# Plasma Lipid Metabolites, Clinical Glycemic Predictors, and Incident Type 2 Diabetes

Arjana Begzati, Karla P. Godinez-Macias, Tao Long, Jeramie D. Watrous, Rafael Moranchel, Edward D. Kantz, Jaakko Tuomilehto, Aki S. Havulinna, Teemu J. Niiranen, Pekka Jousilahti, Veikko Salomaa, Bing Yu, Faye Norby, Casey M. Rebholz, Elizabeth Selvin, Elizabeth A. Winzeler, Susan Cheng, Mona Alotaibi, Ravi Goyal, Trey Ideker, Mohit Jain, and Amit R. Majithia
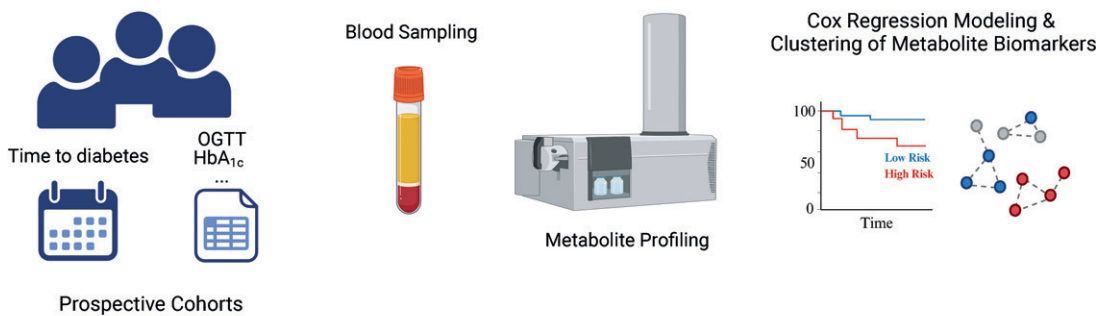
- Diacylglycerols (DAGs) capture type 2 diabetes (T2D) risk information related to 2-hour post-challenge glucose.
- Phosphatidylcholines (PCs) predict incident T2D risk independent of glycemic markers and insulin.
- No significant interaction between race and DAG or PC biomarkers was found.

ARTICLE HIGHLIGHTS

- **Why did we undertake this study?**
  Analysis of metabolite biomarkers of incident type 2 diabetes (T2D) and their relationship to specific etiologies across diverse populations is lacking.

- **What is the specific questions we wanted to answer?**
  Do metabolite biomarkers of T2D risk capture information about specific aspects of T2D etiology, and does this association vary with race?

- **What did we find?**
  Diacylglycerols (DAGs) captured T2D risk information related to 2-h postchallenge glucose. Phosphatidylcholines (PCs) are incident T2D biomarkers independent of glycemic markers and insulin. No significant interaction between race and DAG or PC biomarkers was found.

- **What are the implications of our findings?**
  DAGs and PCs, respectively, capture T2D risk information related to and distinct from conventional glycemic markers. They have potential to be used in ancestrally diverse populations.

Plasma Lipid Metabolites, Clinical Glycemic Predictors, and Incident Type 2 Diabetes

Arjana Begzati,[1] Karla P. Godinez-Macias,[2] Tao Long,[1,3] Jeramie D. Watrous,[1,3] Rafael Moranchel,[1,3] Edward D. Kantz,[1] Jaakko Tuomilehto,[4,5] Aki S. Havulinna,[4,6,7] Teemu J. Niiranen,[4,8,9] Pekka Jousilahti,[4] Veikko Salomaa,[4] Bing Yu,[10] Faye Norby,[11] Casey M. Rebholz,[12] Elizabeth Selvin,[12] Elizabeth A. Winzeler,[2] Susan Cheng,[13] Mona Alotaibi,[1] Ravi Goyal,[1] Trey Ideker,[1] Mohit Jain,[1,3] and Amit R. Majithia[1]

OBJECTIVE

Plasma metabolite profiling has uncovered several nonglycemic markers of incident type 2 diabetes (T2D). We investigated whether such biomarkers provide information about specific aspects of T2D etiology, such as impaired fasting glucose and impaired glucose tolerance, and whether their association with T2D risk varies by race.

RESEARCH DESIGN AND METHODS

Untargeted plasma metabolite profiling was performed of participants in the FINRISK 2002 cohort (n = 7,564). Cox regression modeling was conducted to identify metabolites associated with incident T2D during 14 years of follow-up. Metabolites were clustered into pathways using Gaussian graphical modeling. Clusters enriched for T2D biomarkers were further examined for covariation with fasting plasma glucose (FPG), 2-h postchallenge plasma glucose (2hPG), HbA$_{1c}$, or fasting insulin. Validation analyses and tests of interaction with race were performed in the Atherosclerosis Risk in Communities (ARIC) study.

RESULTS

Two clusters of metabolites, representing diacylglycerols (DAGs) and phosphatidylcholines (PCs), contained the largest number of metabolite associations with incident T2D. DAGs associated with increased T2D incidence (hazard ratio [HR] 1.22; 95% CI 1.14–1.30) independent of FPG, HbA$_{1c}$, and fasting insulin, but not 2hPG. PCs were inversely associated with T2D risk (HR 0.78; 95% CI 0.71–0.85) independent of FPG, 2hPG, HbA$_{1c}$, and fasting insulin. No significant interaction between DAGs or PCs and race was observed.

CONCLUSIONS

Fasting DAGs may capture information regarding T2D risk similar to that represented by 2hPG; PCs may capture aspects of T2D etiology that differ from those represented by conventional biomarkers. The direction of effect and strength of DAG and PC associations with incident T2D are similar across European and African Americans.

