Plasma Lipidome and Prediction of Type 2 Diabetes in the Population-Based Malmö Diet and Cancer Cohort

Céline Fernandez,[1] Michal A. Surma,[2] Christian Klose,[3] Mathias J. Gerl,[3] Filip Ottosson,[1] Ulrika Ericson,[1] Nikolay Oskolkov,[4] Marju Ohro-Melander,[1] Kai Simons,[3] and Olle Melander[1]

Diabetes Care 2020;43:366–373 | https://doi.org/10.2337/dc19-1199

OBJECTIVE

Type 2 diabetes mellitus (T2DM) is associated with dyslipidemia, but the detailed alterations in lipid species preceding the disease are largely unknown. We aimed to identify plasma lipids associated with development of T2DM and investigate their associations with lifestyle.

RESEARCH DESIGN AND METHODS

At baseline, 178 lipids were measured by mass spectrometry in 3,668 participants without diabetes from the Malmö Diet and Cancer Study. The population was randomly split into discovery (n = 1,868, including 257 incident cases) and replication (n = 1,800, including 249 incident cases) sets. We used orthogonal projections to latent structures discriminant analyses, extracted a predictive component for T2DM incidence (lipid-PC_{DM}), and assessed its association with T2DM incidence using Cox regression and lifestyle factors using general linear models.

RESULTS

A T2DM-predictive lipid-PC_{DM} derived from the discovery set was independently associated with T2DM incidence in the replication set, with hazard ratio (HR) among subjects in the fifth versus first quintile of lipid-PC_{DM} of 3.7 (95% CI 2.2–6.5). In comparison, the HR of T2DM among obese versus normal weight subjects was 1.8 (95% CI 1.2–2.6). Clinical lipids did not improve T2DM risk prediction, but adding the lipid-PC_{DM} to all conventional T2DM risk factors increased the area under the receiver operating characteristics curve by 3%. The lipid-PC_{DM} was also associated with a dietary risk score for T2DM incidence and lower level of physical activity.

CONCLUSIONS

A lifestyle-related lipidomic profile strongly predicts T2DM development beyond current risk factors. Further studies are warranted to test if lifestyle interventions modifying this lipidomic profile can prevent T2DM.

Human plasma contains several hundred lipid molecular species, which collectively make up the plasma lipidome (1). Disturbed lipid metabolism is a hallmark of type 2 diabetes mellitus (T2DM). T2DM-associated dyslipidemia is traditionally described in terms of circulating total triacylglycerides (TAGs) and lipoprotein cholesterol levels (i.e., high total TAG and low HDL cholesterol [HDL-C] levels) (2). However, lipidomics-based studies of T2DM have shown that the traditional lipid markers are insufficient

[1]Department of Clinical Sciences, Malmö, Lund University, Lund, Sweden
[2]Łukasiewicz Research Network–PORT Polish Center for Technology Development, Wroclaw, Poland
[3]Lipotype GmbH, Dresden, Germany
[4]Department of Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Lund University, Lund, Sweden

Corresponding author: Céline Fernandez, celine.fernandez@med.lu.se

Received 18 June 2019 and accepted 3 November 2019

This article contains Supplementary Data online at http://care.diabetesjournals.org/lookup/suppl/doi:10.2337/dc19-1199/-/DC1.

to predict disease risk, and more specifically molecular lipid species associated with T2DM risk factors, such as obesity, hyperglycemia, and insulin resistance, have been identified (3,4). In addition, plasma molecular lipid species associated with pre-T2DM or overt T2DM have been discovered (5,6).

Longitudinal studies assessing associations between the plasma lipidome and risk of developing T2DM (7–10) or gestational diabetes mellitus (11) or progressing from gestational diabetes to T2DM (12) have emerged. However, prospective studies of the plasma lipidome in a general population-based setting with many incident cases and long follow-up for the prediction of T2DM are lacking. Phenotypically well-characterized prospective cohorts are necessary to study whether lipid molecular species can be of clinical use for T2DM risk prediction on top of conventional risk factors, including level of glycemia and standard lipids. Identification of such lipids is essential to provide targets for novel lifestyle and pharmacological intervention studies.

Orthogonal projections to latent structures discriminant analysis (OPLS-DA) is a supervised and linear multivariate statistical method for the unbiased selection of independent features predicting events, which uses cross-validation to reduce the likelihood of false-positive results. OPLS partitions the systemic variation in the data into one predictive and one or several orthogonal components (i.e., unrelated to events) (13,14).

In the current study, our aim was to test if plasma lipids, measured by a top-down shotgun lipidomics method, predict T2DM incidence during long-term follow-up in a large population-based prospective cohort study from Malmö, Sweden. To avoid false-positive results, we randomly divided the population into a set for discovery and a set for replication. Because many molecular lipids are highly correlated, we used OPLS-DA for unbiased selection of independent and informative lipid predictors for T2DM. In addition, we investigated if dietary habits and level of leisure time physical activity are related to molecular lipid species levels, in order to pave the way for potential strategies to modulate their levels.

