: current smokers; FS: former smokers; lysoPC: acyl-phosphatidylcholine; NS: never smokers; PC: phosphatidylcholine.

Table 5 Association of reversible metabolites with smoking status change in the prospective dataset (KORA S4 → F4)

	β estimate of smoking status (95% confidence interval)	*P*
Men		
Arginine	-0.12 (-0.18, -0.06)	1.4e^{-04a}
Ornithine	-0.16 (-0.24, -0.08)	2.1e^{-04a}
PC aa C34:1	-0.09 (-0.15, -0.03)	3.3e^{-03a}
PC aa C36:0	0.17 (0.09, 0.25)	6.4e^{-05a}
PC aa C36:1	-0.12 (-0.18, -0.05)	8.5e^{-04a}
PC aa C38:0	0.14 (0.06, 0.22)	3.0e^{-04a}
PC aa C38:3	-0.04 (-0.11, 0.02)	1.7e^{-01}
PC aa C40:4	-0.11 (-0.18, -0.03)	6.0e^{-03}
PC ae C34:3	0.14 (0.06, 0.21)	3.5e^{-04a}
PC ae C38:0	0.13 (0.05, 0.21)	1.8e^{-03a}
PC ae C38:6	0.11 (0.04, 0.18)	1.5e^{-03a}
PC ae C40:6	0.08 (0.01, 0.15)	2.1e^{-02}
lysoPC a C18:2	0.03 (-0.06, 0.11)	5.2e^{-01}
Women		
Carnitine	-0.12 (-0.20, -0.05)	1.4e^{-03}
PC aa C32:1	-0.18 (-0.32, -0.03)	2.1e^{-03}
PC aa C36:1	-0.11 (-0.20, -0.02)	2.0e^{-02}
PC ae C34:3	0.09 (-0.02, 0.19)	0.95
SM (OH) C22.2	0.12 (0.02, 0.22)	1.9e^{-02}

Result of smoking status on metabolite concentrations using linear mixed model for S4 → F4 longitudinal data, adjusted for age, BMI, and alcohol consumption. PC: phosphatidylcholine; aa: diacyl-; ae: acyl-alkyl-; lysoPC: acyl-phosphatidylcholine; SM (OH): hydroxysphingomyeline. [a] FDR<0.05.

vulnerable to lipid peroxidation and influence the risk of different diseases [43,44].

Smoking-related metabolites and cardiovascular disease

The study results implied the potential of metabolomics in revealing the role of an environmental factor, for example a smoking lifestyle, in the pathogenesis and prognosis of CVD.

One study on the peripheral blood metabolite profile showed an association of coronary artery disease and urea cycle-related metabolites, including arginine and glutamate [45], which were also identified in our study as smoking-related metabolites. By scrutinizing the smoking-related metabolites in metabolic pathways, we found further support for the pathophysiological relation between these metabolites and CVD. Previous findings indicated that the glutamate transporter in human lung epithelial cells, encoded by the *SLC7A11* gene, is activated in CS [23,46], which increases the transportation of glutamate and subsequently raises the levels of the downstream metabolites, arginine and ornithine (Figure 6B). The activation of the cysteine-glutamate transporter (encoded by *SLC7A11*) and the increased glutamate level as a response to oxidative stress is also of great importance to endothelial dysfunction involved at all stages of atherosclerotic plaque evolution, which leads to CVD [47,48].

Ether lipid and glycerophospholipid metabolisms are associated with smoking [1,49]. The decreased level of

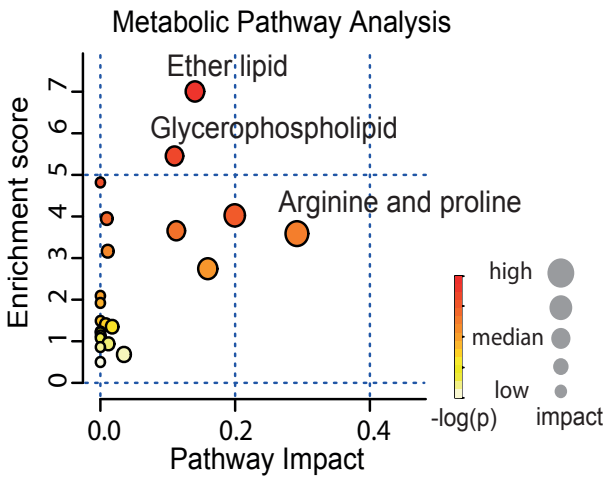

Figure 5 Pathway analyses of smoking-related metabolites. Figure shows enrichment and impact of smoking-related metabolites in Kyoto Encyclopedia of Genes and Genomes pathways. The enrichment scores are shown on y-axis, which was calculated as the negative logarithm of the *P*-value from an enrichment test. The x-axis indicates the structural impact with a score from 0 to 1 of the smoking-related metabolites in the enriched pathways.

lysoPC a C18:2 reflects the inhibition of upstream synthesis and activation of downstream hydrolysis. As shown in Figure 6C, upregulation of *S100A10* and *GPC1* inhibits cytosolic phospholipase A2, which plays a role in the synthesis of lyso-PCs. The lysophospholipase I isoform, which hydrolyses lysoPC into glycerophosphocholine, is upregulated in CS [23]. Interestingly, one recent study showed that a disorder of phosphatidylcholine metabolism would promote CVD [50], which may establish a link between smoking-related phosphatidylcholine variation and cardiovascular events. For example, the phosphatidylcholine hydroperoxide will promote angiogenesis in endothelial cells that are associated with atherosclerotic development [51].

The reversibility of metabolite concentrations in a small time window may reveal a reduced risk of smoking-related diseases after stopping smoking. Concentrations of arginine and glutamate that are associated with both smoking and coronary artery diseases quickly returned to normal levels (within seven years) after smoking cessation, which is in line with epidemiological findings that the smoking effects on CVD are quickly and largely reduced after smoking cessation [8,9,52]. The reversed glutamate level indicates reduced oxidative stress after smoking cessation, and the reversion of arginine and ornithine reflects a reversion of functioning in the urea cycle. Our findings provide metabolic insight into the reduced risk of CVD after smoking cessation and provide support for the remarkable benefits people would gain by stopping smoking.

Concordance of reversibility in metabolic network

The protein-metabolite interaction network shows that the reversibility of metabolite concentrations also coincided with gene expression (Figure 6A). Arginine and glutamate were quickly reversed after smoking cessation, which was in line with the quick reversibility of *SLC7A11* expression. Expression of enzyme coding genes for the hydrolysis of diacyl-PCs and acyl-alkyl-PCs, for instance lysophospholipase, cytosolic phospholipase A2 and S100 calcium binding protein A2, were quickly reversible and smoking-related diacyl-PCs and acyl-alkyl PCs shared the same reverse pattern.

Gender-specific effects of smoking

In this study, we found gender-specific effects of smoking on metabolite profiles (Table S1 in Additional file 1). This result supports the assumption that differences in smoking effects on men and women are not solely based on smoking intensity but are also gender-specific. Glutamate was higher in both male and female CS, however, the levels of arginine and ornithine were only higher in male CS. According to a previous study of the metabolomic and genetic biomarkers on sexual dimorphisms [30], the *CPS1* gene, which regulates the formation of arginine, has a gender-specific manner in certain single nucleotide polymorphisms, with stronger effects in women than in men. The gender-specific genetic effect might cause a lower efficiency in women in regard to the transformation of extra glutamate to citrulline (Figure 6C).

Strengths and limitations

We used a systematic targeted metabolomics approach with 140 metabolites in a large population-based cohort. Analyzing the effects of smoking and smoking cessation in this prospective manner (follow-up of seven years) provides more power to investigate smoking effects by ruling out individual differences. However, our study is based on a limited range and number of metabolites and cannot fully represent the whole metabolome. Thus, an improved metabolomics technique measuring more metabolites is urgently needed for a comprehensive understanding of both reversible and permanent effects of smoking on human metabolism. It would be interesting for future studies to also include data on other environmental factors such as diet and lifestyle, which are known to have effects on the human metabolome [53,54].

Conclusions

Our study shows the power of the metabolomics approach in investigating the molecular signature of lifestyle-related environmental exposures. We demonstrated that smoking is associated with concentration variations in amino acids, ether lipid and glycerophospholipid metabolism at an 'omics' level. The smoking-related changes in the human

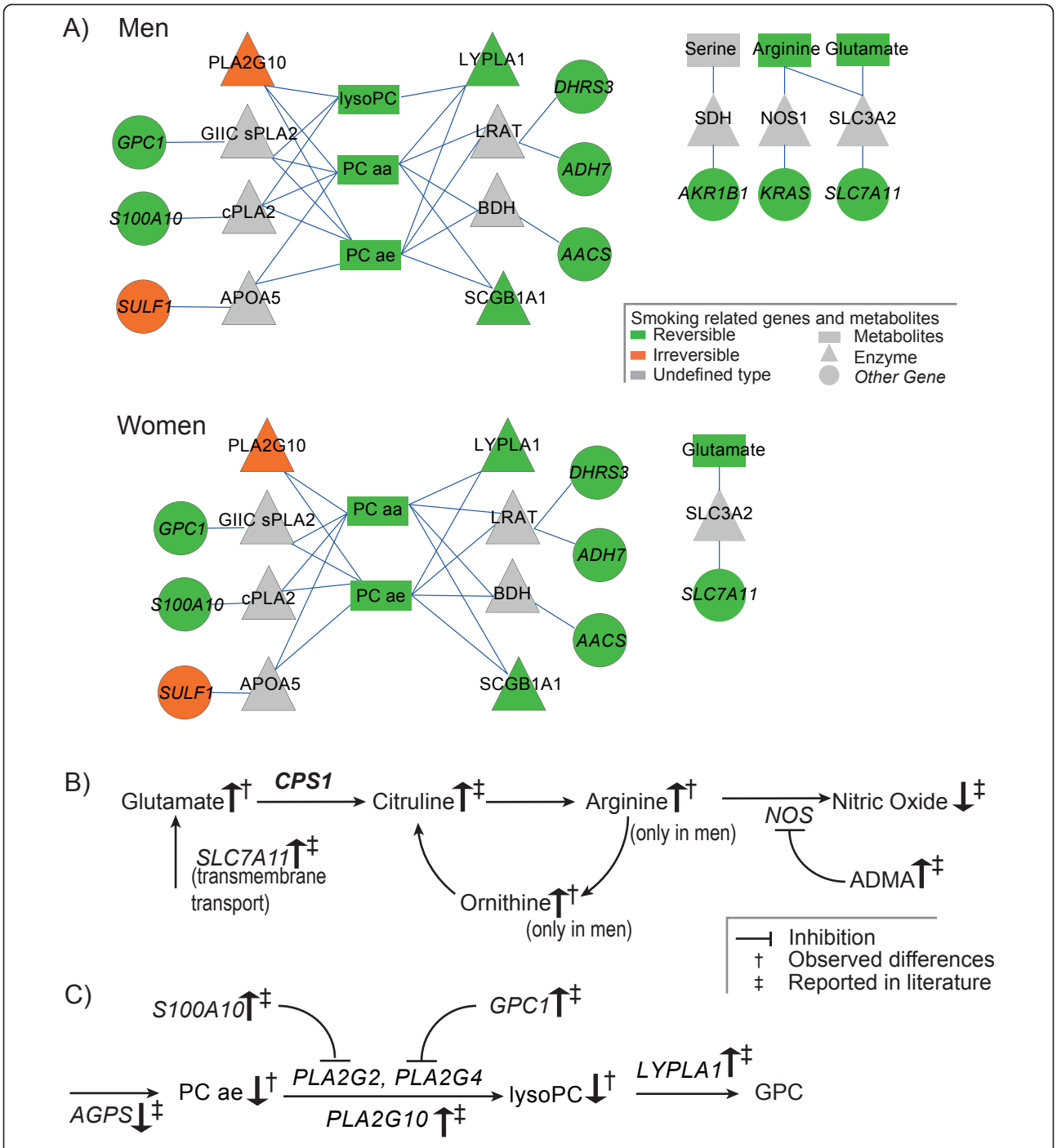

Figure 6 Protein-metabolite networks and pathways of the smoking-related metabolites and genes. **(A)** Network linking metabolites and proteins encoded by smoking-related genes with maximum one intermediate. Node color indicates the reversibility after smoking cessation. **(B, C)** Effects of smoking on arginine and glutamate as well as on lipid metabolism. Metabolites are in regular font, protein coding genes are in italic, gender-specific gene (CPS1) is in bold italic font. aa: diacyl-; ae: acyl-alkyl-; APOA5: apolipoprotein A-V; BDH: 3-hydroxybutyrate dehydrogenase, type 1; cPLA2: cytosolic phospholipase A2; CS: current smokers; FS: former smokers; GIIC sPLA2: phospholipase A2, membrane associated; LRAT: lecithin retinol acyltransferase; LYPLA1: lysophospholipase I; lysoPC: acyl-phosphatidylcholine; NOS1: nitric oxide synthase 1; NS: never smokers; PC: phosphatidylcholine; PLA2G10: group 10 secretory phospholipase A2; SCGB1A1: uteroglobin; SDH: serine dehydratase; SLC3A2: solute carrier family 3 member 2.

serum metabolite profile are reversible after stopping smoking. This indicates the remarkable benefits of smoking cessation and provides a link to CVD benefits. Furthermore, linking metabolomic knowledge to other 'omics' approaches, for example, transcriptomics, may have the potential to identify novel biomarkers as well as new risk assessment tools.

Additional material

Additional file 1: Table S1: Cessation time-related metabolites in FS. FDR was calculated by *P*-value adjusted for the number of smoking-related metabolites with Benjamini-Hochberg method. aa: diacyl-; ae: acyl-alkyl-; C0: carnitine; FS: former smokers; lysoPC: acyl-phosphatidylcholine; PC: phosphatidylcholine; SM (OH): hydroxysphingomyeline.

Additional file 2: Table S2: Enrichment and impact of smoking-related metabolites in Kyoto Encyclopedia of Genes and Genomes pathways. Table shows the enrichment and impact scores of smoking-related metabolites in Kyoto encyclopedia of Genes and Genomes pathways. The pathway analysis consists of enrichment and a structural impact analysis both based on Kyoto Encyclopedia of Genes and Genomes database. The -log (*P*) was considered as the enrichment score. Impact, scored between 0 and 1, indicated the pathway topological importance of the metabolites. In particular, the parameter Total is the total number of compounds in the pathway; the parameter Hits is the actual number of metabolites with significant variations in the pathway; the Raw *P* was the original *P*-value calculated from the enrichment analysis; the FDR was calculated as the *P*-value adjusted using Benjamini-Hochberg method.

Additional file 3: Table S3: Links between smoking-related metabolites, enzymes and genes. The table describes the links showed in Figure 6 of the main text. The smoking-related metabolites, enzymes and genes are listed in the first and second columns. The score of interaction is given according to the definition by the Search Tool for the Retrieval of Interacting Genes/Proteins [1]. A reference for each link and a short description is provided. The Column of reaction shows the possible biochemical reaction of the corresponding link or the type of protein interaction. The enzymes includes, phospholipase A2, membrane associated (GIIC sPLA2), cytosolic phospholipase A2 (cPLA2), group 10 secretory phospholipase A2 (PLA2G10), lysophospholipase I (LYPLA1), apolipoprotein A-V (APOA5), uteroglobin (SCGB1A1), lecithin retinol acyltransferase (LRAT), nitric oxide synthase 1 (NOS1), solute carrier family 3 member 2 (SLC3A2), serine dehydratase (SDH), 3-hydroxybutyrate dehydrogenase, type 1 (BDH). The smoking-related gene/protein includes, S100 calcium binding protein A10 (*S100A10*), glypican 1 (*GPC1*), sulfatase 1 (*SULF1*), alcohol dehydrogenase 7 (*ADH7*), dehydrogenase member 3 (*DHRS3*), aldose reductase (*AKR1B1*), acetoacetyl-CoA synthetase (*AACS*), V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (*KRAS*), solute carrier family 7 (*SLC7A11*) and three enzyme listed above, PLA2G10, LYPLA1, SCGB1A1. The links in the network for male and female CS are combined and listed together. Smoking-related genes are show in italic. aa: diacyl-; ae: acyl-alkyl-; C0: carnitine; lysoPC: acyl-phosphatidylcholine; PC: phosphatidylcholine; SM (OH): hydroxysphingomyeline.

Abbreviations

aa: diacyl-; ae: acyl-alkyl-; BMI: body mass index; CS: current smokers; CVD: cardiovascular disease; FDR: false discovery rate; FS: former smokers; lysoPC: acyl-phosphatidylcholine; NS: never smokers; PC: phosphatidylcholine; SM: sphingomyeline; SM (OH): hydroxysphingomyeline.

Authors' contributions

HEW, KS, JA, TI, AP and RWS initiated and designed the study. CP, WRM, WR, HEW, KHL and JA were involved in and performed the experiment. TX and ZY performed the data analysis. TX, CH, ZY and RWS wrote the manuscript, XD, EB, CP, KP, MJ, YL, HW, FT, JA and AP revised the manuscript. The manuscript has been approved by all authors.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We express our appreciation to all KORA study participants for donating their blood and time. We thank the field staff in Augsburg who conducted the KORA studies. The KORA research platform and the KORA Augsburg studies were initiated and financed by the Helmholtz Zentrum München, which is funded by the German Federal Ministry of Education, Science, Research and Technology and by the State of Bavaria. The KORA study group consists of A. Peters (speaker), J. Heinrich, R. Holle, R. Leidl, C. Meisinger, K. Strauch and their co-workers, who are responsible for the design and conduct of the KORA studies. We thank Julia Scarpa, Katharina Sckell and Arsin Sabunchi for metabolomics measurements performed at the Helmholtz Zentrum München, Genome Analysis Centre, Metabolomics Core Facility. This study was supported in part by a grant from the German Federal Ministry of Education and Research (BMBF) to the German Centre for Diabetes Research (DZD e.V.) and this work was partly supported by the BMBF project 'Metabolomics of ageing' (FKZ: 01DO12030) and EU FP7 grant HEALTH-2009-2.2.1-3/242114 (Project OPTiMiSE).

Author details

[1]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Ingolstädter Landstraße 1, Neuherberg, 85764, Germany. [2]Else Kroener-Fresenius-Center for Nutritional Medicine, University Hospital, Klinikum rechts der Isar, Technische Universität München, Ismaninger Straße 22, Munich, 81675, Germany. [3]Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Yue Yang Road 320, Shanghai, 200031, China. [4]University of Chinese Academy of Sciences, Yuquanlu 19A, Beijing, 100049, China. [5]Institute of Experimental Genetics, Genome Analysis Center, Helmholtz Zentrum München, Ingolstädter Landstraße 1, Neuherberg, 85764, Germany. [6]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Ingolstädter Landstraße 1, Neuherberg, 85764, Germany. [7]Institute of Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University, Moorenstr. 5, Düsseldorf, 40225, Germany. [8]Institute of Epidemiology I, Helmholtz Zentrum München, Ingolstädter Landstraße 1, Neuherberg, 85764, Germany. [9]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Marchioninistr. 15, Munich, 81377, Germany. [10]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald; Ernst-Moritz-Arndt University, Ferdinand-Sauerbruch-Straße, Greifswald, 17475, Germany. [11]Institute of Epidemiology II, Helmholtz Zentrum München, Ingolstädter Landstraße 1, Neuherberg, 85764, Germany. [12]Department of Psychosomatic Medicine and Psychotherapy, University Hospital, Klinikum rechts der Isar, Technische Universität München, Ismaninger Straße 22, Munich, 81675, Germany. [13]Department of Mathematics, Technische Universität München, Boltzmannstraße 3, Garching, 85748, Germany. [14]Faculty of Biology, Ludwig-Maximilians-Universität, Großhaderner Straße 2-4, Planegg-Martinsried, 82152, Germany. [15]Department of Physiology and Biophysics, Weill Cornell Medical College in Qatar (WCMC-Q), PO Box 24144, Doha, Qatar. [16]Chair of Experimental Genetics, Technische Universität München, Ingolstädter Landstraße 1, Neuherberg, 85764, Germany. [17]Hannover Unified Biobank, Hannover Medical School, Carl-Neuberg-Straße 1, Hannover, 30625, Germany. [18]Harvard School of Public Health, Department of Environmental Health, 677 Huntington Avenue, Boston, 02115, USA.

Received: 17 October 2012 Accepted: 4 March 2013
Published: 4 March 2013

References

1. Wang-Sattler R, Yu Y, Mittelstrass K, Lattka E, Altmaier E, Gieger C, Ladwig KH, Dahmen N, Weinberger KM, Hao P, Liu L, Li Y, Wichmann H-E, Adamski J, Suhre K, Illig T: **Metabolic profiling reveals distinct variations linked to nicotine consumption in humans–first results from the KORA study.** *PLoS One* 2008, **3**:e3863.

2. Xavier F, Henn L de A, Oliveira M, Orlandine L: **Smoking and its relation to the histological type, survival, and prognosis among patients with primary lung cancer.** *Sao Paulo Med J* 1996, **114**:1298-1302.
3. **Smoking-attributable mortality, years of potential life lost, and productivity losses–United States, 2000-2004.** *Morbidity and Mortality Weekly Report* 2008, **57**:1226-1228.
4. Bøttcher M, Falk E: **Pathology of the coronary arteries in smokers and non-smokers.** *J Cardiovasc Risk* 1999, **6**:299-302.
5. Howard G, Burke GL, Szklo M, Tell GS, Eckfeldt J, Evans G, Heiss G: **Active and passive smoking are associated with increased carotid wall thickness. The Atherosclerosis Risk in Communities Study.** *Arch Intern Med* 1994, **154**:1277-1282.
6. Libby P: **Inflammation in atherosclerosis.** *Nature* 2002, **420**:868-874.
7. Szmitko PE, Wang C-H, Weisel RD, Jeffries GA, Anderson TJ, Verma S: **Biomarkers of vascular disease linking inflammation to endothelial activation: Part II.** *Circulation* 2003, **108**:2041-2048.
8. Danesh J, Whincup P, Walker M, Lennon L, Thomson A, Appleby P, Gallimore JR, Pepys MB: **Low grade inflammation and coronary heart disease: prospective study and updated meta-analyses.** *BMJ* 2000, **321**:199-204.
9. Danesh J, Muir J, Wong YK, Ward M, Gallimore JR, Pepys MB: **Risk factors for coronary heart disease and acute-phase proteins. A population-based study.** *Eur Heart J* 1999, **20**:954-959.
10. Rigotti NA, Pasternak RC: **Cigarette smoking and coronary heart disease: risks and management.** *Cardiol Clin* 1996, **14**:51-68.
11. Cook DG, Shaper AG, Pocock SJ, Kussick SJ: **Giving up smoking and the risk of heart attacks. A report from The British Regional Heart Study.** *Lancet* 1986, **2**:1376-1380.
12. Aberg A, Bergstrand R, Johansson S, Ulvenstam G, Vedin A, Wedel H, Wilhelmsson C, Wilhelmsen L: **Cessation of smoking after myocardial infarction. Effects on mortality after 10 years.** *Br Heart J* 1983, **49**:416-422.
13. Halpern MT, Gillespie BW, Warner KE: **Patterns of absolute risk of lung cancer mortality in former smokers.** *J Natl Cancer Inst* 1993, **85**:457-464.
14. Ebbert JO, Yang P, Vachon CM, Vierkant RA, Cerhan JR, Folsom AR, Sellers TA: **Lung cancer risk reduction after smoking cessation: observations from a prospective cohort of women.** *J Clin Oncol* 2003, **21**:921-926.
15. Bakhru A, Erlinger TP: **Smoking cessation and cardiovascular disease risk factors: results from the Third National Health and Nutrition Examination Survey.** *PLoS Med* 2005, **2**:e160.
16. De Maat MP, Pietersma A, Kofflard M, Sluiter W, Kluft C: **Association of plasma fibrinogen levels with coronary artery disease, smoking and inflammatory markers.** *Atherosclerosis* 1996, **121**:185-191.
17. Pradhan AD, Manson JE, Rossouw JE, Siscovick DS, Mouton CP, Rifai N, Wallace RB, Jackson RD, Pettinger MB, Ridker PM: **Inflammatory biomarkers, hormone replacement therapy, and incident coronary heart disease: prospective analysis from the Women's Health Initiative observational study.** *JAMA* 2002, **288**:980-987.
18. Yanbaeva DG, Dentener MA, Creutzberg EC, Wesseling G, Wouters EFM: **Systemic effects of smoking.** *Chest* 2007, **131**:1557-1566.
19. Powell CA, Klares S, O'Connor G, Brody JS: **Loss of heterozygosity in epithelial cells obtained by bronchial brushing: clinical utility in lung cancer.** *Clin Cancer Res* 1999, **5**:2025-2034.
20. Wistuba II, Lam S, Behrens C, Virmani AK, Fong KM, LeRiche J, Samet JM, Srivastava S, Minna JD, Gazdar AF: **Molecular damage in the bronchial epithelium of current and former smokers.** *J Natl Cancer Inst* 1997, **89**:1366-1373.
21. Wistuba II, Mao L, Gazdar AF: **Smoking molecular damage in bronchial epithelium.** *Oncogene* 2002, **21**:7298-7306.
22. Guo M, House MG, Hooker C, Han Y, Heath E, Gabrielson E, Yang SC, Baylin SB, Herman JG, Brock MV: **Promoter hypermethylation of resected bronchial margins: a field defect of changes?** *Clin Cancer Res* 2004, **10**:5131-5136.
23. Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A: **Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression.** *Genome Biol* 2007, **8**:R201.
24. 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 A, Suhre K, Li Y, Adamski J, Spector TD, Illig T, Wang-Sattler R: **Human serum metabolic profiles are age dependent.** *Aging cell* 2012, **11**:960-967.
25. Ellis JK, Athersuch TJ, Thomas LD, Teichert F, Pérez-Trujillo M, Svendsen C, Spurgeon DJ, Singh R, Järup L, Bundy JG, Keun HC: **Metabolic profiling detects early effects of environmental and lifestyle exposure to cadmium in a human population.** *BMC Med* 2012, **10**:61.
26. Vulimiri SV, Misra M, Hamm JT, Mitchell M, Berger A: **Effects of mainstream cigarette smoke on the global metabolome of human lung epithelial cells.** *Chem Res Toxicol* 2009, **22**:492-503.
27. Holle R, Happich M, Löwel H, Wichmann HE: **KORA–a research platform for population based health research.** *Gesundheitswesen* 2005, **67(Suppl 1)**:S19-25.
28. Wichmann H-E, Gieger C, Illig T: **KORA-gen–resource for population genetics, controls and a broad spectrum of disease phenotypes.** *Gesundheitswesen* 2005, **67(Suppl 1)**:S26-30.
29. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, Heim K, Campillos M, Holzapfel C, Thorand B, Grallert H, Xu T, Bader E, Huth C, Mittelstrass K, 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, *et al*: **Novel biomarkers for pre-diabetes identified by metabolomics.** *Mol Syst Biol* 2012, **8**:615.
30. Mittelstrass K, Ried JS, Yu Z, 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, **7**:e1002215.
31. 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 FJ, Krumsiek J, Wichmann H-E, Linseisen J: **Body fat free mass is associated with the serum metabolite profile in a population-based study.** *PLoS One* 2012, **7**:e40009.
32. Illig T, Gieger C, Zhai G, Römisch-Margl W, Wang-Sattler R, Prehn C, Altmaier E, Kastenmüller G, Kato BS, Mewes H-W, Meitinger T, Angelis MH de, Kronenberg F, Soranzo N, Wichmann H-E, Spector TD, Adamski J, Suhre K: **A genome-wide perspective of genetic variation in human metabolism.** *Nat Genet* 2010, **42**:137-141.
33. Römisch-Margl W, Prehn C, Bogumil R, Röhring C, Suhre K, Adamski J: **Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics.** *Metabolomics* 2012, **8**:133-142.
34. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Statist Soc B* 1995, **57**:289-300.
35. **NCI Dictionary of Cancer Terms.** [http://www.cancer.gov/dictionary?CdrID = 306510].
36. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C: **The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored.** *Nucleic Acids Res* 2011, **39**:D561-568.
37. Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, Mandal R, Sinelnikov I, Xia J, Jia L, Cruz JA, Lim E, Sobsey CA, Shrivastava S, Huang P, Liu P, Fang L, Peng J, Fradette R, Cheng D, Tzur D, Clements M, Lewis A, De Souza A, Zuniga A, Dawe M, *et al*: **HMDB: a knowledgebase for the human metabolome.** *Nucleic Acids Res* 2009, **37**:D603-610.
38. Dijkstra EW: **A note on two problems in connexion with graphs.** *Numerische Mathematik* 1959, **1**:269-271.
39. 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, **2**:e149.
40. Csardi G, Nepusz T: **The igraph software package for complex network research.** *InterJournal* 2006, **Complex Systems**:1695.
41. Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T: **Cytoscape 2.8: new features for data integration and network visualization.** *Bioinformatics* 2011, **27**:431-432.
42. Xia J, Mandal R, Sinelnikov IV, Broadhurst D, Wishart DS: **MetaboAnalyst 2.0–a comprehensive server for metabolomic data analysis.** *Nucleic Acids Res* 2012, **40**:W127-133.
43. Kris-Etherton PM: **AHA Science Advisory. Monounsaturated fatty acids and risk of cardiovascular disease. American Heart Association. Nutrition Committee.** *Circulation* 1999, **100**:1253-1258.
44. Mozaffarian D, Rimm EB, Herrington DM: **Dietary fats, carbohydrate, and progression of coronary atherosclerosis in postmenopausal women.** *Am J Clin Nutr* 2004, **80**:1175-1184.