[1]Department of Medicine, University of California San Diego, La Jolla, CA

[2]Department of Pediatrics, University of California San Diego, La Jolla, CA

[3]Sapient Bioanalytics, San Diego, CA

[4]Department of Public Health, Finnish Institute for Health and Welfare, Helsinki, Finland

[5]Department of Public Health, University of Helsinki, Helsinki, Finland

[6]Department of Computing, University of Turku, Turku, Finland

[7]Institute for Molecular Medicine Finland, FIMM-HiLIFE, Helsinki, Finland

[8]Division of Medicine, Turku University Hospital, Turku, Finland

[9]Department of Internal Medicine, University of Turku, Turku, Finland

[10]Department of Epidemiology, School of Public Health, University of Texas School of Public Health, Houston, TX

[11]Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN

[12]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

[13]Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA

Corresponding author: Amit R. Majithia, amajithia@ucsd.edu

Type 2 diabetes (T2D) currently affects 500 million adults worldwide and its prevalence is increasing globally (1). Prevention of T2D can be partially achieved by premorbid medical and lifestyle interventions, although these require intensive patient and provider effort (2,3). Furthermore, there is a growing appreciation of heterogeneity in disease manifestation and complications for T2D that are not captured by

conventional glycemic risk factors alone (4,5). Thus, targeting individuals for prevention and individualizing treatment strategies require biomarkers that capture aspects of molecular etiology distinct from commonly used biomarkers. To some extent, dynamic tests such as the oral glucose tolerance tests (OGTTs) can start to distinguish T2D pathophysiology (6). For example, the fasting plasma glucose (FPG) level largely represents the gluconeogenic output of the liver (7), whereas the 2-h postchallenge plasma glucose (2hPG) level represents skeletal muscle insulin sensitivity, because skeletal muscle is the major glucose disposal organ (8). However, OGTTs are not commonly used in the clinic, because they require lengthy patient visits and multiple blood samples.

Additionally, a recently recognized but clinically important source of heterogeneity is race/genetic ancestry. Glycated hemoglobin A_{1c} (HbA$_{1c}$) has long been used as a marker of glycemic control and as a prognosticator of micro- and macrovascular complications (9). Clinical HbA$_{1c}$ cutoff values for treatment intensification were applied across all populations, extrapolating from landmark studies performed primarily with White participants (9). With the availability of frequently sampled glucose by continuous glucose monitoring, it was found that HbA$_{1c}$ values overestimate glucose in African American individuals, potentially leading to overdiagnosis and unnecessary treatment (10). These findings highlight the importance of evaluating race/ancestry in any biomarker discovery effort.

Given that T2D is a disorder of metabolism, metabolites are an important class of molecules from which novel etiologic markers might be identified. With advances in high-throughput metabolomics, it is now possible to measure thousands of metabolites in blood, enabling the identification of novel chemical classes of T2D risk predictors beyond the standard metabolite glucose and its derivatives (11). Identified metabolite predictors of T2D include branched-chain amino acids (12), short- and long-chain acylcarnitines (13), acylglycerols (including diacylglycerols [DAGs]), and phospholipids (including phosphatidylcholines [PCs]) (14). Whether these metabolite classes capture the same aspects of T2D etiology as conventional glycemic markers remains largely unknown.

In this study, we evaluated to what extent metabolite biomarkers of different chemical classes may provide information on T2D risk related to known and potentially novel aspects of T2D etiology. We also determined whether any identified T2D risk association varied between populations of European and African ancestry. Our findings highlight two classes: circulating DAGs and PCs. We find that DAGs capture T2D risk information that is related to 2hPG level, and PCs capture T2D risk information that is independent of glycemic markers and insulin. Furthermore, we observe that both DAGs and PCs have similar T2D risk association in European American (EA) and African American (AA) individuals.

RESEARCH DESIGN AND METHODS

Study Design and Participants
The FINRISK is a population-based prospective study whose main objective is to monitor noncommunicable disease risk factors, health behavior, and their changes in the Finnish population. The FINRISK 2002 cohort is composed of participants aged 25–74 years from Finland. Participants completed questionnaires, underwent physical examination with blood sampling, and were followed up for 14 years through computerized record linkage. In addition, an OGTT was carried out by a subset of participants (aged 44 years or older) after a 12-h fast, as described previously (15). Ascertainment of the T2D cases at baseline and incident cases during follow-up were made using the participants' Finnish personal identification number and information stored in the National Hospital Discharge Register (NHDR), Causes of Death Register (CDR), and the Drug Reimbursement and Drug Purchase Registers (DPR). The ICD codes E10-E14 (ICD-10) or 250*B (ICD-9) were used to extract diabetes diagnoses from the NHDR and CDR. The Anatomical Therapeutic Chemical code A10 or a special reimbursement code for diabetes medications was used to extract diabetes status from the DPR. If medicine purchase was the only available evidence for diabetes diagnosis, three or more purchases were required. Furthermore, to exclude type 1 diabetes cases, participants younger than 30 years who were treated with insulin, or insulin and metformin, and those aged 30–40 years when insulin only was first used, were not categorized as having T2D. To identify individuals who were taking a lipid medication, Anatomical Therapeutic Chemical code C10 was used. Details on the FINRISK study have been described previously (16).

For validation and assessment of a potential race interaction with selected biomarkers, we repeated analyses in the Atherosclerosis Risk in Communities (ARIC) study wherein an overlapping set of metabolites was profiled using a bioactive lipids liquid chromatography-mass spectrometry (LC-MS) profiling method. ARIC is a prospective cohort designed to investigate the causes of atherosclerosis and related outcomes by race, sex, location, and date. Participants aged 45–65 years at baseline were recruited from four U.S. communities (Forsyth County, NC; Jackson, MS; suburbs of Minneapolis, MN; and Washington County, MD) in 1987–1989 and followed up to date. Details on the ARIC study have been previously described (17,18). T2D was defined by the following criteria: FPG ≥126 mg/dL, HbA$_{1c}$ >6.5%, taking glucose-lowering medication, or self-reported diagnosis of diabetes by a doctor.

Plasma Metabolite Profiling
Metabolite signatures in the FINRISK cohort were measured using untargeted LC-MS, as described in the Supplementary Material. Chromatographic peaks were extracted using an image-based deep neural network algorithm (19). Raw intensity values were batch corrected and normalized by multiplying by the cohort median peak intensity and dividing by the plate median peak intensity. Metabolite signatures missing in >50% of FINRISK samples (n = 246) were filtered out; for the remaining metabolite signatures (n = 31,701), missing values (i.e., metabolites whose levels were below the detection limit in certain samples) were imputed by sampling from the uniform distribution of the minimum value of each metabolite and dividing the minimum value by 6. Metabolite signatures were systematically deduplicated to distinguish unique metabolites by leveraging expected charge to mass ratio (m/z) differences for isotopes and adducts, and requiring coelution (i.e., the same retention time), similar chromatographic peak shape, and strong correlation in intensity (Pearson correlation value ≥ 0.75) for a signature to be flagged as a duplicate of another signature. Metabolite signatures were annotated using m/z matching to entries in the Human Metabolome Database (version 5.0) (20).