RESEARCH DESIGN AND METHODS

Study Participants and Data Collection

The Malmö Diet and Cancer–Cardiovascular Cohort (MDC-CC) is a prospective population-based cohort designed to study the epidemiology of carotid artery disease collected from 1991 to 1994 (15). At baseline, all the MDC-CC participants underwent a medical history, physical examination, and laboratory and lifestyle assessment. Of the 5,405 overnight-fasted participants, citrate plasma samples were available in 4,067 subjects for analysis of the lipidome. Of these, a total of 377 participants were further excluded from statistical analyses because of the presence of T2DM (for definition, see below) at baseline examination. Additionally, lipidomic results from 22 samples were excluded because of too many missing values (for threshold, see below). The remaining 3,668 study participants without diabetes were then randomly divided into discovery (n = 1,868, of whom 257 developed incident T2DM during follow-up) and replication (n = 1,800, of whom 249 developed incident T2DM during follow-up) sets using the function "Select a random sample of cases, ~50% of all cases" in SPSS.

Data on baseline levels of BMI, systolic blood pressure (SBP), current smoking, use of antihypertensive treatment, physical activity during leisure time, clinical lipids, and blood glucose were obtained as previously described (16). Physical activity during leisure time was transformed into a score (17). A diet risk score previously shown to predict T2DM (DRS_{DM}) based on four foods consistently associated with T2DM in epidemiological studies (processed meat, sugar-sweetened beverages, whole grain, and coffee) was constructed as previously described and expressed as risk points from 0 to 8. The DRS_{DM} was divided into three groups: low DRS_{DM} (0–2 points), medium DRS_{DM} (3–5 points), and high DRS_{DM} (6–8 points) (18).

T2DM at baseline examination was defined as a fasting whole-blood glucose ≥6.1 mmol/L (corresponding to plasma glucose ≥7.0 mmol/L), self-report of a physician diagnosis, or use of diabetes medication.

All participants provided written informed consent, and the study (LU 51-90) was approved by the Ethics Committee at Lund University.

Follow-Up and End Points Retrieval

Subjects were followed for incidence of T2DM until 31 December 2014. The median follow-up time was 21.2 years (25–75% interquartile range [IQR] 16.9–22.0). T2DM was defined as a fasting plasma glucose ≥7.0 mmol/L, a history of physician diagnosis of T2DM, being on diabetes medication, or having been registered in local or national diabetes registries, as described previously (19).

Analytical Process Design

Samples were divided into analytical batches of 84 samples each. Each batch was accompanied by a set of four blank samples and eight identical control reference samples (human plasma). These control samples in groups of one blank and two reference samples were distributed evenly across each batch, extracted, and processed together with study samples to control for background and intrarun reproducibility. All batches were measured within 4 weeks.

Mass Spectrometry Lipidomics

Lipid extraction of 1 μL of overnight-fasted citrate plasma samples stored at −80°C upon collection, followed by quantitative mass spectrometry–based lipid analysis, was performed at Lipotype GmbH using a high-throughput shotgun lipidomics technology (20).

Postprocessing

Spectra were analyzed with an in house–developed lipid identification software based on LipidXplorer (21). Data postprocessing and normalization were performed using an in house–developed data management system. Only lipid identifications with a signal-to-noise ratio >5 and a signal intensity fivefold higher than in corresponding blank samples were considered for further data analysis.

Batch correction was applied using eight reference samples per 96 wells. Amounts were corrected for analytical drift if the P value of the slope was <0.05, with an R^2 >0.75, and the relative drift was >5%. The median coefficient of lipid molecular species variation as assessed by reference samples was 10.49%. An occupational threshold of 75% was applied to the data, resulting in 178 lipid molecular species for further analysis. Samples with >20% of the 178 lipids missing were removed (22 samples). Missing values were imputed using the

NIPALS algorithm (R 3.4.3, nipals package), a linear and multivariate statistical-based method that can cope with a moderate amount of data missing at random (22).

Statistical Analyses

Statistical analyses were conducted in SPSS (version 25.0) unless otherwise stated. Student *t* tests and Pearson χ^2 tests were used to compare continuous and dichotomous variables, respectively, at baseline examination between the two cohorts. Due to nonnormality, lipid species were transformed with the base 10 log. The lipids data were mean centered and unit variance scaled prior to statistical analysis.

In the discovery set, OPLS-DA was adopted in order to simultaneously analyze correlations between the 178 lipid molecular species and T2DM incidence (13,14). The systematic variation in lipid levels that correlated with T2DM incidence was captured in one lipid predictive component for T2DM incidence (lipid-PC_{DM}), and the variation that did not correlate with T2DM was contained in orthogonal components. The model performance was validated by sevenfold cross-validation. In the next step, data from the replication set were projected on the OPLS-DA model from the discovery set, and lipid loadings from the discovery set were used to calculate a lipid-PC_{DM} in the replication set. Analyses using OPLS-DA were performed in SIMCA, version 14.1 (Sartorius Stedim Biotech, Göttingen, Germany).