45. Shah SH, Bain JR, Muehlbauer MJ, Stevens RD, Crosslin DR, Haynes C, Dungan J, Newby LK, Hauser ER, Ginsburg GS, Newgard CB, Kraus WE: **Association of a peripheral blood metabolic profile with coronary artery disease and risk of subsequent cardiovascular events.** *Circ Cardiovasc Genet* 2010, **3**:207-214.
46. Bridges CC, Kekuda R, Wang H, Prasad PD, Mehta P, Huang W, Smith SB, Ganapathy V: **Structure, function, and regulation of human cystine/glutamate transporter in retinal pigment epithelial cells.** *Invest Ophthalmol Vis Sci* 2001, **42**:47-54.
47. Harrison D, Griendling KK, Landmesser U, Hornig B, Drexler H: **Role of oxidative stress in atherosclerosis.** *Am J Cardiol* 2003, **91**:7-11.
48. Glass CK, Witztum JL: **Atherosclerosis. The road ahead.** *Cell* 2001, **104**:503-516.
49. Palmerini CA, Saccardi C, Ferracci F, Arienti S: **Lipid patterns in the saliva of smoking young adults.** *Hum Exp Toxicol* 2011, **30**:1482-1488.
50. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung Y-M, Wu Y, Schauer P, Smith JD, Allayee H, Tang WHW, DiDonato JA, Lusis AJ, Hazen SL: **Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease.** *Nature* 2011, **472**:57-63.
51. Nakagawa K, Shibata A, Saito T, Sookwong P, Kato S, Tsuduki T, Matsubara K, Miyazawa T: **Phosphatidylcholine hydroperoxide promotes VEGF-induced angiogenesis in endothelial cells and rat aorta ring cultures.** *Biochim Biophys Acta* 2011, **1810**:1205-1211.
52. Danesh J, Collins R, Appleby P, Peto R: **Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease: meta-analyses of prospective studies.** *JAMA* 1998, **279**:1477-1482.
53. Gibney MJ, Walsh M, Brennan L, Roche HM, German B, van Ommen B: **Metabolomics in human nutrition: opportunities and challenges.** *Am J Clin Nutr* 2005, **82**:497-503.
54. Pohjanen E, Thysell E, Jonsson P, Eklund C, Silfver A, Carlsson I-B, Lundgren K, Moritz T, Svensson MB, Antti H: **A multivariate screening strategy for investigating metabolic effects of strenuous physical exercise in human serum.** *J Proteome Res* 2007, **6**:2113-2120.

Pre-publication history
The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1741-7015/11/60/prepub

doi:10.1186/1741-7015-11-60
Cite this article as: Xu *et al.*: Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *BMC Medicine* 2013 **11**:60.

Citation: Transl Psychiatry (2013) 3, e276; doi:10.1038/tp.2013.55

www.nature.com/tp

Alcohol-induced metabolomic differences in humans

M Jaremek[1], Z Yu[1], M Mangino[2], K Mittelstrass[1], C Prehn[3], P Singmann[1], T Xu[1], N Dahmen[4], KM Weinberger[5,6,15], K Suhre[7,8], A Peters[1,9,10], A Döring[11], H Hauner[12], J Adamski[5,13], T Illig[1,14], TD Spector[2] and R Wang-Sattler[1]

Alcohol consumption is one of the world's major risk factors for disease development. But underlying mechanisms by which moderate-to-heavy alcohol intake causes damage are poorly understood and biomarkers are sub-optimal. Here, we investigated metabolite concentration differences in relation to alcohol intake in 2090 individuals of the KORA F4 and replicated results in 261 KORA F3 and up to 629 females of the TwinsUK adult bioresource. Using logistic regression analysis adjusted for age, body mass index, smoking, high-density lipoproteins and triglycerides, we identified 40/18 significant metabolites in males/females with *P*-values $<3.8E-04$ (Bonferroni corrected) that differed in concentrations between moderate-to-heavy drinkers (MHD) and light drinkers (LD) in the KORA F4 study. We further identified specific profiles of the 10/5 metabolites in males/females that clearly separated LD from MHD in the KORA F4 cohort. For those metabolites, the respective area under the receiver operating characteristic curves were 0.812/0.679, respectively, thus providing moderate-to-high sensitivity and specificity for the discrimination of LD to MHD. A number of alcohol-related metabolites could be replicated in the KORA F3 and TwinsUK studies. Our data suggests that metabolomic profiles based on diacylphosphatidylcholines, lysophosphatidylcholines, ether lipids and sphingolipids form a new class of biomarkers for excess alcohol intake and have potential for future epidemiological and clinical studies.

Translational Psychiatry (2013) **3,** e276; doi:10.1038/tp.2013.55; published online 2 July 2013

Introduction

Alcohol consumption is the world's third largest risk factor for disease burden and is associated with diseases, including neuropsychiatric disorders,[1–4] cardiovascular diseases, cirrhosis of the liver, various cancers and fetal alcohol syndrome. Each year, an estimated 2.5 million people die from alcohol-related disease worldwide.[5] Biomarkers for alcohol intake include direct blood alcohol concentration, γ-glutamyltransferase activity, carbohydrate deficient transferrin[6] or mean corpuscular volume of erythrocytes.[7] Nevertheless, further research is needed to understand alcohol-specific metabolic responses and the underlying pathophysiology. For example, identification of potential biomarkers for monitoring of alcohol consumption or determination of pharmacotherapy targets could facilitate early intervention for patients with specific alcohol-related disorders.

Targeted metabolomics is a promising method that can elucidate the effect of alcohol consumption on human metabolism. Metabolites are products of cellular processes, and their levels can be regarded as the ultimate response of biological systems to genetic or environmental changes.[8–11] Recent advances in metabolomic technologies have enabled high-throughput measurement of not only one but several compound classes simultaneously (for example, amino acids, sugars, glycerophospholipids)[12,13] resulting in a fast and more comprehensive identification of candidate biomarkers. As far as we are aware, no large-scale metabolic profiling analyses of humans with alcohol consumption have yet been conducted.

The aims of the underlying study were to (1) investigate the relation of alcohol intake and serum metabolite concentrations in German and UK studies and (2) identify potential biomarkers that could predict high levels of intake.

Materials and methods

KORA F4 study population. Cooperative Health Research in the Region of Augsburg (KORA) is a population-based research platform with subsequent follow-up studies in the fields of epidemiology and health-care research.[14–16] The KORA F4 study is the follow-up of KORA-Survey 4

[1]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany; [2]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; [3]Institute of Experimental Genetics, Genome Analysis Center, Helmholtz Zentrum München, Neuherberg, Germany; [4]Department of Psychiatry and Psychotherapy, University Medical Centre, Mainz, Germany; [5]Biocrates Life Sciences AG, Innrain 66, Innsbruck, Austria; [6]Institute for Electrical and Biomedical Engineering, University for Health Sciences, Medical Informatics and Technology, Eduard Wallnöfer-Zentrum 1, Tirol, Austria; [7]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Neuherberg, Germany; [8]Department of Physiology and Biophysics, Weill Cornell Medical College in Qatar, Doha, Qatar; [9]Institute of Epidemiology II, Helmholtz Zentrum München, Neuherberg, Germany; [10]Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA; [11]Institute of Epidemiology I, Helmholtz Zentrum München, Neuherberg, Germany; [12]Else Kroener-Fresenius-Centre for Nutritional Medicine, Klinikum rechts der Isar, Technische Universität München, München, Germany; [13]Institute of Experimental Genetics, Life and Food Science Center Weihenstephan, Technische Universität München, München, Germany and [14]Hannover Unified Biobank, Hannover Medical School, Hannover, Germany
Correspondence: Professor Dr J Adamski, Institute of Experimental Genetics, Genome Analysis Center, Helmholtz Zentrum München, Neuherberg 85764, Germany or Dr R Wang-Sattler, Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Neuherberg 85764, Germany.
E-mail: adamski@helmholtz-muenchen.de or rui.wang-sattler@helmholtz-muenchen.de
[15]Present address: Bavarian Nordic GmbH, Fraunhoferstr. 13, Martinsried, Germany.
Keywords: alcohol; alcoholism; biomarkers; gender; lipids; metabolomics
Received 19 November 2012; revised 16 May 2013; accepted 25 May 2013

(S4, 1999/2001) conducted in 2006/2008. In all, 3080 individuals participated in the follow-up study. For 3061 individuals, metabolic data was available.[17,18] From 3061 individuals, 1144 males and 946 females aged 32–81 years were selected for further analysis after application of the following exclusion criteria: non-fasting at examination, diabetic, alcohol abstainer, missing data or outliers (that is, extreme low or high values) in metabolite concentration data (see Statistical analysis section for outlier detection calculation). Study participants were categorized according to daily alcohol intake as light drinkers (LD; females $<20\,g\,day^{-1}$ and males $<40\,g\,day^{-1}$) and moderate-to-heavy drinkers (MHD; females $\geqslant 20\,g\,day^{-1}$ and males $\geqslant 40\,g\,day^{-1}$).

KORA F3 replication data set. The KORA F3 study is a follow-up of the KORA-Survey 3 (S3, examined in 1994/95), conducted in 2004/05. The KORA F3 cohort is a 10-year follow-up survey of the KORA S3 survey. A total of 2974 individuals participated in the follow-up. From 2974 individuals, 377 individuals had metabolic data available. In all, 154 males and 107 females aged 55–84 years were selected for further analysis after the application of KORA F4 exclusion criteria. KORA F4 and KORA F3 are two independent cohorts and do not contain common participants and were conducted at different time points.[19,20]

TwinsUK replication data set. The UK Adult Twin Registry (TwinsUK) is a UK-wide twin registry sample of 11 000 adults founded in 1993 with the aim to explore the genetic epidemiology of common adult diseases.[21] A total of 629 individuals aged 23–73 years were selected for analysis after the application of KORA F4 exclusion criteria. For 277 probands, high-density lipoproteins (HDL) data were available.

Ethics statement. Written informed consent has been given by each KORA and TwinsUK participant. The KORA studies, including the protocols for subject recruitment and assessment and the informed consent for participants, were reviewed and approved by the local ethical committee (Bayerische Landesärztekammer). For the TwinsUK study, ethics approval was received from the St Thomas' Hospital Ethics.

Blood sampling. KORA F4 and F3 blood samples for metabolic analysis were collected using the similar collection procedures together with medical examinations described previously.[22–24] KORA F4 blood samples were drawn into serum tubes in the morning between 0800 and 1030 hours after overnight fasting. Tubes were gently inverted twice, followed by 30-min resting at room temperature to obtain complete coagulation. For serum collection, centrifugation of blood was performed for 10 min (2750 *g*, 15 °C). Serum was frozen at −80 °C until execution of metabolic analyses.

In the TwinsUK study, similar collection procedure was used as that in the KORA study. TwinsUK blood samples were taken after at least 6 h of overnight fasting. The samples were immediately inverted three times, followed by 40-min resting at 4 °C to obtain complete coagulation. The samples were then centrifuged for 10 min at 2000 *g*. Serum was removed from the centrifuged brown-topped tubes as the top, yellow, translucent layer of liquid. Four aliquots of 1.5 ml were placed into skirted micro centrifuge tubes and then stored in a −45 °C freezer until sampling.[25]

Metabolite measurements. Metabolomic analysis was performed on 3061 subjects from the KORA F4 study, 377 subjects from the KORA F3 study and 629 TwinsUK study. Comparison of metabolite concentrations (that is, comparison between LD and MHD) was conducted within the same cohort and within the same site of collection. The targeted metabolomic approach was based on flow injection analysis coupled with electrospray ionization tandem mass spectrometry measurements by Absolute*IDQ* p150 assay (BIOCRATES Life Sciences AG, Innsbruck, Austria). The method of Absolute*IDQ* p150 assay has been proven to be in conformance with FDA-Guideline 'Guidance for Industry—Bioanalytical Method Validation (May 2001)',[26] which implies proof of reproducibility within a given error range. The assay procedures of the Absolute*IDQ* p150 kit as well as the metabolite nomenclature have been described in detail previously.[2,27] Data evaluation for quantification of metabolite concentrations and quality assessment is performed with the Met*IQ* software package, which is an integral part of the Absolute*IDQ* kit. Internal standards serve as reference for the calculation of metabolite concentrations. To ensure data quality, each metabolite had to meet the three criteria described previously:[17,19] (1) average value of the coefficient of variance for the metabolite in the three quality controls should be smaller than 25%; (2) 90% of all the measured sample concentrations for the metabolite should be above the limit of detection; and (3) the correlation coefficient between two duplicate measurements of the metabolite in 144 re-measured samples should be above 0.5. In total, 131 metabolites passed the three quality controls, and the final metabolomics data set contained the sum of hexoses (H1), 14 amino acids, 24 acylcarnitines, 13 sphingomyelins, 34 diacylphosphatidylcholines (PCs), 37 acyl-alkyl-phosphatidylcholines and 8 lysophosphatidylcholines (lysoPCs). Supplementary Table S1 summarizes the characteristics of 163 metabolites measured in KORA F4.

Statistical analysis. Statistical analysis was performed with the open source software R (version 2.14.1). To detect outliers, concentrations obtained for the 131 metabolites were first scaled to zero mean and unity s.d. and were projected onto the unit sphere, and Mahalanobis distances for each individual were then calculated using the robust principal components algorithm.[28] Calculations were done separately for males and females. For each group, the mean Mahalanobis distance plus three times variance were defined as the cutoff. Missing values were imputed using the R package 'mice'.[29] Metabolite concentrations were logarithmized for all subsequent analysis steps. Shapiro–Wilk test[30] was applied on single metabolites to check for normal distribution of metabolites in the study population in order to choose proper follow-up tests. Mann–Whitney test[31] was applied for the comparison of two variables not satisfying normal distribution. Fisher's exact test[32] was applied for comparing binomial proportions.

Logistic regression[33] was applied on each of the 131 metabolites to investigate associations of metabolites between MHD and LD. *P*-values were corrected according to the Bonferroni correction, at a level of 3.8E − 4 (for a total use of 131 metabolites at the 5% level). To further select candidate biomarkers, two additional methods were applied:[2,8] the random forest selection[34] and the stepwise selection, which assess the metabolites as a group. Between the two groups, the random forest was first used to select the metabolites among the 30 highest ranking variables of importance score, allowing the best separation of the individuals from different groups. Age, body mass index (BMI), smoking, HDL and triglycerides were also included in this method with all the metabolites. We further selected the metabolites using stepwise selection on the logistic regression model. Metabolites with significantly different concentrations between the compared groups in logistic regression, and which were also selected using random forest, were used in this model along with all the covariates. Akaike's Information Criterion was used to evaluate the performance of these subsets of metabolites used in the models. The model with minimal Akaike's Information Criterion was chosen. The area under the receiver-operating characteristic curves (AUC) was used to evaluate the models.

Heat maps were used to illustrate the trends of metabolite concentrations with increasing alcohol consumption. Alcohol consumption data were split into alcohol consumption categories increasing by 5 g day^{-1}. A matrix of mean metabolite concentrations was calculated for each alcohol consumption category for significant male/female-specific metabolites from logistic regression. In the same procedure, step hierarchical clustering with Euclidean distance was applied on the metabolite concentration matrix to generate a hierarchical dendogram clustering metabolites with similar mean metabolite concentrations. For the meta-analysis of the combined KORA F4 and KORA F3 studies, a fixed effect model was used.

Results

Description of the study populations. Based on previous results from KORA F4, which showed strong metabolomic differences between men and women,[19] we conducted strictly sex-separated analyses. For both sexes, we classified our probands into two groups according to daily alcohol consumption of LD and MHD and compared MHD with LD (Table 1). Alcohol abstainers (ND; defined as alcohol intake of 0 g day^{-1}) were included (view Supplementary Table S3 for description of the ND group, view Supplementary Table S4 for sensitivity analysis). In general, age and BMI was comparable between MHD and LD. A significantly lower age could be observed in MHD of KORA F3 males and TwinsUK participants (*P*-value 1.3E − 02 and 1.6E − 02, respectively). BMI was significantly increased in MHD in male KORA F4 participants (*P*-value 3.3E − 03). The proportion of smokers was significantly higher in MHD in KORA F4 male and TwinsUK female populations (*P*-values 1.0E − 04 and 1.3E − 02, respectively). In all the three studies, there was a significant increase in HDL in MHD compared with LD (*P*-values 7.1E − 12–1.3E − 02). Except in KORA F3, the

Table 1 Basic characteristics of the discovery and replication data sets

	Discovery KORA F4 (n = *2090*)						*Replication KORA F3* (n = *261*)						*Replication TwinsUK* (n = *277*)			
	Male (n = *1144*)			*Female* (n = *946*)			*Male* (n = *154*)			*Female* (n = *107*)			*Female* (n = *277*)[a]			
	LD[b]	*MHD*[c]	P-*value*	*LD*	*MHD*	P-*value*	*LD*	*MHD*	P-*value*	*LD*	*MHD*	P-*value*	*LD*	*MHD*	P-*value*
Participants, *n* (%)	859 (75)	285 (25)	—	712 (75)	234 (25)	—	106 (69)	48 (31)	—	84 (78)	23 (22)	—	250 (90)	27 (10)	—
Age (years) mean ± s.d.	56.2 ± 13.7	57.1 ± 11.6	2.8E − 01[d]	54.5 ± 13.2	55.7 ± 11.5	1.8E − 01[d]	67.9 ± 7.6	64.5 ± 6.8	**1.3E − 02[d]**	65.5 ± 7.4	64.6 ± 7.7	5.3E − 01[d]	48.4 ± 10.2	43.5 ± 9.0	**1.6E − 02[d]**
BMI(kg m^{-2}) mean ± s.d.	27.6 ± 3.8	28.4 ± 4.2	**3.4E − 03[d]**	26.6 ± 4.9	26.3 ± 4.7	3.6E − 01[d]	28.1 ± 3.1	29.1 ± 3.6	1.2E − 01[d]	27.6 ± 3.7	27.5 ± 4.0	9.5E − 01[d]	25.1 ± 4.6	24.7 ± 4.1	8.1E − 01[d]
Smoker, *n* (%)	153 (18)	66 (23)	**1.0E − 04[e]**	103 (14)	44 (19)	1.2E − 01[e]	8 (7)	8 (17)	9.5E − 02[e]	4 (5)	3 (13)	1.7E − 01[e]	67 (27)	14 (52)	**1.3E − 02[e]**
HDL (mg dl^{-1}) mean ± s.d.	49.9 ± 11.3	55.8 ± 13.5	**7.1E − 12[d]**	62.7 ± 13.6	66.0 ± 15.0	**4.4E − 03[d]**	49.2 ± 13.2	59.3 ± 15.3	**1.5E − 04[d]**	65.7 ± 16.7	71.10 ± 16.2	1.4E − 01[d]	58.9 ± 14.5	66.2 ± 13.9	**1.3E − 02[d]**
Triglyceride (mg dl^{-1}) mean ± s.d.	134.5 ± 82.0	154.0 ± 104.9	**3.4E − 02[e]**	103.9 ± 71.4	100.9 ± 53.6	9.7E − 01[e]	188.0 ± 106.7	192.3 ± 130.2	9.6E − 01[e]	156.2 ± 74.1	146.4 ± 85.0	3.0E − 01[e]	99.6 ± 47.3	94.1 ± 71.4	1.1E − 01[d]

Abbreviations: BMI, body mass index; HDL, high-density lipoprotein; LD, light drinkers; MHD, moderate-to-heavy drinkers.
Significant *P*-values are in bold.
[a]629 participants were available for replication analysis; for 277 participants HDL-C and triglyceride data were available. [b]Alcohol consumption < 40g day^{-1} males, < 20g day^{-1} females. [c]Alcohol consumption ⩾ 40 g day^{-1} males, ⩾ 20 g day^{-1} females. [d]Mann–Whitney test (two-sided). [e]Fisher's exact test (two-sided), significance level $\alpha = 0.05$.

mean HDL was increased, but *P*-value was not significant. Significant increase of mean triglyceride concentration could be observed in KORA F4 male MHD only (*P*-value 3.4E − 02).

Analysis of global metabolite concentration differences between MHD and LD. We identified 40 metabolites in males and 18 metabolites in females using logistic regression analysis (adjusted for age, BMI, smoking, HDL and triglycerides) that significantly differed (*P*-value $<3.8E-4$) in concentration between MHD and LD in the KORA F4 study (view Supplementary Table S2 for detailed *P*-values and direction). To illustrate the trend of metabolite levels with increasing 5 g day^{-1} alcohol consumption increments, heat maps were displayed based on normalized mean metabolite residuals for each of the 40/18 male/female metabolites. Hierarchical clustering with Euclidean distance was used in order to find similar metabolite groups. The final clusterogram (display of dendogram and heat map) resulted in two main clusters C1 and C2 both in males and females (Figure 1). C1 consists of metabolites that increase in concentration with increasing alcohol consumption (high in MHD and low in LD). In contrast, C2 consist of metabolites that decrease in concentration with increasing alcohol consumption (low in MHD and high in LD). PC aa Cx:ys, ether lipids (PC ae Cx:ys), lysoPC a Cx:ys and sphingomyelins (SMs) occurred in both males and females. Only the acylcarnitine C16:1 occurred in males. All PC ae Cx:ys and SMs were decreased in MHD in males and females. PC aa Cx:ys were increased in MHD compared with LD in males and females (except PC aa C32:3, which was decreased in MHD in females). All lysoPC a Cx:ys were increased in MHD in males and females (except lysoPC a C17:0).

The logistic regression analysis was based on each single metabolite, and some of these 40/18 male/female metabolites are expected to correlate with each other. To find more specific and independent metabolites that best separate MHD from LD as potential biomarkers for alcohol-consumption, we further applied Random Forest and Stepwise Selection method. Ten metabolites in males (PC aa C32:1, PC aa C36:1, PC aa C36:5, PC aa C40:4, PC ae C40:6, lysoPC a C17:0, lysoPC a C18:1, SM (OH) C22:1, SM (OH) C22:2, SM (OH) C16:1) and five metabolites in females (PC aa C34:1, PC ae C30:2, PC ae C40:4, lysoPC a C16:1, lysoPC a C17:0) were further selected (Figure 1). To evaluate the model of the combination of the 10/5 male/female specific metabolites with covariates (that is, how good does the logistic regression model adjusted for age, BMI, smoking, HDL and triglycerides distinguish between MHD and LD), the AUC was calculated. The AUC value in males was 0.812 and in females 0.679 (Figure 1).

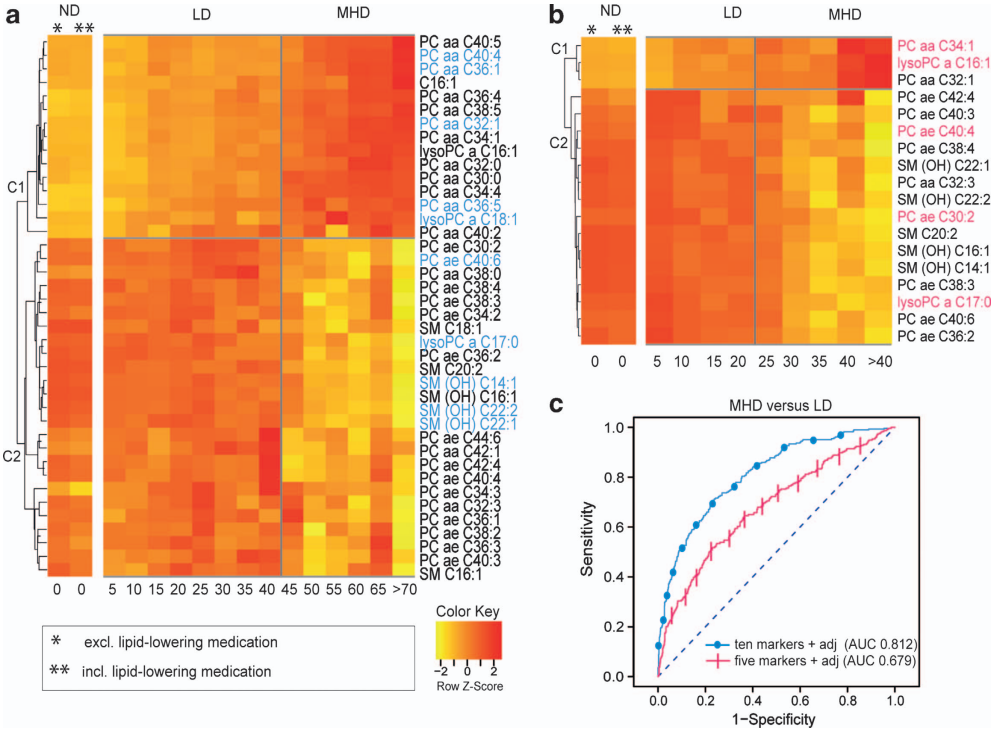

Figure 1 Alcohol-specific metabolomic profiles. Clusterograms show 40 and 18 metabolite concentrations in relation to alcohol consumption in light drinkers (LD) and moderate-to-heavy drinkers (MHD) in (**a**) males and (**b**) females, respectively. The additional two-column clusterogram shows the effect of lipid-lowering medication (that is, statins, fibrates, herbal-based lipid-lowering agents) on metabolite concentrations in non-drinkers (ND). Relative concentration of metabolites are represented by x-fold s.d. from overall mean concentrations for groups of alcohol consumption of 5 g day^{-1}. Horizontal axis displays the alcohol concentration in g day^{-1}, while vertical axis represent hierarchical clustering. The 10/5 most significant metabolites separating MHD from LD in males/females are highlighted in blue and pink. (**c**) Graphic shows receiver operating characteristic (ROC) curves for the set of most significant 10/5 metabolites in males (PC aa C32:1, PC aa C36:1, PC aa C36:5, PC aa C40:4, PC ae C40:6, lysoPC a C17:0, lysoPC a 18:1, SM (OH) C22:1, SM (OH) C22:2, SM (OH) C16:1) and females (PC aa C34:1, PC ae C30:2, PC ae C40:4, lysoPC a C16:1, lysoPC a 17:0). ROC curve displayed as dotted/crossed line represent marker performance in males/females. The area under the ROC curve was calculated for the combined metabolite panel with adjustment for age, body mass index, smoking status, high-density lipoproteins and triglycerides.

Table 2 Results of logistic regression analysis of alcohol-specific metabolites in males

	Discovery KORA F4 (n = *1144*)			*Replication KORA F3* (n = *154*)			*Meta analysis discovery + replication fixed effects* (n = *1298*)		
Metabolite	*LD*[a] *mean ± s.d.* (*μM*)[b]	*MHD*[c] *mean ± s.d.* (*μM*)[b]	P-*value*[d,e]	*LD mean ± s.d.* (*μM*) [b]	*MHD mean ± s.d.* (*μM*) [b]	P-*value*[d,f]	*LD mean ± s.d.* (*μM*) [b]	*MHD mean ± s.d.* (*μM*) [b]	P-*value*[d,f]
PC aa C32:1	19.3 ± 9.2	31.0 ± 16.9	**1.1E − 18**	20.7 ± 9.7	35.0 ± 16.5	**7.3E − 05**	19.4 ± 9.2	31.6 ± 16.8	**8.6E − 23**
PC aa C36:5	28.7 ± 12.9	36.6 ± 17.6	**3.9E − 07**	31.8 ± 19.8	44.0 ± 278	4.9E − 02	29.0 ± 13.9	37.7 ± 19.5	**5.6E − 08**
PC aa C40:4	4.0 ± 1.2	4.9 ± 1.6	**1.2E − 07**	3.7 ± 1.0	4.7 ± 1.5	1.6E − 02	4.0 ± 1.2	4.8 ± 1.6	**2.8E − 09**
PC aa C36:1	51.4 ± 11.9	59.7 ± 16.4	**6.6E − 06**	53.0 ± 11.2	66.6 ± 15.9	**9.9E − 04**	51.5 ± 11.8	60.7 ± 16.5	**9.9E − 09**
lysoPC a C17:0	1.7 ± 0.5	1.5 ± 0.5	**5.4E − 10**	1.6 ± 0.4	1.3 ± 0.4	3.4E − 02	1.7 ± 0.5	1.5 ± 0.5	**4.1E − 11**
lysoPC a C18:1	20.0 ± 5.5	22.7 ± 7.4	**1.6E − 05**	18.5 ± 4.9	20.9 ± 5.0	1.3E − 01	19.9 ± 5.4	22.5 ± 7.1	**3.9E − 06**
PC ae C40:6	4.9 ± 1.2	4.4 ± 1.1	**4.3E − 10**	5.4 ± 1.3	5.0 ± 1.1	3.5E − 02	4.9 ± 1.2	4.5 ± 1.1	**1.5E − 10**
SM(OH) C16:1	3.1 ± 0.7	2.7 ± 0.7	**6.8E − 12**	2.7 ± 0.7	2.4 ± 0.6	**2.0E − 03**	3.0 ± 0.7	2.6 ± 0.7	**1.2E − 13**
SM(OH) C22:1	12.6 ± 2.5	11.4 ± 2.9	**3.4E − 09**	10.0 ± 2.5	9.3 ± 2.1	1.2E − 02	12.3 ± 2.6	11.1 ± 2.9	**4.4E − 10**
SM(OH) C22:2	10.2 ± 2.1	9.1 ± 2.3	**4.0E − 13**	8.1 ± 2.1	7.7 ± 1.7	2.3E − 02	9.9 ± 2.2	8.9 ± 2.3	**1.3E − 13**

Abbreviations: LD, light drinkers; MHD, moderate-to-heavy drinkers.
[a]Alcohol consumption $<40\,g\,day^{-1}$ males, $<20\,g\,day^{-1}$ females. [b]Mean and s.d. of the metabolite concentration from serum. [c]Alcohol consumption $\geqslant 40\,g\,day^{-1}$ males, $\geqslant 20\,g\,day^{-1}$ females. [d]Logistic regression analysis adjusted for age, body mass index, smoking, high-density lipoproteins and triglycerides. [e]Significance level <0.00038 (Bonferroni corrected). [f]Significance level <0.005 males (Bonferroni corrected). Significant *P*-values are represented in bold.