For at least one metabolite signature within each network cluster, the annotation was confirmed using a commercial standard or diagnostic tandem mass spectrometry fragmentation. Metabolites of interest (i.e., DAG and PC lipid species) were measured in the ARIC cohort using LC-MS of bioactive lipids, as described in the Supplementary Material, and manually mapped and extracted from the LC-MS data using m/z, retention time and chromatographic peak shape characteristics.

Statistical Analysis
Cox Regression Analysis
Outlier metabolite intensity values were capped using the 99th percentile of each metabolite. The data were scaled by subtracting the population mean and dividing by the population SD for each metabolite as follows:

$$z = (metabolite\ intensity$$
$$- population\ mean)$$
$$/ population\ standard\ deviation$$

where z is the z score and the value that was used for all downstream analysis, with the exception of the race interaction analysis in ARIC for which metabolite intensity values without scaling were used. To identify metabolites that associate with T2D incidence risk, we performed Cox proportional hazard regressions adjusting for age, sex, and BMI (the basic model), using the function $coxph()$ from the R package $survival$ (21). This analysis was carried out for each of the measured metabolite signatures in FINRISK and prioritized lipid metabolites in ARIC separately. Metabolite signatures from the analysis in FINRISK with P value below the metabolome-wide significance level of 10^{-6} were considered statistically significant and are referred to as biomarkers. For subsequent analyses, Bonferroni correction was used to set the P value threshold.

In a secondary analysis for selected biomarkers, we tested the metabolites for association with T2D incidence when adjusting for FPG, 2hPG, HbA_{1c}, and fasting serum insulin values. The proportional hazards assumption for biomarkers of interest was verified by plotting Schoenfeld residuals. Survival curves were plotted using the function $ggsurvplot()$ from the R package $survminer$ (22).

Gaussian Graphical Modeling Network Creation
Of 31,701 detected metabolite signatures, we filtered out those that were not statistically significant biomarkers of incident T2D ($P > 10^{-6}$) and either present at low intensity in >50% of the samples or were predicted to be isotopes or adducts of other signatures. This resulted in 12,237 metabolite signatures. Unlike for the Cox regression analysis, the levels of these metabolite signatures were first log-transformed before z transformation. To cluster metabolites by chemical pathway, we applied Gaussian graphical modeling (GGM) and estimated partial correlations between metabolite signature pairs (23). The GGM network was created using the R package $GeneNet$ (24). In short, this tool uses a shrinkage-based regularization approach to create a stable estimate of the partial correlation matrix and a mixture model to generate P values for each partial correlation value. We constructed a network with metabolites as nodes, and edges between them were selected from the partial correlation matrix such that any edge had a probability >0.9 (i.e., local false discovery rate < 0.1). Last, we performed cluster detection on the network using the Louvain algorithm implemented in $NetworkX$ (25), with the absolute value of the partial correlations as weights and resolution set to 11. Networks were visualized using $Cytoscape$ (26).

Ethics Approval
The FINRISK 2002 study was approved by the Ethical Committee for Epidemiology and Public Health of the Helsinki University Hospital District on 19 December 2001 (Ref 558/E3/2001). The ARIC study was approved by the institutional review boards of participating institutions and all participants provided written informed consent.

Data and Resource Availability
The data sets generated during and/or analyzed during this study are not publicly available, because they contain protected health information on study participants. Data Tables 1, 2, and 3 are available at https://docs.google.com/spreadsheets/d/e/2PACX-1vRqh2W9AJeVTDNF9wjjmUcnAX20NSodZ-zLAiIN-CRhOYbNGQb52dAHxw8zlqbmGFQ/pubhtml. Additional de-identified data are available from the corresponding author upon reasonable request and with permission of the Finnish Institute for Health and Welfare and the ARIC Coordinating Center.

RESULTS
Cohort Characteristics
From 8,014 participants included in FINRISK 2002, plasma was collected and metabolites were profiled using untargeted LC-MS (Fig. 1A). For this study, we analyzed the plasma metabolome of 7,564 participants free of known T2D at baseline. Of these, 47% of the participants were men. The mean age was 48 years and the mean BMI was 27 kg/m^2. Through a mean of 14 years of follow-up, 656 T2D incident cases were recorded (Supplementary Table 1, Supplementary Fig. 1).

Bioactive lipid data from fasting plasma samples (i.e., ≥8 h self-reported fasting time) from visit 2 of ARIC (1990–1992) from 7,562 participants with nonprevalent T2D were analyzed. The mean age of these participants was 57 years; 44% of the participants were men; and their mean BMI was 27 kg/m^2. Over a mean follow-up time of 12 years (up to visit 5 of the study), 1,211 individuals developed T2D. In the cohort, 6,145 individuals self-identified as EA and 1,417 as AA (Supplementary Table 2).