The associations between lipids and T2DM incidence were investigated using multivariable Cox proportional hazards models. Participants with incomplete data on covariates were excluded. Hazard ratios (HRs) were expressed per SD increment or per increasing quintiles (quintiles two through five compared with quintile one). Model 1 was adjusted for age and sex, and model 2 was further adjusted for BMI, SBP, use of antihypertensive treatment, current smoking, and fasting plasma levels of glucose, triglycerides, LDL cholesterol (LDL-C), and HDL-C at baseline examination. Model 2 was additionally adjusted for family history of diabetes, lipid-lowering medication, leisure time physical activity level, and DRS_{DM}. In model 3, BMI was included as a variable divided into three clinically defined categories ($\leq$25 kg/m^2, normal weight; 25–30 kg/m^2, overweight; $>$30 kg/m^2, obese) instead of as a continuous variable. Normal weight subjects were used as the reference category. In sensitivity analyses, we only included males, females, or participants with overnight-fasted plasma glucose concentration $<$5.6 mmol/L and performed adjustments according to models 1 and 2.

In the replication set, Kaplan-Meier survival curve was used to describe the rate of T2DM incidence over time in quintiles of baseline levels of lipid-PC_{DM}.

The added value of the lipid-PC_{DM} for T2DM risk prediction was assessed by two methods in the replication set. First, receiver operating characteristic (ROC) curves were constructed and Delong test, performed in R 3.4.3 using the pROC package, was used to compare the area under the curve (AUC) of the different models. In the first model (i.e., conventional nonlipid risk factors), the predictive value of age, sex, BMI, SBP, use of antihypertensive treatment, current smoking, and fasting plasma levels of glucose were investigated. In a second model (i.e., all conventional risk factors), fasting plasma levels of triglycerides, LDL-C, and HDL-C were further added. In the last model (i.e., all conventional risk factors and PC_{DM}), the lipid-PC_{DM} was also added.

Second, continuous net reclassification improvement (cNRI) and integrated discrimination improvement (IDI) were calculated using the survIDINRI package in R 3.4.3 (23) in order to compare the predictive capacity of the lipid-PC_{DM} to all conventional risk factors.

In the whole cohort, general linear models were used to study associations between the lipid-PC_{DM} and lifestyle factors (i.e., tertiles intake of energy-adjusted food groups, DRS_{DM}, and level of physical activity during leisure time) adjusting for potential confounders, which were successively added in three models (age, sex, BMI, smoking, alcohol intake, and physical activity).

Bonferroni correction was used to determine significance for multiple tests performed in the discovery set. For all other tests, a two-sided *P* value of $<$0.05 was considered statistically significant.

RESULTS

Baseline Characteristics

The baseline characteristics of the study participants in the discovery and replication sets are listed in Table 1. There

Table 1—Characteristics of the study participants

	Discovery set (n = 1,868)	Replication set (n = 1,800)	P^a
Age (years)	57.50 ± 5.95	57.39 ± 6.04	NS
Sex (% women)	58.7	61.7	NS
Incident T2DM (%)	13.8	13.8	NS
BMI (kg/m^2)	25.38 ± 3.81	25.38 ± 3.64	NS
Fasting glucose (mmol/L)	4.91 ± 0.43	4.91 ± 0.44	NS
Fasting triglycerides (mmol/L)	1.27 ± 0.59	1.26 ± 0.60	NS
Fasting LDL (mmol/L)	4.16 ± 0.97	4.16 ± 0.99	NS
Fasting HDL (mmol/L)	1.41 ± 0.37	1.42 ± 0.37	NS
SBP (mmHg)	141.28 ± 18.53	140.68 ± 18.96	NS
Antihypertensive medication use (%)	14.3	14.8	NS
Lipid-lowering therapy (%)	2.6	1.4	0.012
Smoker (%)	27.6	26.8	NS

Values are displayed as means ± SD or frequency in percent. ns, nonstatistical significant. [a]*P* values were obtained from Student *t* tests or Pearson χ^2 tests for the binary variables.

was no statistical difference for the investigated parameters, including the clinical lipids, between the two study sets except for the intake of lipid-lowering medication.

