



Changes in metabolomics profiles over ten years and subsequent risk of developing type 2 diabetes: Results from the Nurses' Health Study

Clemens Wittenbecher,^{a,b,c,*} Marta Guasch-Ferré,^{a,d} Danielle E. Haslam,^{a,d} Courtney Dennis,^e Jun Li,^a Shilpa N. Bhupathiraju,^{a,d} Chih-Hao Lee,^{a,f} Qibin Qi,^{a,g} Liming Liang,^{h,i} A. Heather Eliassen,^{a,d,h} Clary Clish,^e Qi Sun,^{a,d,h} and Frank B Hu^{a,d,h,*}

^aDepartment of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

^bDepartment of Molecular Epidemiology, German Institute of Human Nutrition Potsdam Rehbruecke, Nuthetal, Germany

^cGerman Center for Diabetes Research (DZD), Neuherberg, Germany

^dChanning Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, MA, USA

^eBroad Institute of MIT and Harvard, Cambridge, MA, USA

^fDepartment of Molecular Metabolism, Harvard T.H. Chan School of Public Health, Boston, MA, USA

^gDepartment of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA

^hDepartment of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

ⁱDepartment of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Summary

Background Metabolomics profiles were consistently associated with type 2 diabetes (T2D) risk, but evidence on long-term metabolite changes and T2D incidence is lacking. We examined the associations of 10-year plasma metabolite changes with subsequent T2D risk.

Methods We conducted a nested T2D case-control study (n=244 cases, n=244 matched controls) within the Nurses' Health Study. Repeated metabolomics profiling (170 targeted metabolites) was conducted in participant blood specimens from 1989/1990 and 2000/2001, and T2D occurred between 2002 and 2008. We related 10-year metabolite changes (Δ -values) to subsequent T2D risk using conditional logistic models, adjusting for baseline metabolite levels and baseline levels and concurrent changes of BMI, diet quality, physical activity, and smoking status.

Findings The 10-year changes of thirty-one metabolites were associated with subsequent T2D risk (false discovery rate-adjusted p-values [FDR]<0.05). The top three high T2D risk-associated 10-year changes were (odds ratio [OR] per standard deviation [SD], 95%CI): Δ isoleucine (2.72, 1.97-3.79), Δ leucine (2.53, 1.86-3.47), and Δ valine (1.93, 1.52-2.44); other high-risk-associated metabolite changes included alanine, tri-/diacylglycerol-fragments, short-chain acyl-carnitines, phosphatidylethanolamines, some vitamins, and bile acids (ORs per SD between 1.31 and 1.82). The top three low T2D risk-associated 10-year metabolite changes were (OR per SD, 95% CI): Δ N-acetylaspartic acid (0.54, 0.42-0.70), Δ C20:0 lysophosphatidylethanolamine (0.68, 0.56-0.82), and Δ C16:1 sphingomyelin (0.68, 0.56-0.83); 10-year changes of other sphingomyelins, plasmalogens, glutamine, and glycine were also associated with lower subsequent T2D risk (ORs per SD between 0.66 and 0.78).

Interpretation Repeated metabolomics profiles reflecting the long-term deterioration of amino acid and lipid metabolism are associated with subsequent risk of T2D.

Copyright © 2021 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: Metabolomics; Type 2 diabetes; Lipids; Amino acids; Repeated measurements; Change analysis; Prospective cohort study

Introduction

Type 2 diabetes (T2D) is a major public health concern, and effective T2D prevention strategies are of the utmost importance.¹⁻⁴ Novel biomarkers, specifically those which detect metabolic perturbations long before disease manifestation, may open new windows for

*Corresponding authors at: Department of Nutrition, Harvard T.H. Chan School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA.

E-mail addresses: cwittenbecher@hsph.harvard.edu (C. Wittenbecher), flu@hsph.harvard.edu (F.B. Hu).

Research in context

Evidence before this study

Based on the evidence from a previous meta-analysis published by our group and literature review of studies published after that (searching PubMed and EMBASE for diabetes [and related terms] AND metabolomics [and related terms]), several classes of plasma metabolites have been consistently associated with incident T2D. Baseline levels of several metabolites available in our study were previously associated with type 2 diabetes risk. For example, the three branched-chain amino acids, alanine, tyrosine, glutamate, phenylalanine, methionine, and lysine were associated with higher type 2 diabetes risk, and glycine and glutamine with lower type 2 diabetes risk. Short-chain acylcarnitines (C4OH, C5, C5DC) were also associated with higher type 2 diabetes risk. Several phosphatidylethanolamines, triglycerides, and diglycerides were associated with higher risk and several lysophospholipids with lower type 2 diabetes incidence. The PREDIMED trial did not find robust associations of one-year metabolite changes with subsequent T2D risk. We did not identify previous publications reporting the associations of metabolite changes over several years with subsequent type 2 diabetes.

Added value of this study

The 10-year changes of thirty-one metabolites were significantly (FDR <0.05, Wald test) associated with subsequent type 2 diabetes risk, after adjusting for baseline metabolite levels, as well as baseline levels and concurrent changes of anthropometric and lifestyle risk factors for type 2 diabetes. Greater 10-year increases of branched-chain amino acids, diglyceride-/triglyceride-fragments, phosphatidylethanolamines, some vitamins, and bile acids were associated with higher type 2 diabetes risk. The 10-year changes of lysophospholipids, sphingomyelins, and plasmalogens were associated with lower type 2 diabetes risk. These results demonstrate for the first time that intraindividual changes of specific lipids and amino acids over a decade are independent molecular markers of subsequent type 2 diabetes risk.

Implications of all the available evidence

Complementary to the extensive evidence on baseline metabolomics profiles and type 2 diabetes risk, our results indicate that repeated high-throughput metabolic profiling captures changes in plasma amino acid and lipid profiles associated with the subsequent risk of developing type 2 diabetes. The dynamic nature of some metabolite-type 2 diabetes risk associations warrants further research into potential biomarker applications, particularly for monitoring disease risk over time and research into the molecular targets of type 2 diabetes prevention.

targeted disease prevention and personalised interventions.

In the past two decades, metabolomics techniques identified numerous plasma metabolites that were robustly associated with T2D risk, including branched-chain amino acids (BCAA), alanine, aromatic amino acids, phospholipids, ceramides, acylcarnitines, and triglycerides. Other metabolites, including glycine, lyso- and ether-phospholipids, and sphingomyelins, have been associated with lower T2D risk.⁵ Relevant data from genetic and experimental studies support potential causal roles for some of these metabolites in the pathophysiology of T2D.^{6–9}

To our knowledge, no study has examined long-term changes in metabolites and subsequent risk of T2D. In the PREDIMED-trial, the association of one-year changes of lipid and acylcarnitine profiles with subsequent T2D risk were mainly in the expected directions but not statistically significant.^{10,11} Besides the limited sample size, the one-year interval between measurements may have been too narrow to capture substantial metabolic deterioration. One-year increases in alanine and BCAA were associated with higher subsequent T2D risk.^{12,13}

Herein, we aimed to examine the relationship of long-term metabolite changes with subsequent T2D risk, based on a T2D case-control study nested within the Nurses' Health Study (NHS) with repeated metabolomics measurements approximately ten years apart. The targeted plasma metabolomics profiles included a broad range of amino acids, lipids, and other organic compounds. The multivariate analyses were adjusted for baseline and concurrent changes in body weight, diet, physical activity, and other lifestyle factors.

Methods

Study design

In 1976, the NHS recruited 121,701 female nurses aged 30–55 years. Ever since, participants have been contacted biennially, assessing lifestyle and health status with validated, self-administered questionnaires. Detailed information on recruitment and follow-up procedures of the NHS has been published.¹⁴

A subset of 32,826 nurses provided blood samples in 1989 or 1990, of whom 18,743 provided a second blood sample in 2000 or 2001. After collection in 15mL heparin tubes, the blood specimens were shipped on an ice pack in a Styrofoam container to the study laboratory by overnight mail (95% were received within 24 h after blood draw).¹⁵ Upon arrival, samples were immediately centrifuged and aliquoted into cryotubes as plasma, buffy coat, and red blood cells and stored in liquid nitrogen (–130°C). Among participants with blood specimens collected at both time points, we designed a

nested case-control study within the NHS, including 244 incident T2D cases and 244 controls (1:1-matching). Inclusion criteria were availability of required blood specimens, no prevalent T2D at baseline, no incident T2D until after the second blood collection, and fasting ≥ 8 hours at blood collection. Matching factors were age, race and ethnicity, and calendar time (± 3 months) of the first blood collection (**Supplemental Figure 1**). Study inception was defined at the first blood collection date. All incident T2D cases were diagnosed between 2002 and 2008. The median follow-up time from the second blood collection to type 2 diabetes diagnosis was 3.9 years (minimum 1.08 years, IQR 2.4–5.3 years, maximum 6.8 years).

We validated the baseline metabolomics-T2D risk associations using a similar case-control sample in NHS with the same inclusion criteria (except that T2D incidence was allowed after the first blood collection) and matching factors described above, including 480 incident T2D cases and 480 controls (1:1-matching). For these participants, blood specimens and metabolomics data were only available at baseline (1989–90), and T2D cases were diagnosed between 1992 and 2008.

Laboratory measurements

For each study participant, two targeted plasma metabolomics profiles based on the two blood specimens from 1989/90 and 2000/01 were generated at the Broad Institute (Cambridge, MA). We used hydrophilic interaction liquid chromatography coupled with positive-ion mode mass spectrometry detection (HILIC-positive method) to measure polar metabolites. Raw data were processed with the TraceFinder 3.3 software (Thermo Fisher Scientific) and Progenesis Q1 (Nonlinear Dynamics, Newcastle upon Tyne, U.K.). The annotation of mass spectra to specific metabolites was confirmed with authentic reference standards or reference samples.¹⁶ Aliquots of the same pooled reference plasma were also interspersed at 20-sample intervals throughout all measurements. Every metabolite peak in the mass spectra was standardised to the nearest reference sample by calculating the ratio between the sample peak area and the reference peak area and multiplying with the median peak area across reference samples. Thereby, the metabolomics data were corrected for temporal drift in instrument sensitivity and scaled between batches. Blood specimens from matched case-control pairs and repeated blood specimens from the same individuals were handled under identical conditions, shipped to the lab in the same box, and analysed in the same batch but in random order.