Table 3 Results from logistic regression analysis of alcohol-specific metabolites in females

	Discovery KORA F4 (n = *946*)			*Replication KORA F3* (n = *107*)			*Meta analysis discovery + replication fixed effect* (n = *1053*)[a]			*Replication TwinsUK* (n = *277*)		
Metabolite	*LD*[b] *mean ± s.d.* (*μM*)[c]	*MHD*[d] *mean ± s.d.* (*μM*)[c]	P-*value*[e,f]	*LD mean ± s.d.* (*μM*)[c]	*MHD mean ± s.d.* (*μM*)[c]	P-*value*[e,g]	*LD mean ± s.d.* (*μM*)[c]	*MHD mean ± s.d.* (*μM*)[c]	P-*value*[e,g]	*LD mean ± s.d.* (*μM*)[c]	*MHD mean ± s.d.* (*μM*)[c]	P-*value*[e,g]
PC aa C34:1	241.1 ± 52.9	259.4 ± 56.3	**1.0E − 04**	274.8 ± 72.8	308.5 ± 73.6	5.8E − 02	244.6 ± 56.2	263.7 ± 59.6	**6.6E − 05**	311.5 ± 98.8	350.4 ± 121.0	**9.4E − 03**[h]
lysoPC a C16:1	2.8 ± 0.9	3.1 ± 1.0	**4.7E − 05**	2.5 ± 1.0	3.04 ± 0.9	5.2E − 02	2.7 ± 0.8	3.1 ± 1.0	**6.2E − 06**	4.21 ± 1.5	4.24 ± 1.7	7.3E − 01
lysoPC a C17:0	1.8 ± 0.5	1.6 ± 0.5	**2.8E − 04**	1.6 ± 0.5	1.4 ± 0.4	2.3E − 02	1.7 ± 0.5	1.6 ± 0.4	**7.7E − 05**	2.3 ± 0.9	2.2 ± 0.7	5.6E − 01
PC ae C30:2	0.17 ± 0.04	0.16 ± 0.04	**4.0E − 05**	0.17 ± 0.03	0.15 ± 0.04	**2.5E − 03**	0.16 ± 0.04	0.15 ± 0.03	**4.2E − 06**	0.2 ± 0.1	0.2 ± 0.1	7.8E − 01
PC ae C40:6	5.4 ± 1.3	5.0 ± 1.2	**2.4E − 07**	6.1 ± 1.5	5.4 ± 1.4	2.2E − 02	5.4 ± 1.3	5.0 ± 1.2	**1.7E − 08**	7.5 ± 2.7	7.4 ± 2.7	9.9E − 01

Abbreviations: LD, light drinkers; MHD, moderate-to-heavy drinkers.
[a]Meta analysis consist of KORA F4 discovery, KORA F3 and TwinsUK replication data sets. [b]Alcohol consumption $<40\,g\,day^{-1}$ males, $<20\,g\,day^{-1}$ females. [c]Mean and s.d. of the metabolite concentration from serum. [d]Alcohol consumption $\geqslant 40\,g\,day^{-1}$ males, $\geqslant 20\,g\,day^{-1}$ females. [e]Logistic regression analysis adjusted for age, body mass index, smoking, high-density lipoproteins and triglycerides. [f]Significance level <0.00038 (Bonferroni corrected). [g]Significance level <0.01 females (Bonferroni corrected). [h]Logistic regression analysis adjusted for age, BMI, smoking with $n = 629$ study participants. Significant *P*-values are represented in bold.

Replication analysis in two independent cohorts. Replication analysis of the most significant 10 alcohol-related metabolites in males and five metabolites in females found in KORA F4 discovery sample was performed in two independent KORA F3 and TwinsUK cohorts (Tables 2 and 3). In males, 3 out of 10 metabolites (that is, PC aa C32:1, PC aa C36:1, SM (OH) C16:1) could be replicated in KORA F3 (Table 2). In females, two out of five metabolites could be replicated (Table 3); one metabolite in KORA F3 (that is, PC ae C30:2) and one metabolite (that is, PC aa C34:1) in TwinsUK. In the TwinsUK population, only females were available for replication analysis. In all, 629 TwinsUK participants met the inclusion criteria and were eligible for the replication analysis; however, only for 277 participants HDL and triglyceride data were available for the same time point. In TwinsUK, we performed the replication analysis using 277 and 629 study participants. In the first replication analysis on 277 participants, logistic regression adjusted for age, BMI and smoking, HDL and triglyceride resulted in no significant *P*-values. When we increased the sample size to 629 and used the logistic regression model adjusted for age, BMI and smoking, the metabolite PC aa C34:1 could be replicated.

Additionally, we pooled data from the KORA F4 discovery and KORA F3 replication samples and conducted a meta-analysis with a fixed effect model in order to investigate the combined effect of alcohol on metabolite concentrations. In the meta-analysis, the replication succeeded for all 10 metabolites in men and 5 metabolites in women. This indicates that due to the small sample size in TwinsUK and KORA F3 cohorts the previous replication could not be achieved for all metabolites. Nevertheless, the trends of metabolite concentrations (as stated by the comparison of means of metabolite concentrations between MHD and LD in Tables 2 and 3) for all 10 and 5 metabolites are consistent with the trends in the discovery across all studies. For example, the metabolite lysoPC a C18:1 was not replicated in KORA F3 and TwinsUK, still the mean metabolite concentration is higher in MHD compared with LD throughout the KORA F4, KORA F3 and TwinsUK studies.

Discussion

In the current study, we used a targeted metabolomics approach and identified, as well as partly replicated, alcohol-related metabolites in German and UK human studies. Our results suggest that alcohol affects mostly the sphingolipid, glycerophospholipid and ether lipid metabolism. A schematic overview of the observed alcohol-specific metabolic

differences and the potential underlying mechanisms is depicted in Figure 2 and are discussed below.

The underlying mechanism for lower sphingomyelin concentrations (SM(OH)C16:1, SM(OH)C22:1, SM(OH)C22:1) in MHD compared with LD could be attributed to acid sphingomyelinase (ASM) activity. ASM catalyzes the hydrolysis of sphingomyelins by cleaving the phosphodiester bond of sphingomyelins generating ceramide and phosphorylcholine,[35,36] which is again reassembled to phosphatidylcholine.[3] Enzymatic dysfunction of ASM results in Niemann–Pick disease A (NPD-A, OMIM 257200) and B (NPD-B, OMIM 607616), a lipid storage disease characterized by accumulation of sphingomyelins within the endo-lysosomal compartment.[37] Interestingly, this mechanism is reciprocal when alcohol is administered. Several studies investigating cellular response to alcohol *in vitro* and *in vivo* have provided evidence that alcohol stimulates the ASM activity leading to accumulation of ceramide and decrease of sphingomyelins.[36,38–41] A recent *in vivo* study on patients with alcohol dependence reported alcohol-induced release of phosphorylcholine from sphingomyelins in the peripheral blood cells confirming alcohol-induced activation of ASM.[42]

There is a direct correlation between PC concentrations and phosphatidylethanol (PEth). PEth is a clinical biomarker of the past 1–2 weeks of moderate-to-heavy alcohol consumption.[43] PEth is a unique phospholipid that is synthesized only in the presence of ethanol and is directly formed from PCs by the enzyme phospholipase D[44–46] that catalyzes the exchange of ethanol for choline in PCs.[46] Different PEth molecular species have a common phosphoethanol head group onto which two fatty acid moieties derived from PCs are attached. A study by Helander and Zheng[47] has shown that PEth-16:0/18:1 (34:1) was the most predominant molecular species accounting for 37% of all PEth species. A recent study by Nalesso *et al.*[48] compared the occurrence of different PEth species between heavy drinkers and social drinkers (defined as daily alcohol intake 60–300 and 0–20 g day^{-1}, respectively). Interestingly, PEth 16:0/18:1 (34:1), PEth 18:0/18:1 (36:1) and PEth 16:0/16:1 (C32:1) were most abundant in heavy drinkers. This may be consistent with our findings in which PC aa C34:1 in female, PC aa C36:1 and PC aa 32:1 in male had higher concentration in MHD compared with that in LD. We hypothesize that concentrations of specific PC species can be used as surrogate biomarkers for PEth to distinguish MHD from LD. However, PEth measurements are out of scope of this study. Dedicated and parallel measurements of PC aa C34:1 and PEth (34:1) would be required in order to investigate whether PC aa C 34:1 can be a substitute PEth (34:1).

lysoPCs are derived from PCs[49] and have been reported to have cytotoxic effects.[50] They accumulate in alcohol-related conditions as in atherosclerosis[51] or ischaemia.[52] LysoPCs originate from several metabolic pathways, as part of the production is attributed to the transesterification of PCs and free cholesterol catalysed by the enzyme lecithin-cholesterol acyltransferase (LCAT), where LCAT hydrolyses the sn-2 acyl group and subsequently transfers and esterifies the fatty acid to free cholesterol.[53] A study by Goto *et al.*,[54] investigating clinical alcoholics, reported an increase of LCAT concentration in individuals with alcohol intake of >30 g day^{-1}. Another metabolic pathway generating lysoPC species is attributed to

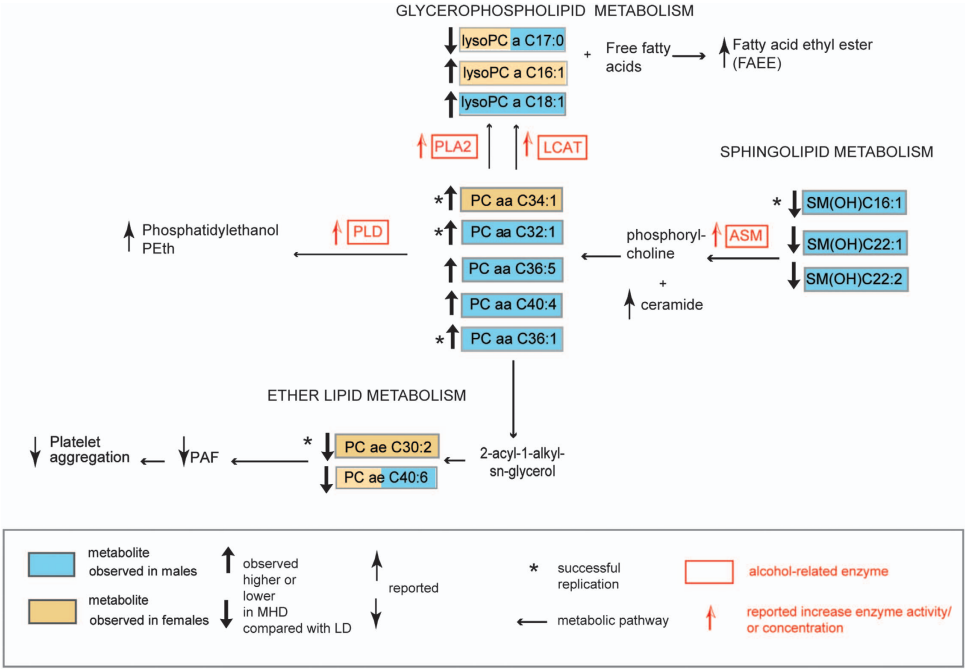

Figure 2 Schematic overview of metabolite concentration differences in moderate-to-heavy drinkers (MHD) compared with light drinkers (LD) in males and females. Ten/five metabolites that best discriminate MHD from LD in males/females are shown. Yellow and blue boxes represent male- and female-specific alcohol-related metabolites identified in this study. Combined yellow-blue boxes represent metabolites identified both in males and females. Bold black arrows represent observed higher or lower of metabolite concentration in MHD compared with LD in the discovery. Replicated metabolites are marked by a star. Thin black arrows represent the higher or lower of alcohol-related analytes in MHD reported in earlier publications. Red boxes represent alcohol-related enzymes and red arrows represent the effect on the respective enzyme activity or concentration reported in previous publications in MHD. ASM, acid sphingomyelinase; LCAT, lecithin-cholesterol acyltransferase; PAF, platelet-activating factor; PLA2, phospholipase A2; PLD, phospholipase D.

the enzyme phospholipase A2, which catalyzes the hydrolysis of an ester bond at the sn-2 position of 1,2-sn-diacylglycerols yielding lysoPCs and free fatty acids,[55] which are esterified into fatty acid ethyl esters that have been reported as alcohol marker to distinguish social from heavy drinkers or alcohol-dependent individuals.[56,57]

Fatty acids with uneven number of carbons (that is, C15:0 and C17:0) are produced by bacterial flora of human intestine.[58] It is known that alcohol acts as a disinfectant which kills bacteria. Thus a possible explanation for the lower concentrations of lysoPC a C17:0 in MHD could be that alcohol consumption leads to the disruption of the respective intestinal bacterial microflora in the gut which thus influences lysoPC a C17:0 levels in human blood. On the other hand, the fatty acid C17:0 is also found in the bacterial flora of ruminants.[59,60] A study by Wolk *et al.*[61] revealed that portions of the fatty acids C15:0 and C17:0 in adipose tissue reflected milk fat consumption in women. An earlier study[62] investigating associations of reported alcohol intake with dietary habits in probands from the EPIC cohort found that alcohol consumers had a lower intake of dairy products than abstainers. This is consistent with another French cohort of the EPIC study,[63] which found that high alcohol intake was associated with lower consumption of dairy products in both genders compared with moderate alcohol consumption. Thus another plausible explanation to the lower concentrations of lysoPC a C17:0 in MHD in our study could be based on lower intake of dairy products. Based on the above findings and explanations, lysoPC a C17:0 might also be a dietary biomarker associated with distinguished dietary behavior of MHD compared with LD rather than a biomarker for alcohol-induced toxic or inflammatory mechanisms.

Ether lipids (for example, PC ae C30:2 and PC ae C40:6) have a role as precursor of platelet-activating factor.[64,65] Platelet-activating factor is an important mediator in hemostasis and has an important role in platelet aggregation (that is, thrombotic effects). A number of studies indicate that ethanol directly affects hemostasis via a number of mechanisms, including platelet aggregation and activation.[66–69] This mechanism is still not fully understood; however, based on our results, it can be hypothesized that reduced platelet-activating factor levels in response to moderate-to-heavy alcohol consumption might form a bottleneck in the process of platelet activation leading to poor platelet aggregation and to alcohol-related hemorrhagic events. This is supported by studies from the United States and Sweden showing that the baseline incidence of acute upper gastrointestinal bleeding increased by threefold as alcohol consumption increased from $\leqslant$1 drink to $>$20 drinks per week.[70]

Conclusion and outlook. Our study provides new insights into the impact of alcohol consumption on human metabolism. Our results suggest that metabolomic profiles based on PCs, lysoPCs, ether lipids and sphingolipids form a new class of biomarkers for alcohol consumption. This may be of great value for the clinical assessment of alcohol use, alcohol-specific disease detection and drug-therapy monitoring. Side effects of alcohol consumption on specific organs as liver could be investigated by future studies using an association study analysing metabolite concentrations in relation to concentrations of liver biomarkers as, for example, γ-glutamyltransferase.[71] The current analysis is based on a targeted metabolomics approach that is limited to a subset of 131 currently known metabolites in human (for example, lipid metabolism, amino acid metabolism). A study using a broader metabolomics approach that quantifies a bigger number of metabolites would be needed to investigate alcohol effects on other areas of metabolism. Further research is needed to elucidate the exact underlying mechanisms. A prospective study in large sample would help validate the predictive potential of these results.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements. We express our appreciation to all KORA and TwinsUK study participants for donating their blood and time. We thank the field staff in Augsburg conducting the KORA studies. The KORA Study group consists of A. Peters (speaker), J. Heinrich, R. Holle, R. Leidl, C. Meisinger, K. Strauch and their co-workers, who are responsible for the design and conduct of the KORA studies. We thank the staff from the Institute of Epidemiology at the Helmholtz Zentrum München and the Genome Analysis Center Metabolomics Platform who helped in the sample logistics, data and straw collection and metabolomic measurements and especially J. Scarpa, K. Sckell, B. Hochstrat, F. Scharl, N. Lindemann, H. Chavez and A. Sabunchi. Part of this project was supported by the German Federal Ministry of Education and Research (BMBF) by funding the project 'Metabolomics of ageing' (FKZ: 01DO12030), the German Center for Diabetes (DZD) and the NGFN, as well as by EU FP7 Grant HEALTH-2009-2.2.1-3/242114 (Project OPTiMiSE). The TwinsUK study was funded by the Wellcome Trust, EU FP7 Grant HEALTH-F2-2008-201865-GEFOS and the FP-5 GenomEUtwin Project (QLG2-CT-2002-01254). The study also received support from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. KS is supported by 'Biomedical Research Program' funds at Weill Cornell Medical College in Qatar, a program funded by the Qatar Foundation. TDS is an NIHR senior Investigator.

1. Kaddurah-Daouk R, Bogdanov MB, Wikoff WR, Zhu H, Boyle SH, Churchill E *et al.* Pharmacometabolomic mapping of early biochemical changes induced by sertraline and placebo. *Translational Psychiatry* 2013; **3**: e223.
2. He Y, Yu Z, Giegling I, Xie L, Hartmann AM, Prehn C *et al.* Schizophrenia shows a unique metabolomics signature in plasma. *Translational Psychiatry* 2012; **2**: e149.
3. Kaddurah-Daouk R, McEvoy J, Baillie RA, Lee D, Yao JK, Doraiswamy PM *et al.* Metabolomic mapping of atypical antipsychotic effects in schizophrenia. *Mol Psychiatry* 2007; **12**: 934–945.
4. Oresic M, Hyotylainen T, Herukka SK, Sysi-Aho M, Mattila I, Seppanan-Laakso T *et al.* Metabolome in progression to Alzheimer's disease. *Translational Psychiatry* 2011; **1**: e57.
5. World Health O. *Global status report on alcohol and health* Geneva, Switzerland, 2011.
6. Kapur A, Wild G, Milford-Ward A, Triger DR. Carbohydrate deficient transferrin: a marker for alcohol abuse. *BMJ* 1989; **299**: 427–431.
7. Peterson K. Biomarkers for alcohol use and abuse—a summary. *Alcohol Res Health* 2004; **28**: 30–37.
8. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y *et al.* Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol* 2012; **8**: 615.
9. Fiehn O. Metabolomics—the link between genotypes and phenotypes. *Plant Mol Biol* 2002; **48**: 155–171.
10. Krug S, Kastenmuller G, Stuckler F, Rist MJ, Skurk T, Sailer M *et al.* The dynamic range of the human metabolome revealed by challenges. *FASEB J* 2012; **26**: 2607–2619.
11. Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wagele B *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 2011; **477**: 54–60.
12. Griffiths WJ, Koal T, Wang Y, Kohl M, Enot DP, Deigner HP. Targeted metabolomics for biomarker discovery. *Angew Chem Int Ed Engl* 2010; **49**: 5426–5445.
13. Dudley E, Yousef M, Wang Y, Griffiths WJ. Targeted metabolomics and mass spectrometry. *Adv Protein Chem Struct Biol* 2010; **80**: 45–83.
14. Holle R, Happich M, Lowel H, Wichmann HE. KORA—a research platform for population based health research. *Gesundheitswesen* 2005; **67**(Suppl 1): S19–S25.
15. Lowel H, Doring A, Schneider A, Heier M, Thorand B, Meisinger C. The MONICA Augsburg surveys—basis for prospective cohort studies. *Gesundheitswesen* 2005; **67**(Suppl 1): S13–S18.

16. Wichmann HE, Gieger C, Illig T. KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* 2005; **67**(Suppl 1): S26–S30.
17. Yu Z, Zhai G, Singmann P, He Y, Xu T, Prehn C *et al.* Human serum metabolic profiles are age dependent. *Aging Cell* 2012; **11**: 960–967.
18. Xu T, Holzapfel C, Dong X, Bader E, Yu Z, Prehn C *et al.* Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. *BMC Med* 2013; **11**: 60.
19. Mittelstrass K, Ried JS, Yu Z, Krumsiek J, Gieger C, Prehn C *et al.* Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet* 2011; **7**: e1002215.
20. Yu Z, Kastenmuller G, He Y, Belcredi P, Moller G, Prehn C *et al.* Differences between human plasma and serum metabolite profiles. *PLoS One* 2011; **6**: e21230.
21. Moayyeri A, Hammond CJ, Valdes AM, Spector TD. Cohort profile: TwinsUK and Healthy Ageing Twin Study. *Int J Epidemiol* 2012; **42**: 76–85.
22. Wang-Sattler R, Yu Y, Mittelstrass K, Lattka E, Altmaier E, Gieger C *et al.* Metabolic profiling reveals distinct variations linked to nicotine consumption in humans—first results from the KORA study. *PLoS One* 2008; **3**: e3863.
23. Illig T, Gieger C, Zhai G, Romisch-Margl W, Wang-Sattler R, Prehn C *et al.* A genome-wide perspective of genetic variation in human metabolism 63. *Nat Genet* 2010; **42**: 137–141.
24. Jourdan C, Petersen AK, Gieger C, Doring A, Illig T, Wang-Sattler R *et al.* Body fat free mass is associated with the serum metabolite profile in a population-based study. *PLoS One* 2012; **7**: e40009.
25. Zhai G, Wang-Sattler R, Hart DJ, Arden NK, Hakim AJ, Illig T *et al.* Serum branched-chain amino acid to histidine ratio: a novel metabolomic biomarker of knee osteoarthritis. *Ann Rheum Dis* 2010; **69**: 1227–1231.
26. Health USDo, Human S, Food, Drug A, Center for Drug E, Research. Guidance for Industry—Bioanalytical Method Validation 2001.
27. Römisch-Margl W, Prehn C, Bogumil R, Röhring C, Suhre K, Adamski J. Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics* 2011; **8**: 133–142.
28. Filzmoser P, Maronna R, Werner M. Outlier identification in high dimensions. *Comput Stat Data Anal* 2008; **52**: 1694–1711.
29. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Software* 2011; **45**: 1–67.
30. Royston P. An extension of Shapiro and Wilk's W test for normality to large samples. *Appl Stat* 1982; **31**: 115–124.
31. Bauer DF. Constructing confidence sets using rank statistics. *J Am Stat Assoc* 1972; **67**: 687–690.
32. Agresti A. *Categorical Data Analysis*. 2nd edn. Chapter 3.5.1. Fisher's Exact Test for 2×2 Tables. Wiley: New York, NY, USA, 2002, pp 91–101.
33. Hastie TJ, Pregibon D. Generalized linear models. In: Chambers JM, Hastie TJ (eds). *Statistical Models in S*. Wadsworth & Brooks/Cole: Pacific Grove, CA, USA, 1992.
34. Breiman L. Random forests. *Machine Learn* 2001; **45**: 5–32.
35. Jenkins RW, Canals D, Hannun YA. Roles and regulation of secretory and lysosomal acid sphingomyelinase. *Cell Signal* 2009; **21**: 836–846.
36. Liu JJ, Wang JY, Hertervig E, Cheng Y, Nilsson A, Duan RD. Activation of neutral sphingomyelinase participates in ethanol-induced apoptosis in Hep G2 cells. *Alcohol Alcohol* 2000; **35**: 569–573.
37. Li Z, Hailemariam TK, Zhou H, Li Y, Duckworth DC, Peake DA *et al.* Inhibition of sphingomyelin synthase (SMS) affects intracellular sphingomyelin accumulation and plasma membrane lipid organization. *Biochim Biophys Acta* 2007; **1771**: 1186–1194.
38. Deaciuc IV, Nikolova-Karakashian M, Fortunato F, Lee EY, Hill DB, McClain CJ. Apoptosis and dysregulated ceramide metabolism in a murine model of alcohol-enhanced lipopolysaccharide hepatotoxicity. *Alcohol Clin Exp Res* 2000; **24**: 1557–1565.
39. Pascual M, Valles SL, Renau-Piqueras J, Guerri C. Ceramide pathways modulate ethanol-induced cell death in astrocytes. *J Neurochem* 2003; **87**: 1535–1545.
40. Saito M, Saito M, Cooper TB, Vadasz C. Ethanol-induced changes in the content of triglycerides, ceramides, and glucosylceramides in cultured neurons. *Alcohol Clin Exp Res* 2005; **29**: 1374–1383.
41. Viktorov AV, Yurkiv VA. Effects of ethanol and lipopolysaccharide on the sphingomyelin cycle in rat hepatocytes. *Bull Exp Biol Med* 2008; **146**: 753–755.
42. Reichel M, Greiner E, Richter-Schmidinger T, Yedibela O, Tripal P, Jacobi A *et al.* Increased acid sphingomyelinase activity in peripheral blood cells of acutely intoxicated patients with alcohol dependence. *Alcohol Clin Exp Res* 2010; **34**: 46–50.
43. Stewart SH, Reuben A, Brzezinski WA, Koch DG, Basile J, Randall PK *et al.* Preliminary evaluation of phosphatidylethanol and alcohol consumption in patients with liver disease and hypertension. *Alcohol Alcohol* 2009; **44**: 464–467.
44. Gustavsson L, Alling C. Formation of phosphatidylethanol in rat brain by phospholipase D. *Biochem Biophys Res Commun* 1987; **142**: 958–963.
45. Kobayashi M, Kanfer JN. Phosphatidylethanol formation via transphosphatidylation by rat brain synaptosomal phospholipase D. *J Neurochem* 1987; **48**: 1597–1603.
46. Mueller GC, Fleming MF, LeMahieu MA, Lybrand GS, Barry KJ. Synthesis of phosphatidylethanol—a potential marker for adult males at risk for alcoholism. *Proc Natl Acad Sci USA* 1988; **85**: 9778–9782.
47. Helander A, Zheng Y. Molecular species of the alcohol biomarker phosphatidylethanol in human blood measured by LC-MS. *Clin Chem* 2009; **55**: 1395–1405.
48. Nalesso A, Viel G, Cecchetto G, Mioni D, Pessa G, Favretto D *et al.* Quantitative profiling of phosphatidylethanol molecular species in human blood by liquid chromatography high resolution mass spectrometry. *J Chromatogr A* 2011; **1218**: 8423–8431.
49. Croset M, Brossard N, Polette A, Lagarde M. Characterization of plasma unsaturated lysophosphatidylcholines in human and rat. *Biochem J* 2000; **345**(Pt 1): 61–67.
50. Weltzien HU. Cytolytic and membrane-perturbing properties of lysophosphatidylcholine. *Biochim Biophys Acta* 1979; **559**: 259–287.
51. Matsumoto T, Kobayashi T, Kamata K. Role of lysophosphatidylcholine (LPC) in atherosclerosis. *Curr Med Chem* 2007; **14**: 3209–3220.
52. Wang J, Wang H, Han H, Zhang Y, Yang B, Nattel S *et al.* Phospholipid metabolite 1-palmitoyl-lysophosphatidylcholine enhances human ether-a-go-go-related gene (HERG) K(+) channel function. *Circulation* 2001; **104**: 2645–2648.
53. Holleboom AG, Kuivenhoven JA, Vergeer M, Hovingh GK, van Miert JN, Wareham NJ *et al.* Plasma levels of lecithin:cholesterol acyltransferase and risk of future coronary artery disease in apparently healthy men and women: a prospective case-control analysis nested in the EPIC-Norfolk population study. *J Lipid Res* 2010; **51**: 416–421.
54. Goto A, Sasai K, Suzuki S, Fukutomi T, Ito S, Matsushita T *et al.* Plasma concentrations of LPL and LCAT are in putative association with females and alcohol use which are independent negative risk factors for coronary atherosclerosis among Japanese. *Clin Chim Acta* 2003; **329**: 69–76.
55. Caro AA, Cederbaum AI. Role of cytochrome P450 in phospholipase A2- and arachidonic acid-mediated cytotoxicity. *Free Radic Biol Med* 2006; **40**: 364–375.
56. Wurst FM, Alexson S, Wolfersdorf M, Bechtel G, Forster S, Alling C *et al.* Concentration of fatty acid ethyl esters in hair of alcoholics: comparison to other biological state markers and self reported-ethanol intake. *Alcohol Alcohol* 2004; **39**: 33–38.
57. Salem RO, Refaai MA, Cluette-Brown JE, Russo JW, Laposata M. Fatty acid ethyl esters in liver and adipose tissues as postmortem markers for ethanol intake. *Clin Chem* 2001; **47**: 722–725.
58. Hopkins MJ, Sharp R, Macfarlane GT. Age and disease related changes in intestinal bacterial populations assessed by cell culture, 16S rRNA abundance, and community cellular fatty acid profiles. *Gut* 2001; **48**: 198–205.
59. Smedman AE, Gustafsson IB, Berglund LG, Vessby BO. Pentadecanoic acid in serum as a marker for intake of milk fat: relations between intake of milk fat and metabolic risk factors. *Am J Clin Nutr* 1999; **69**: 22–29.
60. Wu Z, Palmquist DL. Synthesis and biohydrogenation of fatty acids by ruminal microorganisms in vitro. *J Dairy Sci* 1991; **74**: 3035–3046.
61. Wolk A, Vessby B, Ljung H, Barrefors P. Evaluation of a biological marker of dairy fat intake. *Am J Clin Nutr* 1998; **68**: 291–295.
62. Ruf T, Nagel G, Altenburg HP, Miller AB, Thorand B. Food and nutrient intake, anthropometric measurements and smoking according to alcohol consumption in the EPIC Heidelberg study. *Ann Nutr Metab* 2005; **49**: 16–25.
63. Kesse E, Clavel-Chapelon F, Slimani N, van Liere M. Do eating habits differ according to alcohol consumption? Results of a study of the French cohort of the European Prospective Investigation into Cancer and Nutrition (E3N-EPIC). *Am J Clin Nutr* 2001; **74**: 322–327.
64. Nagan N, Zoeller RA. Plasmalogens: biosynthesis and functions. *Prog Lipid Res* 2001; **40**: 199–229.
65. Snyder F. Platelet-activating factor: the biosynthetic and catabolic enzymes. *Biochem J* 1995; **305**(Pt 3): 689–705.
66. Dimmitt SB, Rakic V, Puddey IB, Baker R, Oostryck R, Adams MJ *et al.* The effects of alcohol on coagulation and fibrinolytic factors: a controlled trial. *Blood Coagul Fibrinolysis* 1998; **9**: 39–45.
67. Hillbom M, Neiman J. Platelet thromboxane formation capacity after ethanol withdrawal in chronic alcoholics. *Haemostasis* 1988; **18**: 170–178.
68. Mikhailidis DP, Jenkins WJ, Barradas MA, Jeremy JY, Dandona P. Platelet function defects in chronic alcoholism. *Br Med J (Clin Res Ed)* 1986; **293**: 715–718.
69. Salem RO, Laposata M. Effects of alcohol on hemostasis. *Am J Clin Pathol* 2005; **123**(Suppl): S96–105.
70. Kaufman DW, Kelly JP, Wiholm BE, Laszlo A, Sheehan JE, Koff RS *et al.* The risk of acute major upper gastrointestinal bleeding among users of aspirin and ibuprofen at various levels of alcohol consumption. *Am J Gastroenterol* 1999; **94**: 3189–3196.
71. Ruckert IM, Heier M, Rathmann W, Baumeister SE, Doring A, Meisinger C. Association between markers of fatty liver disease and impaired glucose regulation in men and women from the general population: the KORA-F4-study. *PLoS One* 2011; **6**: e22932.