Lipid Species and Risk of T2DM in the MDC-CC Discovery Set

In the MDC-CC discovery set with 1,868 participants, 257 developed new-onset T2DM during a median follow-up time of 21.3 years (IQR 17.2–22.0). Out of 178 individual lipid species analyzed (Supplementary Table 1), 58 associated with T2DM incidence at Bonferroni significance ($P < 2.8E-4$) in Cox proportional hazards models adjusted for age and sex (Supplementary Fig. 1*A* and Supplementary Table 2). Increased baseline levels of several lipid species belonging to the TAG and diacylglyceride (DAG) lipid classes (HR/SD change of lipid levels) associated with increased risk of incident T2DM ($1.25 \leq$ HR ≤ 1.61), whereas increased levels of ether phospatidylcholine (PC O-) and of lipid species containing a fatty acyl chain with C18:2 (except one, DAG 16:0;0_18:2;0) associated with decreased risk of future T2DM ($0.64 \leq$ HR ≤ 0.80).

Because individual lipid species are highly correlated, in a next step, we used OPLS-DA to simultaneously study the association between all 178 lipid species and T2DM incidence, which resulted in one lipid predictive component for T2DM incidence (lipid-PC_{DM}) as well as two orthogonal components (Supplementary Fig. 2*A*). The goodness of fit (R2Y) and the predictive ability (Q2Y) of the OPLS-DA model were 12% and 8%, respectively. A total of 77 lipids had significant loadings ($P < 0.05$) and contributed to the lipid-PC_{DM}, explaining 5.3% of the 178 lipids proportion of variance (R2X = 0.053). Results from the multivariate analysis were concordant with the Cox regression analyses, with 49 lipids overlapping between the two analyses. Furthermore, as in the Cox regression models, lipids belonging to the TAG and DAG lipid classes had positive loadings whereas lipids belonging to the PC O- and C18:2-containing lipid species had generally negative loadings (Supplementary Table 3). More specifically, TAG 48:0;0 and LysoPC 18:2;0 (LPC 18:2;0) displayed the strongest positive and negative associations, respectively, with T2DM incidence in terms of HR effect size and loadings.

Next, we tested if the generated lipid-PC_{DM} associated with T2DM incidence in Cox regression models. A higher value of the lipid-PC_{DM} (HR/SD increment of the PC_{DM}) was associated with increased risk of T2DM after age and sex adjustment (model 1) (HR 1.93 [95% CI 1.77–2.10], $P = 4.50E-49$) (Table 2). The association remained significant after further adjustment for BMI, SBP, use of antihypertensive treatment, current smoking, and fasting plasma levels of glucose, triglycerides, LDL-C, and HDL-C at baseline (model 2), with a multivariable-adjusted HR of 1.74 (95% CI 1.58–1.93, $P = 4.07E-27$) (Table 2). The HR was unchanged after additional adjustment for family history of diabetes, intake of lipid-lowering therapy, leisure time physical activity level, and DRS_{DM} (Supplementary Table 4). There was a linear relationship between the lipid-PC_{DM} and the risk of incident T2DM in both model 1 and 2 (*P* for trend respectively equal to 4.48E−27 or 5.00E−14), and the top versus bottom quintile of the lipid-PC_{DM} in the fully adjusted model was associated with an HR of 5.12 (95% CI 2.92–8.96) (Table 2).

Validation of Results in the MDC-CC Replication Set

In the MDC-CC replication set with 1,800 individuals, 249 developed new-onset T2DM during a median follow-up time of 21.2 years (IQR 16.8–22.0). As compared with the discovery set, all 58 individual lipids were consistently significantly related to T2DM incidence in the replication set at $P < 4.0E-3$ (Supplementary Fig. 1*B* and Supplementary Table 2).

In the next step, data from the replication set were projected on the OPLS-DA model from the discovery set (Supplementary Fig. 2*B*). Loadings from the discovery set were used to calculate a lipid-PC_{DM} in the replication set. The cumulative incidence rate of T2DM according to quintiles of the calculated lipid-PC_{DM} was higher for subjects in the upper quintiles (Supplementary Fig. 3). A higher value of the lipid-PC_{DM} (HR/SD increment of PC_{DM}) was associated with increased risk of T2DM after age and sex adjustment (model 1) (HR 1.89 [95% CI 1.72–2.07], $P = 4.07E-42$) (Table 2). The association remained significant after full adjustment for the traditional risk factors for T2DM

Table 2—Incidence of T2DM in relation to baseline levels of the lipid predictive component in the discovery ($n = 1,868$) and the replication ($n = 1,800$) set