All blood specimens used in this study were analysed in consecutive batches. The lab personnel was blinded to the case-control status of the blood specimen donors. We removed all metabolites with coefficients of variation above 0.25 among reference samples or with a

missing data proportion of 25% or higher from the metabolomics data. One-hundred-seventy known metabolites passed these quality control criteria.

Covariables

All covariables were derived from validated self-reported questionnaires. Body weight and smoking status were assessed biennially, and physical activity was assessed every four years. Body mass index (BMI) was calculated from weight (kg) divided by height in square meters. Smoking status was categorised into *never smokers*, *past smokers*, and *current smokers*. We estimated metabolic equivalents (METs) in hours per week from recreational and leisure time activities. Diet quality was measured with the Alternative Healthy Eating Index (AHEI). The AHEI comprises 11 health-related foods and nutrients and, for each, assigning 10 points to participants with the most beneficial and 0 points to those with the most adverse habitual consumption level.¹⁷ The underlying habitual consumption frequencies were assessed with a validated 131-item semiquantitative food-frequency questionnaire (FFQ) administered every four years. We used dietary data from the years 1990 and 1998. These self-reported lifestyle and anthropometric variables were validated and showed good to excellent agreement with gold-standard measurements.^{18–22} For analyses of single metabolite values, the closest available covariable data before blood collection was used for adjustment.

Type 2 diabetes ascertainment

T2D incidence was detected based on self-reported diagnosis and confirmed by a validated supplementary questionnaire.²³ Before 1998, confirmation of T2D incidence relied on the National Diabetes Data Group criteria, and from 1998 onwards on the American Diabetes Association diagnostic criteria. Validation studies in the NHS have demonstrated the validity of the supplementary questionnaires to adjudicate T2D diagnosis, showing that more than 97% of participants with self-reported T2D detected by questionnaires were re-confirmed through medical record review by endocrinologists blinded to questionnaire information.^{23,24}

Statistical analysis

Among the quality-controlled metabolites, most had complete data. However, missing values were present in 23 metabolites, from which five metabolites had between 1% and 5% missing values, and 18 metabolites had less than 1% missing values (**Supplemental Table 1**). Missing values were imputed with half the minimally measured value. For the 10-year metabolite change analyses, we first log-transformed the metabolite values and then calculated the 10-year change on the log-scale by subtracting the first from the second measurement and transformed the results (Δ -values) into Z-scores

(mean=0, SD=1). We also constructed, for each metabolite, five balanced categories according to the quintiles of the metabolite change distribution in the study sample. For analyses of metabolomics data from a single time point, we applied inverse normal transformation, which generates a rank-based standard normal distribution (mean=0, SD=1), minimising outliers' possible influence and generating comparable effect sizes between metabolites.

For descriptive statistics, we examined the distribution of non-normal continuous variables according to the median and interquartile range and the distribution of approximately normally distributed continuous variables according to mean and standard deviation. Categorical variables were described as frequencies across categories. The normally distributed, inverse normal transformed metabolite values were used to estimate Pearson correlation coefficients between the two metabolite measurements approximately 10 years apart. Based on the inverse normal transformed metabolite values, we also estimated partial correlations for each metabolite pair adjusted for all other metabolites, using the *ppcor* R-package.²⁵

We applied the causal structure learning PC-algorithm²⁶ to the metabolomics data to estimate and graphically represent the metabolites' conditional dependencies. First, we generated a conditional dependency network where edges between pairs of metabolites represented associations that were robust against adjustment for any set of other metabolites for each of the three metabolomics experiments (first and second metabolomics profiles in the case-control samples with repeated metabolomics measurements and metabolomics data in the validation sample with a single baseline profile). Then, we retained only edges detected in all three metabolomics experiments because robust edges across subgroups are likely to be biologically meaningful.^{27,28} We used the Louvain algorithm of the *igraph* R-package (<http://igraph.org/>) to detect densely connected clusters within the metabolomics network.

Associations of metabolites with T2D risk were estimated in conditional logistic regression models using the *clogit* function of the *survival* R-package (<https://CRAN.R-project.org/package=survival>), stratifying by the matched case-control pairs. We used the Efron approximation²⁹ for partial likelihood estimators. In terms of covariable adjustments, all models were conditioned on age, race/ethnicity, and time of blood collection by the matched case-control design. The matched design was broken only for the quintile-based categorical models, where age was modelled as a continuous covariable. The AHEI [points], BMI [kg/m²], and METS [hours/week] were included as continuous covariables. Smoking status was modelled as category-based dummy variables according to the above-defined smoking categories. All change analyses were also adjusted for baseline metabolite levels (1989/90) and concurrent changes in covariables. In secondary analyses, statistical

interactions between the first (1989/90) and the second (2000/01) metabolite measurement on T2D risk were tested by including a multiplicative term (metabolite₁*metabolite₂) along with the main metabolite terms and covariables in the logistic regression models described above.

We considered two-tailed p-values below 0.05 (Wald test) statistically significant. According to the number of metabolites, all p-values were adjusted for multiple testing by controlling the false discovery rate (FDR).³⁰ We used SAS version 9.4 (SAS Institute, NC, USA) to prepare datasets and R version 3.6.0 (R Foundation for Statistical Computing, Vienna, Austria) for all descriptive and inferential statistical analyses.

Because of participant confidentiality and privacy concerns, data from the Nurses' Health Study are not made publicly available. All data and statistical programs are permanently stored on the research computing cluster of the Channing Division of Network Medicine, Brigham and Women's Hospital. Requests to access these data should be made via: <https://www.nurseshealthstudy.org/researchers>.

Ethics statement

The study protocol was approved by the institutional review boards (IRBs) of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and the IRBs allowed participants' completion of questionnaires to be considered as implied consent.

Role of funders

This work was supported by research grants UM1 CA186107, R01 CA49449, DK112940, and DK119268 from the National Institutes of Health. CW was supported by the German Research Foundation's (DFG) individual fellowship #WI132/1-1 and the Boston Nutrition Obesity Research Center (P30 DK46200). MG-F was supported by the American Diabetes Association grant #1-18-PMF-029. DEH was supported by the National Institutes of Health T32-CA009001. Dr. Li was supported by NIDDK K99 DK122128 and the Boston Nutrition Obesity Research Center (P30 DK46200). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. None of the funding sources played a role in the design, collection, analysis, or interpretation of the data or decision to submit the manuscript for publication.

Results

Descriptive statistics

Participants' median ages were 54 and 65 years at the first and second blood collection, respectively. Compared to participants with incident T2D, we observed

higher diet quality (AHEI score), physical activity, proportion of never smokers, and alcohol consumption, and lower median BMI among controls. Moreover, the favourable changes of these lifestyle factors were observed among controls, except that participants quit smoking at a similar rate among cases and controls. More cases than controls had first-degree relatives with T2D (Family history of T2D) (Table 1).

The distributions of 170 quality-controlled, known metabolites are summarised in Supplemental Table 1. The 10-year changes were mostly moderate (median +4%, IQR: -1% to +9% in controls, median +5%, IQR -1% to +11% in cases), and for all the metabolites the repeated measurements were positively correlated [controls' median Spearman correlation coefficient across metabolites $r = 0.45$, IQR: 0.38 to 0.53; cases' median $r = 0.43$, IQR: 0.36 to 0.52 (Supplemental Table 2)]. Correlation and partial correlation structures between different metabolites were consistent between time points (Supplemental Figures 2 and 3) and comparable with correlations and partial correlations between the 10-year metabolite changes (Supplemental Figure 4).

Metabolite changes over ten years and subsequent T2D risk

We examined the T2D risk-association of intraindividual metabolite changes over 10 years, matched for age, race/ethnicity, time of blood sampling, and adjusted for the metabolite's baseline level, as well as baseline values and 10-year changes of BMI, physical activity (METS-h/week), diet quality (AHEI score), and smoking status. Thirty-one 10-year metabolite changes showed a nominally significant T2D risk trend across quintile-based categories ($P_{trend} < 0.05$) and were significant after multiple testing correction in the continuous analysis (FDR < 0.05 , Wald test) (Table 2 and Supplemental Tables 3 and 4). The strongest T2D risk-associated metabolite changes were 10-year increases in BCAA levels [OR_Δ per SD (95%CI)]: Disoleucine 2.72 (1.95, 3.76), Δleucine 2.53 (1.85, 3.45), and Δvaline 1.93 (1.53, 2.43) (Figure. 1). Among the other 10-year amino acid changes, Δalanine and Δmethionine were associated with higher T2D risk, and ΔN-acetylaspartic acid, Δglutamine, and Δglycine with lower T2D risk. Greater 10-year increases in short-chain acylcarnitines (ΔC₄-OH carnitine, ΔC₅ carnitine), phosphatidylethanolamines (ΔC_{36:4} PE, ΔC_{40:6} PE, ΔC_{38:6} PE, ΔC_{36:2} PE, ΔC_{34:2} PE, ΔC_{38:4} PE), vitamins (Δretinol, Δpantothenate), bile acids (Δglycocholate, Δglycodeoxycholate / Δglycochenodeoxycholate), and diacylglycerol (DG) / triacylglycerol (TG)-fragment (ΔC_{36:2} DG/TG, ΔC_{34:1} DG/TG, ΔC_{34:2} DG/TG) were associated with higher subsequent T2D risk. Smaller 10-year decreases in lysophospholipids (ΔC_{20:0} LPE, ΔC_{16:1} LPC plasmalogen), sphingomyelins (ΔC_{16:1} SM, ΔC_{16:0} SM, ΔC_{14:0} SM, ΔC_{18:2} SM), and plasmalogens (ΔC_{36:3}