Supplementary Information accompanies the paper on the Translational Psychiatry website (http://www.nature.com/tp)

Effects of Metformin on Metabolite Profiles and LDL Cholesterol in Patients With Type 2 Diabetes

DOI: 10.2337/dc15-0658

Tao Xu,[1,2] Stefan Brandmaier,[1,2] Ana C. Messias,[3] Christian Herder,[4,5] Harmen H.M. Draisma,[6,7,8] Ayse Demirkan,[9,10] Zhonghao Yu,[1,2] Janina S. Ried,[11] Toomas Haller,[12] Margit Heier,[2] Monica Campillos,[13] Gisela Fobo,[13] Renee Stark,[14] Christina Holzapfel,[15] Jonathan Adam,[1,2] Shen Chi,[1,2] Markus Rotter,[1,2] Tommaso Panni,[2] Anne S. Quante,[11,16] Ying He,[17,18] Cornelia Prehn,[19] Werner Roemisch-Margl,[13] Gabi Kastenmüller,[13] Gonneke Willemsen,[6,7] René Pool,[6,7] Katarina Kasa,[9] Ko Willems van Dijk,[10] Thomas Hankemeier,[20] Christa Meisinger,[2] Barbara Thorand,[2] Andreas Ruepp,[13] Martin Hrabé de Angelis,[21,22,23] Yixue Li,[17,18] H.-Erich Wichmann,[2,16,24] Bernd Stratmann,[25] Konstantin Strauch,[11,16] Andres Metspalu,[12] Christian Gieger,[1,2] Karsten Suhre,[13,26,27] Jerzy Adamski,[19,22,23] Thomas Illig,[1,28,29] Wolfgang Rathmann,[30] Michael Roden,[4,5,31] Annette Peters,[1,2,23,32] Cornelia M. van Duijn,[8,33] Dorret I. Boomsma,[6,7] Thomas Meitinger,[34,35] and Rui Wang-Sattler[1,2,23]

[1]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany
[2]Institute of Epidemiology II, Helmholtz Zentrum München, Neuherberg, Germany
[3]Institute of Structural Biology, Helmholtz Zentrum München, Neuherberg, Germany
[4]Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University, Düsseldorf, Germany

Corresponding author: Rui Wang-Sattler, rui.wang-sattler@helmholtz-muenchen.de.

Received 30 March 2015 and accepted 24 June 2015.

This article contains Supplementary Data online at http://care.diabetesjournals.org/lookup/suppl/doi:10.2337/dc15-0658/-/DC1.

T.X., S.B., A.C.M., C.He., H.H.M.D., and A.D. contributed equally. C.M.v.D. and D.I.B. contributed equally.

OBJECTIVE

Metformin is used as a first-line oral treatment for type 2 diabetes (T2D). However, the underlying mechanism is not fully understood. Here, we aimed to comprehensively investigate the pleiotropic effects of metformin.

RESEARCH DESIGN AND METHODS

We analyzed both metabolomic and genomic data of the population-based KORA cohort. To evaluate the effect of metformin treatment on metabolite concentrations, we quantified 131 metabolites in fasting serum samples and used multivariable linear regression models in three independent cross-sectional studies (n = 151 patients with T2D treated with metformin [mt-T2D]). Additionally, we used linear mixed-effect models to study the longitudinal KORA samples (n = 912) and performed mediation analyses to investigate the effects of metformin intake on blood lipid profiles. We combined genotyping data with the identified metformin-associated metabolites in KORA individuals (n = 1,809) and explored the underlying pathways.

RESULTS

We found significantly lower (P < 5.0E-06) concentrations of three metabolites (acyl-alkyl phosphatidylcholines [PCs]) when comparing mt-T2D with four control groups who were not using glucose-lowering oral medication. These findings were controlled for conventional risk factors of T2D and replicated in two independent studies. Furthermore, we observed that the levels of these metabolites decreased significantly in patients after they started metformin treatment during 7 years' follow-up. The reduction of these metabolites was also associated with a lowered blood level of LDL cholesterol (LDL-C). Variations of these three metabolites were significantly associated with 17 genes (including *FADS1* and *FADS2*) and controlled by AMPK, a metformin target.

CONCLUSIONS

Our results indicate that metformin intake activates AMPK and consequently suppresses FADS, which leads to reduced levels of the three acyl-alkyl PCs and LDL-C. Our findings suggest potential beneficial effects of metformin in the prevention of cardiovascular disease.

Type 2 diabetes (T2D) is a chronic disease with diminished response to insulin and relative insulin deficiency (1). Patients with T2D mostly take metformin as first-line oral treatment to lower their glucose levels and to improve insulin sensitivity (2). Despite metformin's use as an antihyperglycemic agent for more than 50 years, its

primary mode of action is not yet completely understood (3). Inside a cell, metformin apparently inhibits complex I of the mitochondrial electron transport chain and thereby reduces the cellular energy status and upregulates the cytoplasmic 5′-AMPK pathway (3). Activated AMPK stimulates catabolic processes (glycolysis and fatty acid oxidation) and inhibits anabolic pathways (gluconeogenesis and fatty acid synthesis). So far, six metformin targets are documented in the DrugBank (4) database, including the AMPK complex and five metformin transporters. Furthermore, metformin was reported to have several possible pleiotropic effects, resulting in reduced risks for both cancer (5) and cardiovascular disease (CVD) (6), as well as reduced levels of LDL cholesterol (LDL-C) (7,8).

Metabolomic studies have detected metabolite profile changes during the development of T2D (9–12) and identified concentration differences caused by various physiological and environmental factors such as age (13), sex (14), smoking status (15), and alcohol consumption (16). Several metabolomic studies attempted to unravel the physiological effects of metformin (17–21). However, they either used technologies covering only small sets of metabolites or examined relatively few participants (e.g., 20 healthy volunteers [18], 15 patients [17,19], 31 patients [20], and 24 patients treated with glipizide and 23 patients with metformin [21]). As interindividual genetic variations contribute to diverse metabolite profiles and different drug responses, combining metabolomics and genomics may help to understand the mechanisms underlying the action of medications (22–25).

In this study, we discovered metformin treatment–associated metabolites in the Cooperative Health Research in the Region of Augsburg (KORA) cohort (26,27). We confirmed our finding in longitudinal KORA data and replicated them in two independent studies: the Erasmus Rucphen Family study (ERF) (28) and the Netherlands Twin Register (NTR) (29). The biologically relevant pathways for the identified metabolites and their associated genes were further analyzed in organ-specific protein-metabolite interaction networks (30,31). Additionally, we assessed the effects of metformin treatment on LDL-C levels.

RESEARCH DESIGN AND METHODS

An overview of the analysis work flow is shown in Fig. 1.

Ethics Statement

All participants gave written informed consent. The KORA study was approved by the ethics committee of the Bavarian Medical Association, Germany; the ERF study by the medical ethics board of the Erasmus MC Rotterdam, the Netherlands; and the NTR study by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam, the Netherlands.

KORA Cohort

KORA is a population-based cohort study conducted in Southern Germany (26). The baseline survey 4 (KORA S4) consists of 4,261 individuals (aged 25–74 years) examined between 1999 and 2001. During the years 2006–2008, 3,080 participants took part in the follow-up survey 4 (KORA F4). Clinical data for each participant were retrieved from medical records. Based on physician-validated and self-reported diagnosis (9,26), fasting glucose and 2-h postglucose load, and information on medications (Table 1), we excluded *1*) patients suffering from type 1 and steroid-induced diabetes (n = 9), *2*) patients with T2D treated with both metformin and insulin (n = 15), *3*) patients taking glucose-lowering oral medication other than metformin (n = 25), and *4*) patients lacking clear information on treatment (n = 1). Furthermore, participants with overnight nonfasting blood samples (n = 16) or isolated impaired fasting glucose (n = 112) were excluded. We previously showed that impaired fasting glucose and impaired glucose tolerance (IGT) should be considered two different phenotypes (9). In KORA F4, we focused on five groups: *1*) patients

[5]German Center for Diabetes Research, Düsseldorf, Germany
[6]Department of Biological Psychology, Faculty of Psychology and Education, VU University Amsterdam, Amsterdam, the Netherlands
[7]EMGO Institute for Health and Care Research, Amsterdam, the Netherlands
[8]Neuroscience Campus Amsterdam, Amsterdam, the Netherlands
[9]Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands
[10]Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands
[11]Institute of Genetic Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany
[12]Estonian Genome Center, University of Tartu, Tartu, Estonia
[13]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Neuherberg, Germany
[14]Institute of Health Economics and Health Care Management, Helmholtz Zentrum München, Neuherberg, Germany
[15]Else Kroener-Fresenius-Center for Nutritional Medicine, Faculty of Medicine, Technische Universität München, Munich, Germany
[16]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany
[17]Shanghai Center for Bioinformation Technology, Shanghai, China
[18]Bioinformatics Center, Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China
[19]Institute of Experimental Genetics, Genome Analysis Center, Helmholtz Zentrum München, Neuherberg, Germany
[20]Faculty of Science, Leiden Academic Centre for Drug Research, Analytical BioSciences, the Netherlands
[21]Institute of Experimental Genetics, Helmholtz Zentrum München, Neuherberg, Germany
[22]Chair of Experimental Genetics, Center of Life and Food Sciences Weihenstephan, Technische Universität München, Freising, Germany
[23]German Center for Diabetes Research, Neuherberg, Germany
[24]Institute of Medical Statistics and Epidemiology, Technische Universität München, Munich, Germany
[25]Heart and Diabetes Center NRW, Diabetes Center, Ruhr-University Bochum, Bad Oeynhausen, Germany
[26]Faculty of Biology, Ludwig-Maximilians-Universität, Planegg-Martinsried, Germany
[27]Department of Physiology and Biophysics, Weill Cornell Medical College in Qatar, Education City, Qatar Foundation, Doha, Qatar
[28]Hannover Unified Biobank, Hannover Medical School, Hannover, Germany
[29]Institute for Human Genetics, Hannover Medical School, Hanover, Germany
[30]Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University, Düsseldorf, Germany
[31]Department of Endocrinology and Diabetology, Medical Faculty, Düsseldorf, Germany
[32]Department of Environmental Health, Harvard School of Public Health, Boston, MA
[33]Center for Medical Systems Biology, Leiden, the Netherlands
[34]Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany
[35]Institute of Human Genetics, Technische Universität München, Munich, Germany

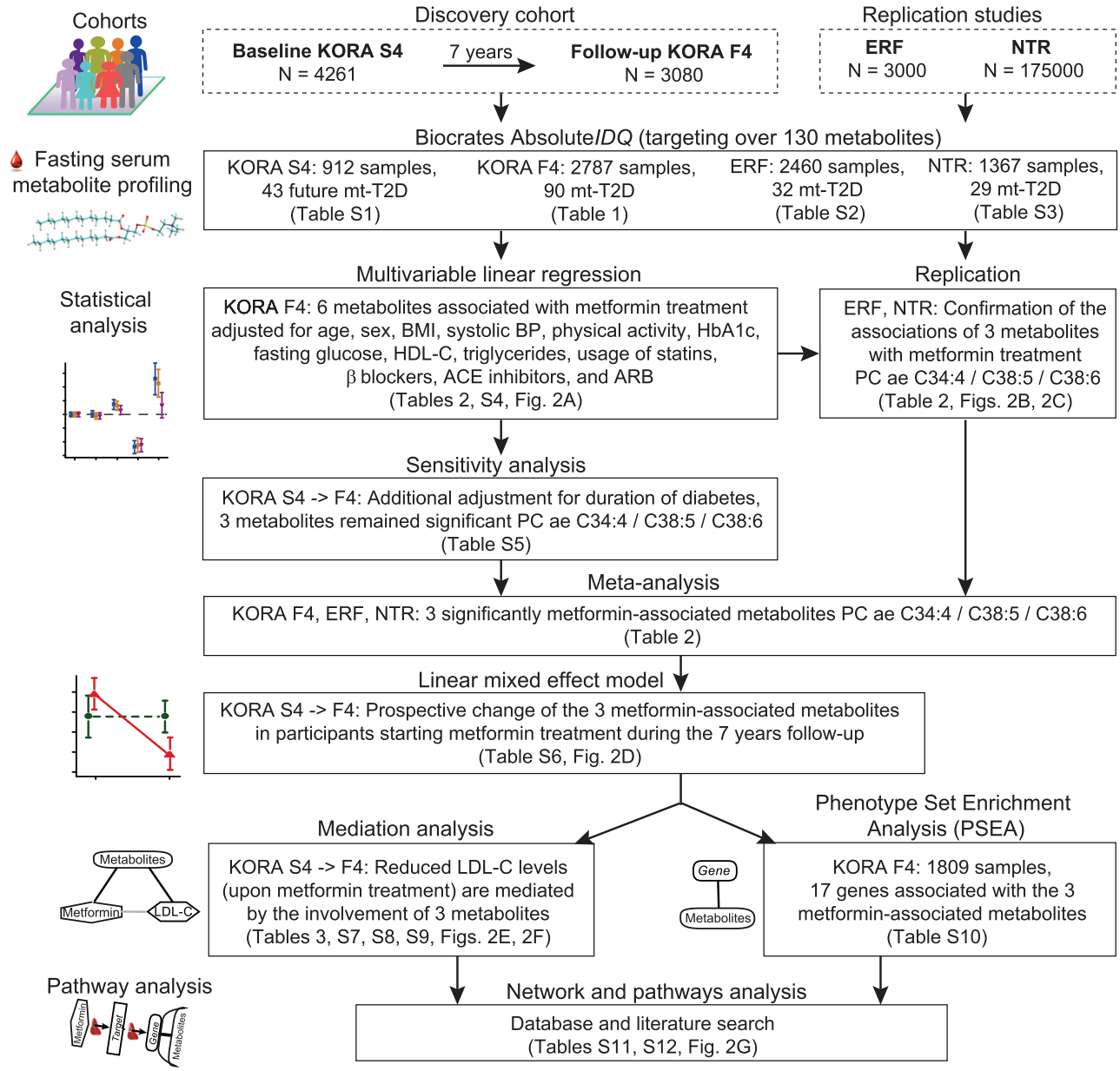

Figure 1—Flowchart of the study design.

with metformin-treated T2D (mt-T2D), *2*) patients with T2D with insulin treatment (it-T2D), *3*) patients with T2D without glucose-lowering treatment (non–antidiabetes drug treated [ndt-T2D]), *4*) participants with prediabetes with IGT, and *5*) healthy individuals with normal glucose tolerance (NGT) (Table 1).

Replication Studies

The ERF includes 3,000 living descendants of 22 couples who had at least six children baptized in the community church around 1850–1900. The participants are not selected based on any disease or other outcome. Details about the genealogy of the population have previously been provided (28).

The NTR recruits twins and their family members to study the causes of individual differences in health, behavior, and lifestyle. Participants are followed longitudinally; details about the cohort have previously been published (29). A subsample of unselected twins and their family members has taken part in the NTR-Biobank (32) in which biological samples, including DNA and RNA, were collected in a standardized manner after overnight fasting.

Duration of diabetes and 2-h postglucose levels were not available in either the ERF or NTR study. The diagnosis of patients with diabetes in both ERF and NTR studies was based on self-report. Owing to the limited number of it-T2D patients in these two replication studies ($n = 3$ and $n = 9$, respectively), this group is not included in the statistical analyses in these two replication studies.

Initially, we had contacted a third potential replication study, the Estonian Genome Center of the University of Tartu (EGCUT). However, only two mt-T2D participants with available metabolomics data were available in this cohort; results from the EGCUT study are therefore not shown.

Table 1—Characteristics of the KORA F4 cross-sectional study population

Clinical parameters	NGT	IGT	ndt-T2D	mt-T2D	it-T2D
n	2,129	375	169	90	24
Age, years	52.8 (12.6)	63.9 (11.0)	66.3 (9.7)	66.8 (8.7)	69.2 (9.8)
Male	46	49	62	59	54
BMI, kg/m^2	26.6 (4.3)	29.7 (4.9)	30.8 (4.4)	31.7 (5.4)	32.2 (5.9)
Waist, cm	90.5 (12.9)	99.7 (14.3)	104.6 (11.4)	106.3 (1.27)	107.2 (12.4)
Physical activity, >1 h per week	58	50	47	33	17
High alcohol intake†	17	17	18	20	8
Smoker	21	8	12	13	8
Systolic BP, mmHg	119.1 (17.4)	127.5 (18.5)	133.7 (18.6)	131.3 (18.9)	135.6 (22.7)
HDL-C, mg/dL	57.6 (14.4)	54.1 (14.0)	47.8 (12.1)	50.6 (10.5)	48.0 (9.6)
LDL-C, mg/dL	134.9 (34.3)	143.7 (35.4)	138.5 (36.5)	122.9 (29.0)	120.0 (31.6)
Triglycerides, mg/dL	110.6 (73.0)	146.0 (86.2)	175.1 (127.0)	174.4 (132.2)	142.1 (73.2)
HbA_{1c}, %	5.4 (0.3)	5.6 (0.3)	6.3 (0.9)	6.9 (1.1)	7.3 (1.1)
HbA_{1c}, mmol/mol	36 (3.3)	38 (3.3)	45 (9.8)	52 (12.0)	56 (12.0)
Fasting glucose, mg/dL	91.7 (7.6)	100.1 (10.6)	125.7 (29.1)	144.1 (37.1)	141.9 (39.0)
2-h postglucose load, mg/dL	97.7 (20.8)	161.7 (17.1)	214.5 (50.7)¥	—	—
Time since diagnosis, years	—	—	1.0 (3.1)#	7.7 (7.1)	16.7 (7.4)
Insulin, μIU/mL	6.9 (25.9)	13.1 (64.0)	16.6 (30.1)	10.4 (10.4)	32.2 (77.8)
Statin usage	8	16	24	38	33
β-Blocker usage	12	31	43	41	63
ACE inhibitor usage	8	21	31	43	58
ARB usage	6	9	15	13	8
Metformin usage	0	0	0	100	0
Insulin therapy	0	0	0	0	100

Percentages of individuals or means (SD) are shown for each variable and each group (NGT, IGT, ndt-T2D, mt-T2D, and it-T2D). †≥20 g/day for women; ≥40 g/day for men. ¥*n* = 121. #For newly diagnosed T2D patients (*n* = 112), years since T2D diagnosis was defined as 0.

Blood Sampling

In the KORA cohort study, blood was drawn into *S*-Monovette serum tubes (Sarstedt AG & Co., Nümbrecht, Germany) in the morning between 8:00 A.M. and 10:30 A.M. after at least 8 h of fasting. Tubes were gently inverted twice, followed by 30 min resting at room temperature to obtain complete coagulation. For serum collection, blood was centrifuged at 2,750 g at 15°C for 10 min. Serum was filled into synthetic straws, which were stored in liquid nitrogen (−196°C) until the metabolomics analyses (9,23).

In the ERF and NTR, the overnight fasting serum samples were drawn for metabolite profiling. Details about the sampling in these two cohorts were described in previous publications (28,32).

Metabolomics Measurement

The serum samples from participants in the baseline KORA S4 and follow-up KORA F4 study were measured with the Absolute*IDQ*p180 and Absolute*IDQ*p150 kits (Biocrates Life Sciences AG, Innsbruck, Austria), respectively. The assay procedures were previously described in detail (27). For KORA S4 and F4, identical quality-control procedures (9,13), which are explained in details in our previous publications, were used. In KORA F4, 131 metabolites of the initially targeted 163 metabolites passed all quality-control criteria: hexose (H1), 24 acylcarnitines, 14 amino acids, 13 sphingomyelines, 34 phosphatidylcholines (PCs), diacyl (aa), 37 PCs acyl-alkyl (ae), and 8 lysoPCs. In total, 124 metabolites overlapped between KORA S4 and F4, including H1, 21 acylcarnitines, 14 amino acids, 13 sphingomyelines, 33 PC aas, and 34 PC aes, as well as 8 lysoPCs.

The metabolite measurements for both replication studies (ERF and NTR) were performed using the same platform (Absolute*IDQ*p150 kit) as in the KORA F4 study. Additionally, in ERF, PC ae C36:4, PC ae C38:5, and PC ae C38:6 were measured in the full set of serum samples by a targeted liquid chromatography–mass spectrometry method. The measurement is performed on a UPLC-ESI-Q-TOF (Agilent 6530; Agilent Technologies, San Jose, CA) mass spectrometer using reference mass correction. Chromatographic separation was achieved on an ACQUITY UPLC HSS T3 column (1.8 μm, 2.1 * 100 mm) with a flow of 0.4 mL/min over a 16-min gradient. The metabolites were detected in full scan in the positive-ion mode. The raw data were processed using Agilent MassHunter Quantitative Analysis software (version B.04.00; Agilent Technologies).

Measured concentration values of all analyzed metabolites are reported in micromolar (μM) and were natural-log transformed, and the distributions were subsequently standardized with mean of zero and an SD of 1 for all analyses unless otherwise indicated.

Single Nucleotide Polymorphism Genotyping, Imputation, and Genes

In KORA F4, we carried out genotyping using the Affymetrix 6.0 GeneChip array (Affymetrix, Santa Clara, CA). Imputation was performed with Impute (http://mathgen.stats.ox.ac.uk/impute/), version 0.4.2 (reference HapMap phase 2, release 22). We only used autosomal single nucleotide polymorphisms (SNPs) with a

minor allele frequency >5%, call rate >95%, and imputation quality >0.4. For the phenotype set enrichment analysis (PSEA), we only mapped those SNPs to a gene that were either in its transcribed region or in its flanking region (110 kb upstream, 40 kb downstream). Gene information was downloaded from the UCSC (University of California, Santa Cruz) genome browser (http://genome.ucsc.edu). The SNP gene mapping was described in detail previously (25). In total, 20,801 genes were analyzed.

Statistical Analysis

To evaluate the effect of metformin treatment on metabolites, we used multivariable linear regression models with the metabolite concentration values as outcome and the grouping variable as predictor. Each metabolite was assessed individually. To include potential confounders, we adjusted for two sets of covariates: *1*) age and sex as the crude model and *2*) age, sex, BMI, physical activity, alcohol intake, smoking, systolic blood pressure (BP), levels of HDL cholesterol (HDL-C), triglycerides, HbA_{1c}, and fasting glucose, as well as the use of statins, β-blockers, ACE inhibitors, and angiotensin receptor blockers (ARBs) as the full model (Table 1). To account for multiple testing, we used Bonferroni correction and considered only those metabolites with a $P < 0.05/131 = 3.8E-04$ to be statistically significantly different in KORA F4. Meta-analysis of the three studies was performed using random effect models, using a restricted maximum-likelihood estimator.

In the KORA S4 to F4 longitudinal study, we used linear mixed-effect models. We adjusted for the two sets of covariates as described above while assigning a random offset to each of the individual participant in the longitudinal study. Additionally, using linear regression models on the KORA data set, including two time points (S4 n = 1,335 and F4 n = 2,763) (9), we calculated the residues of the metabolite concentrations adjusted for age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP, HDL-C, triglyceride, fasting glucose, and HbA_{1c}. The significance of the changes in the metabolite concentrations between the two time points (S4 and F4) was tested using a linear mixed-effect model with the covariates at two time points.

PSEA is a gene-based approach to analyze the associations of genome-wide SNP data with multiple phenotypes in a combined way (25). The significance of enrichment was calculated based on 10,000 permutations (limited by computational restrictions), while setting the significance level at $P < 1.0E-04$ (lowest possible P value owing to the permutation number).

Mediation analysis (33) was conducted to model the identified metabolites as mediators for the association between metformin treatment and LDL-C and total cholesterol in the longitudinal KORA data. The mediation effects of each single metabolite and their summed concentration were tested with crude and fully adjusted multivariable linear regression models.

All statistical analyses were performed in R (version 3.0.1 [http://cran.r-project.org/]).

Pathway Analysis

With use of a bioinformatical approach, a network was constructed by retrieving pairwise connections between candidate metabolites, PSEA-identified genes, intermediate proteins, and known metformin target genes (9,31). Information on protein-protein interactions was extracted from STITCH (30). Known metformin target genes were retrieved from the DrugBank (4). In our network, we only considered the shortest paths (allowing one intermediate protein, confidence score >0.7) connecting the protein encoded by the genes identified in PSEA with the metformin target genes.

RESULTS

Metabolite Profiles in Three Cohorts

We quantified >130 metabolites in fasting serum samples from the KORA S4 and F4, ERF, and NTR studies (Fig. 1). The discovery study, KORA F4, includes 2,129 NGT, 375 IGT, 169 ndt-T2D, 90 mt-T2D, and 24 it-T2D subjects (characteristics shown in Table 1). In the longitudinal study, we used samples from 912 participants without metformin treatment at baseline (KORA S4); 43 of them were treated with metformin at follow-up (KORA F4 [Supplementary Table 1]). In reference to the two replication cohorts, ERF contained 29 ndt-T2D and 32 mt-T2D patients (characteristics shown in Supplementary Table 2), while NTR included 73 ndt-T2D and 29 mt-T2D patients (characteristics shown in Supplementary Table 3).

In general, patients with T2D in the three studies were older and more frequently men, with higher BMI, and took more nonantihyperglycemic medications than the participants without diabetes. Among the five groups in KORA F4, people on statin treatment had significantly lower LDL-C levels than those who were not taking statins (Supplementary Fig. 1). When comparing mt-T2D with ndt-T2D, lower levels of LDL-C were observed both in the cross-sectional (KORA F4, ERF, and NTR) and in the longitudinal KORA studies. Following the 43 patients, who started metformin treatment after the baseline, we did not observe significant changes in the levels of HbA_{1c} and fasting glucose but, however, observed significant changes for LDL-C and total cholesterol (Supplementary Table 1).

Metabolites Associated With Metformin Treatment

We found six metabolites including three acyl-alkyl PCs, two diacyl (aa) PCs, and one amino acid to have significantly lower concentrations in the 90 mt-T2D patients compared with the 169 ndt-T2D individuals in KORA F4 (Table 2). For example, for the metabolite PC ae C36:4, we observed that the fully adjusted effect estimate was −0.66 with P = 4.92E-07; i.e., the PC ae C36:4 level in the mt-T2D group was 0.66 SD lower than the ndt-T2D group.

We further investigated whether the observed differences are specifically for metformin treatment or just reflect the progress of T2D in general. The concentrations of the six metabolites are significantly lower in mt-T2D than in the NGT and IGT groups (Supplementary Table 4). In contrast, none of the six metabolites showed a significantly different concentration in the pairwise comparisons among the four groups without metformin treatment, i.e., NGT, IGT, ndt-T2D, and it-T2D (Supplementary Table 4).

For sensitivity analysis, we tested the associations of the six metabolites after adding the duration of T2D to the fully adjusted model. The three acyl-alkyl PCs (PC ae C36:4, PC ae C38:5, and PC ae C38:6), which are composed of at least one polyunsaturated fatty acid (PUFA), remained significantly different in the

Table 2—Metabolites associated with metformin treatment in patients with T2D

	Discovery KORA		Replication ERF		Replication NTR		Meta-analysis	
Metabolite	Effect estimate (95% CI)	*P*	Effect estimate (95% CI)	*P*	Effect estimate (95% CI)	*P*	Effect estimate (95% CI)	*P*
Crude model								
PC ae C36:4	−0.66 (−0.90, −0.43)	**1.07E-07**	−1.03 (−1.36, −0.69)	**2.65E-08**	−0.97 (−1.38, −0.56)	**9.21E-06**	−0.86 (−1.11, −0.60)	**5.24E-11**
PC ae C38:5	−0.65 (−0.90, −0.41)	**2.92E-07**	−1.04 (−1.37, −0.71)	**1.94E-08**	−0.72 (−1.12, −0.32)	**5.75E-04**	−0.79 (−1.04, −0.55)	**2.86E-10**
PC ae C38:6	−0.58 (−0.82, −0.35)	**2.39E-06**	−0.76 (−1.11, −0.42)	**3.53E-05**	−0.61 (−1.02, −0.19)	**4.44E-03**	−0.63 (−0.81, −0.45)	**4.11E-12**
PC aa C36:0	−0.50 (−0.75, −0.25)	**9.53E-05**	−0.30 (−0.78, 0.18)	2.31E-01	−0.54 (−1.08, 0.01)	5.43E-02	−0.47 (−0.68, −0.26)	**7.48E-06**
PC aa C38:0	−0.73 (−0.97, −0.49)	**1.48E-08**	−0.31 (−0.88, 0.27)	2.98E-01	−0.56 (−1.12, −0.01)	**4.49E-02**	−0.64 (−0.87, −0.40)	**8.22E-08**
Ornithine	−0.57 (−0.80, −0.33)	**4.42E-06**	−0.03 (−0.18, 0.11)	6.49E-01	0.01 (−0.45, 0.46)	9.78E-01	−0.21 (−0.59, 0.17)	2.72E-01
Full model								
PC ae C36:4	−0.66 (−0.91, −0.41)	**4.92E-07**	−1.08 (−1.33, −0.83)	**8.37E-05**	−0.84 (−1.27, −0.42)	**1.67E-04**	−0.79 (−1.02, −0.55)	**5.27E-11**
PC ae C38:5	−0.62 (−0.88, −0.36)	**3.81E-06**	−1.11 (−1.36, −0.85)	**6.18E-05**	−0.62 (−1.04, −0.20)	**4.65E-03**	−0.73 (−0.99, −0.47)	**3.73E-08**
PC ae C38:6	−0.58 (−0.82, −0.34)	**4.94E-06**	−0.74 (−1.00, −0.48)	**6.15E-03**	−0.49 (−0.91, −0.07)	**2.15E-02**	−0.58 (−0.78, −0.39)	**2.96E-09**
PC aa C36:0	−0.57 (−0.84, −0.30)	**4.25E-05**	−0.37 (−0.62, −0.12)	1.41E-01	−0.31 (−0.86, 0.24)	2.68E-01	−0.49 (−0.71, −0.27)	**9.79E-06**
PC aa C38:0	−0.68 (−0.94, −0.42)	**5.25E-07**	−0.35 (−0.64, −0.05)	2.47E-01	−0.34 (−0.90, 0.22)	2.32E-01	−0.56 (−0.82, −0.30)	**1.85E-05**
Ornithine	−0.58 (−0.80, −0.19)	**3.82E-05**	−0.01 (−0.09, 0.07)	8.80E-01	−0.13 (−0.62, 0.36)	6.02E-01	−0.25 (−0.62, 0.13)	2.00E-01

Effect estimates and *P* values for the comparison between mt-T2D and ndt-T2D were calculated using multivariable linear regression analysis adjusted for crude (age and sex) and full model, which includes age, sex, BMI, physical activity, alcohol intake, smoking, systolic BP, HDL-C, triglycerides, HbA_{1c}, fasting glucose, and use of statins, β-blockers, ACE inhibitors, and ARBs. Lipid side chain composition is abbreviated as Cx:y, where x denotes the number of carbons in the side chain and y the number of double bonds. Statistically significant *P* values appear in boldface.

comparison between mt-T2D and ndt-T2D ($P < 3.8E-04$) (Fig. 2*A*), whereas the other three metabolites were not significantly different anymore (Supplementary Table 5). After adjustment of the full model for *1*) waist, *2*) LDL-C, and *3*) the combination of LDL-C and insulin, the effect estimates of the six metabolites were almost unchanged (Supplementary Table 5).