	Discovery set		Replication set	
	Model 1 ($n = 257/1,610$)	Model 2 ($n = 246/1,571$)	Model 1 ($n = 249/1,550$)	Model 2 ($n = 243/1,517$)
Overall				
HR[a] (95% CI)	1.93 (1.77–2.10)	1.74 (1.58–1.93)	1.89 (1.72–2.07)	1.57 (1.40–1.76)
P value	<0.001	<0.001	<0.001	<0.001
Quintiles				
HR[b] Q1	1.0 (referent)	1.0 (referent)	1.0 (referent)	1.0 (referent)
HR[b] Q2	1.30 (0.68–2.45)	1.11 (0.57–2.13)	1.45 (0.79–2.66)	1.32 (0.72–2.43)
HR[b] Q3	2.72 (1.55–4.76)	2.02 (1.13–3.63)	2.04 (1.16–3.61)	1.23 (0.68–2.21)
HR[b] Q4	3.64 (2.11–6.26)	2.41 (1.35–4.29)	3.41 (2.00–5.82)	2.05 (1.18–3.56)
HR[b] Q5	8.30 (5.00–13.81)	5.12 (2.92–8.96)	7.90 (4.80–13.00)	3.49 (2.01–6.05)
P[c] for trend	<0.001	<0.001	<0.001	<0.001

Model 1 is adjusted for age and sex. Model 2 is adjusted for age, sex, BMI, SBP, use of antihypertensive treatment, current smoking, and fasting plasma levels of glucose, triglycerides, LDL-C, and HDL-C at baseline. Q, quintile. [a]HRs are expressed per 1-SD increment of the lipid predictive component. 95% CIs of the HRs are reported in parentheses. [b]HRs are expressed per quintile of the lipid predictive component, using quintile 1 as reference. 95% CIs of the HRs are reported in parentheses. [c]*P* for trend across quintiles of the lipid predictive component in a linear regression model.

(model 2), with a multivariable-adjusted HR of 1.57 (95% CI 1.40–1.76, P = 3.76E−15) (Table 2). The HR was unchanged after additional adjustment for family history of diabetes, lipid-lowering medication, leisure time physical activity level, and DRS_{DM} (Supplementary Table 4). There was also a linear relationship in the replication set between the lipid-PC_{DM} and the risk of incident T2DM in both model 1 and 2 (P for trend respectively equal to 5.88E−26 or 2.95E−08), and the top versus bottom quintile of PC_{DM} in the fully adjusted model was associated with an HR of 3.49 (95% CI 2.01–6.05) (Table 2). The standardized lipid-PC_{DM} was more strongly associated with incident T2DM than the standardized BMI in model 2 (HR 1.57 [95% CI 1.40–1.76], P = 3.76E−15, vs. HR 1.21 [95% CI 1.08–1.36], P = 0.001). Furthermore, being in the fourth or fifth quintile of lipid-PC_{DM} was associated with a higher T2DM risk than being obese in all models (Table 3 and Supplementary Table 5).

In sensitivity analyses, the lipid-PC_{DM} significantly associated with T2DM incidence in males, females, and participants with a normal fasted plasma glucose concentration (Supplementary Table 6). The effect size of the lipid-PC_{DM} association with T2DM incidence was even strengthened when removing participants with impaired fasting glucose and adjusting for all traditional risk factors (HR 2.01 [95% CI 1.63–2.47], P = 5.15E−11).

Next, we tested whether the lipidomic-based lipid measurements contributed to T2DM risk discrimination. Whereas clinical lipids did not improve risk discrimination on top of conventional non-lipid risk factors in ROC curves (AUC = 0.775 [95% CI 0.744–0.806] vs. 0.770 [95% CI 0.738–0.801], P = 0.176), adding the lipid-PC_{DM} to the conventional risk factors for T2DM (including clinical lipids) significantly improved the AUC by 3% (AUC = 0.807 [95% CI 0.778–0.837] vs. 0.775 [95% CI 0.744–0.806], P < 0.001) (Fig. 1). Furthermore, estimates of IDI and cNRI for 18-year T2DM event rates increased when adding the lipid-PC_{DM} on top of all the traditional risk factors for T2DM (IDI = 0.037 [95% CI 0.02–0.07], P < 0.001; cNRI = 0.17 [95% CI 0.06–0.27], P = 0.013).

Association of Diet Pattern and Physical Activity with the Plasma Lipidome

In the whole study cohort of 3,668 participants, we investigated if the lipid-PC_{DM} associated with the DRS_{DM}, which is based upon intake of whole grain, coffee, sugar-sweetened beverages, and processed meat (18). The DRS_{DM} (high- vs. low-risk diet) was also associated with increased risk of T2DM in our cohort after adjustment for age and sex (HR 1.73 [95% CI 1.29–2.30], P = 2.12E−04). There was a positive association between the lipid-PC_{DM} and the DRS_{DM} when adjusting for age and sex (β = 0.12, P trend <0.001), which remained unchanged when further adjusting for BMI and lifestyle factors. All components of the DRS_{DM}, except for whole grain intake, contributed significantly to this association (Supplementary Table 7). The lipid-PC_{DM} was positively associated with total dairy intake (β = 0.05, P trend <0.01) but was not found to associate with intake of fish and shellfish, nor with intake of fruits and vegetables. Level of leisure time physical activity was negatively correlated with the lipid-PC_{DM} (β = 0.09, P trend <0.001), and the association remained significant when further adjusting for BMI and lifestyle factors (Supplementary Table 7).