	First metabolomics profile [1990]		Second metabolomics profile [2000]		Change [2000 vs. 1990]*	
	Controls [n=248]	Cases [n=248]	Controls [n=244]#	Cases [n=244]	ΔControls	ΔCases
Age [years]	54.0 [§] (49.3, 58.7)	53.9 (49.5, 58.7)	65.3 (60.4, 69.8)	64.8 (60.3, 70)	11.2 (10.9-11.4)	11.2 (10.8-11.4)
Body mass index (kg/m ²)	24.0 (22.0, 26.8)	26.7 (23.9, 29.9)	25.9 (22.7, 29.1)	28.8 (25.5, 33.1)	1.2 (0.0-2.4)	1.8 (0.4-3.5)
Physical activity [METS-h/week]	9.1 (3.1, 21.3)	8.4 (2.5, 21.6)	12.9 (5.1, 21.5)	10.2 (3.6, 23.2)	2.1 (-5.0-10.0)	0 (-5.1-8.0)
Diet quality (AHEI)	52.4 (46.8, 59.5)	51.0 (43.6, 58.4)	53.9 (46.5, 60.5)	50.2 (44.2, 58.1)	0.3 (-5.1-6.5)	-0.1 (-5.5-5.7)
Alcohol [g/day]	1.3 (0.0, 5.5)	0.9 (0.0, 3.7)	1.5 (0.0, 6.3)	0.9 (0.0, 5.2)	0.0 (-0.8, 1.1)	0.0 (0.0, 0.9)
Smoking						
Current	13%	12%	6%	5%	-7%	-7%
Former	35%	40%	42%	46%	7%	6%
Never	52%	47%	52%	48%	0	1% [§]
Family history of T2D (one or more prevalent T2D case among parents and siblings)						
Yes	25%	40%	-	-	-	-

Table 1: Baseline characteristics of participants in the type 2 diabetes case-control study with repeated metabolomics profiles, nested within the Nurses' Health Study.

Controls: women without incident type 2 diabetes during follow-up.
 Cases: women with incident type 2 diabetes during follow-up.
 * Change in participant characteristics between first (1989/90) and second (2000/01) blood collection. Covariable changes: Body mass index, 2000 vs. 1990; Alternative Healthy Eating Index (AHEI), 1998 vs. 1990; METS, 2000 vs. 1988; smoking status, 2000 vs. 1990.
 # Four case-control pairs were without valid metabolomics data for the second time point and excluded from all analyses that included the second metabolomics profile.
 § Median (IQR), all such values.
 § Diet quality was measured with the Alternative Healthy Eating Index (AHEI).
 § Higher proportion of Never Smokers due to the exclusion of four case-control pairs.

	Categorical analysis					Continuous analysis		
	Q1	Q2	Q3	Q4	Q5	P-	OR (95%CI)	FDR
Metabolite		OR (95%CI)	OR (95%CI)	OR (95%CI)	OR (95%CI)	Trend	OR (95%CI)	FDR
Δisoleucine	ref.	1.68 (1.12, 2.56)	1.95 (1.26, 3.01)	2.23 (1.44, 3.48)	3.53 (2.23, 5.54)	2E-07	2.72 (1.97, 3.79)	3E-07
Δleucine	ref.	1.27 (0.84, 1.93)	1.40 (0.92, 2.14)	1.82 (1.15, 2.89)	2.56 (1.70, 3.89)	4E-06	2.53 (1.86, 3.47)	4E-07
Δvaline	ref.	1.35 (0.84, 2.16)	1.72 (1.10, 2.70)	1.70 (1.05, 2.76)	2.61 (1.64, 4.17)	5E-05	1.93 (1.52, 2.44)	3E-06
Δretinol	ref.	1.36 (0.88, 2.11)	1.31 (0.83, 2.05)	1.75 (1.12, 2.74)	2.23 (1.40, 3.54)	8E-04	1.82 (1.44, 2.30)	3E-05
ΔN-acetylaspartic acid	ref.	0.92 (0.64, 1.34)	0.70 (0.45, 1.11)	0.56 (0.36, 0.88)	0.55 (0.35, 0.88)	0.002	0.54 (0.42, 0.70)	1E-04
ΔC4-OH carnitine	ref.	1.38 (0.86, 2.23)	1.75 (1.08, 2.87)	2.27 (1.41, 3.64)	2.53 (1.57, 4.10)	2E-05	1.84 (1.41, 2.40)	2E-04
Δalanine	ref.	1.67 (1.13, 2.44)	1.39 (0.93, 2.07)	1.62 (1.03, 2.53)	2.18 (1.37, 3.48)	0.011	1.68 (1.31, 2.16)	9E-04
ΔC20:0 LPE	ref.	1.00 (0.70, 1.44)	0.85 (0.57, 1.28)	0.54 (0.35, 0.84)	0.63 (0.41, 0.96)	0.003	0.68 (0.56, 0.82)	0.001
Δglycocholate	ref.	1.43 (0.93, 2.19)	1.48 (0.97, 2.25)	1.67 (1.10, 2.55)	1.67 (1.01, 2.73)	0.016	1.58 (1.26, 2.00)	0.001
ΔC16:1 SM	ref.	0.72 (0.49, 1.00)	0.67 (0.45, 1.00)	0.97 (0.66, 1.44)	0.39 (0.24, 0.64)	0.004	0.68 (0.56, 0.83)	0.002
ΔC5 carnitine	ref.	1.21 (0.78, 1.87)	1.13 (0.70, 1.82)	1.95 (1.25, 3.05)	2.14 (1.37, 3.33)	1E-04	1.57 (1.24, 1.99)	0.002
Δmethionine	ref.	1.11 (0.72, 1.71)	1.13 (0.75, 1.70)	1.14 (0.71, 1.80)	1.92 (1.24, 2.96)	0.015	1.55 (1.24, 1.95)	0.002
ΔC36:2 from DG/TG	ref.	1.48 (0.97, 2.23)	1.20 (0.78, 1.83)	1.28 (0.84, 1.97)	1.82 (1.16, 2.87)	0.049	1.62 (1.25, 2.07)	0.003
ΔC16:0 SM	ref.	0.85 (0.59, 1.25)	0.73 (0.50, 1.06)	0.73 (0.50, 1.06)	0.64 (0.42, 1.00)	0.025	0.66 (0.52, 0.82)	0.003
ΔC34:1 from DG/TG	ref.	1.34 (0.86, 2.09)	1.86 (1.22, 2.85)	1.65 (1.06, 2.57)	1.99 (1.23, 3.21)	0.002	1.68 (1.26, 2.24)	0.005
Δpantothenate	ref.	1.38 (0.87, 2.16)	1.58 (1.02, 2.48)	1.34 (0.80, 2.21)	1.80 (1.09, 2.99)	0.008	1.55 (1.21, 1.97)	0.005
ΔC38:6 PE	ref.	1.11 (0.70, 1.75)	1.12 (0.72, 1.73)	1.68 (1.10, 2.60)	1.42 (0.90, 2.26)	0.019	1.43 (1.17, 1.76)	0.005
ΔC36:4 PE	ref.	1.31 (0.84, 2.03)	1.36 (0.88, 2.11)	1.52 (0.97, 2.36)	1.72 (1.09, 2.70)	0.015	1.45 (1.17, 1.78)	0.005
ΔC40:6 PE	ref.	1.14 (0.73, 1.77)	1.26 (0.84, 1.87)	1.35 (0.89, 2.05)	1.82 (1.19, 2.78)	0.007	1.48 (1.19, 1.86)	0.005
ΔC36:3 PE plasmalogen	ref.	0.76 (0.52, 1.11)	0.72 (0.48, 1.07)	0.58 (0.38, 0.89)	0.68 (0.46, 1.01)	0.017	0.74 (0.62, 0.88)	0.006
ΔC16:1 LPC plasmalogen	ref.	0.83 (0.55, 1.24)	0.78 (0.52, 1.18)	0.64 (0.40, 1.0)	0.64 (0.42, 0.99)	0.028	0.68 (0.54, 0.85)	0.007
ΔC36:2 PE	ref.	1.43 (0.95, 2.16)	1.22 (0.79, 1.86)	1.35 (0.87, 2.10)	2.32 (1.45, 3.72)	0.011	1.42 (1.15, 1.76)	0.008
ΔC34:3 PC plasmalogen	ref.	0.70 (0.47, 1.03)	0.59 (0.39, 0.88)	0.58 (0.38, 0.88)	0.63 (0.42, 0.93)	0.019	0.70 (0.55, 0.89)	0.020
ΔC34:2 PE	ref.	1.67 (1.06, 2.58)	1.31 (0.83, 2.06)	1.57 (1.01, 2.43)	2.05 (1.28, 3.32)	0.012	1.38 (1.11, 1.70)	0.020
ΔC34:2 from DG/TG	ref.	1.20 (0.76, 1.86)	1.48 (0.96, 2.24)	1.43 (0.93, 2.21)	1.73 (1.10, 2.77)	0.011	1.46 (1.13, 1.89)	0.022
Δglutamine	ref.	0.80 (0.55, 1.17)	0.67 (0.44, 1.01)	0.50 (0.32, 0.77)	0.69 (0.45, 1.07)	0.015	0.74 (0.61, 0.91)	0.027
ΔC38:4 PE	ref.	1.12 (0.71, 1.73)	1.43 (0.94, 2.19)	1.17 (0.76, 1.82)	1.80 (1.15, 2.84)	0.021	1.31 (1.09, 1.59)	0.027
ΔC18:2 SM	ref.	0.81 (0.56, 1.18)	0.70 (0.48, 1.05)	0.64 (0.43, 0.94)	0.64 (0.41, 0.99)	0.006	0.78 (0.65, 0.93)	0.028
Δglycine	ref.	0.84 (0.58, 1.21)	0.55 (0.37, 0.83)	0.64 (0.43, 0.95)	0.52 (0.34, 0.79)	6E-04	0.73 (0.58, 0.91)	0.030
ΔC14:0 SM	ref.	0.90 (0.61, 1.32)	0.83 (0.56, 1.24)	0.93 (0.62, 1.41)	0.58 (0.36, 0.92)	0.045	0.77 (0.64, 0.93)	0.030
ΔGDX / GCDC	ref.	1.72 (1.10, 2.65)	1.72 (1.10, 2.67)	1.42 (0.88, 2.28)	1.97 (1.19, 3.28)	0.044	1.35 (1.08, 1.67)	0.033

Table 2: Associations of 10-year-plasma metabolite changes (2000/01 vs. 1989/90) with subsequent risk of developing type 2 diabetes (2002-2008).