Identification of Metabolite Biomarkers of Incident T2D
Of the 31,701 metabolite signatures measured in the FINRISK cohort, 171 signatures were metabolome-wide, statistically significant ($P < 10^{-6}$) biomarkers for T2D incidence in a Cox regression model adjusted for age, sex, and BMI (Fig. 1B, Data Table 1). The biomarkers span a wide range of chemical classes, including polar small molecules, polar lipids, and nonpolar lipids (Fig. 1C). We observed a number of known and previously reported predictors of increased T2D risk, including glucose (hazard ratio [HR] 1.29; 95% CI 1.20–1.39), lactate (HR 1.22; 95% CI 1.15–1.30) (27), and several amino acids, including glutamate (HR 1.42; 95% CI 1.34–1.51) (28), threonine (HR 1.25; 95% CI 1.19–1.31), proline (HR 1.26; 95% CI 1.17–1.35) (29), and arginine (HR 1.19; 95% CI 1.11–1.27) (30). We also detected the known inverse risk marker linoleoylglycerophosphocholine (HR 0.77; 95% CI 0.70–0.85) (31) (Data Table 1).
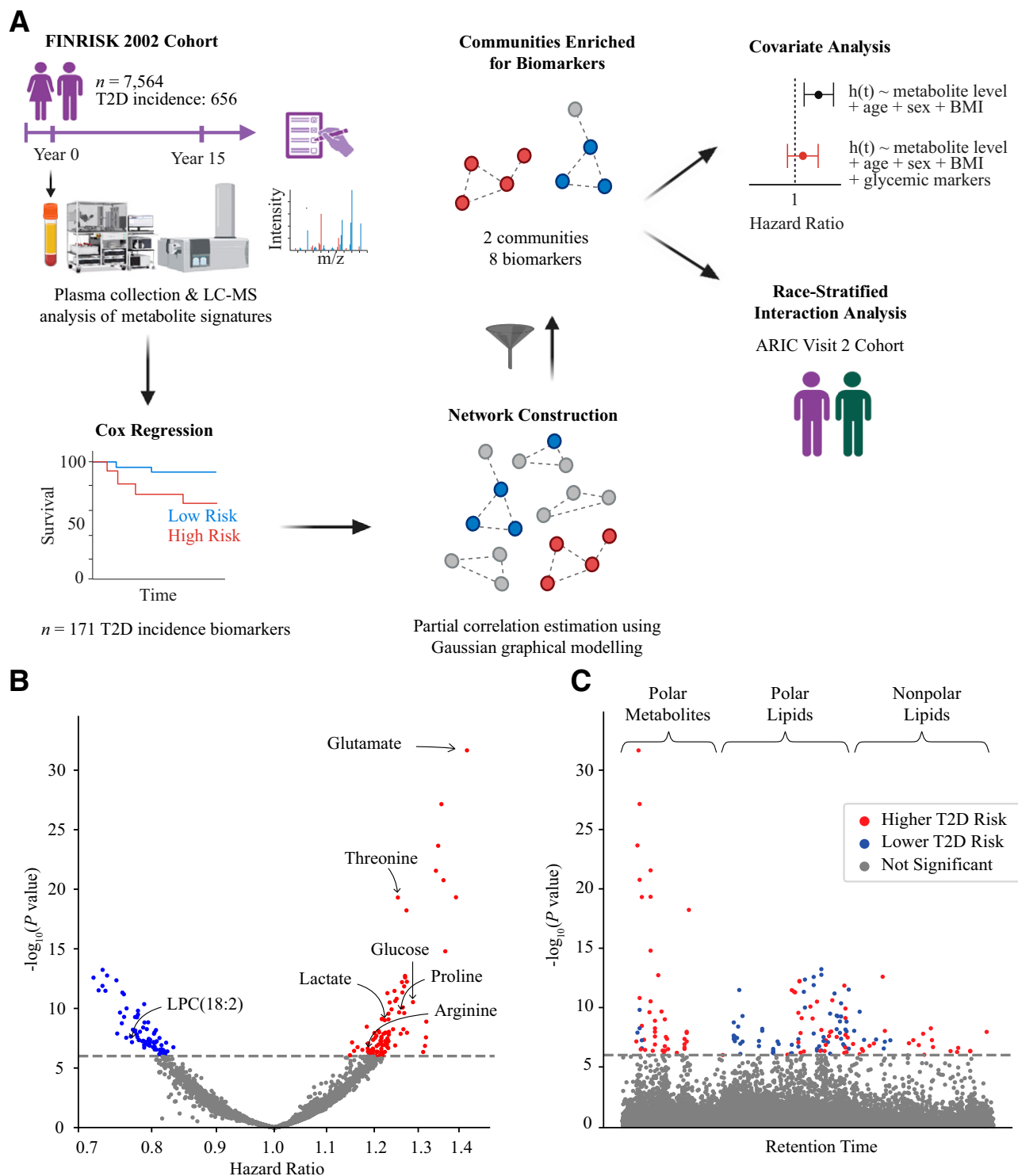
Figure 1—Study overview and identification of metabolite biomarkers of incident T2D. *A*: Analysis overview showing metabolite profiling in FINRISK; Cox regression modeling of T2D risk; network construction using all measured metabolites; covariate analysis in FINRISK; and race-stratified interaction analysis in ARIC on selected metabolite biomarkers. *B*: Metabolite signatures significantly associated with increased (red) or decreased (blue) risk for incident T2D in FINRISK, as determined by Cox regression modeling with adjustment for age, sex, and BMI. Previously reported incident T2D predictors are labeled. HRs are per 1 SD increase in metabolite level and plotted on a logarithmic scale. *C*: Metabolite signatures ordered by chromatographic retention time, highlighting the chemical diversity of the incident T2D biomarkers. Dashed line: the metabolome-wide significance threshold ($P < 10^{-6}$).