Table 3—Incidence of T2DM in relation to baseline levels of the lipid predictive component in the replication set (n = 1,800), adjusting for traditional risk factors, including BMI categories

	Model 3 (n = 243 of 1,517)
BMI (kg/m^2) categories[a]	
≤25	1.0 (referent)
25–30	1.03 (0.76–1.41)
>30	1.79 (1.23–2.62)
Quintiles of lipid predictive component	
Q1	1.0 (referent)
Q2	1.33 (0.72–2.44)
Q3	1.27 (0.71–2.29)
Q4	2.18 (1.25–3.80)
Q5	3.72 (2.15–6.45)

Data are displayed as HR (95% CI). In model 3, the following independent variables were simultaneously entered: age, sex, categories of BMI, SBP, use of antihypertensive treatment, current smoking, and fasting plasma levels of glucose, triglycerides, LDL-C, and HDL-C at baseline. Q, quintile. [a]Number of individuals per BMI category is 904, 717, and 179, repectively.

CONCLUSIONS

Here, we identified individual molecular lipid species as well as a pattern of lipids (i.e., lipid-PC_{DM}) prospectively associated with T2DM risk during long-term follow-up in a general population. Our results are robust as they were very similar in the replication set as compared with the discovery set. The lipid-PC_{DM} was more strongly related to T2DM than BMI was, both in terms of its effect size and statistical significance. As BMI is considered a key clinical risk factor for T2DM, this result demonstrates the clinical importance of the lipid-PC_{DM}. Moreover, clinical lipids used in many T2DM-predictive algorithms were outperformed by the detailed lipid measurement contained within the lipid-PC_{DM}. Last, we identified associations between the lipid-PC_{DM} and dietary intake, including a previously T2DM-associated dietary pattern, as well as physical activity.

We identified several individual lipids prospectively associated with T2DM and replicated all associations in the replication set. Among the lipids of interest, several ether lipids displayed a significant inverse association with T2DM incidence, which is in accordance with previous results from the Prevención con Dieta Mediterránea (PREDIMED) trial case-cohort study of T2DM incidence (8). Previous studies also found that ether lipid levels were lower in individuals with obesity (4), prediabetes, and T2DM (6). This is of interest because of the suggested role for plasmalogens, the most common form of ether lipids, as sacrificial antioxidants, sparing the oxidation of other lipids, and as scavengers of reactive oxygen species (24). Plasmalogens might also modulate inflammation by acting as storage depots for arachidonic acid and docosahexaenoic acid (25).

We also found that several lysophospholipids (i.e., LPC and lysophosphatidylethanolamine [LPE]) were inversely