Shown are 10-year metabolite changes with nominally significant trend test in the categorical analysis ($P < 0.05$) and a significant multiple testing-corrected T2D risk association in the continuous analysis ($FDR < 0.05$, Wald test). Case-control sample with repeated metabolomics profiles (1989/90 and 2000/01) and subsequent type 2 diabetes incidence (between 2002 and 2008), $n = 244$ cases and $n = 244$ controls, 1:1 matched for age, race/ethnicity, and time of blood draw.

OR (95%CI): odds ratio (95% confidence interval) per one standard deviation higher levels from a conditional logistic regression model, adjusted for baseline metabolite level, BMI [kg/m^2], diet quality [AHEI-points], physical activity [METs-h/week], Δ BMI (2000 vs. 1990), Δ AHEI (1998 vs. 1990), Δ METS (2000 vs. 1988), and smoking status [3 categories: never, current, past] at both blood sampling time points.

Q1-Q5: The case-control sample was categorised according to the quintiles of the metabolite distribution; for the categorical analysis, models were not conditioned on matched case-control pair but adjusted for age, all other adjustments remaining the same.

FDR: false discovery rate-controlled p-value (Wald test), adjusted for testing 170 known metabolites.

Δ : 10-year changes (1989/90 vs. 2000/01).

Abbreviations - DG: diacylglycerol, TG: triacylglycerol, PE: phosphatidylethanolamine, LPE: lysophosphatidylethanolamine, PC: phosphatidylcholine, LPC: lysophosphatidylcholine, SM: sphingomyelin, GDX: glycodeoxycholate, GCDC: glycochenodeoxycholate.

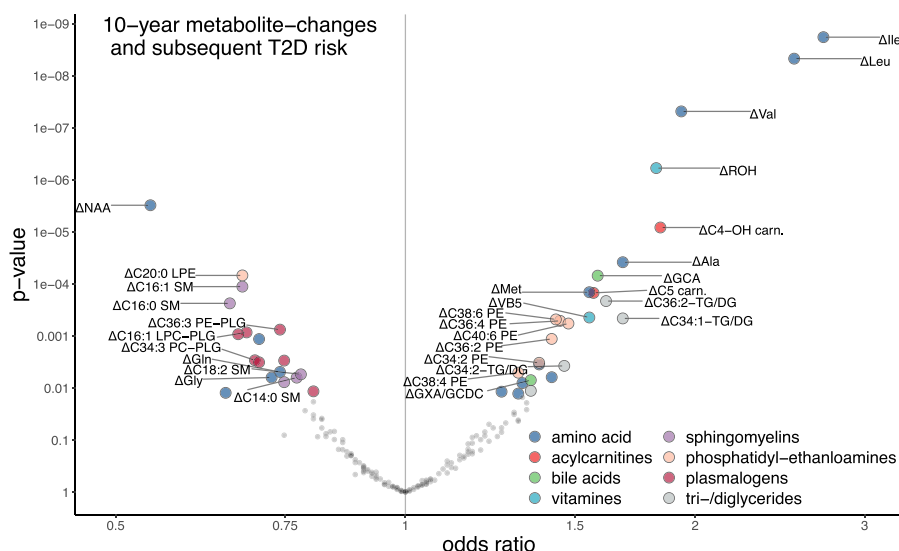


Figure 1. Association of 10-year metabolite changes with T2D risk. Colored and labeled dots are T2D-associated 10-year metabolite changes (continuous analysis FDR < 0.05 [Wald test] and categorical analysis P < 0.05); small grey dots are non-T2D-associated metabolite changes (continuous analysis FDR > 0.05). Odds ratio (conditional logistic regression) per one standard deviation higher 10-year metabolite change, matched for age, race/ethnicity, and time of blood draw, and adjusted for baseline metabolite level, BMI [kg/m²], diet quality [AHEI-points], physical activity [METS-h/week], Δ BMI (2000 vs. 1990), Δ AHEI (1998 vs. 1990), Δ METS (2000 vs. 1988), and smoking status [3 categories; never, current, past] at both blood sampling time points. Case-control study with 244 incident type 2 diabetes cases and 244 matched controls.

Abbreviations - Ala: alanine, carn.: carnitine, DG: diglyceride, GCA: glycocholate, GCDC: glycochenodeoxycholate, Gln: glutamine, Gly: glycine, GXA: glycodeoxycholate, Ile: isoleucine, Leu: leucine, LPE: lysophosphatidyl-ethanolamine, Met: methionine, NAA: N-acetylaspartic acid, PE: phosphatidylethanolamine, PLG: plasmalogen, Pro: proline, ROH: retinol, SM: sphingomyelin, TG: triglyceride, Val: valine, VB5: pantothenate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

PE plasmalogen, Δ C_{34:3} PC plasmalogen) were associated with lower subsequent T2D risk (Figure 1).

Co-occurrence T2D risk-associated metabolite changes in network clusters

We constructed a data-driven conditional independence network and identified densely connected metabolite clusters (Supplemental Figure 5). Higher T2D risk-associated 10-year changes co-occurred in clusters of BCAA & other amino acids, DG/TG-fragments, some vitamins, and bile acids (Figure 2). Inverse T2D-risk associations co-occurred in clusters of plasmalogens and sphingomyelins. Three clusters (glycine & other amino acids, glycerophospholipids, and short-chain FA-containing acylcarnitines & derivatives) included both, metabolite changes associated with higher and other metabolite changes associated with lower T2D risk. For example, in a glycerophospholipid cluster, phosphatidylethanolamine changes were associated with higher T2D risk, while lysophospholipid changes were inversely associated with T2D risk. Overrepresentation analysis with knowledge-based metabolite sets supported the enrichment of significantly T2D-associated

metabolites changes in metabolite sets related to BCAA catabolism and bile acid synthesis (Supplemental Figure 6 and Supplemental Table 5).

Metabolite-T2D risk associations across samples and time points

We also analysed the T2D risk associations of single metabolites with and without a 10-year lag time (time-span between first blood collection [1989/90] and first incident type 2 diabetes case) separately in age-matched models adjusted for BMI, physical activity (METS-h/week), diet quality (AHEI score), and smoking status. We further verified these associations in a validation sample with no lag time (n=960, T2D incidence over sixteen years). Twenty-five metabolites were associated with subsequent T2D with and without a lag time and in the validation sample (FDR < 0.05 [Wald test] in all analyses) (Table 3). When comparing the FDR with and without a 10-year lag (Figure 3), the associations of three BCAA, alanine, and glycine with T2D risk were markedly stronger in the no-lag analysis, though highly significant at both times. Several DG/TG-fragments and aromatic amino acids were significantly (FDR <

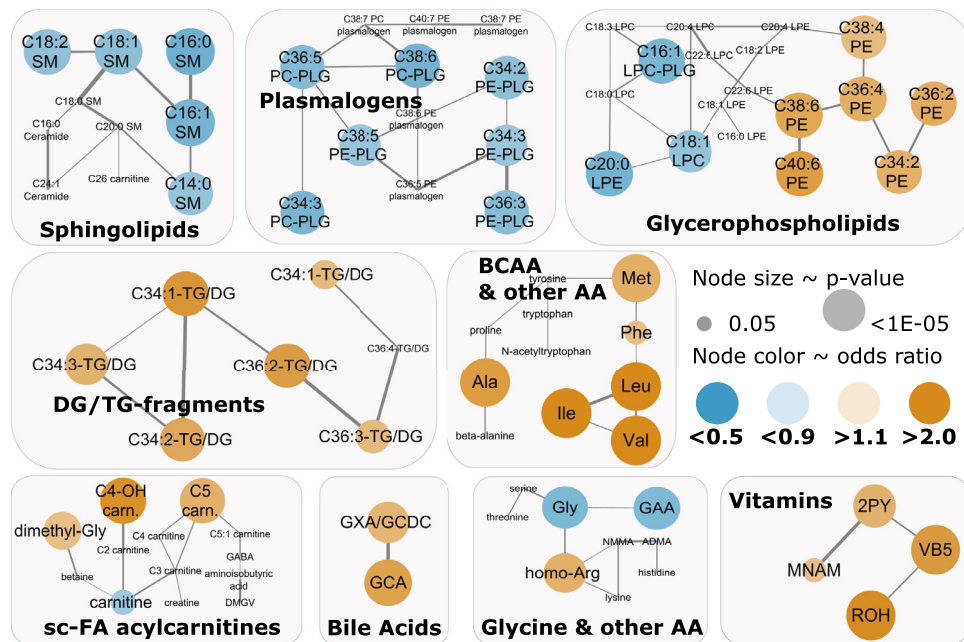


Figure 2. Co-occurrence of T2D-associated 10-year plasma metabolite changes in metabolomics network clusters. Type 2 diabetes-associated ($P < 0.05$, Wald test) metabolites are color-coded: Orange - high type 2 diabetes risk, Blue - low type 2 diabetes risk. The data-driven conditional independence network was constructed with the causal structure learning PC-algorithm; clusters were detected with the Louvain algorithm. Only the clusters with either three or more than 50% type 2 diabetes-associated 10-year metabolite changes are displayed. Odds ratio (conditional logistic regression) per one standard deviation higher 10-year metabolite change, matched for age, race/ethnicity, and time of blood draw, and adjusted for baseline metabolite level, BMI [kg/m^2], diet quality [AHEI-points], physical activity [METS-h/week], ΔBMI (2000 vs. 1990), ΔAHEI (1998 vs. 1990), ΔMETS (2000 vs. 1988), and smoking status [3 categories; never, current, past] at both blood sampling time points. Case-control study with 244 incident type 2 diabetes cases and 244 matched controls.