Replication and Meta-analysis

For the three acyl-alkyl PCs, we observed consistent results in both replication studies (ERF and NTR); i.e., significantly lower levels were observed in mt-T2D patients compared with ndt-T2D individuals ($P < 0.05$) (Table 2 and Fig. 2*B* and *C*). Additionally, a meta-analysis of the three studies (KORA F4, ERF, and NTR) yielded significant results for the three replicated metabolites ($P < 3.8E-04$) (Table 2). We refer to these three highly intercorrelated metabolites, which are not associated with fasting glucose or HbA_{1c}, as metformin associated in the following paragraphs (Supplementary Fig. 2).

In the longitudinal examination, we found significantly decreased levels of the three metformin-associated metabolites in patients who underwent metformin treatment during the follow-up ($P < 3.8E-04$ using the fully adjusted model) (Supplementary Table 6). Consistent results for the three acyl-alkyl PCs were observed in a sensitivity analysis with a subgroup of 55 ndt-T2D patients at the KORA S4, of whom 19 were ndt-T2D patients and 36 were mt-T2D patients in KORA F4 in the fully adjusted model ($P < 0.05$) (Fig. 2*D* and Supplementary Table 6). These prospective findings confirmed our observations in the cross-sectional study.

Relationship Between Metformin Treatment, the Three Metabolites, and LDL-C Levels

To investigate a potentially mediating effect of the three acyl-alkyl PCs on the associations between metformin treatment and lipid profiles, we explored the prospective data of 912 KORA participants (Supplementary Table 1). We found that metformin treatment accounts for a significant decrease of LDL-C and total cholesterol levels, while its influence on HDL-C and triglycerides was not significant in both crude and

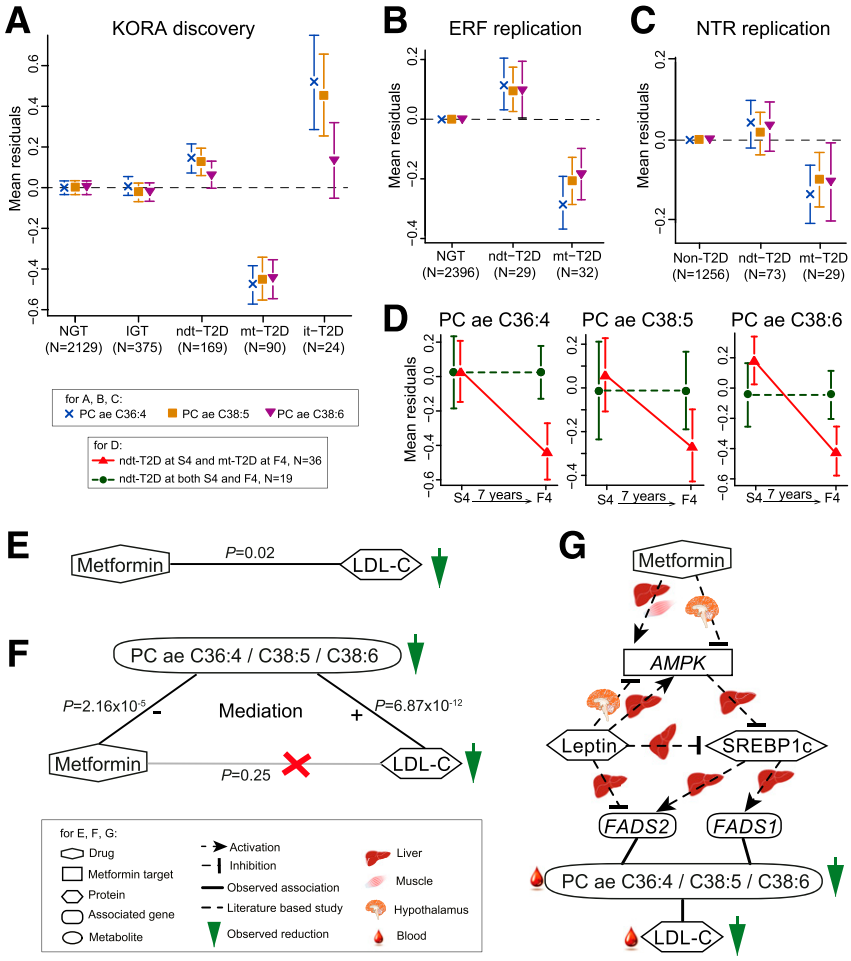

Figure 2—Differences in metabolite concentrations, mediation effect, and organ-specific pathways. Mean residuals of the concentrations (with SEs) of three identified acyl-alkyl PC metabolites for the NGT, IGT, ndt-T2D, mt-T2D, and it-T2D groups derived in cross-sectional analysis of the KORA F4 are shown in *A*. The mean residuals of the same metabolites in ERF are illustrated for the NGT, ndt-T2D, and mt-T2D groups in *B* and in NTR for the non-T2D, ndt-T2D, and mt-T2D groups in *C*, respectively. *D* refers to the longitudinal setting of the KORA study and shows the mean residuals of the concentrations (with SEs) of the three metabolites with respect to changes within the 7 years between baseline and follow-up study when people were treated with metformin. Residuals were calculated from linear regression model with the full adjustment. *E*: The association between metformin and LDL-C without consideration of the three metformin-associated metabolites. *F*: The results of the mediation analysis; the red cross indicates that the direct association between metformin and LDL-C is not significant anymore. *G*: An overview of the involved pathways. The connections indicated by liver, hypothalamus, muscle, and blood show organ specificity between genes, pathway-related proteins, and metformin drug targets as well as metformin. The metabolites (ellipses) were connected to metformin treatment (straight side hexagons) through genes (rounded rectangles), proteins (hexagons), and metformin targets (rectangles). The activation or inhibition is indicated. Plus or minus symbol next to the line indicate positive or negative association. For further information, see Table 3 and Supplementary Tables 8 and 12.

fully adjusted models ($P < 0.05$) (Fig. 2*E* and Supplementary Table 7). In particular, metformin was associated with a decrease in LDL-C levels of 11.83 mg/dL. We therefore focused on the analysis of LDL-C and total cholesterol.

After adding the three metabolites to the full model, the direct association between metformin treatment and LDL-C levels was not significant anymore ($P = 0.25$) (Fig. 2*F* and Supplementary Table 8*A*). Based on longitudinal analysis, we found consistent results as reported above (Table 2); i.e., significantly reduced levels of the three metabolites in the metformin-treated patients were observed (e.g., for the summed metabolite concentration $P = 2.16E-05$) (Fig. 2*F* and Supplementary Table 8*B*). Furthermore, we found significant positive associations between LDL-C and each of the three metabolites, as well as their summed concentration after adjusting for metformin treatment (e.g., for the summed metabolite concentration $P = 6.87E-12$) (Fig. 2*F* and Supplementary Table 8*C*). This means that these associations of the metformin-associated metabolites with LDL-C are independent of metformin treatment. Finally, for each of the three metformin-associated metabolites (and their summed concentration), the mediation effects on the association between metformin treatment and the LDL-C levels were significant in both models (Table 3). For instance, the summed concentration of the metabolites mediates 3.43 mg/dL reduction in LDL-C level, which accounts for 29% of the total effect of metformin on LDL-C (Table 3).

To rule out the potential effect of statin intake, we performed a sensitivity analysis by excluding individuals taking statin at baseline KORA S4 and/or follow-up F4. The mediation effects of the summed concentration were also significant for the associations between metformin and LDL-C level (Supplementary Table 9*A*). However, although the crude and full model showed similarly significant mediation effects for total cholesterol (Supplementary Table 8*D* and *E* and Table 3), after excluding statin users from the analysis, the effects on total cholesterol were not significant anymore with respect to the fully adjusted model ($P < 0.05$) (Supplementary Table 9*B*).

Seventeen Genes Are Linked to Metformin-Associated Metabolites and Pathway Analysis

To identify genes associated with the three metabolites, we applied PSEA on these metabolites in a subset of KORA F4 individuals (n = 1,809) with available genotyping data and metabolite profiles. We found 17 genes with an enrichment of SNPs in their transcribed or flanking region ($P < 1.0E-4$) (Supplementary Table 10). These genes belong to five clusters, one of them containing 12 genes located on chromosome 11. A literature search revealed disease phenotypes associated with these 17 genes. Six genes, namely, *FADS1*, *FADS2*, *FADS3*, *MYRF*, *BEST1* and *RAB3IL1*, are associated with T2D or its comorbidities, including retinopathy and

Table 3—Mediation effects of the three metabolites for the association between metformin treatment and reduction of LDL-C and total cholesterol

	Crude model			Full model		
	Effect estimate (95% CI)	*P*	Explained effect (%)	Effect estimate (95% CI)	*P*	Explained effect (%)
LDL-C						
PC ae C36:4	−3.05 (−4.38, −1.71)	2.21E-04	25.74	−2.51 (−3.71, −1.31)	1.33E-03	21.22
PC ae C38:5	−2.94 (−4.21, −1.67)	2.65E-04	24.82	−2.36 (−3.50, −1.22)	1.94E-03	19.97
PC ae C38:6	−5.25 (−8.11, −2.40)	1.34E-05	44.40	−4.02 (−6.59, −1.44)	4.55E-04	33.95
Summed concentration†	−4.37 (−6.37, −2.37)	1.52E-05	36.92	−3.43 (−5.19, −1.67)	2.95E-04	28.99
Total cholesterol						
PC ae C36:4	−5.00 (−7.77, −2.23)	2.63E-05	26.1	−3.08 (−4.72, −1.45)	7.42E-04	26.5
PC ae C38:5	−4.99 (−7.71, −2.26)	2.33E-05	26.0	−2.77 (−4.25, −1.30)	1.38E-03	23.8
PC ae C38:6	−6.99 (−11.86, −2.13)	8.96E-06	36.4	−4.16 (−6.98, −1.35)	5.11E-04	35.8
Summed concentration†	−6.63 (−10.55, −2.71)	2.76E-06	34.6	−3.87 (−6.05, −1.69)	2.41E-04	33.3

The estimates of the mediation effects and *P* values were calculated using the longitudinal (KORA S4→F4) mediation analysis adjusted for the crude and full model. The mediation effects for the three metformin-associated metabolites and the summed concentration are shown. †The summed concentration refers to the overall concentration of the three metabolites (PC ae C36:4, PC ae C38:5, and PC ae C38:6).

coronary artery diseases (for references, see Supplementary Table 10).

To explore potentially related pathways, we used a bioinformatics approach, integrating the 17 identified genes with 6 known metformin target genes (4) into a protein-protein interaction network (9,30,31). For 3 of the 17 genes, there was no record for *Homo sapiens* in the STITCH (30); therefore, we investigated the interaction of the remaining 14 genes with the 6 metformin targets (Supplementary Table 11). *AMPK* was found to be linked to *FADS1* and *FADS2* through interacting proteins (leptin and sterol regulatory element–binding protein 1c [SREBP1c]). A manual evaluation of these interactions in a literature research showed organ specificity, mainly referring to liver and hypothalamus (Fig. 2*G*). The AMPK complex is inhibited by leptin and metformin in the hypothalamus, whereas it is activated by metformin and leptin in the liver. (References for each interaction are provided in the Supplementary Table 12).

CONCLUSIONS

We found significant concentration differences for three metabolites (PC ae C36:4, PC ae C38:5, and PC ae C38:6) in the blood of patients with T2D under metformin treatment and replicated them in two independent studies. We identified SNP variations in 17 genes (including *FADS1* and *FADS2*) that were associated with the three metabolites. Based on these genes, we built an interaction network to investigate the underlying mechanisms of metformin treatment and identified the organ-specific AMPK pathway. We further found that the reduced LDL-C levels in metformin-treated patients with T2D were mediated partially by the three acyl-alkyl PCs. Sensitivity analyses were performed to consider the duration of diabetes and statin use.

The levels of metabolites depend on multiple modifiable factors, such as lifestyle and environment (9–11,13–16). We therefore considered a number of confounding effects, e.g., physiological parameters (age, sex, BMI, and systolic BP), lifestyle (physical activity, alcohol intake, and smoking), glucose levels (HbA_{1c} and fasting glucose), lipid levels (HDL-C and triglycerides), and medication usage (statins, β-blockers, ACE inhibitors, and ARBs). Additionally, intermediates or end products of metabolism are influenced by underlying genetic factors (23,24). In our study, phenotypes and genotypes are available for each person (n = 1,809); we thus used phenotype set enrichment analysis (25). Our combined analysis of genetic and metabolomic data enabled us to identify genes associated with the three metabolites and supported the identification of an organ-specific pathway. The observation of significantly lower levels of the three metformin-associated metabolites (polyunsaturated acyl-alkyl PCs) in the mt-T2D patients can be explained by metformin's effects on AMPK in the liver (Fig. 2*G* and Supplementary Table 12). In the hepatocyte, metformin increases the AMP-to-ATP ratio and thus leads to the activation of AMPK. Activated AMPK blocks SREBP1c, a transcription factor controlling enzymes involved in the fatty acid synthesis and inhibiting the synthesis of FADS1 and FADS2 (22). This results in a reduced synthesis of unsaturated fatty acids and consequently lower acyl-alkyl PCs concentrations. Leptin occupies a central position in the network (Fig. 2*G*) and affects the FADS complex via three different interactions. In the liver, leptin not only activates AMPK, thereby suppressing SREBP1c and downregulating FADS1 and FADS2, but can also directly inhibit both SREBP1c and FADS2 (34). Metformin and leptin exert opposite effects in the hypothalamus and in the liver (for references, see Supplementary Table 12), but further studies are required to better understand the organ-specific metformin effects in humans.

Recently, clinical practice guidelines have recommended the usage of metformin as first-line therapy in patients with T2D with heart failure (1,2). Our observation of lower blood levels of LDL-C in metformin-treated patients points toward a beneficial effect of metformin for the prevention of CVD. A meta-analysis of randomized clinical trials shows that metformin treatment results in lowered LDL-C levels in newly diagnosed T2D patients (8). Similar results were also reported in patients without T2D in an epidemiological study (7). Here, we observed that metformin treatment leads to lowered LDL-C levels, an effect mediated most likely through metformin-mediated reduction of FADS

activity and consequently reduction of the levels of PUFA, namely arachidonic acid (35). It has been suggested that lower levels of arachidonic acid leads to an increased membrane fluidity, thus increasing LDL-C receptor recycling (35). This hypothesis is especially strong, given that genetic variants assigned to lower activity of FADS1 and -2 were significantly associated with lower LDL-C levels (36). While certain PCs can indeed exert antidiabetic effects (37), further mechanistic studies are required to test whether lowering of these circulating lipids contributes directly to the prevention of CVDs or merely by its antidiabetic effect (1).

Beyond its common antihyperglycemic action and its effect in lowering LDL-C, metformin can potentially reduce the risk of cancer mortality and diminish the progression of cancer (38). In the current study, we have found the three metformin-associated metabolites significantly associated with two genes, *FEN1* and *C20orf94*, which are involved in DNA repair (39,40). This may partly explain that metformin has been shown to influence the prevalence of different types of carcinoma, such as gastrointestinal cancers (39) and leukemia (40).

The strength of our study is that we used three independent cohort studies to discover and replicate our observations. Importantly, all results presented in this study were independent of physiological parameters, lifestyle, glucose levels, lipid levels, and medication. We combined metabolomics and genomics data, broad literature research, and organ-specific information from animal studies to deepen the insight into the underlying mechanisms.

Our findings are limited by the observational nature of cohort studies, and the applied methods, such as the mediation analysis, are of purely statistical character, but they offer the opportunity to raise new questions for experimental confirmation studies, such as randomized controlled clinical trials to investigate, for instance, the effect of metformin on blood lipid levels of patients without diabetes.

In the present studies (KORA, ERF, NTR), the duration of T2D is based on self-reported information. Moreover, neither data on the dosage nor data on duration and compliance of the metformin treatment were available. Furthermore, it has to be mentioned that the degree of diabetes severity presumably discriminates the different groups of patients (ndt-T2D, mt-T2D, and it-T2D), which is reflected by their HbA_{1c} and fasting glucose values (Table 1). Although the investigated metabolite panel does not represent the whole human metabolome, the comprehensive analysis of >130 metabolites from different classes represents a considerable improvement compared with previous technologies.

We found three metformin-associated metabolites, which showed no overlap with the findings of previous studies (17–21). This is likely to result from the use of different sampling matrices (plasma vs. serum), unmeasured metabolites (asymmetric dimethylarginine), or study design (glipizide treatment). Additional, our study considered considerably more potential cofounding effects in a comparably larger number of individuals than previous studies (17–21).

In conclusion, we observed that metformin treatment reduced levels of the three acyl-alkyl PC metabolites in patients with T2D. This change in the metabolic profiles may mediate lowered blood levels of LDL-C. The underlying mechanism is most likely the metformin-induced activation of AMPK and the consequent suppression of SREBP1c and FADS, which leads to reduced levels of PUFA and LDL-C. Our findings suggest a pharmaco-epidemiologic mechanism by which metformin may exert beneficial effects to prevent CVD. More importantly, our study suggests a novel approach to identify pleiotropic effects of medication using multilevel omics data.

Acknowledgments. The authors express their appreciation to all KORA study participants for donating their blood and time. The authors thank the field staff in Augsburg conducting the KORA studies. The authors thank the staff from the Institute of Epidemiology at the Helmholtz Zentrum München and the Genome Analysis Center Metabolomics Platform, who helped in the sample logistics, data and straw collection, and metabolomic measurements, and especially J. Scarpa, K. Faschinger, F. Scharl, N. Lindemann, H. Chavez, A. Sabunchi, A. Schneider, A. Ludolph, S. Jelic, and B. Langer. The authors are grateful to all general practitioners involved in ERF for their contributions. The authors thank all participants in the NTR.

Funding. The KORA study was initiated and financed by the Helmholtz Zentrum München–German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by The State of Bavaria. Furthermore, KORA research was supported within The Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. Part of this project was supported by The European Community's Seventh Framework Programme grants HEALTH-2009-2.2.1-3/242114 (Project OPTiMiS) and HEALTH-2013-2.4.2-1/602936 (Project CarTarDis). The German Diabetes Center is funded by the German Federal Ministry of Health (Berlin, Germany) and The Ministry of Innovation, Science and Research of the State of North Rhine-Westphalia (Düsseldorf, Germany). The diabetes part of the KORA F4 study was funded by a grant from the German Research Foundation (DFG) RA 459/3-1. This study was supported in part by a grant from The BMBF to the German Center for Diabetes Research (DZD e. V.). W.R.-M. is funded by The German Federal Ministry of Education and Research grant 03IS2061B (project Gani_Med). K.Su. is supported by Biomedical Research Program funds at Weill Cornell Medical College in Qatar, a program funded by the Qatar Foundation. The EGCUT received support from The European Community's Seventh Framework Programme grant BBMRI-LPC 313010, targeted financing from Estonian Government IUT20-60 and IUT24-6, Estonian Research Roadmap through the Estonian Ministry of Education and Research (3.2.0304.11-0312), the Center of Excellence in Genomics (EXCEGEN), and Development Fund from the University of Tartu (SP1GVARENG), and from an EFSD New Horizons grant. This work was also supported by the U.S. National Institutes of Health (R01DK075787). The ERF was supported by grants from the Netherlands Organisation for Scientific Research (NWO), and Erasmus MC and the Centre for Medical Systems Biology (CMSB). Telomere length assessment was supported through funds from The European Community's Seventh Framework Programme (FP7/2007-2013), grant agreement HEALTH-F4-2007-201413 (ENGAGE). Research for the NTR was funded by The Netherlands Organisation for Scientific Research (NWO) (MagW/ZonMW grants 904-61-090, 985-10-002, 904-61-1, 480-04-004, 400-05-717; Addiction-31160008 Middelgroot-911-09-032; and Spinozapremie 56-464-14192), the Center for Medical Systems Biology (CSMB, NWO Genomics), NBIC/BioAssist/RK(2008.024), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI –NL) (184·021·007), the VU University's Institute for Health and Care Research (EMGO+), and The European Community's Seventh Framework Program (FP7/2007-2013), grant HEALTH-F4-2007-201413 (ENGAGE).

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. T.X., S.B., A.C.M., C.He., H.H.M.D., A.D., Z.Y., J.S.R., and T.Hal. analyzed the data and interpreted the results. T.X., S.B., A.C.M., C.He., M.Rod., T.M., and R.W.-S. wrote the manuscript. M.H., R.S., C.Ho., J.Adam, S.C., M.Rot., T.P., Y.H., G.K., C.M., B.T., A.R., M.H.d.A., Y.L., H.-E.W., B.S., A.M., C.G., K.St., and W.R. assisted in manuscript generation and revision. M.C., G.F., A.S.Q., G.W., R.P., K.K., and K.W.v.D. interpreted the results. C.P., W.R.-M., K.Su., T.Han., J.Adams., and T.I. performed the metabolic profiling. T.I., A.P., C.M.v.D., D.I.B.,

T.M., and R.W.-S. conceived and designed the study. R.W.-S. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Foretz M, Guigas B, Bertrand L, Pollak M, Viollet B. Metformin: from mechanisms of action to therapies. Cell Metab 2014;20:953–966
2. ADA. Executive summary: Standards of medical care in diabetes–2013. Diabetes Care 2013; 36(Suppl. 1):S4–S10
3. Madiraju AK, Erion DM, Rahimi Y, et al. Metformin suppresses gluconeogenesis by inhibiting mitochondrial glycerophosphate dehydrogenase. Nature 2014;510:542–546
4. Law V, Knox C, Djoumbou Y, et al. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res 2014;42:D1091–D1097
5. Decensi A, Puntoni M, Goodwin P, et al. Metformin and cancer risk in diabetic patients: a systematic review and meta-analysis. Cancer Prev Res (Phila) 2010;3:1451–1461
6. Norwood DK, Chilipko AA, Amin SM, Macharia D, Still KL. Evaluating the potential benefits of metformin in patients with cardiovascular disease and heart failure. Consult Pharm 2013;28:579–583
7. Glueck CJ, Fontaine RN, Wang P, et al. Metformin reduces weight, centripetal obesity, insulin, leptin, and low-density lipoprotein cholesterol in nondiabetic, morbidly obese subjects with body mass index greater than 30. Metabolism 2001; 50:856–861
8. Wulffelé MG, Kooy A, de Zeeuw D, Stehouwer CD, Gansevoort RT. The effect of metformin on blood pressure, plasma cholesterol and triglycerides in type 2 diabetes mellitus: a systematic review. J Intern Med 2004;256:1–14
9. Wang-Sattler R, Yu Z, Herder C, et al. Novel biomarkers for pre-diabetes identified by metabolomics. Mol Syst Biol 2012;8:615
10. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. Nat Med 2011;17:448–453
11. Floegel A, Stefan N, Yu Z, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. Diabetes 2013;62:639–648
12. Newgard CB, An J, Bain JR, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. Cell Metab 2009;9: 311–326
13. Yu Z, Zhai G, Singmann P, et al. Human serum metabolic profiles are age dependent. Aging Cell 2012;11:960–967
14. Mittelstrass K, Ried JS, Yu Z, et al. Discovery of sexual dimorphisms in metabolic and genetic biomarkers. PLoS Genet 2011;7:e1002215
15. Xu T, Holzapfel C, Dong X, et al. Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. BMC Med 2013;11:60
16. Jaremek M, Yu Z, Mangino M, et al. Alcohol-induced metabolomic differences in humans. Transl Psychiatry 2013;3:e276
17. Marchetti P, Masiello P, Benzi L, et al. Effects of metformin therapy on plasma amino acid pattern in patients with maturity-onset diabetes. Drugs Exp Clin Res 1989;15:565–570
18. Cai S, Huo T, Li N, Xiong Z, Li F. Lysophosphatidylcholine–biomarker of Metformin action: studied using UPLC/MS/MS. Biomed Chromatogr 2009;23:782–786
19. Huo T, Cai S, Lu X, Sha Y, Yu M, Li F. Metabonomic study of biochemical changes in the serum of type 2 diabetes mellitus patients after the treatment of metformin hydrochloride. J Pharm Biomed Anal 2009;49:976–982
20. Asagami T, Abbasi F, Stuelinger M, et al. Metformin treatment lowers asymmetric dimethylarginine concentrations in patients with type 2 diabetes. Metabolism 2002;51:843–846
21. Zhang Y, Hu C, Hong J, et al. Lipid profiling reveals different therapeutic effects of metformin and glipizide in patients with type 2 diabetes and coronary artery disease. Diabetes Care 2014;37:2804–2812
22. Gieger C, Geistlinger L, Altmaier E, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. PLoS Genet 2008;4:e1000282
23. Illig T, Gieger C, Zhai G, et al. A genome-wide perspective of genetic variation in human metabolism. Nat Genet 2010;42:137–141
24. Suhre K, Shin SY, Petersen AK, et al.; CARDIoGRAM. Human metabolic individuality in biomedical and pharmaceutical research. Nature 2011;477:54–60
25. Ried JS, Döring A, Oexle K, et al. PSEA: Phenotype Set Enrichment Analysis–a new method for analysis of multiple phenotypes. Genet Epidemiol 2012;36:244–252
26. Rathmann W, Kowall B, Heier M, et al. Prediction models for incident type 2 diabetes mellitusin the older population: KORA S4/F4 cohort study. Diabet Med 2010;27:1116–1123
27. Jourdan C, Petersen AK, Gieger C, et al. Body fat free mass is associated with the serum metabolite profile in a population-based study. PLoS One 2012;7:e40009
28. Aulchenko YS, Heutink P, Mackay I, et al. Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet 2004; 12:527–534
29. Boomsma DI, de Geus EJ, Vink JM, et al. Netherlands Twin Register: from twins to twin families. Twin Res 2006;9:849–857
30. Kuhn M, Szklarczyk D, Pletscher-Frankild S, et al. STITCH 4: integration of protein-chemical interactions with user data. Nucleic Acids Res 2014;42:D401–D407
31. Lechner M, Höhn V, Brauner B, et al. CIDeR: multifactorial interaction networks in human diseases. Genome Biol 2012;13:R62
32. Willemsen G, de Geus EJ, Bartels M, et al. The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. Twin Res 2010;13:231–245
33. Bauer DJ, Preacher KJ, Gil KM. Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. Psychol Methods 2006;11:142–163
34. Lim CT, Kola B, Korbonits M. AMPK as a mediator of hormonal signalling. J Mol Endocrinol 2010;44:87–97
35. Sone Y, Kido T, Ainuki T, et al. Genetic variants of the fatty acid desaturase gene cluster are associated with plasma LDL cholesterol levels in Japanese males. J Nutr Sci Vitaminol (Tokyo) 2013;59:325–335
36. Global Lipids Genetics Consortium, Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet 2013;45:1274–1283
37. Lee JM, Lee YK, Mamrosh JL, et al. A nuclear-receptor-dependent phosphatidylcholine pathway with antidiabetic effects. Nature 2011;474:506–510
38. Pernicova I, Korbonits M. Metformin–mode of action and clinical implications for diabetes and cancer. Nat Rev Endocrinol 2014;10:143–156
39. Liu L, Zhou C, Zhou L, et al. Functional FEN1 genetic variants contribute to risk of hepatocellular carcinoma, esophageal cancer, gastric cancer and colorectal cancer. Carcinogenesis 2012;33:119–123
40. Meissner B, Bartram T, Eckert C, et al. Frequent and sex-biased deletion of SLX4IP by illegitimate V(D)J-mediated recombination in childhood acute lymphoblastic leukemia. Hum Mol Genet 2014;23:590–601

Jonathan Adam,[1,2] Stefan Brandmaier,[1,2] Jörn Leonhardt,[3] Markus F. Scheerer,[4,5] Robert P. Mohney,[6] Tao Xu,[1,2] Jie Bi,[7] Markus Rotter,[1,2] Martina Troll,[1,2] Shen Chi,[1,2] Margit Heier,[2] Christian Herder,[5,8] Wolfgang Rathmann,[5,9] Guido Giani,[9] Jerzy Adamski,[4,5,10,11] Thomas Illig,[12,13] Konstantin Strauch,[14,15] Yixue Li,[7] Christian Gieger,[1,2,5] Annette Peters,[1,2,5,16] Karsten Suhre,[3,17,18] Donna Ankerst,[19] Thomas Meitinger,[20,21] Martin Hrabě de Angelis,[4,5,11] Michael Roden,[5,22,23] Susanne Neschen,[4,5] Gabi Kastenmüller,[3] and Rui Wang-Sattler[1,2,5]

Metformin Effect on Nontargeted Metabolite Profiles in Patients With Type 2 Diabetes and in Multiple Murine Tissues

Diabetes 2016;65:3776–3785 | DOI: 10.2337/db16-0512

Metformin is the first-line oral medication to increase insulin sensitivity in patients with type 2 diabetes (T2D). Our aim was to investigate the pleiotropic effect of metformin using a nontargeted metabolomics approach. We analyzed 353 metabolites in fasting serum samples of the population-based human KORA (Cooperative Health Research in the Region of Augsburg) follow-up survey 4 cohort. To compare T2D patients treated with metformin (mt-T2D, *n* = 74) and those without antidiabetes medication (ndt-T2D, *n* = 115), we used multivariable linear regression models in a cross-sectional study. We applied a generalized estimating equation to confirm the initial findings in longitudinal samples of 683 KORA participants. In a translational approach, we used murine plasma, liver, skeletal muscle, and epididymal adipose tissue samples from metformin-treated *db/db* mice to further corroborate our findings from the human study. We identified two metabolites significantly (*P* < 1.42E-04) associated with metformin treatment. Citrulline showed lower relative concentrations and an unknown metabolite X-21365 showed higher relative concentrations in human serum when comparing mt-T2D with ndt-T2D. Citrulline was confirmed to be significantly (*P* < 2.96E-04) decreased at 7-year follow-up in patients who started metformin treatment. In mice, we validated significantly (*P* < 4.52E-07) lower citrulline values in plasma, skeletal muscle, and adipose tissue of metformin-treated animals but not in their liver. The lowered values of citrulline we observed by using a nontargeted approach most likely resulted from the pleiotropic effect of metformin on the interlocked urea and nitric oxide cycle. The translational data derived from multiple murine tissues corroborated and complemented the findings from the human cohort.