Clustering of T2D Risk Biomarkers

To evaluate if these T2D risk biomarkers clustered into metabolic pathways, we clustered all measured metabolites using GGM (23). Of the 171 identified T2D biomarkers metabolites, 119 were connected to other metabolites in the GGM network. We applied cluster detection to disambiguate the entire GGM network into
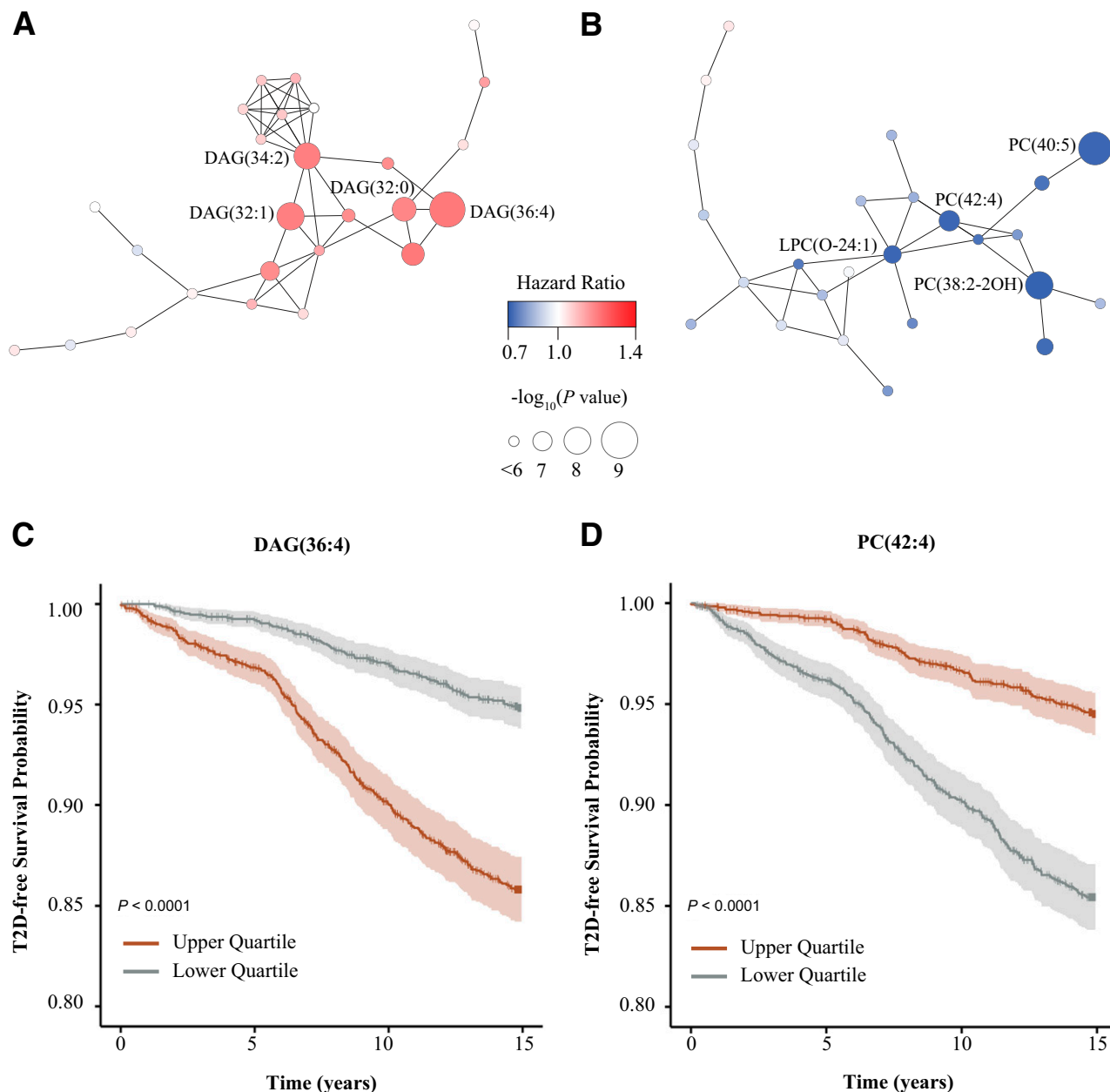
Figure 2—Clusters enriched for T2D biomarkers from the GMM network of measured metabolite signatures. *A*: Cluster of DAGs associated with increased risk for incident T2D. *B*: Cluster of PCs associated with decreased risk for incident T2D. Edges represent statistically significant partial correlations between metabolite signatures. Kaplan-Meier curves for samples with DAG(36:4) (*C*) and PC(42:4) (*D*) levels in the upper vs. lower quartile. *P* values for the Kaplan-Meier curves were determined using the log-rank test.

657 clusters of chemically related metabolites. The 119 T2D risk biomarkers were distributed among 44 clusters (Supplementary Fig. 2, Data Table 2). We focused attention on the DAG and PC clusters, which composed the top two clusters that contained the most unique T2D biomarkers (*n* = 4 DAGs and *n* = 4 PCs; Fig. 2*A* and *B*).

Four unique metabolites in the DAG cluster—DAG(32:0), DAG(32:1), DAG(34:2), and DAG(36:4)—were associated with increased T2D risk, with HRs ranging between 1.20 and 1.22 (Fig. 2*C*, Table 1,

Supplementary Fig. 3). The cluster composed of PCs associated with decreased T2D incidence with HRs ranging between 0.78 and 0.79 (Fig. 2*D*). The specific PC species we found include LPC(O–24:1), PC(38:2-2OH), PC acyl-alkyl (40:5), and PC(42:4). To our knowledge, LPC(O–24:1) and PC(38:2-2OH) have not been linked to T2D incidence previously. These lipid biomarkers replicated in the independent ARIC cohort at a multiple hypothesis corrected significance threshold ($P < 6 \times 10^{-3}$ [0.05/8]) (Table 1).

Relationship of DAGs and PCs With Conventional T2D Biomarkers

To assess whether these DAG and PC biomarkers provide information about specific aspects of T2D etiology, we evaluated the association of DAGs and PCs with incident T2D in models corrected for conventional T2D risk markers. We recomputed HR estimates for the DAGs and PCs in Cox regression models adjusted for age, sex, BMI, and FPG, 2hPG, HbA_{1c}, or fasting serum insulin levels. The association of DAGs with T2D incidence remained significant in

Table 1—Cox regression statistics for DAG and PC biomarkers with incident T2D

Biomarker	FINRISK HR	FINRISK P value	ARIC HR	ARIC P value	Biomarker-race interaction P value
DAG(32:0)	1.20	2.00×10^{-8}	1.12	2.66×10^{-5}	0.5
DAG(32:1)	1.21	7.62×10^{-9}	1.18	1.67×10^{-12}	0.8
DAG(34:2)	1.21	1.03×10^{-8}	1.20	8.95×10^{-15}	0.6
DAG(36:4)	1.22	7.65×10^{-10}	1.16	8.23×10^{-11}	0.8
LPC(O–24:1)	0.79	1.17×10^{-7}	0.78	2.44×10^{-13}	0.7
PC(38:2-2OH)	0.78	7.06×10^{-9}	0.81	6.69×10^{-11}	0.3
PC(40:5)	0.79	1.52×10^{-9}	0.78	2.00×10^{-15}	0.4
PC(42:4)	0.79	5.22×10^{-8}	0.78	1.05×10^{-12}	0.8

models adjusted for FPG, HbA$_{1c}$, or fasting insulin levels but was attenuated when adjusted for 2hPG value (Fig. 3A, Data Table 3). The species from the PC cluster remained significant markers of T2D incidence when adjusted for FPG, 2hPG, HbA$_{1c}$, or fasting insulin concentration. Notably, their association was partially attenuated by the adjustment for 2hPG concentration, especially for PC(38:2-2OH) (Fig. 3B).