Abbreviations - 2PY: N1-methyl-2-pyridone-5-carboxamide, Ala: alanine, Arg: arginine, carn.: carnitine, DG: diglyceride, GAA: guanidoacetic acid, GABA: gamma-aminobutyric acid, Galn: galactosamine, GCA: glycocholate, GCDC: glycochenodeoxycholate, Gln: glutamine, Gly: glycine, GXA: glycodeoxycholate, Ile: isoleucine, Leu: leucine, LPE: lysophosphatidylethanolamine, Met: methionine, MNAM: 1-methyl-nicotinamide; NAA: N-acetylaspartic acid, PE: phosphatidylethanolamine, Phe: phenylalanine, PLG: plasmalogen, Pro: proline, ROH: retinol, SM: sphingomyelin, TG: triglyceride, Val: valine, VB5: pantothenate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

0.05, Wald test) associated with T2D risk in the analyses with and without a lag time, with comparable significance levels in both analyses. Uric acid was the only consistent and validated T2D risk-associated metabolite with a substantially stronger T2D risk association in the 10-year lag analysis.

Another group of validated metabolite-T2D risk associations was only evident in the no-lag analysis, including acylcarnitine C4-OH, bile acids, some phosphatidylethanolamines (all associated with higher T2D risk), as well as N-acetylaspartic acid, several plasmalogens, and sphingomyelins (lower T2D risk). Allo-isoleucine was the only validated T2D risk-associated metabolite in the 10-year-lag-analysis that was non-significant in the no-lag-analysis. However, the p-values (0.04 at T1 and 0.07 at T2 [Wald tests]) and effect estimates (1.34 [1.08, 1.64] at T1 and 1.2 [1.02, 1.42] at T2) for the association of allo-isoleucine with T2D at both

timepoints were comparable (Figure 3 and Supplemental Tables 6-8).

Secondary analyses

The plasma level of uric acid in 1990 was significantly associated with T2D risk ($\text{FDR} < 0.05$, Wald test) after adjusting for the second measurements ten years later, and T2D risk associations of the early measurements of another twenty-four metabolites (including BCAA and most lipid markers in Table 3) were suggestive ($P < 0.05$, Wald test) when adjusted for the second measurement (Supplemental Table 9). These analyses suggest that even with a second metabolomics profile relatively close to T2D incidence, the historical metabolomics data (i.e., change of metabolites over 10 years) convey additional T2D risk information. We detected significant ($\text{FDR} < 0.05$, Wald test) multiplicative

Metabolite	Case-control sample with repeated metabolomics profiles*				Validation sample [§]	
	First profile (1989-90) 10-year lag to T2D		Second profile (2000-01) No lag		Single profile (1989-90) No lag	
	OR (95%CI)	FDR	OR (95%CI)	FDR	OR (95%CI)	FDR
leucine	1.88 (1.50, 2.36)	3E-06	2.80 (2.13, 3.68)	1E-11	1.99 (1.60, 2.47)	2E-08
isoleucine	2.05 (1.61, 2.61)	8E-07	2.80 (2.13, 3.66)	1E-11	2.12 (1.69, 2.66)	3E-09
valine	1.60 (1.31, 1.94)	9E-05	2.12 (1.70, 2.62)	5E-10	2.18 (1.72, 2.74)	3E-09
alanine	1.36 (1.13, 1.65)	0.009	1.82 (1.51, 2.19)	1E-08	1.79 (1.48, 2.18)	9E-08
glycine	0.69 (0.57, 0.84)	0.002	0.58 (0.48, 0.71)	9E-07	0.63 (0.53, 0.73)	2E-07
C36:2 from DG/TG	1.57 (1.26, 1.97)	0.001	1.93 (1.52, 2.45)	2E-06	1.60 (1.37, 1.88)	1E-07
C34:1 from DG/TG	1.67 (1.35, 2.05)	7E-05	2.08 (1.59, 2.7)	2E-06	2.08 (1.71, 2.52)	4E-11
C34:2 from DG/TG	1.70 (1.36, 2.12)	1E-04	1.90 (1.48, 2.42)	6E-06	1.82 (1.52, 2.18)	3E-09
C34:3 PC plasmalogen	0.66 (0.55, 0.79)	1E-04	0.59 (0.48, 0.72)	9E-06	0.65 (0.55, 0.77)	9E-06
C38:5 from DG/TG	1.52 (1.24, 1.87)	9E-04	1.67 (1.35, 2.03)	1E-05	1.63 (1.39, 1.93)	1E-07
C5 carnitine	1.32 (1.11, 1.58)	0.011	1.60 (1.32, 1.93)	1E-05	1.40 (1.19, 1.67)	4E-04
C34:3 from DG/TG	1.57 (1.25, 1.95)	0.001	1.68 (1.37, 2.08)	1E-05	1.60 (1.34, 1.89)	1E-06
C40:6 PE	1.36 (1.13, 1.65)	0.009	1.65 (1.35, 2.01)	1E-05	1.54 (1.30, 1.80)	3E-06
homoarginine	1.36 (1.13, 1.63)	0.009	1.62 (1.32, 1.99)	3E-05	1.40 (1.17, 1.70)	0.002
C34:1 from DG/TG	1.55 (1.23, 1.95)	0.002	1.62 (1.32, 1.98)	3E-05	1.55 (1.33, 1.82)	7E-07
C36:3 from DG/TG	1.65 (1.28, 2.11)	0.001	1.63 (1.33, 2.02)	4E-05	1.43 (1.24, 1.65)	1E-05
phenylalanine	1.52 (1.25, 1.85)	5E-04	1.43 (1.21, 1.70)	2E-04	1.35 (1.14, 1.60)	0.003
C36:4 from DG/TG	1.62 (1.29, 2.02)	6E-04	1.51 (1.23, 1.83)	3E-04	1.39 (1.20, 1.61)	1E-04
urate	1.82 (1.49, 2.24)	8E-07	1.38 (1.15, 1.65)	0.002	1.48 (1.25, 1.73)	2E-05
C4 carnitine	1.27 (1.08, 1.49)	0.025	1.35 (1.14, 1.59)	0.002	1.23 (1.05, 1.45)	0.034
C3 carnitine	1.35 (1.14, 1.61)	0.005	1.38 (1.15, 1.66)	0.003	1.22 (1.05, 1.43)	0.029
tyrosine	1.31 (1.08, 1.59)	0.040	1.34 (1.12, 1.61)	0.005	1.65 (1.35, 2.01)	9E-06
N4-acetylcytidine	1.38 (1.14, 1.68)	0.009	1.28 (1.07, 1.53)	0.018	1.28 (1.11, 1.49)	0.003
DMGV	1.36 (1.13, 1.63)	0.009	1.23 (1.04, 1.46)	0.035	1.45 (1.23, 1.70)	9E-05
beta-alanine	1.34 (1.10, 1.60)	0.018	1.26 (1.04, 1.53)	0.047	1.39 (1.20, 1.62)	1E-04

Table 3: Metabolites that were significantly associated with type 2 diabetes-risk in three distinct metabolomics profiles from two case-control samples, nested within the Nurses' Health Study.

* Type 2 diabetes case-control study with repeated metabolomics profiles: first metabolomics profile 1989/90 (n= 248 case-control pairs); second metabolomics profile in 2000/01 (n= 244 case-control pairs); type 2 diabetes incidence between 2002 and 2008.

§ Type 2 diabetes case-control study with one metabolomics profile in 1989/90 (n= 480 case-control pairs); type 2 diabetes incidence between 1992 and 2008. OR (95%CI): odds ratio (95% confidence interval) per one standard deviation higher levels from a conditional logistic regression model, matched for age, race/ethnicity, time of blood draw, and adjusted for BMI [kg/m²], diet quality [AHEI-points], physical activity [METs-h/week], and smoking status [3 categories; never, current, past], assessed at the time of metabolomics profiling. FDR: false discovery rate-controlled p-value (Wald test), adjusted for testing 170 metabolites. Abbreviations - DG: diacylglycerol; DMGV: dimethylguanidino valerate; LPE: lysophosphatidylethanolamine; PC: phosphatidylcholine; PE: phosphatidylethanolamine, SM: sphingomyelin; T2D: type 2 diabetes; TG: triacylglycerol.

interactions on T2D risk between the first and second measurements of carnitine, acetaminophen, and N-acetylleucine, none of which was highlighted in [Table 2 \(Supplemental Table 10\)](#). These analyses suggest that the risk associations of the second measurement (2000/01) do not depend on baseline levels (1989/90) for most metabolites. We detected no statistically significant associations of 10-year metabolite changes with participant age on T2D risk (all FDR > 0.05 [Wald test]; data not shown). These interaction analyses suggest that the 10-year metabolite change-associated T2D risk estimates are consistent across the study subpopulations. To account for dietary retinol sources, we adjusted for baseline levels and concurrent changes in habitual red meat consumption and use of vitamin supplements and the results did not appreciably change. ([Supplemental Table 11](#)).