Metformin became the first-line choice for treatment of type 2 diabetes (T2D) in the course of the UK Prospective Diabetes Study (UKPDS) (1). Additionally, metformin has

[1]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany
[2]Institute of Epidemiology II, Helmholtz Zentrum München, Neuherberg, Germany
[3]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Neuherberg, Germany
[4]Institute of Experimental Genetics, Helmholtz Zentrum München, Neuherberg, Germany
[5]German Center for Diabetes Research (DZD), Neuherberg, Germany
[6]Metabolon, Inc., Durham, NC
[7]Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China
[8]Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University, Düsseldorf, Germany
[9]Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University, Düsseldorf, Germany
[10]Institute of Experimental Genetics, Genome Analysis Center, Helmholtz Zentrum München, Neuherberg, Germany
[11]Institute of Experimental Genetics, Center of Life and Food Sciences Weihenstephan, Technische Universität München, Freising, Germany
[12]Hannover Unified Biobank, Hannover Medical School, Hannover, Germany
[13]Institute for Human Genetics, Hannover Medical School, Hannover, Germany
[14]Institute of Genetic Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany
[15]Genetic Epidemiology, Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität München, München, Germany
[16]Department of Environmental Health, Harvard School of Public Health, Boston, MA
[17]Faculty of Biology, Ludwig-Maximilians-Universität, Planegg-Martinsried, Germany
[18]Department of Physiology and Biophysics, Weill Cornell Medical College in Qatar (WCMC-Q), Education City–Qatar Foundation, Doha, Qatar
[19]Lehrstuhl für Mathematische Modelle Biologischer Systeme, Technische Universität München, Garching, Germany

been reported to have other pleiotropic effects; e.g., it reduces insulin resistance (2), improves the uptake of glucose in muscle (3,4), reduces the risk for cancer (5), and lowers the values of LDL cholesterol (LDL-C) (6). The underlying mechanism of the reduction of LDL-C is, at least in part, due to the activation of the AMPK in the liver (7). Apart from that, AMPK affects several processes such as nitric oxide (NO) production by endothelial NO synthase (eNOS) (8), which is also stimulated by metformin (9). However, the mode of action of metformin is not completely understood (10–12).

Our previous study was based on a targeted metabolomics approach to explore the effects of metformin on lipid profiles in the population-based KORA (Cooperative Health Research in the Region of Augsburg) cohort (6,13,14). Irving et al. (15) recently reported decreased levels of arginine and citrulline as an effect of insulin sensitizer therapy in 12 metformin- and pioglitazone-treated individuals and 13 placebo-treated control subjects. Nontargeted metabolomic measurements have been applied to investigate hyperglycemia (16,17) and the effects of metformin treatment in individuals without diabetes (18). However, none of the previous nontargeted metabolomics studies investigated metformin treatment in patients with T2D.

In this study we focused on serum metabolites associated with metformin treatment based on a nontargeted approach in a human population from the KORA cohort. A cross-link from human to mice was corroborated in multiple tissues (plasma, liver, skeletal muscle, and epididymal adipose tissue) from a mouse study. Biologically relevant pathways for the identified metabolites were analyzed using bioinformatical approaches.

RESEARCH DESIGN AND METHODS

Ethics Statement

All participants gave written informed consent. The KORA study was approved by the ethics committee of the Bavarian Medical Association, Munich, Germany.

Approval for Mouse Study

Within this study, all mice were bred and housed in a temperature- and humidity-controlled environment in compliance with Federation of European Laboratory Animal Science Associations protocols. Animal experiments were approved by the District Government of Upper Bavaria (Regierung von Oberbayern, Gz.55.2–1-54–2531–70–07, 55.2–1-2532–153–11).

KORA Cohort

KORA is a population-based cohort study conducted in southern Germany (14). The baseline survey 4 (KORA S4) consists of 4,261 individuals (aged 25–74 years) examined between 1999 and 2001. During the years of 2006 to 2008, 3,080 individuals took part in the follow-up survey 4 (KORA F4). Clinical data for each participant were retrieved from medical records. On the basis of fasting glucose, 2-h postglucose load, and physician-validated and self-reported diagnoses, KORA participants were classified according to the World Health Organization diagnostic criteria. A further grouping of patients with T2D was based on information on medication (19,20) (Table 1). Only participants with metabolite measurements were included in the present analysis (Metabolon, n = 1,768 in KORA F4). We excluded *1*) participants with overnight nonfasting serum samples (n = 8), *2*) patients suffering from type 1 diabetes and drug-induced (e.g., via steroids) diabetes (n = 6), *3*) T2D patients treated with insulin (n = 16) or both insulin and metformin (n = 13), and *4*) patients taking glucose-lowering oral medication other than metformin (n = 17). Furthermore, participants with isolated impaired fasting glucose (IFG) (n = 77) were excluded. We have previously shown that IFG and impaired glucose tolerance (IGT) should be considered as two different phenotypes (21).

In KORA F4, we focused on four groups: *1*) participants with normal glucose tolerance (NGT), *2*) individuals with prediabetes with IGT, *3*) T2D patients without glucose-lowering treatment (non-antidiabetes drug treated, ndt-T2D), and *4*) metformin-treated T2D (mt-T2D) patients (Table 1).

The same exclusion and classification criteria were used in the longitudinal analyses. We only considered participants with metabolite measurements in both studies (KORA S4 and F4, n = 818), and we excluded at both time points *1*) participants with overnight nonfasting serum samples (n = 88), which included patients suffering from type 1 diabetes or drug-induced diabetes, *2*) participants taking oral glucose-lowering medication other than metformin (n = 11), *3*) participants undergoing insulin treatment (n = 3), and *4*) participants with a missing diabetes status (n = 33). The remaining 683 participants were ndt-T2D individuals with prediabetes and healthy control subjects at KORA S4, 37 of whom started metformin treatment at KORA F4.

[20]Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany
[21]Institute of Human Genetics, Technische Universität München, München, Germany
[22]ShanghaiTech University, Shanghai, China
[23]Department of Endocrinology and Diabetology, Medical Faculty, Düsseldorf, Düsseldorf, Germany

Corresponding author: Rui Wang-Sattler, rui.wang-sattler@helmholtz-muenchen.de.

Received 21 April 2016 and accepted 1 August 2016.

This article contains Supplementary Data online at http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db16-0512/-/DC1.

G.K. and R.W.-S. contributed equally to this study.

M.F.S. is currently affiliated with Diabetes Medical Department, AstraZenenca GmbH, Wedel, Germany. S.C. is currently affiliated with ShanghaiTech University, Shanghai, China. S.N. is currently affiliated with Sanofi Deutschland GmbH, R&D Diabetes Research & Translational Medicine, Frankfurt am Main, Germany.

See accompanying article, p. 3537.

Table 1—Characteristics of the KORA F4 cross-sectional study samples (n = 1,604)

Clinical parameters	NGT	IGT	ndt-T2D	mt-T2D
n	1,143	272	115	74
Age, years	58.9 (8.5)	63.8 (8.1)	65.1 (7.1)	66.2 (7.5)
Male, %	45	50	61	58
BMI, kg/m^2	27.0 (4.3)	29.8 (4.7)	31.0 (4.8)	32.0 (5.6)
Waist, cm	92 (12.9)	100.2 (14.3)	104.9 (12.1)	106.7 (13.0)
Physical activity, % >1 h per week	63	54	48	36
High alcohol intake, %*	20	18	25	19
Smoker, %	18	8	12	15
Systolic BP, mmHg	121.6 (17.7)	129.0 (19.0)	134.3 (19.2)	130.7 (18.2)
HDL-C, mg/dL	58.9 (14.8)	54.4 (14.2)	49.6 (11.5)	50.4 (9.6)
LDL-C, mg/dL	140.7 (34.3)	144.5 (36.2)	136.6 (36.0)	124.1 (28.3)
Total cholesterol, mg/dL	223.4 (37.7)	226.4 (41.5)	214.1 (37.0)	203.8 (37.8)
Triglycerides, mg/dL	118.5 (84.6)	150.7 (89.0)	172.5 (128.7)	177.5 (140.5)
HbA_{1c}, %	5.4 (0.3)	5.6 (0.3)	6.3 (0.9)	6.8 (1.1)
HbA_{1c}, mmol/mol	35.7 (3.2)	38.1 (3.9)	44.91 (9.9)	51.2 (11.6)
Fasting glucose, mg/dL	93.2 (7.5)	101.2 (10.6)	126.6 (30.5)	142.0 (36.0)
2-h postglucose load, mg/dL	100.8 (20.6)	162.6 (17.5)	216.0 (50.7)†	
Time since diagnosis, years			1.4 (2.6)‡	7.4 (6.6)
Insulin, μIU/mL	7.1 (26.1)	10.1 (10.9)	14.3 (14.3)	11.6 (11.0)
Leptin, ng/mL	17.2 (19.2)	24.2 (21.3)	26.8 (21.2)	27.8 (25.0)
Statin usage, %	12	15	28	36
β-Blocker usage, %	16	31	43	38
ACE inhibitor usage, %	10	21	32	45
ARB usage, %	8	10	17	14
Insulin therapy, %	0	0	0	0
Metformin usage, %	0	0	0	100
Parental T2D, %	28	30	37	49

KORA F4 study characteristics (including solely subjects with available Metabolon measurements). Percentages of individuals or means (SD) are shown for each variable and each group (NGT, IGT, ndt-T2D, and mt-T2D). *≥20 g/day for women; ≥40 g/day for men. †n = 81. ‡For newly diagnosed T2D patients (n = 74), years since T2D diagnosis was defined as 0.

The data from KORA S4 and F4, including metabolite concentrations with clinical phenotypes, are available upon request through the platform KORA-PASST (project application self-service tool) (www.helmholtz-muenchen.de/kora-gen).

Blood Sampling

In the KORA cohort study, blood was drawn into Monovette serum tubes (Sarstedt AG & Co., Nümbrecht, Germany) in the morning between 8:00 A.M. and 10:30 A.M. after at least 8 h of fasting. Tubes were gently inverted twice, followed by resting 30 min at room temperature to obtain complete coagulation. For serum collection, blood was centrifuged at 2,750*g* at 15°C for 10 min. Serum was filled into synthetic straws, which were stored in liquid nitrogen (−196°C) until metabolomic analyses.

Nontargeted Metabolite Profiling

The serum samples from participants of KORA S4 and F4 were measured with the Metabolon analytical system (Metabolon, Inc., Durham, NC). Metabolon applied a nontargeted semiquantitative liquid chromatography–tandem mass spectrometry (LC-MS/MS) and gas chromatography–mass spectrometry (GC-MS) platform for the identification of structurally named and unknown molecules (22,23). We measured 363 (including 109 unknown) metabolites in fasting serum samples from KORA S4. In the 7-year KORA F4, 353 metabolites (including 107 unknown) were determined (24).

In this study, we applied the same criteria for quality control as described by Albrecht et al. (25). In brief, metabolites with more than 20% missing values were excluded, as were samples with more than 10% missing metabolites (25). All normalized relative ion counts were log transformed, and the remaining data were imputed with Multivariate Imputation by Chained Equations (MICE) (26). We used 363 metabolites in KORA S4 and 353 metabolites in KORA F4 (Supplementary Table 1). The number of overlapping metabolites in KORA S4 and

F4 was 312. Metabolite names were used according to Shin et al. (27); however, the identity of metabolite ID M32654 and the molecule "3-dehydrocarnitine*" could not be confirmed. We therefore used the name X-21365 (Supplementary Table 1).

Each metabolite was standardized with a mean of zero and an SD of one in each study after the exclusion of nonfasting participants.

Metformin Mouse Intervention Study

Pharmacological studies were conducted in 20 male 8-week-old diabetic BKS.Cg-*Dock7*$^{m+/+}$ *Lepr*^{db}/J (*db/db*) mice that were bred and housed in a temperature- and humidity-controlled environment in compliance with Federation of European Laboratory Animal Science Associations protocols. To exclude estrous cycle–related influences on metabolic parameters, only male mice were included in this study. From age 3 weeks, all mice were fed a high-fat diet (S0372-E010; ssniff Spezialdiäten, Soest, Germany) containing (gm%) palm fat (13.5), sunflower oil (13.5), starch (30), saccharose (10), casein (20), lignocellulose (5), mineral+vitamin mix (5+2), safflower oil (0.5), and linseed oil (0.5) to manifest a uniform diabetic phenotype. Animals received either vehicle (5% solutol/95% hydroxyethylcellulose) without (n = 10) or with metformin (300 mg/kg; Sigma Aldrich, Taufkirchen, Germany; n = 10) via gavage once daily between 5:00 and 6:00 P.M. before dark-phase onset (6:00 P.M.) for 14 days. At 18 ± 2 h after the last treatment, 4-h fasted mice were sacrificed with an isoflurane overdose, and organs and blood were immediately collected (4). Murine plasma was prepared from whole blood by centrifugation at 4°C, and tissues were freeze-clamped; both were stored at −80°C until further analyses. All samples were measured with the Metabolon analytical system. Metabolites with more than 20% missing values were excluded, as were samples with more than 10% missing metabolites (25). All normalized relative ion counts were log transformed, and the remaining data were imputed with MICE (26). Linear regression was done on metabolite values for metformin-treated mice as the cases as well as for the nonmetformin-treated, vehicle-gavaged mice as the controls. A metabolomics examination was done for plasma, liver, skeletal muscle, and epididymal adipose tissue (Table 5 and Fig. 1*B*).

Statistical Analysis

To evaluate the effect of metformin treatment on certain metabolites, multivariable linear regression models were conducted with the relative metabolite concentration values as outcome and the grouping variable as predictor. Each metabolite was assessed individually. To consider potential risk factors and confounding parameters with known effect on metabolite profiles (6,13,28–32), two models were used: *1*) adjusted for age and sex as the crude model and *2*) adjusted for age, sex, BMI, physical activity, high alcohol intake, smoking status, systolic blood pressure (BP), HbA_{1c}, fasting glucose, HDL cholesterol (HDL-C), and triglycerides as well as the use of statins, β-blockers, ACE inhibitors, and angiotensin receptor blockers (ARB) as the full model. The association of conventional risk factors of T2D as well as other population characteristics with metformin treatment was calculated via χ^2 test for categorical variables. Shapiro–Wilk test was applied to test continuous variables for normal distribution ($P \leq 0.05$ for nonnormally distributed variables, $P > 0.05$ for normally distributed variables), followed by Student's t test for normally distributed continuous variables and Wilcoxon test for nonnormally distributed continuous variables.

To account for multiple testing for the linear models, Bonferroni correction was applied, and only metabolites with $P < 0.05/353 = 1.42E\text{-}04$ were considered to be statistically significantly different in KORA F4. In addition, we calculated the adjusted P value with the false discovery rate (FDR) using the Benjamini-Hochberg method, which is not as stringent as the Bonferroni correction. For the full linear models, participants were excluded because of missing information of considered confounders. This led to 1,138 NGT (after exclusion of five individuals because of missing confounding information), 272 IGT, 114 ndt-T2D (after exclusion of one individual because of missing confounding information), and 70 mt-T2D (after exclusion of four individuals because of missing confounding information) participants.

In the KORA S4 to F4 longitudinal study (S4 → F4), generalized estimating equations (GEE) were used to validate the significant metabolites in both crude and full models.

All statistical analyses were performed in R (version 3.2.2) (33).

Pathway Analysis

Pathways were explored using databases, considering tissue and organ specificity. The link from observed significant metabolites to the interacting enzymes was drawn using the Human Metabolome Database (34). Protein–protein interactions were analyzed with the Search Tool for the Retrieval of Interacting Genes (35) and the Kyoto Encyclopedia of Genes and Genomes (36). To consider drug-related effects of metformin on certain targets, we used DrugBank (37). The link between metformin targets and the protein network was analyzed using the Kyoto Encyclopedia of Genes and Genomes (36).

RESULTS

Population Characteristics of Human and Mouse Studies

On the basis of the available nontargeted metabolomic profiles, our human discovery study, KORA F4, includes 1,143 NGT, 272 IGT, 115 ndt-T2D, and 74 mt-T2D participants (Table 1). Among the four groups, mt-T2D patients were the oldest, were more frequently men, and had the highest values of HbA_{1c}, fasting glucose, triglycerides, BMI, and waist circumference (Table 1).

The longitudinal KORA study includes samples of 683 participants without metformin treatment at baseline,

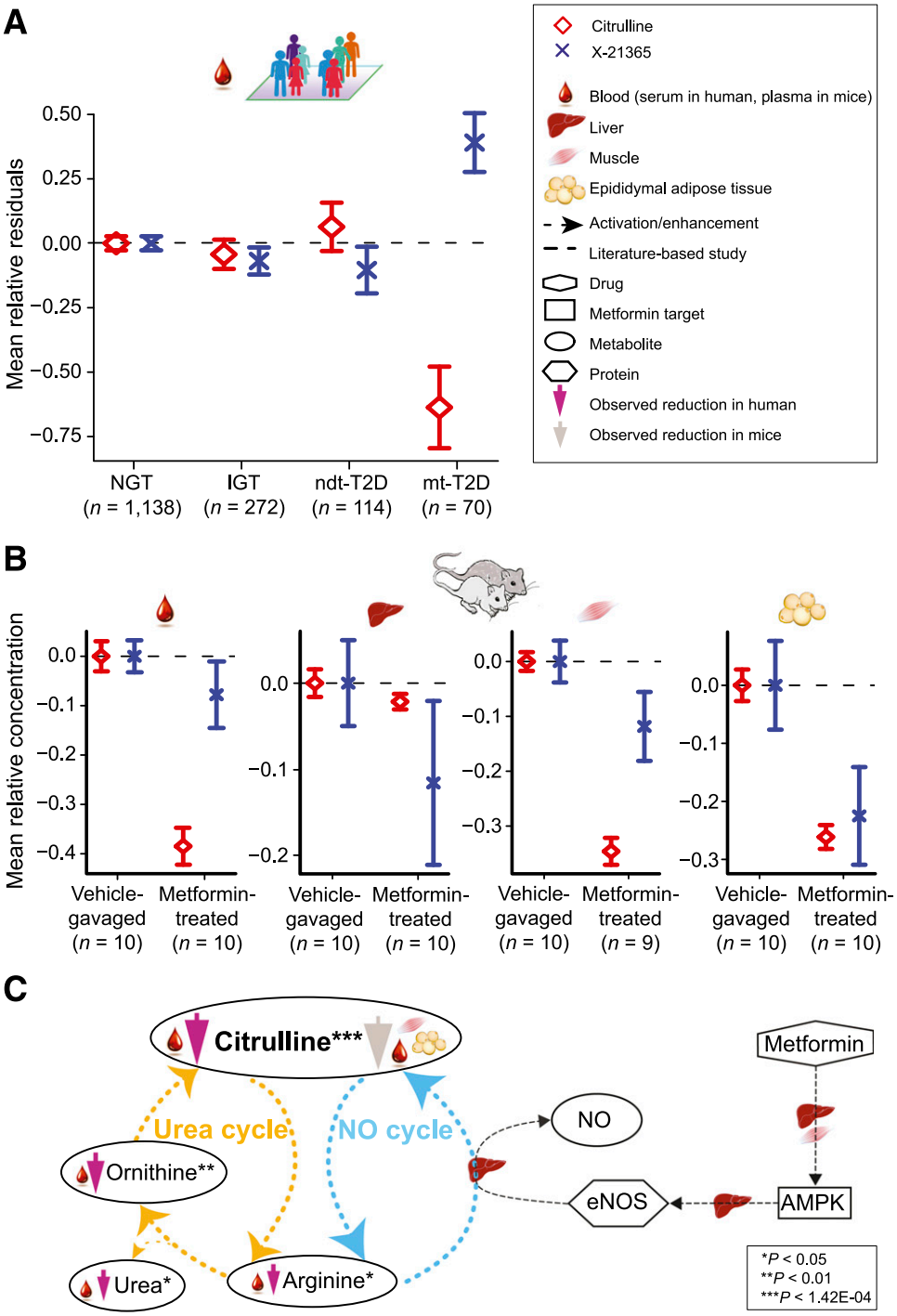

Figure 1—Differences in relative metabolite concentrations in a human study, in a mouse study, and in organ-specific pathways. *A*: Mean relative residuals of the concentrations (with SEM) of two metabolites for the NGT, IGT, ndt-T2D, and mt-T2D groups derived in cross-sectional analysis of KORA F4. Residuals were calculated from a linear regression model with full adjustments. *B*: Mean relative concentrations (with SEM) of two metabolites in four different mouse tissues (plasma, liver, skeletal muscle, and epididymal adipose tissue). *C*: The connections indicated by liver, muscle, and blood (plasma and serum) show organ specificity between metabolites, pathway-related proteins, metformin targets, and metformin. The metabolites (ellipses) were connected to metformin treatment (straight-sided hexagons), proteins (hexagons), and metformin targets (rectangles). The activation/stimulation is indicated with arrows. For further information, see Tables 3–5 and Supplementary Tables 2 and 3.

37 of whom were treated with metformin in the 7-year follow-up (Table 2).

From the metformin-treated mice, we obtained 10 samples for plasma, liver, and epididymal adipose tissue and 9 samples for skeletal muscle. In the same amount of vehicle-gavaged control mice, we obtained 10 samples for plasma, liver, epididymal adipose tissue, and skeletal muscle.

Table 2—Characteristics of the KORA S4 → F4 prospective study samples (n = 683)

	KORA S4 w/o metformin§ → KORA F4 w/o metformin§			KORA S4 w/o metformin§ → KORA F4 w/ metformin		
Clinical parameters	S4	F4	*P* value	S4	F4	*P* value
n	646	646		37	37	
Age, years	61.4 (4.2)	68.5 (4.2)		63.4 (3.8)	70.4 (3.9)	
Male, %	51	51		54	54	
BMI, kg/m^2	27.9 (3.9)	28.2 (4.2)	0.2	32.8 (4.3)	32 (4.5)	0.49‖
Waist, cm	93.9 (11.0)	96.8 (11.9)	1.72E-05	106.3 (11.3)	106.7 (12.3)	0.87‖
Physical activity, % >1 h per week	47	57	5.47E-04	32	43	0.47
High alcohol intake, %*	20	19	0.65	27	16	0.4
Smoker, %	13	8	9.00E-03	14	8	0.71
Systolic BP, mmHg	132.1 (18.7)	128.2 (19.6)	3.12E-04	144.9 (18.1)	131.6 (18.3)	2.43E-03‖
HDL-C, mg/dL	59.1 (16.3)	56.7 (14.2)	0.02	53.4 (11.8)	52.4 (7.9)	0.69‖
LDL-C, mg/dL	154.5 (40.8)	142.5 (36.9)	1.86E-08	143.5 (37.6)	123.5 (23.6)	0.02
Total cholesterol, mg/dL	245.5 (42.0)	225.2 (40.7)	2.2E-16	234.1 (41.0)	202.1 (33.6)	5.20E-04
Triglycerides, mg/dL	129.9 (76.4)	132.7 (83.4)	0.56	170.6 (169)	154.5 (159.2)	0.79
HbA_{1c}, %	5.6 (0.3)	5.6 (0.5)	0.53	6.4 (0.9)	6.6 (0.7)	0.14
HbA_{1c}, mmol/mol	37.6 (3.7)	38.0 (5.7)	0.53	46.8 (10.3)	48.3 (7.8)	0.14
Fasting glucose, mg/dL	99.7 (10.9)	100.7 (17.6)	0.53	130.7 (30.1)	129.6 (28.3)	0.9
2-h postglucose load, mg/dL	115.1 (37.1)	126.9 (40.8)	8.34E-09	205.5 (76.8)		
Statin usage, %	10	22	8.05E-10	8	29	0.04
β-Blocker usage, %	17	31	5.16E-09	16	29	0.27
ACE inhibitor usage, %	8	21	2.72E-11	16	51	3.18E-03
ARB usage, %	3	12	3.67E-10	3	14	0.2
Insulin therapy, %	0	0		0	0	
Metformin usage, %	0	0		0	100	
Parental T2D, %	25	25		47	47	

Percentages of individuals or means (SD) of participants (with available Metabolon measurements for KORA S4 and F4) are shown for each variable and each group. w/o, without; w/, with. *>40 g/day in men; >20 g/day in women. §Includes participants with NGT, isolated IFG, IGT, and ntd-T2D. ‖Normally distributed (every other distribution is not normally distributed).

Two Metabolites Are Associated With Metformin Treatment in a Human Cross-sectional Study

Two out of the 353 used metabolites (citrulline and X-21365) were found to be significantly ($P < 1.42E-04$) associated with metformin treatment when comparing mt-T2D with ndt-T2D patients in the cross-sectional KORA F4 study (Table 3 and Fig. 1*A*). Using multivariable linear regression models, we detected negative β-estimates for both the crude (β = –0.75, P = 2.31E-05) and full adjustment (β = –0.79, P = 2.54E-05) for

Table 3—Two human serum metabolites significantly associated with metformin treatment in a cross-sectional analyses (KORA F4)

	Crude linear model mt-T2D (n = 74) vs. ndt-T2D (n = 115)			Full linear model mt-T2D (n = 70)¶ vs. ndt-T2D (n = 114)¶		
Metabolite	β (95% CI) per SD	*P* value	FDR	β (95% CI) per SD	*P* value	FDR
Citrulline	−0.75 (−1.09, −0.41)	**2.31E-05**	**2.83E-04**	−0.79 (−1.15, −0.43)	**2.54E-05**	**2.83E-04**
X-21365	0.67 (0.38, 0.96)	**7.54E-06**	**1.42E-04**	0.65 (0.34, 0.97)	**5.20E-05**	**1.42E-04**

Estimates (β) and P values for the comparison of 189 participants (74 mt-T2D and 115 ndt-T2D) were calculated using linear regression analysis with the crude and full adjustments. Because of missing confounding information, the models with full adjustment were based on fewer participants. Significant metabolites are highlighted in boldface type with respect to Bonferroni correction ($P < 0.05/353 = 1.42E-04$) or the FDR. ¶After exclusion of individuals because of missing confounding information.

citrulline. Hence, the relative concentration of citrulline is significantly lower in mt-T2D compared with ndt-T2D patients. By contrast, the relative concentration of X-21365 was significantly higher in mt-T2D patients than in ndt-T2D patients (Table 3 and Fig. 1*A*). When applying the FDR, no additional associations were found to be significant in both crude and full models (Supplementary Table 3). When applying a significance cutoff of $P < 0.05$ to the comparison of mt-T2D with ndt-T2D for the models with crude and full adjustment, 44 additional metabolites were found, including ornithine, arginine, and urea (Supplementary Table 3).

Five additional pairwise comparisons between the four groups (NGT, IGT, ndt-T2D, mt-T2D) confirmed that these two metabolites are specific for metformin treatment and not due to the progression of the disease. The relative concentration of citrulline was significantly lower in the mt-T2D than in the NGT and IGT groups, whereas the concentration of X-21365 was significantly higher (Fig. 1*A* and Supplementary Table 2). Consistently, neither of the two metabolites showed a significantly different relative concentration in the pairwise comparisons among the three groups without metformin treatment, i.e., NGT, IGT, and ndt-T2D (Fig. 1*A* and Supplementary Table 2).

The Spearman correlation coefficient between the two metabolites was low ($r = 0.06$). We observed similar associations between the two metabolites with a number of risk factors of T2D ($-0.19 < r < 0.19$) when considering 189 individuals with ndt-T2D and mt-T2D in KORA F4 (Supplementary Fig. 1).

Metformin Treatment Is Associated With Decreased Blood Citrulline Values in a Human Longitudinal Cohort

The two metformin-associated metabolites were further investigated in the prospective KORA study. In 37 patients who started metformin treatment during the 7-year follow-up, citrulline was found to be significantly (Bonferroni cutoff for two identified metabolites $P < 0.05/2 = 0.025$) decreased in longitudinal in both the crude (β = –0.67, P = 2.03E-05) and the full model (β = –0.61, 2.96E-04, Table 4). In the same group, X-21365 was significantly increased in the crude (β = 0.41, P = 5.62E-03) but not in the full model (β = 0.16, P = 0.374, Table 4).

Lower Citrulline Relative Concentrations in Plasma, Skeletal Muscle, and Epididymal Adipose Tissue Confirmed in Metformin-Treated Mice

We observed significantly lower plasma citrulline relative concentrations in *db/db* mice following daily, subchronic metformin treatment compared with the vehicle-gavaged control mice (β = –0.39, P = 2.56E-07, Table 5 and Fig. 1*B*), which is consistent with the results observed in humans. In addition, we found significantly lower values of citrulline in both skeletal muscle (β = –0.35, P = 1.79E-09) and epididymal adipose tissue (β = –0.26, P = 4.52E-07). However, citrulline values in the liver did not differ between the metformin-treated and vehicle-gavaged non–metformin-treated *db/db* animals (β = –0.02, P = 0.258, Table 5 and Fig. 1*B*). Significantly different relative concentrations of X-21365 were not found in plasma, skeletal muscle, epididymal adipose tissue, or liver of metformin-treated mice when compared with the controls (Table 5 and Fig. 1*B*).