We further examined if the DAG and PC associations with T2D incidence were related to serum lipid measurements, because DAGs especially are in the biosynthetic pathway for triacylglycerols (TAGs; commonly measured in serum triglycerides). Again, we computed HR estimates for DAGs and PCs with incident T2D in Cox regression models, this time including serum TAGs, HDL cholesterol, or total cholesterol

as covariates (Supplementary Table 3). We found that adjustment for HDL or total cholesterol had minimal effect on the hazard estimates for DAGs and PCs. Correction for TAGs attenuated the strength of association of DAGs with incident T2D, but three of four DAG species remained significant. Finally, we found that use of lipid-lowering medication was associated with increases in plasma levels of two of
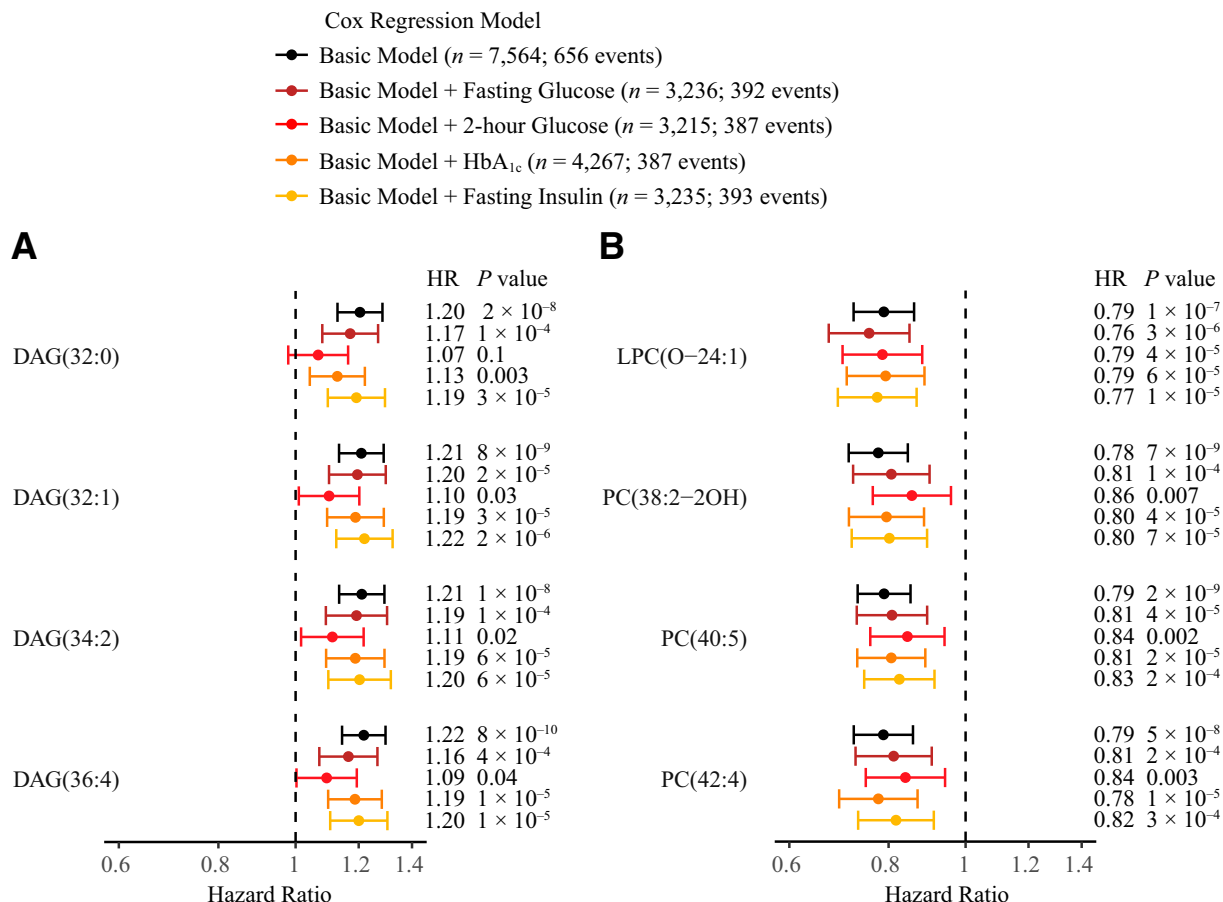

Cox Regression Model
- Basic Model (n = 7,564; 656 events)
- Basic Model + Fasting Glucose (n = 3,236; 392 events)
- Basic Model + 2-hour Glucose (n = 3,215; 387 events)
- Basic Model + HbA$_{1c}$ (n = 4,267; 387 events)
- Basic Model + Fasting Insulin (n = 3,235; 393 events)

A

	HR	P value
DAG(32:0)	1.20	2×10^{-8}
	1.17	1×10^{-4}
	1.07	0.1
	1.13	0.003
	1.19	3×10^{-5}
DAG(32:1)	1.21	8×10^{-9}
	1.20	2×10^{-5}
	1.10	0.03
	1.19	3×10^{-5}
	1.22	2×10^{-6}
DAG(34:2)	1.21	1×10^{-8}
	1.19	1×10^{-4}
	1.11	0.02
	1.19	6×10^{-5}
	1.20	6×10^{-5}
DAG(36:4)	1.22	8×10^{-10}
	1.16	4×10^{-4}
	1.09	0.04
	1.19	1×10^{-5}
	1.20	1×10^{-5}

B

	HR	P value
LPC(O−24:1)	0.79	1×10^{-7}
	0.76	3×10^{-6}
	0.79	4×10^{-5}
	0.79	6×10^{-5}
	0.77	1×10^{-5}
PC(38:2−2OH)	0.78	7×10^{-9}
	0.81	1×10^{-4}
	0.86	0.007
	0.80	4×10^{-5}
	0.80	7×10^{-5}
PC(40:5)	0.79	2×10^{-9}
	0.81	4×10^{-5}
	0.84	0.002
	0.81	2×10^{-5}
	0.83	2×10^{-4}
PC(42:4)	0.79	5×10^{-8}
	0.81	2×10^{-4}
	0.84	0.003
	0.78	1×10^{-5}
	0.82	3×10^{-4}

Figure 3—Cox regression statistics with covariate analysis. Associations between each biomarker of the DAG cluster (A) and PC cluster (B) and incident T2D in a Cox regression model adjusted for age, sex, and BMI only (basic model), or including correction for fasting plasma glucose, 2-h postchallenge plasma glucose, HbA$_{1c}$, and fasting insulin values, respectively. HRs are per 1 SD increase in metabolite level and plotted on a logarithmic scale.

four DAG species [but had no impact on association with incident T2D (Supplementary Table 4)].

Associations of DAGs and PCs With Incident T2D in EA and AA Participants

To investigate whether the identified associations of DAGs and PCs with incident T2D differed by race, we performed subgroup analysis in the ARIC cohort stratified by race. We identified race-specific differences in metabolite distributions: median levels of both DAGs and PCs were higher in EA compared with AA individuals [with the exception of PC(38:2-2OH); Supplementary Fig. 4]. To assess if these distributional differences influenced the strength and directionality of the association with incident T2D, we performed race-stratified Cox regression analysis examining the interaction between race and metabolite for each DAG and PC. No significant interactions were identified (Table 1), suggesting that the associations of these metabolites with incident T2D are relevant to both EA and AA individuals.