Discussion

In a case-control study of middle- to older-aged women nested within the NHS, we examined long-term metabolite changes and subsequent T2D incidence. The 10-year changes of thirty-one metabolites were significantly (FDR < 0.05, Wald test) associated with subsequent T2D risk, adjusted for the baseline metabolite level, baseline and changes of BMI and T2D-related lifestyle factors. Greater 10-year increases of BCAA, phosphatidylethanolamines, some vitamins, bile acids, and DG/TG-fragments were associated with higher T2D risk. The 10-year changes of lysophospholipids, sphingomyelins, and plasmalogens were associated with lower T2D risk. To our knowledge, this is the first study to assess long-term changes in repeated metabolomics profiles with subsequent T2D incidence. We also observed and validated twenty-five T2D risk-associated metabolites in

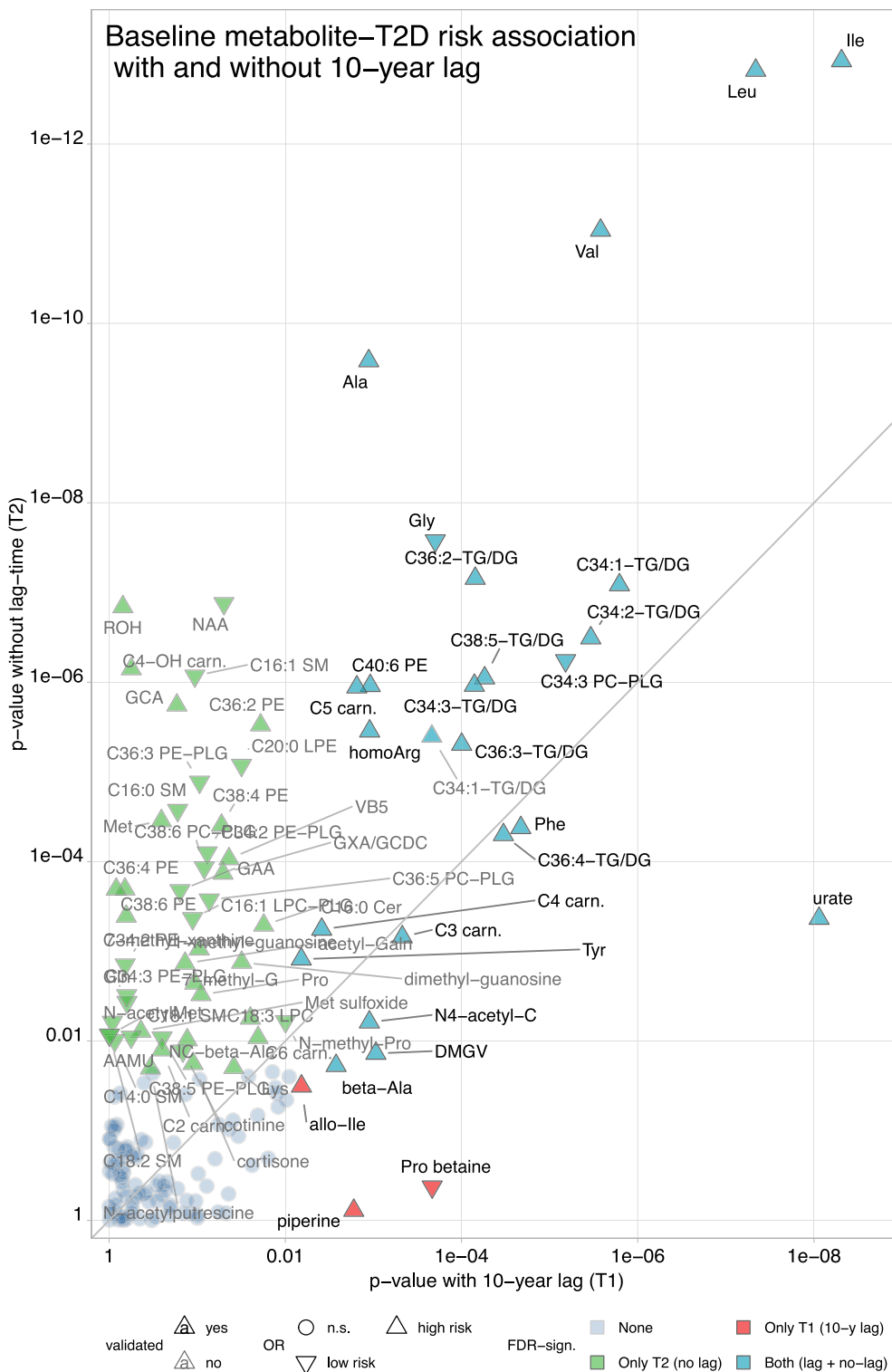


Figure 3. Significance levels of metabolite-type 2 diabetes associations with and without over ten-year lag before disease incidence. P-values (Wald test) are from a conditional logistic regression model, matched for age, race/ethnicity, and time of blood draw, and adjusted for BMI [kg/m²], diet quality [AHEI-points], physical activity [METS-h/week], and smoking status [3 categories; never, current, past], assessed at the time of metabolomics profiling. T1: 1989/90; T2: 2001/02; type 2 diabetes incidence between 2002 and 2008. Case-control study with 244 type 2 diabetes cases and 244 matched controls.

the 10-year lag analyses. Most of these metabolites were previously linked to T2D risk, but we are not aware of analyses restricted to participants with >10 years lag time to T2D incidence.

BCAAs were among the most robust metabolomics T2D risk markers in earlier studies.^{31–36} Mendelian randomization⁶ and pharmacological intervention³⁷ studies support adverse effects of BCAA on glucose homeostasis. Herein, we show, for the first time, that long-term BCAA increases were associated with higher subsequent T2D risk. Furthermore, the 10-year increases of short-chain acylcarnitines (ΔC_4OH , ΔC_5), BCAA catabolism products, were associated with higher T2D risk. The 10-year increases of alanine and methionine were also associated with higher T2D risk and glycine's and glutamine's 10-year changes with lower T2D risk. All these associations are consistent with previously reported associations of the metabolites' baseline levels with T2D risk^{31–35,38–44} and with our analysis of single metabolite measurements with subsequent T2D risk, including the 10-year lag analyses. Our observations indicate perturbed amino acid metabolism preceding T2D incidence for more than a decade and further deterioration of amino acid metabolism over time conferring additional T2D risk.

Several prospective cohort studies reported associations of complex lipids with T2D risk. Baseline levels of DG, TG, and phosphatidylethanolamines were previously related to higher T2D risk.^{43,45} We extended this evidence by showing that several phosphatidylethanolamines' and DG/TG-fragments' greater 10-year increases were associated with subsequently higher T2D risk. Conversely, smaller 10-year decreases of various sphingomyelins, lysophospholipids, and plasmalogens were associated with lower T2D risk, consistent with our findings and previous reports of plasmalogens',¹⁰ lysophospholipids',^{10,43} and sphingomyelins'^{10,40} baseline level-associations with lower T2D risk.

Furthermore, we observed higher subsequent T2D risk with greater 10-year increases of retinol, pantothenate, and bile acid plasma levels. Other studies found T2D risk associations with baseline bile acid levels.^{46–48} The evidence on plasma retinol levels and T2D risk is inconsistent,^{49,50} and little information on pantothenate-related T2D risk is available. The lack of effect attenuation after adjusting for primary dietary sources of retinol (red meat intake and use of vitamin supplements) suggests that the T2D risk association may reflect perturbed lipid metabolism rather than dietary

intake. The novel associations of 10-year N-acetylaspartic acid changes with lower T2D risk aligns with our recent report of a metabolic signature of walnut consumption that contained acetylaspartic acid, which was associated with lower T2D risk in the PREDIMED-trial.⁵¹ However, direct external evidence on N-acetylaspartic acid and T2D risk is lacking.

Our network analyses showed that the T2D risk-associated metabolite changes tended to co-occur in clusters of interdependent metabolites. Notably, for the lipid clusters, the network analysis suggested that upregulation (TG-/DG-fragments, phosphatidylethanolamines) or downregulation (sphingomyelins, lysophospholipids, plasmalogens) of groups of interrelated lipids is related to T2D risk. The enrichment analysis with knowledge-based metabolite sets suggested the overrepresentation of T2D-related 10-year metabolite changes in metabolites linked to BCAA degradation and bile acid biosynthesis. Further exploration of the interdependencies of long-term metabolite changes in relation to subsequent T2D risk is warranted.

Our study further elucidates the metabolic alterations that precede T2D incidence. We observed that plasma levels of BCAA were elevated more than 10 years before T2D onset and further increased before T2D diagnosis. High plasma BCAA levels may reflect impaired BCAA oxidation in hepatic and adipose tissue.⁸ However, dietary energy and animal protein overload may also contribute to high BCAA plasma levels,^{13,52,53} with possible adverse metabolic consequences.^{54,55} The T2D risk-related increase in short-chain acylcarnitines may reflect upregulated but incomplete BCAA oxidation in skeletal muscle, competing with mitochondrial fatty acid oxidation.^{6,8,56} Fatty acid overload may also contribute to the T2D-related increase in plasma levels of phosphatidylethanolamines and TG/DG-fragments.⁵⁶ The T2D risk-related decreases of sphingomyelins, plasmalogens, and lysophospholipids may reflect substrate rechanneling into other lipid classes.