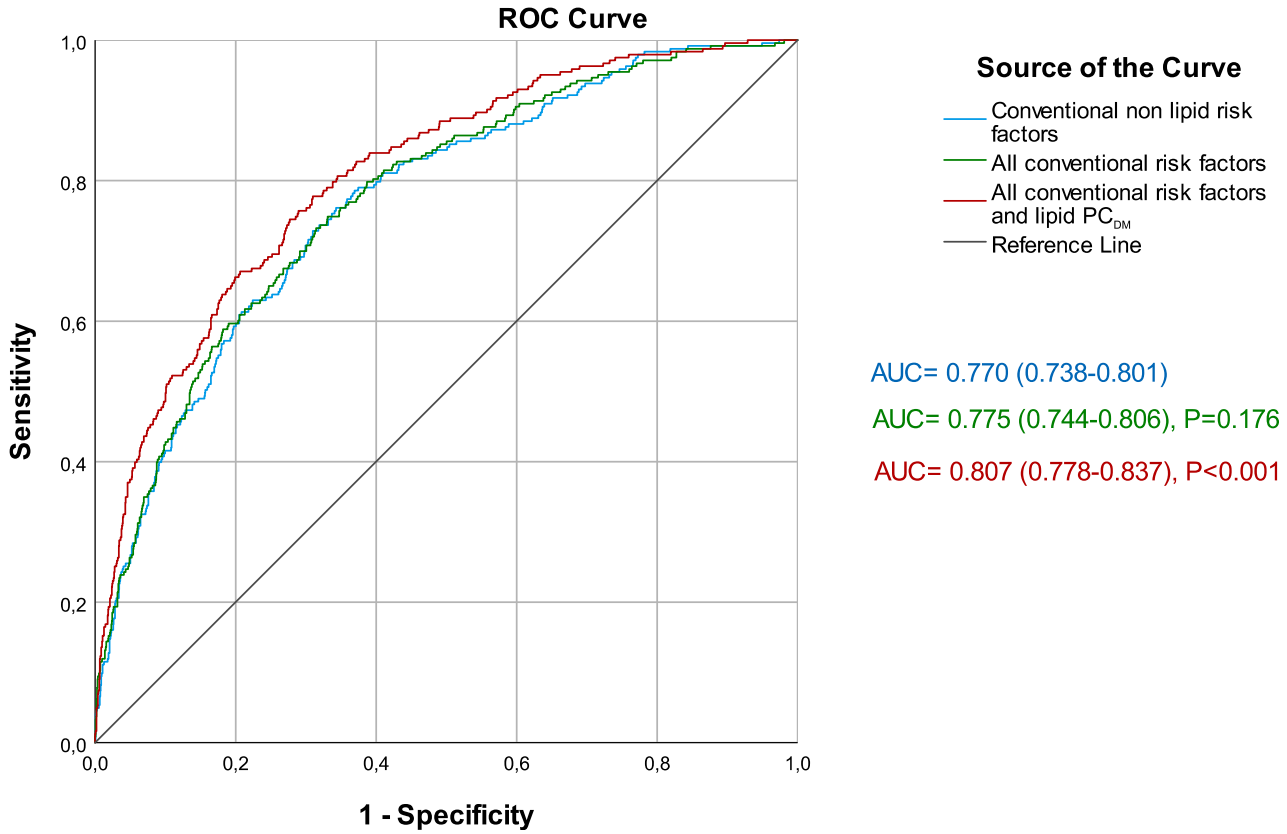

Figure 1—ROC curves showing the different AUCs for the prediction of T2DM calculated for different models in the replication set (n = 1,800). In the first model (i.e., conventional nonlipid risk factors), the predictive value of age, sex, BMI, SBP, use of antihypertensive treatment, current smoking, and fasting plasma levels of glucose was investigated. In a second model (i.e., all conventional risk factors), fasting plasma levels of triglycerides, LDL-C, and HDL-C were further added. In the third and last model (i.e., all conventional risk factors and PC_{DM}), the lipid PC_{DM} was also added.

related with T2DM risk, including LPC 18:2;0, which is in accordance with previous studies (8,10). Furthermore, lower LPC levels have been previously shown to associate with an increased risk of cardiovascular disease (26) and appear to play a role in inflammation and atherogenesis (27). We found that a decreased plasma level of C18:2 fatty acid (linoleic acid) containing complex lipids is associated with an increased risk of T2DM, which we have also previously shown to occur in overt T2DM in the MDC-CC reexamination cohort (5).

Our results also confirm previous studies that have established that DAG and TAG, including TAG 48:0;0, are associated with a higher risk of T2DM (6,8,9). Overall, our study with >20 years of follow-up repeats several findings obtained in case-control studies with shorter follow-up times (i.e., 4 years for the PREDIMED trial [8] and 12 years for the Framingham Heart Study [9]). This indicates that changes in the plasma lipidome occur several years before T2DM onset, which could give any primary preventive treatment targeting the lipid/lipid metabolic pathway enough time to interfere with the T2DM pathophysiological process.

Currently, BMI and presence of obesity is the key risk factor used for T2DM prediction in the clinic. In our study, the lipid-PC_{DM} was more strongly related to T2DM than BMI both in terms of its effect size and statistical significance even when entered into the same model, suggesting it identifies individuals at high T2DM risk independently of BMI (regardless of whether lipid-PC_{DM} is causally related to T2DM or not). Thus, individuals with high lipid-PC_{DM}, irrespective of their BMI, might be candidates for intensive lifestyle preventive actions in order to prevent T2DM in a similar way as individuals with high BMI are treated today. Furthermore, while clinical lipids used in T2DM prediction models, i.e., total TAG and HDL-C, did not improve T2DM risk prediction on top of nonlipid traditional risk factors, the lipid-PC_{DM} improved risk prediction by 3% beyond all risk factors. The lipid-PC_{DM} added value for T2DM risk prediction over all traditional risk factors was confirmed using IDI and cNRI. This highlights the need of a refinement of the concept of "dyslipidemia" (referring to low HDL-C and high TAG) related to the metabolic syndrome (28), as a key role of the dyslipidemia component within the metabolic syndrome is to signal increased risk of T2DM and thus leads to initiation of preventive actions. Collectively, the superior role of lipid-PC_{DM} in T2DM prediction over both obesity and clinical lipids suggests it would bring clinical value if introduced on top of currently used T2DM risk factors.

Plasma lipid levels could be influenced by dietary intake (29,30) and physical activity (31). The finding that the lipid-PC_{DM} associated with a diet score previously shown to be linked to elevated T2DM risk (18), although the effect size of the association was relatively weak, suggests that diet modification targeted at components of that score, i.e., increased coffee intake and decreased sugar-sweetened beverages and processed meat, might be interesting candidates

for future controlled lifestyle intervention trials. Moreover, based on our results, one can speculate that the plasma level of the lipid-PC_{DM} could be a useful tool for monitoring adherence to lifestyle therapy in high-risk subjects for T2DM.