DISCUSSION

We found significantly lower values of citrulline and significantly higher values of X-21365 in the serum of T2D patients who underwent metformin treatment compared with the nontreated patients. Additionally, using longitudinal settings, we observed that the values of citrulline significantly decreased in patients after they started metformin treatment during the follow-up. A mouse intervention study using metformin confirmed the lower values of citrulline in plasma, as well as in skeletal muscle and epididymal adipose tissue, but not in liver. Citrulline is a nonproteinogenic amino acid, the product of anabolic and the substrate of catabolic processes (38,39). It is synthesized from arginine by releasing NO, which is involved in the regulation of numerous processes in the nervous system, the immune system, and the cardiovascular system (8). Additionally, citrulline is produced from ornithine in the urea cycle (38). We observed ornithine, urea, and arginine to be lowered in human serum (Fig. 1*C*). Consistently, in our previous study, which was based on a targeted metabolomics approach, ornithine was found to be significantly lower in the metformin-treated T2D patients of the

Table 4—Citrulline remains significantly associated with metformin treatment in human serum in a longitudinal analysis (KORA S4 → F4)

	Crude GEE model mt-T2D (n = 37) vs. nonmetformin-treated (n = 646) participants§			Full GEE model mt-T2D (n = 33)¶ vs. nonmetformin-treated (n = 629)¶ participants§		
Metabolite	β (95% CI) per SD	P value	FDR	β (95% CI) per SD	P value	FDR
Citrulline	−0.67 (−0.98, −0.36)	**2.03E-05**	**1.76E-03**	−0.61 (−0.94, −0.28)	**2.96E-04**	**3.21E-04**
X-21365	0.41 (0.12, 0.69)	**5.62E-03**	0.011	0.14 (−0.17, 0.45)	0.374	0.024

GEE model with crude and full adjustment was used to assess the associations between metformin treatment and metabolite serum values in the longitudinal study of 683 participants with no antidiabetes medical treatment at KORA S4. Of these participants, 37 started metformin treatment after KORA S4. Because of missing confounding information, the models with full adjustment were based on fewer participants. Significant metabolites are highlighted in boldface type with respect to Bonferroni correction ($P < 0.05/2 = 0.025$) and the FDR. §Includes participants with NGT, isolated IFG, IGT, and ntd-T2D. ¶After exclusion of individuals because of missing confounding information.

KORA F4 study. Citrulline was not measured in the targeted panel we used (6).

Metformin activates AMPK in the liver and muscle (7,40). AMPK in turn may stimulate eNOS by its phosphorylation (8,41), which suggests a consequent increase of the NO production in the NO cycle (Fig. 1*C*). It is known that elevated production of NO is reflected by increased values of citrulline in urine (42), as citrulline can be used as a surrogate marker for NO (43). The decreased values of citrulline and its precursors in blood, skeletal muscle, and epididymal adipose tissue, as were observed in our study, are most likely due to an accountable, increased excretion of this metabolite. However, urine samples were not available in this study. To confirm this assumption, further studies are necessary.

Furthermore, the lower values of citrulline and arginine we observed are likely to be a consequence of the activation of eNOS. In the NO cycle, eNOS catalyzes the reaction from arginine to citrulline, thereby releasing NO (9,38). NO in turn has beneficial cardiovascular effects. The reason is that NO influences smooth muscles and activates their relaxation (44). This underlies the clinical practice guidelines, which have recommended the use of metformin as first-line therapy in T2D patients with cardiovascular disease, mainly in patients with observed reduced NO levels (45). Additional intake of citrulline to compensate for the lower values of citrulline and arginine might even increase the beneficial effects of metformin on cardiovascular disease (46).

Additionally, citrulline is synthesized in the urea cycle, which is strongly interlocked with the NO cycle (Fig. 1*C*). In mammals, both cycles primarily take place in the liver, but they also take place in the kidney (47). The same accounts for the NO cycle, in which arginine also plays an important role. In fact, similar effects of metformin on the urea and NO cycle were mentioned by Irving et al. (15). Their study design focused on plasma samples of 25 male overweight or obese participants. Furthermore, all 12 metformin-treated participants were additionally treated with pioglitazone (15). Our findings in multiple tissues of mice that were exclusively treated with metformin and in serum of 189 T2D patients enable a deeper understanding of the underlying mode of action for metformin.

The observation that the citrulline values are not affected in the liver of metformin-treated mice is presumably a consequence of the hepatic localization of the consecutive production of citrulline in both the NO and the urea cycle (38), which conserves a state of equilibrium. This is in line with observations in a recent study (18). Furthermore, significantly decreased ornithine values were found in plasma of individuals without diabetes (18).

Apart from the NO and urea cycles, there are additional physiological processes that produce citrulline. The metabolite is also synthesized from other amino acids. Examples of such precursors are glutamine, which is converted in the enterocytes, proline, and glutamate (38). However, we did not observe any significant concentration difference for these metabolites in our human cohort.

Table 5—Metabolites significantly associated with metformin treatment in mouse models

	In plasma		In liver		In skeletal muscle		In adipose tissue	
Metabolite	β (95% CI) per SD	*P* value	β (95% CI) per SD	*P* value	β (95% CI) per SD	*P* value	β (95% CI) per SD	*P* value
Citrulline	−0.39 (−0.49, −0.28)	**2.56E-07**	−0.02 (−0.06, 0.02)	0.258	−0.35 (−0.41, −0.28)	**1.79E-09**	−0.26 (−0.33, −0.19)	**4.52E-07**
X-21365	−0.08 (−0.24, 0.08)	0.311	−0.12 (−0.34, 0.11)	0.295	−0.12 (−0.27, 0.04)	0.126	−0.23 (−0.46, 0.01)	0.063

Estimates (β) and *P* values for the comparison between metformin-treated ($n = 10$, in skeletal muscle [$n = 9$]) and nontreated mice ($n = 10$) sacrificed at 4 h after the last treatment. Significant metabolites are highlighted in boldface type ($P < 0.05$).

X-21365 was not found to be significantly higher in the fully adjusted longitudinal analyses of the KORA S4 → F4 cohort, although it was significant in both cross-sectional analyses and in the longitudinal analyses with crude adjustment. In mice, we did not observe significant differences of X-21365 in any of the examined tissues. Recent advances in the identification of metabolites spectra suggest that this unknown metabolite (X-21365) might be 5-trimethylaminovalerate and therefore closely related to the gut microbiome, which is in line with a recent study (48). Additional studies using both blood and stool samples have to be conducted to confirm this.

The values of metabolites in humans of the KORA study are influenced by multiple factors such as age, sex, BMI, lifestyle, clinical measurements, and medication (6,13,28–32). We therefore considered these factors in the models underlying our cross-sectional discovery and longitudinal investigations in a human cohort. Considering the mouse study, there was no need for a comparable adjustment, as the animals were kept under strict laboratory conditions.

Because of the physiological similarity, we used data from a mouse study not only to corroborate our findings in humans but to extend our investigations on other tissues. However, our findings are limited by the comparison of metabolic analytes in two different blood matrices and species: human serum and mouse plasma. In theory, the analytical method could be affected by the difference in matrix, and delicate analytes could deteriorate during the prolonged preparation time of serum compared with that of plasma. Therefore, a direct comparison between the matrices serum and plasma has limitations (49). With respect to this, we compared the serum metabolites only within humans, the plasma metabolites only within mice, and each mice tissue separately (50). The observational nature of cohort studies and the applied methods are of purely statistical character, yet still they offer the opportunity to identify unknown coherences and to design study settings to confirm underlying mechanisms. Because of the fact that NO is below the mass cutoff imposed on the instruments, our investigations did not contain measurements of this chemical compound. Nevertheless, our observations suggest further investigations with a specific design to address the involvement of the NO and urea cycle in metformin treatment.

In summary, we observed that serum values of citrulline were reduced under metformin treatment in human patients with T2D and, in a translational approach, also in plasma, skeletal muscle, and epididymal adipose tissue of diabetic mice. The underlying mechanism is most likely the metformin-induced activation of AMPK and its consequent increase of eNOS activity, which is linked to citrulline by the NO cycle.

Acknowledgments. The authors express their appreciation to all KORA study participants for donating their blood and time. The authors thank the field staff in Augsburg, Germany, for conducting the KORA studies. The authors are grateful to the staff (Julia Scarpa, Katharina Faschinger, Franziska Scharl, Nadine Lindemann, Humberto Chavez, and Arsin Sabunchi) from the Institute of Epidemiology at the Helmholtz Zentrum München and the Genome Analysis Center Metabolomics Platform who helped with the sample logistics, data and straw collection, and metabolomic measurements. Additionally, the authors thank the staff (especially Andrea Schneider, Anja Ludolph, Sladjana Jelic, and Birgit Langer) from the Institute of Genetic Epidemiology at the Helmholtz Zentrum München and the platform KORA project application self-service tool (KORA.PASST) for their help with KORA data logistics. The authors also thank the people of the Institute of Diabetes and Regeneration Research (Anett Seelig and Jürgen Schultheiß) and the Institute of Experimental Genetics (Moya Wu), as well as the animal caretaker staff of the German Mouse Clinic, for excellent technical assistance.

Funding. The KORA study was initiated and financed by the Helmholtz Zentrum München German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) and by the State of Bavaria. Furthermore, KORA research was supported by the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. Part of this project was supported by European Union Seventh Framework Programme grants HEALTH-2009-2.2.1-3/242114 (Project OPTiMiSE) and HEALTH-2013-2.4.2-1/602936 (Project CarTarDis). The German Diabetes Center is funded by the German Federal Ministry of Health (Bundesministerium für Gesundheit, Berlin, Germany) and the Ministry of Innovation, Science and Research of the State of North Rhine-Westphalia (Düsseldorf, Germany). The diabetes part of the KORA F4 study was funded by a grant from the Deutsche Forschungsgemeinschaft (DFG; RA 459/3-1). This study was supported in part by a grant from the Bundesministerium für Bildung und Forschung to the German Center for Diabetes Research (DZD) and by the research consortium "Systems Biology of Metabotypes" (SysMBo grant 0315494A). K.Su. is supported by Biomedical Research Program funds at Weill Cornell Medical College in Qatar, a program funded by the Qatar Foundation.

Duality of Interest. M.F.S. was employed at Helmholtz Zentrum München GmbH during the execution of this study. M.F.S. is currently an employee of the Diabetes Medical Department of AstraZeneca GmbH (Wedel, Germany); however, the company was not involved in work related to data and manuscript generation. R.P.M. is an employee of Metabolon, Inc. S.N. was employed by the Helmholtz Zentrum München GmbH during the execution of this study. S.N. is currently an employee of Sanofi Deutschland GmbH; however, the company was not involved in work related to data and manuscript generation. No other potential conflicts of interest relevant to this article were reported.

Author Contributions. J.Adam, S.B., and R.W.-S. wrote the manuscript. J.L., M.F.S., R.P.M., M.Rot., M.T., S.C., C.H., Y.L., D.A., T.M., M.Rod., and S.N. assisted in manuscript generation. J.Adam, J.L., T.X., and J.B analyzed the data and interpreted the results. M.F.S., R.P.M., M.H., C.H., W.R., G.G., J.Adams., T.I., K.St., C.G., A.P., K.Su., M.H.d.A., S.N., and G.K. performed the experiments, including metabolic profiling. G.K. and R.W.-S. conceived and designed the study. R.W.-S. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Turner R; UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 1998;352:837–853

2. Dunaif A. Insulin resistance and the polycystic ovary syndrome: mechanism and implications for pathogenesis. Endocr Rev 1997;18:774–800

3. Giugliano D, De Rosa N, Di Maro G, et al. Metformin improves glucose, lipid metabolism, and reduces blood pressure in hypertensive, obese women. Diabetes Care 1993;16:1387–1390

4. Neschen S, Scheerer M, Seelig A, et al. Metformin supports the antidiabetic effect of a sodium glucose cotransporter 2 inhibitor by suppressing endogenous glucose production in diabetic mice. Diabetes 2015;64:284–290

5. Decensi A, Puntoni M, Goodwin P, et al. Metformin and cancer risk in diabetic patients: a systematic review and meta-analysis. Cancer Prev Res (Phila) 2010;3:1451–1461
6. Xu T, Brandmaier S, Messias AC, et al. Effects of metformin on metabolite profiles and LDL cholesterol in patients with type 2 diabetes. Diabetes Care 2015; 38:1858–1867
7. Zhou G, Myers R, Li Y, et al. Role of AMP-activated protein kinase in mechanism of metformin action. J Clin Invest 2001;108:1167–1174
8. Chen ZP, Mitchelhill KI, Michell BJ, et al. AMP-activated protein kinase phosphorylation of endothelial NO synthase. FEBS Lett 1999;443:285–289
9. Davis BJ, Xie Z, Viollet B, Zou MH. Activation of the AMP-activated kinase by antidiabetes drug metformin stimulates nitric oxide synthesis in vivo by promoting the association of heat shock protein 90 and endothelial nitric oxide synthase. Diabetes 2006;55:496–505
10. Konopka AR, Esponda RR, Robinson MM, et al. Hyperglucagonemia mitigates the effect of metformin on glucose production in prediabetes. Cell Reports 2016;15:1394–1400
11. Huo T, Cai S, Lu X, Sha Y, Yu M, Li F. Metabonomic study of biochemical changes in the serum of type 2 diabetes mellitus patients after the treatment of metformin hydrochloride. J Pharm Biomed Anal 2009;49:976–982
12. Walford GA, Davis J, Warner AS, et al. Branched chain and aromatic amino acids change acutely following two medical therapies for type 2 diabetes mellitus. Metabolism 2013;62:1772–1778
13. Brandmaier S, Xu T, Illig T, Suhre K, Adamski J, Wang-Sattler R. Response to comment on Xu et al. Effects of metformin on metabolite profiles and LDL cholesterol in patients with type 2 diabetes. Diabetes Care 2015;38:1858–1867. Diabetes Care 2015;38:e216–e217
14. Holle R, Happich M, Löwel H, Wichmann HE.; MONICA/KORA Study Group. KORA–a research platform for population based health research. Gesundheitswesen 2005;67(Suppl. 1):S19–S25
15. Irving BA, Carter RE, Soop M, et al. Effect of insulin sensitizer therapy on amino acids and their metabolites. Metabolism 2015;64:720–728
16. Menni C, Fauman E, Erte I, et al. Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. Diabetes 2013;62:4270–4276
17. Suhre K, Meisinger C, Döring A, et al. Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. PLoS One 2010;5:e13953
18. Rotroff DM, Oki NO, Liang X, et al. Pharmacometabolomic assessment of metformin in non-diabetic, African Americans. Front Pharmacol 2016;7:135
19. Rathmann W, Kowall B, Heier M, et al. Prediction models for incident type 2 diabetes mellitus in the older population: KORA S4/F4 cohort study. Diabet Med 2010;27:1116–1123
20. Meisinger C, Strassburger K, Heier M, et al. Prevalence of undiagnosed diabetes and impaired glucose regulation in 35-59-year-old individuals in Southern Germany: the KORA F4 Study. Diabet Med 2010;27:360–362
21. Wang-Sattler R, Yu Z, Herder C, et al. Novel biomarkers for pre-diabetes identified by metabolomics. Mol Syst Biol 2012;8:615
22. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. Anal Chem 2009;81:6656–6667
23. Ohta T, Masutomi N, Tsutsui N, et al. Untargeted metabolomic profiling as an evaluative tool of fenofibrate-induced toxicology in Fischer 344 male rats. Toxicol Pathol 2009;37:521–535
24. Suhre K, Shin S-Y, Petersen A-K, et al.; CARDIoGRAM. Human metabolic individuality in biomedical and pharmaceutical research. Nature 2011;477:54–60
25. Albrecht E, Waldenberger M, Krumsiek J, et al. Metabolite profiling reveals new insights into the regulation of serum urate in humans. Metabolomics 2014; 10:141–151
26. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45:1–67
27. Shin S-Y, Fauman EB, Petersen A-K, et al.; Multiple Tissue Human Expression Resource (MuTHER) Consortium. An atlas of genetic influences on human blood metabolites. Nat Genet 2014;46:543–550
28. Yu Z, Zhai G, Singmann P, et al. Human serum metabolic profiles are age dependent. Aging Cell 2012;11:960–967
29. Xu T, Holzapfel C, Dong X, et al. Effects of smoking and smoking cessation on human serum metabolite profile: results from the KORA cohort study. BMC Med 2013;11:60
30. Jaremek M, Yu Z, Mangino M, et al. Alcohol-induced metabolomic differences in humans. Transl Psychiatry 2013;3:e276
31. Altmaier E, Fobo G, Heier M, et al. Metabolomics approach reveals effects of antihypertensives and lipid-lowering drugs on the human metabolism. Eur J Epidemiol 2014;29:325–336
32. Jourdan C, Petersen A-K, Gieger C, et al. Body fat free mass is associated with the serum metabolite profile in a population-based study. PLoS One 2012;7:e40009
33. R Development Core Team. R: *A Language and Environment for Statistical Com*puting. Vienna, Austria. R Foundation for Statistical Computing, 2009
34. Wishart DS, Tzur D, Knox C, et al. HMDB: the Human Metabolome Database. Nucleic Acids Res 2007;35:D521–D526
35. Szklarczyk D, Franceschini A, Kuhn M, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011;39:D561–D568
36. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 2014;42:D199–D205
37. Law V, Knox C, Djoumbou Y, et al. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res 2014;42:D1091–D1097
38. Curis E, Nicolis I, Moinard C, et al. Almost all about citrulline in mammals. Amino Acids 2005;29:177–205
39. Baur H, Stalon V, Falmagne P, Luethi E, Haas D. Primary and quaternary structure of the catabolic ornithine carbamoyltransferase from Pseudomonas aeruginosa. Extensive sequence homology with the anabolic ornithine carbamoyltransferases of Escherichia coli. Eur J Biochem 1987;166:111–117
40. Musi N, Hirshman MF, Nygren J, et al. Metformin increases AMP-activated protein kinase activity in skeletal muscle of subjects with type 2 diabetes. Diabetes 2002;51:2074–2081
41. Zhang Y, Lee TS, Kolb EM, et al. AMP-activated protein kinase is involved in endothelial NO synthase activation in response to shear stress. Arterioscler Thromb Vasc Biol 2006;26:1281–1287
42. Wanchu A, Khullar M, Sud A, Deodhar SD, Bambery P. Elevated urinary nitrite and citrulline levels in patients with rheumatoid arthritis. Inflammopharmacology 1999;7:155–161
43. Sud A, Khullar M, Wanchu A, Bambery P. Increased nitric oxide production in patients with systemic sclerosis. Nitric Oxide 2000;4:615–619
44. Conti V, Russomanno G, Corbi G, Izzo V, Vecchione C, Filippelli A. Adrenoreceptors and nitric oxide in the cardiovascular system. Front Physiol 2013;4:321
45. Tessari P, Cecchet D, Cosma A, et al. Nitric oxide synthesis is reduced in subjects with type 2 diabetes and nephropathy. Diabetes 2010;59:2152–2159
46. Romero MJ, Platt DH, Caldwell RB, Caldwell RW. Therapeutic use of citrulline in cardiovascular disease. Cardiovasc Drug Rev 2006;24:275–290
47. Husson A, Brasse-Lagnel C, Fairand A, Renouf S, Lavoinne A. Argininosuccinate synthetase from the urea cycle to the citrulline-NO cycle. Eur J Biochem 2003; 270:1887–1899
48. Zhernakova A, Kurilshikov A, Bonder MJ, et al.; LifeLines cohort study. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science 2016;352:565–569
49. Becker S, Kortz L, Helmschrodt C, Thiery J, Ceglarek U. LC-MS-based metabolomics in the clinical laboratory. J Chromatogr B Analyt Technol Biomed Life Sci 2012;883-884:68–75
50. Yu Z, Kastenmüller G, He Y, et al. Differences between human plasma and serum metabolite profiles. PLoS One 2011;6:e21230

Molecular Systems Biology 8; Article number 615; doi:10.1038/msb.2012.43
Citation: ***Molecular Systems Biology*** **8:615**

www.molecularsystemsbiology.com

molecular systems biology

Novel biomarkers for pre-diabetes identified by metabolomics

Rui Wang-Sattler[1,28,*], Zhonghao Yu[1,28], Christian Herder[2,28], Ana C Messias[3,28], Anna Floegel[4], Ying He[5,6], Katharina Heim[7], Monica Campillos[8], Christina Holzapfel[1,9], Barbara Thorand[10], Harald Grallert[1], Tao Xu[1], Erik Bader[1], Cornelia Huth[10], Kirstin Mittelstrass[1], Angela Döring[11], Christa Meisinger[10], Christian Gieger[12], Cornelia Prehn[13], Werner Roemisch-Margl[8], Maren Carstensen[2], Lu Xie[5], Hisami Yamanaka-Okumura[14], Guihong Xing[15], Uta Ceglarek[16], Joachim Thiery[16], Guido Giani[17], Heiko Lickert[18], Xu Lin[19], Yixue Li[5,6], Heiner Boeing[4], Hans-Georg Joost[4], Martin Hrabé de Angelis[13,20], Wolfgang Rathmann[17], Karsten Suhre[8,21,22], Holger Prokisch[7], Annette Peters[10], Thomas Meitinger[7,23], Michael Roden[2,24], H-Erich Wichmann[11,25], Tobias Pischon[4,26], Jerzy Adamski[13,20] and Thomas Illig[1,27]

[1] Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany, [2] German Diabetes Center, Institute for Clinical Diabetology, Leibniz Center for Diabetes Research at Heinrich Heine University, Düsseldorf, Germany, [3] Institute of Structural Biology, Helmholtz Zentrum München, Neuherberg, Germany, [4] Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany, [5] Shanghai Center for Bioinformation Technology, Shanghai, China, [6] Key Lab of Systems Biology, Bioinformatics Center, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China, [7] Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany, [8] Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Neuherberg, Germany, [9] Else Kroener-Fresenius-Center for Nutritional Medicine, University Hospital 'Klinikum rechts der Isar', Technische Universität München, Munich, Germany, [10] Institute of Epidemiology II, Helmholtz Zentrum München, Neuherberg, Germany, [11] Institute of Epidemiology I, Helmholtz Zentrum München, Neuherberg, Germany, [12] Institute of Genetic Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany, [13] Genome Analysis Center, Institute of Experimental Genetics, Helmholtz Zentrum München, Neuherberg, Germany, [14] Department of Clinical Nutrition, Institute of Health Biosciences, University of Tokushima Graduate School, Tokushima, Japan, [15] Benxi Diabetes Clinic, Benxi Central Hospital, Benxi, China, [16] Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University Hospital Leipzig, Leipzig, Germany, [17] German Diabetes Center, Institute of Biometrics and Epidemiology, Leibniz Center for Diabetes Research at Heinrich Heine University, Düsseldorf, Germany, [18] Institute of Diabetes and Regeneration Research, Helmholtz Zentrum München, Neuherberg, Germany, [19] Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China, [20] Chair of Experimental Genetics, Technische Universität München, Munich, Germany, [21] Faculty of Biology, Ludwig-Maximilians-Universität, Planegg-Martinsried, Germany, [22] Department of Physiology and Biophysics, Weill Cornell Medical College in Qatar (WCMC-Q), Doha, Qatar, [23] Department of Metabolic Diseases, University Hospital Düsseldorf, Düsseldorf, Germany, [24] Klinikum rechts der Isar, Technische Universität München, Munich, Germany, [25] Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany, [26] Molecular Epidemiology Group, Max Delbrueck Center for Molecular Medicine (MDC), Berlin-Buch, Germany and [27] Hannover Unified Biobank, Hannover Medical School, Hannover, Germany
[28]These authors contributed equally to this work
* Corresponding author. Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, 85764 Munich-Neuherberg, Germany. Tel.: +49 89 3187 3978; Fax: +49 89 3187 2428; E-mail: rui.wang-sattler@helmholtz-muenchen.de

Received 13.6.12; accepted 15.8.12

Type 2 diabetes (T2D) can be prevented in pre-diabetic individuals with impaired glucose tolerance (IGT). Here, we have used a metabolomics approach to identify candidate biomarkers of pre-diabetes. We quantified 140 metabolites for 4297 fasting serum samples in the population-based Cooperative Health Research in the Region of Augsburg (KORA) cohort. Our study revealed significant metabolic variation in pre-diabetic individuals that are distinct from known diabetes risk indicators, such as glycosylated hemoglobin levels, fasting glucose and insulin. We identified three metabolites (glycine, lysophosphatidylcholine (LPC) (18:2) and acetylcarnitine) that had significantly altered levels in IGT individuals as compared to those with normal glucose tolerance, with *P*-values ranging from 2.4×10^{-4} to 2.1×10^{-13}. Lower levels of glycine and LPC were found to be predictors not only for IGT but also for T2D, and were independently confirmed in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam cohort. Using metabolite–protein network analysis, we identified seven T2D-related genes that are associated with these three IGT-specific metabolites by multiple interactions with four enzymes. The expression levels of these enzymes correlate with changes in the metabolite concentrations linked to diabetes. Our results may help developing novel strategies to prevent T2D.
Molecular Systems Biology **8:** 615; published online 25 September 2012; doi:10.1038/msb.2012.43
Subject Categories: metabolic and regulatory networks; molecular biology of disease
Keywords: early diagnostic biomarkers; IGT; metabolomics; prediction; T2D

Introduction

Type 2 diabetes (T2D) is defined by increased blood glucose levels due to pancreatic β-cell dysfunction and insulin resistance without evidence for specific causes, such as autoimmune destruction of pancreatic β-cells (Krebs *et al*, 2002; Stumvoll *et al*, 2005; Muoio and Newgard, 2008). A state

of pre-diabetes (i.e., impaired fasting glucose (IFG) and/or impaired glucose tolerance (IGT)) with only slightly elevated blood glucose levels may precede T2D for years (McGarry, 2002; Tabak *et al*, 2012). The development of diabetes in pre-diabetic individuals can be prevented or delayed by dietary changes and increased physical activity (Tuomilehto *et al*, 2001; Knowler *et al*, 2002). However, no specific biomarkers that enable 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 (Shaham *et al*, 2008; Newgard *et al*, 2009; Zhao *et al*, 2010; Pietilainen *et al*, 2011; Rhee *et al*, 2011; Wang *et al*, 2011; Cheng *et al*, 2012; Goek *et al*, 2012). Altered metabolite levels may serve as diagnostic biomarkers and enable preventive action. Previous cross-sectional metabolomics studies of T2D were either based on small sample sizes (Shaham *et al*, 2008; Wopereis *et al*, 2009; Zhao *et al*, 2010; Pietilainen *et al*, 2011) or did not consider the influence of common risk factors of T2D (Newgard *et al*, 2009). Recently, based on prospective nested case–control studies with relative 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, using various comprehensive large-scale approaches, we measured metabolite concentration profiles (Yu *et al*, 2012) in the population-based (Holle *et al*, 2005; Wichmann *et al*, 2005) Cooperative Health Research in the Region of Augsburg (KORA) baseline (survey 4 (S4)) and follow-up (F4) studies (Rathmann *et al*, 2009; Meisinger *et al*, 2010; Jourdan *et al*, 2012). The results of these cross-sectional and prospective studies allowed us to (i) reliably identify candidate biomarkers of pre-diabetes and (ii) build metabolite–protein networks to understand diabetes-related metabolic pathways.

Results

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 1A). 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) (WHO, 1999; Rathmann *et al*, 2009; Meisinger *et al*, 2010; Supplementary Table S1). 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 1A; study characteristics are shown in Table I). Of the 1010 individuals in a fasting state who participated at baseline and follow-up surveys (Figure 1B, study characteristics of the KORA F4 survey are shown in Supplementary Table S2), 876 of them were non-diabetic at baseline. Out of these, about 10% developed T2D (i.e., 91 incident T2D) (Figure 1C). From the 641 individuals with NGT at baseline, 18% developed IGT (i.e., 118 incident IGT) 7 years later (Figure 1D). The study characteristics of the prospective KORA S4→F4 are shown in Table II.

Analyses strategies

We first screened for significantly differed metabolites concentration 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. Finally, the relevance of the identified metabolites was further investigated with our bioinformatical analysis of protein-metabolite interaction networks and gene expression data.

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 2A; odds ratios (ORs) and P-values are shown in Table III). 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; Supplementary Table S3), 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 (Supplementary Table S4A).

To investigate whether HbA_{1c}, 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 2B). 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 HbA_{1c}, fasting glucose and fasting insulin levels (r values are shown in Supplementary Table S3). Only nine metabolite concentrations significantly differed between IGT and NGT individuals (Table III; Supplementary Table S4B). These metabolites therefore

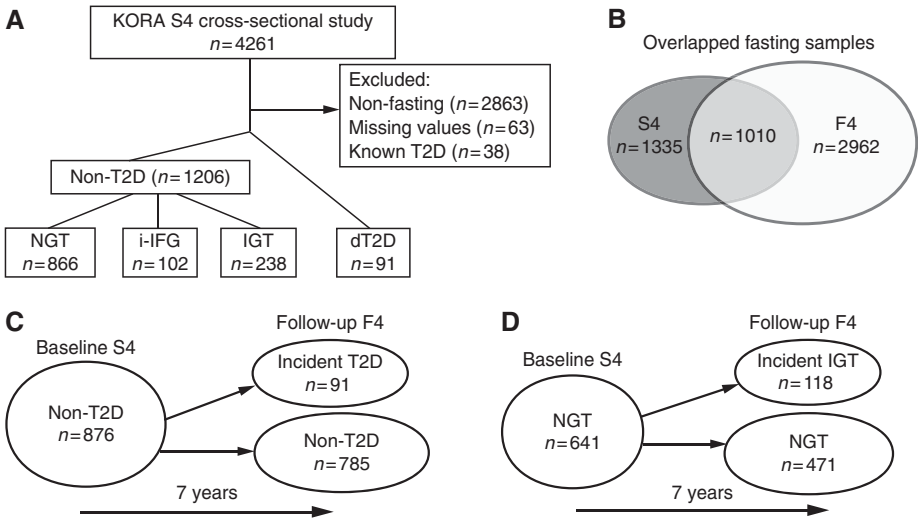

Figure 1 Population description. Metabolomics screens in the KORA cohort, at baseline S4 (**A**), overlapped between S4 and F4 (**B**) and prospective (**C**, **D**). Participant numbers are shown. Normal glucose tolerance (NGT), isolated impaired fasting glucose (i-IFG), impaired glucose tolerance (IGT), type 2 diabetes mellitus (T2D) and newly diagnosed T2D (dT2D). Non-T2D individuals include NGT, i-IFG and IGT participants.