Glycine is utilised in several key metabolic adaptation pathways, including glutathione synthesis, regulation of one-carbon metabolism, porphyrin and carnitine synthesis, and conjugation of bile acids and other acyl moieties.⁵⁷ Genetic evidence links circulating glycine levels to ammonia detoxification via the urea cycle, suggesting that glycine depletion may reflect competitive inhibition of its synthesis by upregulated amino acid catabolism.⁵⁸ However, genetics also suggest that low glycine levels may reflect progressive insulin resistance

Abbreviations - 2PY: N1-methyl-2-pyridone-5-carboxamide, AAMU: 5-Acetylamino-6-amino-3-methyl uracil, Ala: alanine, Arg: arginine, C: cytidine, carn.: carnitine, DG: diglyceride, DMGV: dimethylguanidino valerate, G: guanine, GAA: guanidoacetic acid, GABA: gamma-aminobutyric acid, Galn: galactosamine, GCA: glycocholate, GCDC: glycochenodeoxycholate, Gln: glutamine, Gly: glycine, GXA: glycodeoxycholate, Ile: isoleucine, Leu: leucine, LPE: lysophosphatidylethanolamine, Met: methionine, MNAM: 1-methylnicotinamide; NAA: N-acetylaspartic acid, PE: phosphatidylethanolamine, Phe: phenylalanine, PLG: plasmalogen, Pro: proline, ROH: retinol, SM: sphingomyelin, TG: triglyceride, Tyr: tyrosine, Val: valine, VB5: pantothenate.

rather than be a primary etiological factor in type 2 diabetes development.⁵⁹ Increasing alanine levels may reflect retrograde nitrogen transport from muscle to the liver (Cori cycle). The interaction of these metabolic processes with hyperinsulinemia is likely bidirectional. The metabolic perturbations impose a constant challenge on the insulin-dependent regulation of glucose homeostasis, and in turn, hyperinsulinemia and progressing insulin resistance lead to further dysregulation of amino acid and lipid metabolism. Many details of the mechanisms and modes of interaction in this vicious cycle remain elusive. However, our findings underpin that repeated metabolomics profiles reflect a progressive deterioration of amino acid and lipid metabolism that ultimately culminates in T2D incidence.

A critical criterion to use biomarkers to monitor disease risk in a disease prevention setting is that biomarker changes indicate changes in subsequent disease risk. For example, the broad clinical application of LDL-cholesterol as a biomarker to monitor the risk of CVD is supported by evidence that lowering LDL levels is related to lower CVD risk.⁶⁰ The past two decades of metabolomics research have established links between baseline lipid and amino acid metabolites and future T2D onset.⁵ However, comprehensive analyses of long-term metabolite changes in relation to T2D risk are lacking. Hence, our study fills a critical gap in the evidence for possible applications of metabolomics profiling as potential biomarkers to monitor T2D risk.

We presented results from the first study with repeated metabolomics profiles after several years and subsequent T2D incidence, nested within a well-characterised prospective cohort. However, our study had limitations. We emphasised results significant after multiple testing-correction in each of the three metabolomics profiles, increasing the likelihood of type II errors. Therefore, we provided all metabolite-T2D associations across all metabolomics profiles in the supplement, facilitating future replication studies. The metabolomics platform that generated our data targeted cationic metabolites, including amino acids and lipid species. Repeated metabolomics profiling of other molecular classes such as metabolites involved in carbohydrate metabolism and broader lipid profiling may reveal additional T2D-associated metabolite changes. Our observational analyses cannot prove causal relationships. Moreover, direct evidence on the tissues of origin and the contributing metabolic processes to plasma metabolite levels is scarce. Our findings' biological interpretation is, therefore, speculative. However, adjusting for baseline metabolite levels should block the influence of unmeasured baseline-confounders in the change analysis, and the association of metabolite changes with subsequent T2D risk constitutes a special case of temporality. Therefore, our study substantially strengthens the evidence base on the molecular mechanisms of T2D aetiology in humans. Our models were

not adjusted for clinical markers of glucose homeostasis. Studies that relate metabolite changes to established T2D pathogenic mechanisms are warranted, examining, for example, the relationship with change in markers of insulin resistance, insulin secretion, blood levels of glucose, HbA_{1c}, HDL-cholesterol, triglycerides, and hepatic enzymes. Our study population was limited to white women. Studies of the T2D risk related to long-term metabolite changes in men and other races and ethnicities are warranted.

The present study provides evidence that long-term changes in several plasma metabolite levels are associated with subsequent T2D risk, independent of the baseline metabolite levels. Additionally, we showed that metabolomics profiles reflect the molecular disease predisposition more than a decade before T2D diagnosis. Our observations indicate that high-throughput metabolic profiling captures progressive perturbations of amino acid and lipid metabolism underlying early T2D aetiology. These findings contribute to the accumulating evidence suggesting metabolomics profiling as a critical tool in T2D risk prevention research. Specifically, our observation that the change in several metabolite levels was associated with subsequent T2D risk suggests potential metabolomics applications for tracking T2D risk change over time. Further research is warranted to elucidate the determinants of the T2D-associated metabolite changes and possible interplay with established T2D risk biomarkers.

Funding

This work was supported by research grants [UM1 CA186107](#), [R01 CA49449](#), [DK112940](#), and [DK119268](#) from the National Institutes of Health. CW was supported by the German Research Foundation's (DFG) individual fellowship [#WI15132/1-1](#) and the [Boston Nutrition Obesity Research Center \(P30 DK46200\)](#). MG-F was supported by the American Diabetes Association grant [#1-18-PMF-029](#). DEH was supported by the National Institutes of Health [T32-CA009001](#). Dr. Li was supported by NIDDK [K99 DK122128](#) and the Boston Nutrition Obesity Research Center (P30 DK46200). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. None of the funding sources played a role in the design, collection, analysis, or interpretation of the data or decision to submit the manuscript for publication.

Data sharing statement

Data described in the manuscript, codebook, and analytic code will not be made publicly available. Information on the procedures to obtain and access data from the Nurses' Health Study are described at <http://www.nurseshealthstudy.org/researchers> (email: nhsaccess@channing.harvard.edu).

Contributors

FBH and QS acquired the financial support for the project leading to this publication. AHE, QS, and FBH were involved in NHS sample and data collection. CD and CBC conducted the laboratory metabolomics measurements and curated the raw metabolomics data. CW and FBH conceptualised the study design and methodology. LL and FBH advised the statistical analysis. MGF supported the data analysis. CW conducted the formal data analysis, designed tables and figures, and drafted the manuscript. All authors contributed to interpreting the results, critically revised the manuscript for important intellectual content, and approved the final version of the manuscript. The corresponding authors attest that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Declaration of Competing Interest

SNB is a scientific consultant for LayerIV for work that is not related to this manuscript. The other authors have no competing interests.

Acknowledgements

We thank the participants and staff of the Nurses' Health Study for their valuable contributions. We thank Dr. Fabian Eichmann for supporting the data visualisation and Fenglei Wang for reviewing the R programs.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2021.103799.

References

- Federation. ID. IDF Diabetes Atlas. Brussels, Belgium 2019. Available from: <https://www.diabetesatlas.org>.
- Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol* 2018;14(2):88–98.
- Vijan S. Type 2 diabetes. *Ann Intern Med* 2019;171(9):1tc65–80.
- Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. *Lancet* 2017;389(10085):2239–51.
- Guasch-Ferré M, Hruby A, Toledo E, Clish CB, Martínez-González MA, Salas-Salvadó J, et al. Metabolomics in prediabetes and diabetes: a systematic review and meta-analysis. *Diabetes Care* 2016;39(5):833–46.
- Lotta LA, Scott RA, Sharp SJ, Burgess S, Luan Ja, Tillin T, et al. Genetic predisposition to an impaired metabolism of the branched-chain amino acids and risk of type 2 diabetes: a Mendelian randomisation analysis. *PLoS Med* 2016;13(11):e1002179.
- Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, 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(4):311–26.
- Arany Z, Neinst M. Branched chain amino acids in metabolic disease. *Curr Diabetes Rep* 2018;18(10):76.
- Merino J, Leong A, Liu CT, Porneala B, Walford GA, von Grotthuss M, et al. Metabolomics insights into early type 2 diabetes pathogenesis and detection in individuals with normal fasting glucose. *Diabetologia* 2018;61(6):1315–24.
- Razquin C, Toledo E, Clish CB, Ruiz-Canela M, Dennis C, Corella D, et al. Plasma lipidomic profiling and risk of type 2 diabetes in the PREDIMED trial. *Diabetes Care* 2018;41(12):2617–24.
- Guasch-Ferré M, Ruiz-Canela M, Li J, Zheng Y, Bulló M, Wang DD, et al. Plasma acylcarnitines and risk of type 2 diabetes in a mediterranean population at high cardiovascular risk. *J Clin Endocrinol Metab* 2019;104(5):1508–19.
- Guasch-Ferré M, Santos JL, Martínez-González MA, Clish CB, Razquin C, Wang D, et al. Glycolysis/gluconeogenesis- and tricarboxylic acid cycle-related metabolites, mediterranean diet, and type 2 diabetes. *Am J Clin Nutr* 2020;111(4):835–44.
- Ruiz-Canela M, Guasch-Ferré M, Toledo E, Clish CB, Razquin C, Liang L, et al. Plasma branched chain/aromatic amino acids, enriched Mediterranean diet and risk of type 2 diabetes: case-cohort study within the PREDIMED Trial. *Diabetologia* 2018;61(7):1560–71.
- Bao Y, Bertola ML, Lenart EB, Stampfer MJ, Willett WC, Speizer FE, et al. Origin, methods, and evolution of the three Nurses' Health Studies. *Am J Public Health* 2016;106(9):1573–81.
- Zhang X, Tworoger SS, Eliassen AH, Hankinson SE. Postmenopausal plasma sex hormone levels and breast cancer risk over 20 years of follow-up. *Breast Cancer Res Treat* 2013;137(3):883–92.
- Townsend MK, Clish CB, Kraft P, Wu C, Souza AL, Deik AA, et al. Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. *Clinical chemistry* 2013;59(11):1657–67.
- Chiuvè SE, Fung TT, Rimm EB, Hu FB, McCullough ML, Wang M, et al. Alternative dietary indices both strongly predict risk of chronic disease. *J Nutr* 2012;142(6):1009–18.
- Al-Delaimy WK, Willett WC. Measurement of tobacco smoke exposure: comparison of toenail nicotine biomarkers and self-reports. *Cancer Epidemiol Biomarkers Prev* 2008;17(5):1255–61.
- Giovannucci E, Colditz G, Stampfer MJ, Rimm EB, Litin L, Sampson L, et al. The assessment of alcohol consumption by a simple self-administered questionnaire. *Am J Epidemiol* 1991;133(8):810–7.
- Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985;122(1):51–65.
- Wolf AM, Hunter DJ, Colditz GA, Manson JE, Stampfer MJ, Corsano KA, et al. Reproducibility and validity of a self-administered physical activity questionnaire. *Int J Epidemiol* 1994;23(5):991–9.
- Rimm EB, Stampfer MJ, Colditz GA, Chute CG, Litin LB, Willett WC. Validity of self-reported waist and hip circumferences in men and women. *Epidemiology* 1990;1(6):466–73.
- Hu FB, Manson JE, Stampfer MJ, Colditz G, Liu S, Solomon CG, et al. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N Engl J Med* 2001;345(11):790–7.
- Manson JE, Rimm EB, Stampfer MJ, Colditz GA, Willett WC, Krolewski AS, et al. Physical activity and incidence of non-insulin-dependent diabetes mellitus in women. *Lancet* 1991;338(8770):774–8.
- Kim S. ppcor: an R package for a fast calculation to semi-partial correlation coefficients. *Commun Stat Appl Methods* 2015;22(6):665.
- Maathuis MH, Colombo D, Kalisch M, Bühlmann P. Predicting causal effects in large-scale systems from observational data. *Nat Methods* 2010;7(4):247–8.
- Iqbal K, Dietrich S, Wittenbecher C, Krumsiek J, Kühn T, Lacruz ME, et al. Comparison of metabolite networks from four German population-based studies. *Int J Epidemiol* 2018;47(6):2070–81.
- Wittenbecher C. Linking whole-grain bread, coffee, and red meat to the risk of type 2 diabetes [Monograph]. Potsdam: Potsdam University; 2017.
- Hertz-Picciotto I, Rockhill B. Validity and efficiency of approximation methods for tied survival times in Cox regression. *Biometrics* 1997;53(3):1151–6.
- Benjamini Y, Hochberg Y. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J Educ Behav Stat* 2000;25(1):60–83.
- Stancáková A, Civelek M, Saleem NK, Soininen P, Kangas AJ, Cederberg H, et al. Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 Finnish men. *Diabetes* 2012;61(7):1895–902.
- Ferrannini E, Natali A, Camastra S, Nannipieri M, Mari A, Adam KP, et al. Early metabolic markers of the development of dysglycemia and type 2 diabetes and their physiological significance. *Diabetes* 2013;62(5):1730–7.
- Tillin T, Hughes AD, Wang Q, Würtz P, Ala-Korpela M, Sattar N, et al. Diabetes risk and amino acid profiles: cross-sectional and prospective analyses of ethnicity, amino acids and diabetes in a South