CONCLUSIONS

In this study, we applied high-throughput, untargeted metabolomics to two large prospective cohorts (FINRISK 2002 and ARIC) to identify and replicate metabolites associated with incident T2D. We identified 171 metabolite biomarkers, which clustered into several chemical classes ranging from polar small molecules to lipids, including two classes of lipids, DAGs and PCs, which were respectively associated with an increased and decreased T2D incidence (32,33). Through covariate analysis, we showed that several species of DAGs captured similar information as 2hPG values, whereas several PCs were associated with T2D incidence independently from conventional glycemic markers, indicating that they marked a distinct aspect of T2D etiology. Finally, we found that the hazard conferred by these DAG and PC lipids for T2D was similar for EA and AA individuals.

The molecular physiology of DAGs has been studied with respect to the development of insulin resistance as activating ligands of protein kinase C (PKC) (34). DAG-mediated activation of PKC-ε and PKC-θ (in the liver and skeletal muscle, respectively) inhibits insulin receptor activity in the liver (35) and decreases insulin-stimulated glucose transport activity in muscle (36). In muscle, this molecular pathophysiology aligns with our finding that circulating DAGs capture similar information as 2hPG concentration, which is largely determined by muscle glucose disposal (8). Thus, we would propose that, in humans, DAGs cause muscle-specific insulin resistance via the activation of PKC-θ, leading to impaired muscle glucose disposal as a potential mechanism for T2D risk. Interestingly, although our hazard models showed that DAGs remained significantly associated with T2D when corrected for FPG and, thus, are independent, the signal was attenuated, which is consistent with a partial relationship with hepatic glucose production (Fig. 3A). Our data suggest that, in humans, circulating DAGs reflect skeletal muscle insulin resistance to a greater extent than hepatic insulin resistance.

By contrast, the molecular mechanism underlying the protective effects of PCs on T2D is less clear. One possible mechanism is through a reduction of inflammation in metabolic tissues such as adipose tissue (37) or pancreatic islets (38), which have been shown to be associated with insulin resistance and T2D in model systems. PCs have anti-inflammatory properties (39), but further work is needed to substantiate if amelioration of adipose tissue and pancreatic islet inflammation could contribute to the protection against T2D. Interestingly, insulin secretion in people with T2D increased in a small study of recombinant HDL cholesterol infusion (40). HDL particles have a high component of PCs in their lipid membrane, but our association of PCs and incident T2D was independent of serum HDL (Supplementary Table 3).

Our study has several strengths, including a robust follow-up, diverse cohorts, depth of clinical characterization, and unbiased metabolomic profiling. The population-based FINRISK 2002 discovery cohort includes a long follow-up period (average follow-up of 14 years) with few censored individuals over time. The cohort also has a rich set of conventional glycemic markers ascertained, including FPG, 2hPG, serum insulin from an OGTT, and HbA$_{1c}$ measurements. The ARIC replication and follow-up cohort contained individuals who were different demographically, environmentally, and ancestrally, being drawn from U.S. communities.

Finally, in contrast to previous studies of metabolites associated with T2D (12–14), we profiled a breadth of metabolites with diverse chemistries in the FINRISK cohort (n = 31,701 metabolite signatures).

Limitations of our study include heterogeneity between discovery and replication cohorts, and the semiquantitative nature of the LC-MS metabolomics platform. One important difference between the discovery (FINRISK) and replication (ARIC) is the duration of fasting. Whereas all ARIC samples selected for this study were taken after a minimum 8 h of fasting, only 12% of FINRISK samples met this criterion (however, 71% of participants had fasted for at least 5 h; Supplementary Fig. 5). Nevertheless, the replication of the biomarkers of interest indicates that fasting did not severely compromise our findings. Furthermore, although we tried to distinguish incident T2D from type 1 diabetes in the FINRISK cohort, it was not possible to ensure that this distinction was made without some overlap in both of our study cohorts. It is also important to note that metabolite levels measured using untargeted metabolomics are semiquantitative, meaning that only relative levels are determined. Thus, although they can provide information on aspects of T2D etiology, as we have presented, they cannot be immediately translated to clinical use, for which absolute quantification and diagnostic test evaluation would be required. This would be the subject of future work to develop these DAG and PC metabolites into clinically suitable biomarkers.

equity in the company. J.T. owns shares in Orion Pharma, Aktivolabs Ltd, and Digostics Ltd. T.J.N. has received speaking honoraria from Servier Finland, AstraZeneca, and Orion Corporation. No other potential conflicts of interest relevant to this article were reported.

Author Contributions. A.B., K.P.G.-M., T.I., M.J., and A.R.M. contributed to conceptualization of the study. A.B., K.P.G.-M., and A.R.M. performed analysis and wrote the manuscript. J.D.W. and R.M. developed the untargeted LC-MS method and collected the metabolomics data. E.D.K. created software for the analysis of the LC-MS data. T.L., J.D.W., J.T., A.S.H., T.J.N., P.J., V.S., B.Y., F.N., C.M.R., E.S., E.A.W., S.C., M.A., R.G., M.J., and A.R.M. provided guidance. All authors reviewed the manuscript and approved the submission. M.J. and A.R.M. acquired funding, supervised work, and are the guarantors of this work and, as such, have full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Prior Presentation. Parts of this study were presented at the 83rd Scientific Session of the American Diabetes Association, San Diego, CA, 23–26 June 2023; and at the Keystone Symposium "Cardiometabolic Diseases: The Role of Ethnic Diversity in Precision Medicine," Vancouver, BC, Canada, 4–7 February 2024.

Handling Editors. The journal editor responsible for overseeing the review of the manuscript was Matthew C. Riddle.