Table I Characteristics of the KORA S4 cross-sectional study sample

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 (% >1 h per week)	46.7	35.3	39.9	36.3
Alcohol intake[a] (%)	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

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 ± s.d. are given for each variable and each group (NGT, i-IFG, IGT and dT2D).
[a]⩾20 g/day for women; ⩾40 g/day for men.

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), $P = 8.6 \times 10^{-5}$), LPC (18:2) (OR = 0.58 (0.46–0.72), $P = 2.1 \times 10^{-6}$) and acetylcarnitine C2 (OR = 1.38 (1.16–1.64), $P = 2.4 \times 10^{-4}$) (Figure 2C). Similar results were observed in the follow-up KORA F4 study (Supplementary Figure S1). 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), $P = 9.3 \times 10^{-8}$; LPC (18:2), OR = 0.47 (0.38–0.57), $P = 2.1 \times 10^{-13}$; and C2, OR = 1.33 (1.17–1.49), $P = 4.9 \times 10^{-6}$) (Supplementary Table S5).

Predict 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

Table II Characteristics of the KORA S4→F4 prospective study samples

	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 (% >1 h per week)	52.9	43.2	52.2	58.2
Alcohol intake[a] (%)	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

BP, blood pressure; HDL, high-density lipoprotein; LDL, low-density lipoprotein.
Percentages of individuals or means ± s.d. are given for each variable and each group.
[a]⩾20 g/day for women; ⩾40 g/day for men.

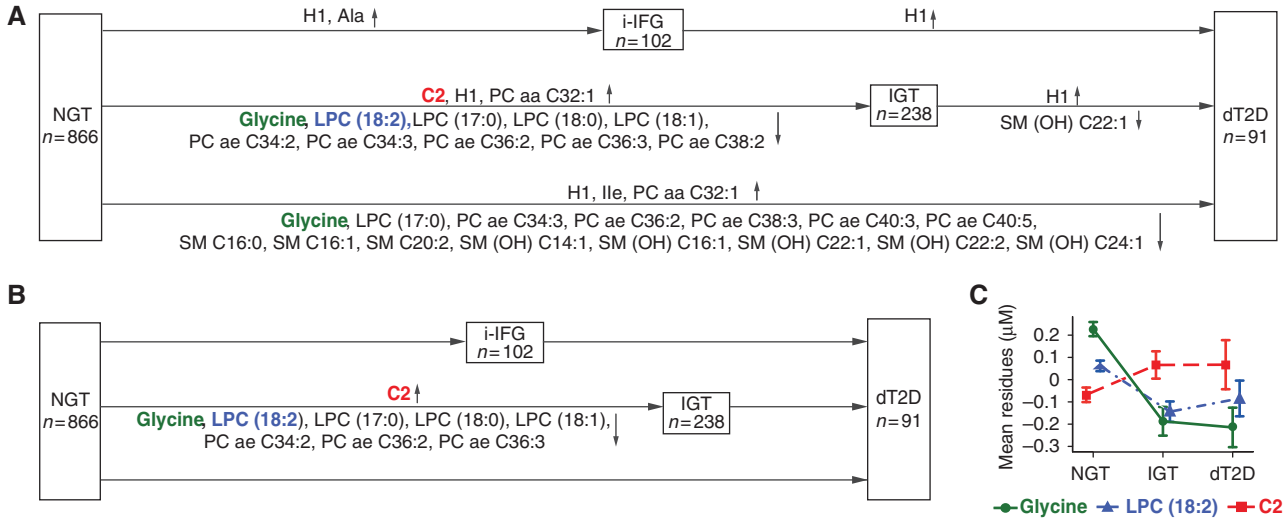

Figure 2 Differences in metabolite concentrations from cross-sectional analysis of KORA S4. Plots (**A**, **B**) show the names of metabolites with significantly different concentrations in multivariate logistic regression analyses (after the Bonferroni correction for multiple testing with $P<3.6\times10^{-4}$) 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 HbA_{1c}, fasting glucose and fasting insulin (model 2). Residuals were calculated from linear regression model (formula: T2D status ~ metabolite concentration + model 2). For further information, see Supplementary Table S4.

using the prospective KORA S4→F4 cohort (Table II). 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 IV; Supplementary Table S6). 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 (s.d.) 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), $P=1.8\times10^{-5}$), compared with 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 (Supplementary Table S6). When the three metabolites were added to the fully

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

Metabolite	Model 1		Model 2	
	OR (95% CI), per s.d.	*P*-value	OR (95% CI), per s.d.	*P*-value
238 IGT versus 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 versus 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 versus 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 versus 866 NGT				
Glycine	0.75 (0.57–0.98)	3.9E-02	0.62[a]	1.0E + 00
LPC (18:2)	0.99 (0.77–1.26)	9.6E-01	0.79[a]	1.0E + 00
C2	1.2 (0.99–1.46)	5.9E-02	0.18[a]	1.0E + 00
91 dT2D versus 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

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

Table IV Prediction of IGT and T2D in the KORA cohort

Model	Glycine	LPC (18:2)	C2	Glycine, LPC (18:2), C2
(A) Metabolite as continuous variable (n = 589)				
Per s.d.	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)				
First quartile	1.0 (reference)	1.0 (reference)	1.0 (reference)	1.0 (reference)
Second 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)
Third 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)
Fourth 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 s.d.	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[a] (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

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.
[a]β Estimate indicates the future difference in the glucose tolerance corresponding to the one s.d. differences in the normalized baseline metabolite concentration.

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 (Supplementary Figure S2; Supplementary Table S7). Thus, this provides an improved prediction of IGT and T2D as compared with T2D risk indicators.

Baseline metabolite concentrations correlate with future glucose tolerance

We next investigated the associations between baseline metabolite concentrations and 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 (Supplementary Table S8). As expected, a very significant association ($P = 1.5 \times 10^{-22}$) was observed for hexose H1 in model 1, while no significance ($P = 0.12$) was observed for it in the fully adjusted model 2 (Supplementary Table S8).

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 Supplementary Table S9). Furthermore, none of these five amino acids showed associations with 2-h glucose values in the cross-sectional KORA S4 study (Supplementary Table S8).

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 Supplementary Table S10; ORs and *P*-values are given in Supplementary Table S11). 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* (2011). In fact, the 91 matched controls include about 50% pre-diabetes individuals, which is significantly higher than the general population (about 15%).

Replication in the cross-sectional EPIC-Potsdam cohort

Metabolomics data from serum samples of a randomly drawn EPIC-Potsdam subcohort ($n = 2500$) were used for replication. Glycine (OR = 0.60 (0.47–0.77), $P = 7.4 \times 10^{-5}$) and LPC (18:2) (OR = 0.79 (0.63–0.98), $P = 0.037$) were replicated when 133 T2D patients were compared with 1253 individuals with NGT at baseline (Supplementary Table S12). However, acetylcarnitine C2 (OR = 0.98 (0.81–1.19), $P = 0.858$) could not be replicated when T2D patients were compared with NGT individuals, since the IGT participants were not available in the data set. The absolute levels of these three metabolites were in a similar range, with only slight differences that were due probably to the differences of the two cohorts or to potential batch effects of metabolomics measurements (Supplementary Tables S12 and S15). Thus, these data therefore provide an independent validation of the metabolomics study.

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 (Wishart *et al*, 2009; Szklarczyk *et al*, 2011). In all, 7 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 3A;

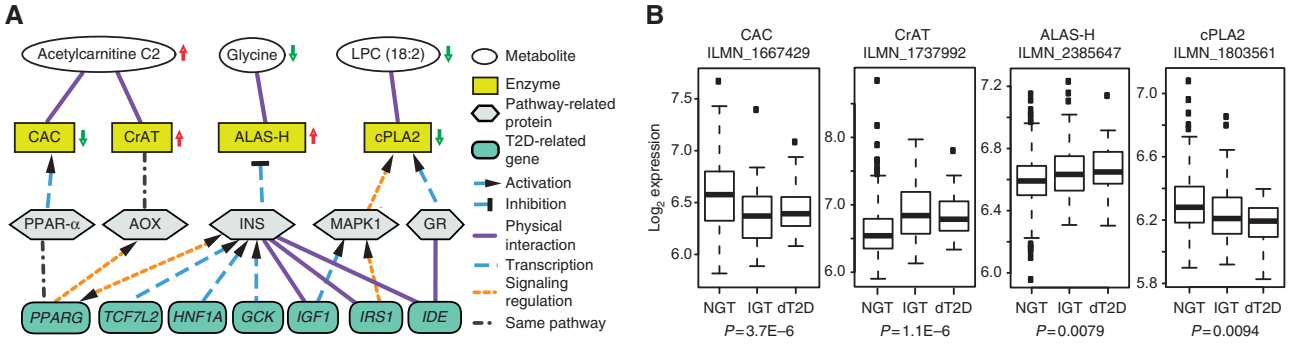

Figure 3 Three candidate metabolites for IGT associated with seven T2D-related genes. (**A**) Metabolites (white), enzymes (yellow), pathway-related proteins (gray) and T2D-related genes (blue) are represented with ellipses, rectangles, polygons and rounded rectangles, respectively. Arrows next to the ellipses and rectangles indicate altered metabolite concentrations in persons with IGT as compared with 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 (gray dot and dash line). The activation or inhibition is indicated. For further information, see Supplementary Table S12. (**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.

the list of 46 genes is shown in Supplementary Table S13). 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 (Supplementary Table S14).

Gene expression analysis in whole-blood samples of participants from the KORA S4 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 3B). 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.

Discussion

Using a cross-sectional approach (KORA S4, F4), we analyzed 140 metabolites and identified three (glycine, LPC (18:2) and C2) which 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, HbA_{1c}, 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.

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 7 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 nested/selected case–control samples (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 (Supplementary Figure S1D). 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.

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.

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 (McGarry, 2002; Stumvoll *et al*, 2005; Faerch *et al*, 2009; Tabak *et al*, 2009), 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.

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 individuals with IGT and T2D, most probably due to an activation of the peroxisome proliferator activated receptor alpha (PPAR-α) 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 β-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 (Morino *et al*, 2005; Szendroedi *et al*, 2007). Together, these findings reflect an important role of increased cellular lipid metabolites and impaired mitochondrial β-oxidation in the development of insulin resistance (McGarry, 2002; Szendroedi *et al*, 2007; Koves *et al*, 2008).

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.

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.

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.

Materials and methods

Ethics statement

Written informed consent was obtained from each KORA and EPIC-Potsdam participant. The KORA and EPIC-Potsdam studies were approved by the ethics committee of the Bavarian Medical Association and the Medical Society of the State of Brandenburg, respectively.

Sample source and classification

The KORA surveys are population-based studies conducted in the city of Augsburg and the surrounding towns and villages (Holle *et al*, 2005; Wichmann *et al*, 2005). KORA is a research platform in the field of epidemiology, health economics and health-care research. Four surveys were conducted with 18 079 participants recruited from 1984 to 2001. The S4 consists of 4261 individuals (aged 25–74 years) examined from 1999 to 2001. From 2006 to 2008, 3080 participants (with an age range of 32–81) took part in an F4 survey. Ascertainments of anthropometric measurements and personal interviews, as well as laboratory measurements of persons, from the KORA S4/F4 have been described elsewhere (Rathmann *et al*, 2009; Meisinger *et al*, 2010; Jourdan *et al*, 2012).

Sampling

In the KORA cohort, blood was drawn into S-Monovette® serum tubes (SARSTEDT AG & Co., Nümbrecht, Germany) in the morning between 0800 and 1030 h after at least 8 h of fasting. Tubes were gently inverted twice, followed by 30 min resting at room temperature, to obtain complete coagulation. For serum collection, blood was centrifuged at 2750 *g* at 15°C for 10 min. Serum was filled into synthetic straws, which were stored in liquid nitrogen until the metabolic analyses were conducted.

Metabolite measurements and exclusion of metabolites

For the KORA S4 survey, the targeted metabolomics approach was based on measurements with the Absolute*IDQ*™ p180 kit (BIOCRATES Life Sciences AG, Innsbruck, Austria). This method allows simultaneous quantification of 188 metabolites using liquid chromatography and flow injection analysis–mass spectrometry. The assay procedures have been described previously in detail (Illig *et al*, 2010; Römisch-Margl *et al*, 2011). 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 two criteria: (1) the coefficient of variance (CV) for the metabolite in the total 110 reference samples had to be smaller than 25%. In total, seven outliers were removed because their concentrations were larger than the mean plus 5 × s.d.; (2) 50% of all measured sample concentrations for the metabolite should be above the limit of detection (LOD), which is defined as 3 × median of the three zero samples. In total, 140 metabolites passed the quality controls (Supplementary Table S15): one hexose (H1), 21 acylcarnitines, 21 amino acids, 8 biogenic amines, 13 sphingomyelins (SMs), 33 diacyl (aa) phosphatidylcholines (PCs), 35 acyl-alkyl (ae) PCs and 8 lysoPCs. Concentrations of all analyzed metabolites are reported in μM.

Measurements of the 3080 KORA F4 samples and the involved cleaning procedure have already been described in detail (Mittelstrass *et al*, 2011; Yu *et al*, 2012).

Gene expression analysis

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.

Data availability

Metabolite concentrations of Glycine, LPC (18:2) and C2 with T2D status in the KORA S4 and F4 are provided (Supplementary Table S16). Additional data from the KORA S4 and F4 studies, including the metabolite concentrations and the gene expression with clinical phenotypes used in this study, are available upon request from KORA-gen (http://epi.helmholtz-muenchen.de/kora-gen). Requests should be sent to kora-gen@helmholtz-muenchen.de and are subject to approval by the KORA board to ensure that appropriate conditions are met to preserve patient privacy. Formal collaboration and co-authorship with members of the KORA study is not an automatic condition to obtain access to the data published in the present paper. More general information about KORA, including S4 and F4 study design and clinical variables, can be found at http://epi.helmholtz-muenchen.de/kora-gen/seiten/variablen_e.php and http:/helmholtz-muenchen.de/en/kora-en/information-for-scientists/current-kora-studies.

Statistical analysis

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

Multivariate logistic regression and linear regression

In multivariate logistic regression analysis, 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 rates from multiple comparisons, the cutoff point for significance was calculated according to the Bonferroni correction, at a level of 3.6×10^{-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 lysoPCs, 33 diacyl PCs, 35 acyl-alkyl PCs and 13 SMs), this correction was 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.

For linear regression analyses, β estimates were calculated from the concentration of each metabolite and the 2-h glucose value. The concentration of each metabolite was log-transformed and normalized to have a mean of zero and an s.d. of one. Various risk factors in the logistic regression were added as covariates, and the same significance level (3.6×10^{-4}) was adopted.

Combination of metabolites

To obtain the combined scores of 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.

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.

Random forest, stepwise selection methods and candidate biomarker selection

To select candidate biomarkers, we applied two additional methods: the random forest selection (Breiman, 2001) and the stepwise selection, which assess the metabolites as a group.

Between two groups, the supervised classification method of random forest was first used to select the metabolites among the 30 highest ranking variables of importance score, allowing the best separation of the individuals from different groups. T2D risk indicators were also included in this method with all the metabolites.

We further selected the metabolites using stepwise selection on the logistic regression model. Metabolites with significantly different concentrations between the compared groups in logistic regression, and which were also selected using random forest, were used in this model along with all the risk indicators. 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. The AUC was used to evaluate the models.

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 (Kastenmuller *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 1 intermediate protein (other proteins) through STRING protein functional interaction and optimized by eliminating edges with a STRING score of <0.7 and undirected paths. The subnetworks were connected by the shortest path from metabolites to T2D-related genes.

Replication

The EPIC-Postdam is part of the multicenter EPIC study (Boeing *et al*, 1999; Riboli *et al*, 2002). It was drawn from the general adult population in Potsdam and surrounding areas and consists of 27 548 participants recruited from 1994 to 1998 (Boeing *et al*, 1999). At baseline, participants underwent anthropometric and BP measurements, completed an interview on prevalent diseases, a questionnaire on socioeconomic and lifestyle factors and submitted a validated food frequency questionnaire. Follow-up questionnaires were administered every 2–3 years (Bergmann *et al*, 1999).

From the EPIC-Potsdam population, a substudy of 2500 participants was randomly selected from all participants who had provided blood samples at baseline ($n = 26\,444$). The substudy had a limited number of fasting samples available. Therefore, non-fasting samples were also considered. Out of the substudy, 814 participants were excluded because of missing information on relevant covariates or missing fasting samples. Individuals with NGT and T2D were determined according to HbA_{1c} categories defined by the American Diabetes Association in 2010 (ADA, 2010).

In the EPIC-Potsdam study, 30 ml of blood was drawn by qualified medical staff during the baseline examination, immediately fractionated into serum, plasma, buffy coat and erythrocytes and aliquoted into straws. The blood samples were stored in liquid nitrogen (at -196°C) until the metabolic analyses.

Metabolite measurements for the EPIC-Potsdam samples were performed using the same kit and the same method as for the KORA F4 samples (Floegel *et al*, 2011).

Calculations were performed using the Statistical Analysis System (SAS), Version 9.2 (SAS Institute, Inc., Cary, NC, USA).

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We express our appreciation to all KORA and EPIC-Potsdam study participants for donating their blood and time. We thank the field staff in Augsburg who conducted the KORA studies. The KORA group consisted of HE Wichmann (speaker), A Peters, C Meisinger, T Illig, R Holle and J John, as well as their co-workers, and they were responsible for the design and conduction of the studies. We thank all the staff of the Institute of Epidemiology, Helmholtz Zentrum München, and the Genome Analysis Center, as well as the Metabolomic Platform, who helped in the sample logistics, the metabolite profiling assays and the genetic expression analyses, especially A Sabunchi, H Chavez, B Hochstrat, F Scharl, N Lindemann and J Scarpa. We thank M Sattler, W Mewes, VA Raker and J Mendes for comments and suggestions. This study was supported in part by a grant from the German Federal Ministry of Education and Research (BMBF) to the German Center for Diabetes Research (DZD e.V). In addition, this work was partly supported by the BMBF project 'Metabolomics of ageing' (FKZ: 01DO12030) and Project 'SysMBo: Systems Biology of Metabotypes' (FKZ: 0315494A). Further support for this study was obtained from the Federal Ministry of Health (Berlin, Germany), the Ministry of Innovation, Science, Research and Technology of the state North-Rhine Westphalia (Düsseldorf, Germany) and the Federal Ministry of Education, Science, Research and Technology (NGFN-Plus AtheroGenomics/01GS0423; Berlin, Germany). The KORA research platform and the KORA Augsburg studies are financed by the Helmholtz Zentrum München, German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education, Science, Research and Technology and by the State of Bavaria. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions: RWS, ZY, CHe, KS, HP, AP, TM, HEW, TP, JA and TI designed the research. RWS, CHe, CP, WRM, MC, KH and HP performed the experiments. RWS, ZY, CHe, ACM, AF, YH, KH, MC, CHo, BT, HG, TX, EB, AD, KM, HYO, YL, LX, KS, AP, HP, TM, MR, HEW, TP, JA and TI analyzed the data. RWS, ZY, CHe, ACM, AF, YH, CHo, HP, TM, AP, MR, TP and JA wrote the paper.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

ADA (2010) Executive summary: standards of medical care in diabetes–2010. *Diabetes Care* **33**(Suppl 1)**:** S4–S10

Adams SH, Hoppel CL, Lok KH, Zhao L, Wong SW, Minkler PE, Hwang DH, Newman JW, Garvey WT (2009) Plasma acylcarnitine profiles suggest incomplete long-chain fatty acid beta-oxidation and altered tricarboxylic acid cycle activity in type 2 diabetic African-American women. *J Nutr* **139:** 1073–1081

Altmaier E, Kastenmuller G, Romisch-Margl W, Thorand B, Weinberger KM, Illig T, Adamski J, Doring A, Suhre K (2011) Questionnaire-based self-reported nutrition habits associate with serum metabolism as revealed by quantitative targeted metabolomics. *Eur J Epidemiol* **26:** 145–156

Bergmann MM, Bussas U, Boeing H (1999) Follow-up procedures in EPIC-Germany–data quality aspects. European Prospective Investigation into Cancer and Nutrition. *Ann Nutr Metab* **43:** 225–234

Bishop DF (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

Boeing H, Wahrendorf J, Becker N (1999) EPIC-Germany–A source for studies into diet and risk of chronic diseases. European Investigation into Cancer and Nutrition. *Ann Nutr Metab* **43:** 195–204

Bolstad BM, Irizarry RA, Astrand M, Speed TP (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

Cheng S, Rhee EP, Larson MG, Lewis GD, McCabe EL, Shen D, Palma MJ, Roberts LD, Dejam A, Souza AL, Deik AA, Magnusson M, Fox CS, O'Donnell CJ, Vasan RS, Melander O, Clish CB, Gerszten RE, Wang TJ (2012) Metabolite profiling identifies pathways associated with metabolic risk in humans. *Circulation* **125:** 2222–2231

Faerch K, Vaag A, Holst JJ, Hansen T, Jorgensen 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. *Diabetes Care* **32:** 439–444

Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, Joost HG, Boeing H, Pischon T (2011) Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. *PLoS ONE* **6:** e21103

Goek ON, Doring A, Gieger C, Heier M, Koenig W, Prehn C, Romisch-Margl W, Wang-Sattler R, Illig T, Suhre K, Sekula P, Zhai G, Adamski J, Kottgen A, Meisinger C (2012) Serum metabolite concentrations and decreased GFR in the general population. *Am J Kidney Dis* **60:** 197–206

Holle R, Happich M, Lowel H, Wichmann HE (2005) KORA–a research platform for population based health research. *Gesundheitswesen* **67:** S19–S25

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 BS, Mewes HW, Meitinger T, de Angelis MH, Kronenberg F, Soranzo N, Wichmann HE, Spector TD, Adamski J, Suhre K (2010) A genome-wide perspective of genetic variation in human metabolism. *Nat Genet* **42:** 137–141

Jourdan C, Petersen AK, Gieger C, Doring A, Illig T, Wang-Sattler R, Meisinger C, Peters A, Adamski J, Prehn C, Suhre K, Altmaier E, Kastenmuller G, Romisch-Margl W, Theis FJ, Krumsiek J, Wichmann HE, Linseisen J (2012) Body fat free mass is associated with the serum metabolite profile in a population-based study. *PLoS ONE* **7:** e40009

Kastenmuller G, Romisch-Margl W, Wagele B, Altmaier E, Suhre K (2011) metaP-server: a web-based metabolomics data analysis tool. *J Biomed Biotechnol* **2011:** 1–7, pii: 839862

Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* **346:** 393–403

Koves TR, Ussher JR, Noland RC, Slentz D, Mosedale M, Ilkayeva O, Bain J, Stevens R, Dyck JR, Newgard CB, Lopaschuk GD, Muoio DM (2008) Mitochondrial overload and incomplete fatty acid oxidation contribute to skeletal muscle insulin resistance. *Cell Metab* **7:** 45–56

Krebs M, Krssak M, Bernroider E, Anderwald C, Brehm A, Meyerspeer M, Nowotny P, Roth E, Waldhausl W, Roden M (2002) Mechanism of amino acid-induced skeletal muscle insulin resistance in humans. *Diabetes* **51:** 599–605

Malipa AC, Meintjes RA, Haag M (2008) Arachidonic acid and glucose uptake by freshly isolated human adipocytes. *Cell Biochem Funct* **26:** 221–227

McGarry JD (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 SE, 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. *Diabet Med* **27:** 360–362

Mittelstrass K, Ried JS, Yu Z, 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 (2011) Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet* **7:** e1002215

Morino K, Petersen KF, Dufour S, Befroy D, Frattini J, Shatzkes N, Neschen S, White MF, Bilz S, Sono S, Pypaert M, Shulman GI (2005) Reduced mitochondrial density and increased IRS-1 serine phosphorylation in muscle of insulin-resistant offspring of type 2 diabetic parents. *J Clin Invest* **115:** 3587–3593

Muoio DM, Newgard CB (2008) Mechanisms of disease: molecular and metabolic mechanisms of insulin resistance and beta-cell failure in type 2 diabetes. *Nat Rev Mol Cell Biol* **9:** 193–205

Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, Haqq AM, Shah SH, Arlotto M, Slentz CA, Rochon J, Gallup D, Ilkayeva O, Wenner BR, Yancy Jr WS, Eisenson H, Musante G, Surwit RS, Millington DS, Butler MD *et al* (2009) A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab* **9:** 311–326

Noland RC, Koves TR, Seiler SE, Lum H, Lust RM, Ilkayeva O, Stevens RD, Hegardt FG, Muoio DM (2009) Carnitine insufficiency caused by aging and overnutrition compromises mitochondrial performance and metabolic control. *J Biol Chem* **284:** 22840–22852

Phillips JD, Kushner JP (2005) Fast track to the porphyrias. *Nat Med* **11:** 1049–1050

Pietilainen KH, Rog T, Seppanen-Laakso T, Virtue S, Gopalacharyulu P, Tang J, Rodriguez-Cuenca S, Maciejewski A, Naukkarinen J, Ruskeepaa AL, Niemela PS, Yetukuri L, Tan CY, Velagapudi V, Castillo S, Nygren H, Hyotylainen T, Rissanen A, Kaprio J, Yki-Jarvinen H *et al* (2011) Association of lipidome remodeling in the adipocyte membrane with acquired obesity in humans. *PLoS Biol* **9:** e1000623

Pontiroli AE, 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 Diabetol* **41:** 70–76

Primrose S, Draper J, Elsom R, Kirkpatrick V, Mathers JC, Seal C, Beckmann M, Haldar S, Beattie JH, Lodge JK, Jenab M, Keun H, Scalbert A (2011) Metabolomics and human nutrition. *Br J Nutr* **105:** 1277–1283

Rathmann W, Kowall B, Heier M, Herder C, Holle R, Thorand B, Strassburger K, Peters A, Wichmann HE, Giani G, Meisinger C (2010) Prediction models for incident type 2 diabetes mellitusin the older population: KORA S4/F4 cohort study. *Diabet Med* **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

Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, Yang E, Farrell L, Fox CS, O'Donnell CJ, Carr SA, Vasan RS, Florez JC, Clish CB, Wang TJ, Gerszten RE (2011) Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J Clin Invest* **121:** 1402–1411

Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, Charrondiere UR, Hemon B, Casagrande C, Vignat J, Overvad K, Tjonneland A, Clavel-Chapelon F, Thiebaut A, Wahrendorf J, Boeing H, Trichopoulos D, Trichopoulou A, Vineis P, Palli D *et al* (2002) European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* **5:** 1113–1124

Römisch-Margl W, Prehn C, Bogumil R, Roehring C, Suhre KJA (2012) Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics* **8:** 133–142

Shaham O, Wei R, Wang TJ, Ricciardi C, Lewis GD, Vasan RS, Carr SA, Thadhani R, Gerszten RE, Mootha VK (2008) Metabolic profiling of the human response to a glucose challenge reveals distinct axes of insulin sensitivity. *Mol Syst Biol* **4:** 214

Stumvoll M, Goldstein BJ, van Haeften TW (2005) Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* **365:** 1333–1346

Szendroedi J, Schmid AI, 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 LJ, von Mering C (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* **39:** D561–D568

Tabak AG, Herder C, Rathmann W, Brunner EJ, Kivimaki M (2012) Prediabetes: a high-risk state for diabetes development. *Lancet* **379:** 2279–2290

Tabak AG, Jokela M, Akbaraly TN, Brunner EJ, Kivimaki M, Witte DR (2009) Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. *Lancet* **373:** 2215–2221

Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, Ilanne-Parikka P, Keinanen-Kiukaanniemi S, Laakso M, Louheranta A, Rastas M, Salminen V, Uusitupa M (2001) Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med* **344:** 1343–1350

Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, O'Donnell CJ, Carr SA, Mootha VK, Florez JC, Souza A, Melander O, Clish CB, Gerszten RE (2011) Metabolite profiles and the risk of developing diabetes. *Nat Med* **17:** 448–453

WHO (1999) Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Report of a WHO Consultation, Geneva, pp 59

Wichmann HE, Gieger C, Illig T (2005) KORA-gen–resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67**(Suppl 1)**:** S26–S30

Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, Mandal R, Sinelnikov I, Xia J, Jia L, Cruz JA, Lim E, Sobsey CA, Shrivastava S, Huang P, Liu P *et al* (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res* **37:** D603–D610

Wopereis S, Rubingh CM, van Erk MJ, Verheij ER, van Vliet T, Cnubben NH, Smilde AK, van der Greef J, van Ommen B, Hendriks HF (2009) Metabolic profiling of the response to an oral glucose tolerance test detects subtle metabolic changes. *PLoS ONE* **4:** e4525

Yu Z, Zhai G, Singmann P, He Y, Xu T, Prehn C, Romisch-Margl W, Lattka E, Gieger C, Soranzo N, Heinrich J, Standl M, Thiering E, Mittelstrass K, Wichmann HE, Peters A, Suhre K, Li Y, Adamski J, Spector TD *et al* (2012) Human serum metabolic profiles are age dependent. *Aging Cell* (e-pub ahead of print 26 July 2012; doi:10.1111/j.1474-9726.2012.00865.x)

Zhao X, Fritsche J, Wang J, Chen J, Rittig K, Schmitt-Kopplin P, Fritsche A, Haring HU, Schleicher ED, Xu G, Lehmann R (2010) Metabonomic fingerprints of fasting plasma and spot urine reveal human pre-diabetic metabolic traits. *Metabolomics* **6:** 362–374