- Asian and European cohort from the SABRE (Southall And Brent Revisited) Study. *Diabetologia* 2015;58(5):968–79.
- 34 Palmer ND, Stevens RD, Antinozzi PA, Anderson A, Bergman RN, Wagenknecht LE, et al. Metabolomic profile associated with insulin resistance and conversion to diabetes in the insulin resistance atherosclerosis study. *J Clin Endocrinol Metab* 2015;100(3):E463–8.
- 35 Ahola-Olli AV, Mustelin L, Kalimeri M, Kettunen J, Jokelainen J, Auvinen J, et al. Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. *Diabetologia* 2019;62(12):2298–309.
- 36 Lu Y, Wang Y, Liang X, Zou L, Ong CN, Yuan J-M, et al. Serum amino acids in association with prevalent and incident type 2 diabetes in a Chinese population. *Metabolites* 2019;9(1):14.
- 37 White PJ, McGarrah RW, Grimsrud PA, Tso SC, Yang WH, Halde- man JM, et al. The BCKDH kinase and phosphatase integrate BCAA and lipid metabolism via regulation of ATP-citrate lyase. *Cell Metab* 2018;27(6):1281–93.e7.
- 38 Guasch-Ferré M, Santos JL, Martínez-González MA, Clish CB, Raz- quin C, Wang D, et al. Glycolysis/gluconeogenesis- and tricarboxylic acid cycle-related metabolites, Mediterranean diet, and type 2 diabetes. *Am J Clin Nutrition* 2020;111(4):835–44.
- 39 Wang TJ, Larson MG, Vasani RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011;17(4):448–53.
- 40 Floegel A, Stefan N, Yu Z, Mühlenbruch K, Drogan D, Joost HG, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* 2013;62(2):639–48.
- 41 Gängler S, Waldenberger M, Artati A, Adamski J, Van Bolhuis JN, Sørgerød EP, et al. Exposure to disinfection byproducts and risk of type 2 diabetes: a nested case–control study in the HUNT and life- lines cohorts. *Metabolomics* 2019;15(4):60.
- 42 Seppälä I, Oksala N, Jula A, Kangas AJ, Soininen P, Hutri-Kähönen N, et al. The biomarker and causal roles of homoarginine in the development of cardiometabolic diseases: an observational and Mendelian randomization analysis. *Sci Rep* 2017;7(1):1130.
- 43 Chen Z-Z, Liu J, Mormingstar J, Heckman-Stoddard BM, Lee CG, Dagogo-Jack S, et al. Metabolite profiles of incident diabetes and heterogeneity of treatment effect in the diabetes prevention program. *Diabetes* 2019;68(12):2337–49.
- 44 Guasch-Ferré M, Ruiz-Canela M, Li J, Zheng Y, Bulló M, Wang DD, et al. Plasma acylcarnitines and risk of type 2 diabetes in a Mediter- ranean population at high cardiovascular risk. *J Clin Endocrinol Metab* 2019;104(5):1508–19.
- 45 Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J Clin Invest* 2011;121(4):1402–11.
- 46 De Mello VD, Paananen J, Lindström J, Lankinen MA, Shi L, Kuusisto J, et al. Indolepropionic acid and novel lipid metabolites are associated with a lower risk of type 2 diabetes in the Finnish diabetes prevention study. *Sci Rep* 2017;7:46337.
- 47 Fall T, Salihovic S, Brandmaier S, Nowak C, Ganna A, Gustafsson S, et al. Non-targeted metabolomics combined with genetic analyses identifies bile acid synthesis and phospholipid metabolism as being associated with incident type 2 diabetes. *Diabetologia* 2016;59(10):2114–24.
- 48 Lu J, Wang S, Li M, Gao Z, Xu Y, Zhao X, et al. Association of serum bile acids profile and pathway dysregulation with the risk of developing diabetes among normoglycemic Chinese adults: find- ings from the 4C study. *Diabetes Care* 2020;dc200884.
- 49 Kim M, Jee SH, Kim M, Yoo HJ, Kang M, Kim J, et al. Serum vita- min A-related metabolite levels are associated with incidence of type 2 diabetes. *Diabetes Metab* 2017;43(3):287–91.
- 50 Blondin SA, Yeung EH, Mumford SL, Zhang C, Browne RW, Wac- tawski-Wende J, et al. Serum retinol and carotenoids in association with biomarkers of insulin resistance among premenopausal women. *ISRN Nutr* 2013;2013:619516.
- 51 Guasch-Ferré M, Hernández-Alonso P, Drouin-Chartier JP, Ruiz- Canela M, Razquin C, Toledo E, et al. Walnut consumption, plasma metabolomics, and risk of type 2 diabetes and cardiovascular dis- ease. *J Nutr* 2021;151(2):303–11.
- 52 Tobias DK, Clish C, Mora S, Li J, Liang L, Hu FB, et al. Dietary intakes and circulating concentrations of branched-chain amino acids in relation to incident type 2 diabetes risk among high-risk women with a history of gestational diabetes mellitus. *Clin Chem* 2018;64(8):1203–10.
- 53 Zheng Y, Li Y, Qi Q, Hruby A, Manson JE, Willett WC, et al. Cumu- lative consumption of branched-chain amino acids and incidence of type 2 diabetes. *Int J Epidemiol* 2016;45(5):1482–92.
- 54 White PJ, Lapworth AL, An J, Wang L, McGarrah RW, Stevens RD, et al. Branched-chain amino acid restriction in Zucker-fatty rats improves muscle insulin sensitivity by enhancing efficiency of fatty acid oxidation and acyl-glycine export. *Mol Metab* 2016;5(7):538–51.
- 55 Jang C, Oh SF, Wada S, Rowe GC, Liu L, Chan MC, et al. A branched-chain amino acid metabolite drives vascular fatty acid transport and causes insulin resistance. *Nat Med* 2016;22(4):421–6.
- 56 Newgard CB. Interplay between lipids and branched-chain amino acids in development of insulin resistance. *Cell Metab* 2012: 15.
- 57 Adeva-Andany M, Souto-Adeva G, Ameneiros-Rodríguez E, Fernán- dez-Fernández C, Donapetry-García C, Domínguez-Montero A. Insulin resistance and glycine metabolism in humans. *Amino Acids* 2018;50(1):11–27.
- 58 Alves A, Bassot A, Bulteau AL, Pirola L, Morio B. Glycine metabo- lism and its alterations in obesity and metabolic diseases. *Nutrients* 2019;11(6).
- 59 Wittemans LBL, Lotta LA, Oliver-Williams C, Stewart ID, Surendran P, Karthikeyan S, et al. Assessing the causal association of glycine with risk of cardio-metabolic diseases. *Nat Commun* 2019;10(1):1060.
- 60 Silverman MG, Ference BA, Im K, Wiviott SD, Giugliano RP, Grundy SM, et al. Association between lowering LDL-C and cardio- vascular risk reduction among different therapeutic interventions: a systematic review and meta-analysis. *JAMA* 2016;316(12):1289–97.