References

1. Magliano DJ, Boyko EJ. IDF Diabetes Atlas 10th edition Scientific Committee. IDF Diabetes Atlas. International Diabetes Federation; 2021
2. Tuomilehto J, Lindström J, Eriksson JG, et al.; Finnish Diabetes Prevention Study Group. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. N Engl J Med 2001;344:1343–1350
3. Knowler WC, Barrett-Connor E, Fowler SE, et al.; Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002;346:393–403
4. Wagner R, Heni M, Tabák AG, et al. Pathophysiology-based subphenotyping of individuals at elevated risk for type 2 diabetes. Nat Med 2021;27:49–57
5. Ahlqvist E, Storm P, Käräjämäki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol 2018;6:361–369
6. Kaga H, Tamura Y, Takeno K, et al. Shape of the glucose response curve during an oral glucose tolerance test is associated with insulin clearance and muscle insulin sensitivity in healthy non-obese men. J Diabetes Investig 2020;11:874–877
7. Rothman DL, Magnusson I, Katz LD, Shulman RG, Shulman GI. Quantitation of hepatic glycogenolysis and gluconeogenesis in fasting

humans with 13C NMR. Science 1991;254:573–576
8. DeFronzo RA, Gunnarsson R, Björkman O, Olsson M, Wahren J. Effects of insulin on peripheral and splanchnic glucose metabolism in noninsulin-dependent (type II) diabetes mellitus. J Clin Invest 1985;76:149–155
9. Nathan DM, Genuth S, Lachin J, et al.; Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med 1993;329:977–986
10. Karter AJ, Parker MM, Moffet HH, Gilliam LK. Racial and ethnic differences in the association between Mean glucose and hemoglobin A1c. Diabetes Technol Ther 2023;25:697–704
11. Trifonova O, Lokhov P, Archakov A. Postgenomics diagnostics: metabolomics approaches to human blood profiling. Omics 2013;17:550–559
12. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. Nat Med 2011;17:448–453
13. Guasch-Ferré M, Ruiz-Canela M, Li J, et al. Plasma acylcarnitines and risk of type 2 diabetes in a Mediterranean population at high cardiovascular risk. J Clin Endocrinol Metab 2019;104:1508–1519
14. Zhong J, Cheung CYY, Su X, et al. Specific triacylglycerol, diacylglycerol, and lyso-phosphatidylcholine species for the prediction of type 2 diabetes: a ∼ 16-year prospective study in Chinese. Cardiovasc Diabetol 2022;21:234–211
15. Saaristo T, Peltonen M, Lindström J, et al. Cross-sectional evaluation of the Finnish Diabetes Risk Score: a tool to identify undetected type 2 diabetes, abnormal glucose tolerance and metabolic syndrome. Diab Vasc Dis Res 2005;2:67–72
16. Borodulin K, Tolonen H, Jousilahti P, et al. Cohort profile: the National FINRISK Study. Int J Epidemiol 2018;47:696–696i
17. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) study: design and objectives. Am J Epidemiol 1989;129:687–702
18. Wright JD, Folsom AR, Coresh J, et al. The ARIC (Atherosclerosis Risk in Communities) study: JACC Focus Seminar 3/8. J Am Coll Cardiol 2021;77:2939–2959
19. Kantz ED, Tiwari S, Watrous JD, Cheng S, Jain M. Deep neural networks for classification of LC-MS spectral peaks. Anal Chem 2019;91:12407–12413
20. Wishart DS, Tzur D, Knox C, et al. HMDB: the Human Metabolome Database. Nucleic Acids Res 2007;35:D521–6
21. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. Springer Science & Business Media; 2013. 356 p
22. survminer: Drawing Survival Curves using "ggplot2." R package survminer, version 0.4.9. 2021. Accessed 11 July 2024. Available from https://CRAN.R-project.org/package=survminer
23. Krumsiek J, Suhre K, Illig T, Adamski J, Theis FJ. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. BMC Syst Biol 2011;5:21–16
24. Schäfer J, Strimmer K. A shrinkage approach to large-scale covariance matrix estimation and

implications for functional genomics. Stat Appl Genet Mol Biol 2005;4:Article32
25. NetworkX. In Encyclopedia of Social Network Analysis and Mining. Alhajj R, Rokne J, Eds. Springer; 2018:1608–1608. Available from https://dx.doi.org/10.1007/978-1-4939-7131-2_100771
26. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–2504
27. Ohlson LO, Larsson B, Björntorp P, et al. Risk factors for type 2 (non-insulin-dependent) diabetes mellitus. Thirteen and one-half years of follow-up of the participants in a study of Swedish men born in 1913. Diabetologia 1988;31:798–805
28. Liu X, Zheng Y, Guasch-Ferré M, et al. High plasma glutamate and low glutamine-to-glutamate ratio are associated with type 2 diabetes: case-cohort study within the PREDIMED trial. Nutr Metab Cardiovasc Dis 2019;29:1040–1049
29. Chen Y, Wang N, Dong X, et al. Associations between serum amino acids and incident type 2 diabetes in Chinese rural adults. Nutr Metab Cardiovasc Dis 2021;31:2416–2425
30. Mirmiran P, Bahadoran Z, Gaeini Z, Azizi F. Habitual intake of dietary ʟ-arginine in relation to risk of type 2 diabetes: a prospective study. BMC Endocr Disord 2021;21:113–117
31. Vangipurapu J, Fernandes Silva L, Kuulasmaa T, Smith U, Laakso M. Microbiota-related metabolites and the risk of type 2 diabetes. Diabetes Care 2020;43:1319–1325
32. Morze J, Wittenbecher C, Schwingshackl L, et al. Metabolomics and type 2 diabetes risk: an updated systematic review and meta-analysis of prospective cohort studies. Diabetes Care 2022;45:1013–1024
33. 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
34. Erion DM, Shulman GI. Diacylglycerol-mediated insulin resistance. Nat Med 2010;16:400–402
35. Samuel VT, Liu Z-X, Wang A, et al. Inhibition of protein kinase Cε prevents hepatic insulin resistance in nonalcoholic fatty liver disease. J Clin Invest 2007;117:739–745
36. Yu C, Chen Y, Cline GW, et al. Mechanism by which fatty acids inhibit insulin activation of insulin receptor substrate-1 (IRS-1)-associated phosphatidylinositol 3-kinase activity in muscle. J Biol Chem 2002;277:50230–50236
37. Burhans MS, Hagman DK, Kuzma JN, Schmidt KA, Kratz M. Contribution of adipose tissue inflammation to the development of type 2 diabetes mellitus. Compr Physiol 2018;9:1
38. Marchetti P. Islet inflammation in type 2 diabetes. Diabetologia 2016;59:668–672
39. Treede I, Braun A, Sparla R, et al. Anti-inflammatory effects of phosphatidylcholine. J Biol Chem 2007;282:27155–27164
40. Drew BG, Duffy SJ, Formosa MF, et al. High-density lipoprotein modulates glucose metabolism in patients with type 2 diabetes mellitus. Circulation 2009;119:2103–2111