The main strength of this study is its large size in terms of number of subjects and prospective end points during long-term follow-up, which permitted us to perform a robust replication of the findings. The very similar and significant findings in both the discovery and replication sets greatly decrease the risk of false-positive results, which are very common in smaller "omics" studies. Moreover, we based our conclusions solely on the replication set, thus minimizing the risk of, for example, inflation of effect sizes and significance levels. On the other hand, both the discovery and replication sets were derived from the same South Swedish Caucasian population, and it would be important to test our findings in other populations with, for example, different lifestyle habits and race for external validation. Another strength of this study comes from the detailed clinical and lifestyle information available in the cohort, although gestational diabetes history was missing. Last, the broad coverage of the lipidome provided a detailed insight into the metabolic changes preceding T2DM, even though we cannot exclude that because of the long sample storage time, some of the lipids have been partially degraded. However, if this would be the case, it should bias our results toward no findings.

We do acknowledge that despite extensive adjustments and the prospective study design, we cannot prove causality between the plasma lipidome and T2DM, given the observational nature of the study. This may not necessarily limit its role as a risk biomarker. However, causality needs to be proven if T2DM risk–associated lipids should be targets for novel preventive drug therapies and lifestyle interventions.

Our results indicate that the plasma lipidome contains information beyond known predictors of T2DM, such as obesity and clinical lipids, which could be used for patients' risk stratification. Furthermore, our results encourage controlled lifestyle intervention trials aimed at testing if modifying the plasma lipidome may reduce the risk of T2DM.

Funding. C.F. was supported by the Albert Påhlsson Foundation, the Crafoord Research Foundation, the Ernhold Lundström Research Foundation, the Royal Physiographic Society of Lund, and the Åke Wiberg Foundation. C.F., M.O.-M., and O.M. additionally acknowledge support from the Swedish Foundation for Strategic Research (IRC15-0067). M.O.-M. was supported by the European Research Council (Consolidator grant 649021), the Swedish Research Council, the Swedish Heart and Lung Foundation, Region Skåne, the Swedish Diabetes Foundation, the Novo Nordisk Foundation, and the Albert Påhlsson Foundation. O.M. was supported by the Knut and Alice Wallenberg Foundation, the Göran Gustafsson Foundation, the Swedish Heart and Lung Foundation, the Swedish Research Council, the Novo Nordisk Foundation, Region Skåne, and Skåne University Hospital.

Duality of Interest. M.A.S., C.K., and K.S. are shareholders of Lipotype GmbH. M.J.G. is an employee of Lipotype GmbH. K.S. is the chief executive officer of Lipotype GmbH. No other potential conflicts of interest relevant to this article were reported.

Author Contributions. C.F. drafted the manuscript. C.F., F.O., and N.O. contributed to the data analyses. C.F., K.S., and O.M. contributed to study concept and design. M.A.S., C.K., M.J.G., U.E., and M.O.-M. contributed to acquisition of data. C.F., M.A.S., C.K., M.J.G., F.O., U.E., N.O., M.O.-M., K.S., and O.M. provided intellectual contributions to drafting and/or revising the manuscript and approved the final version. C.F. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Data and Resource Availability. The MDC data discussed in this article will be made available to appropriate academic parties based on a written application to the MDC steering committee; please contact Anders Dahlin (anders.dahlin@med.lu.se).

References

1. Quehenberger O, Armando AM, Brown AH, et al. Lipidomics reveals a remarkable diversity of lipids in human plasma. J Lipid Res 2010;51:3299–3305
2. Taskinen MR, Borén J. New insights into the pathophysiology of dyslipidemia in type 2 diabetes. Atherosclerosis 2015;239:483–495
3. Huynh K, Barlow CK, Jayawardana KS, et al. High-throughput plasma lipidomics: detailed mapping of the associations with cardiometabolic risk factors. Cell Chem Biol 2019;26:71–84.e4
4. Weir JM, Wong G, Barlow CK, et al. Plasma lipid profiling in a large population-based cohort. J Lipid Res 2013;54:2898–2908
5. Kjellqvist S, Klose C, Surma MA, et al. Identification of shared and unique serum lipid profiles in diabetes mellitus and myocardial infarction. J Am Heart Assoc 2016;5: e004503
6. Meikle PJ, Wong G, Barlow CK, et al. Plasma lipid profiling shows similar associations with prediabetes and type 2 diabetes. PLoS One 2013;8:e74341
7. Mamtani M, Kulkarni H, Wong G, et al. Lipidomic risk score independently and cost-effectively predicts risk of future type 2 diabetes: results from diverse cohorts. Lipids Health Dis 2016;15:67
8. Razquin C, Toledo E, Clish CB, et al. Plasma lipidomic profiling and risk of type 2 diabetes in the PREDIMED trial. Diabetes Care 2018;41:2617–2624
9. Rhee EP, Cheng S, Larson MG, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. J Clin Invest 2011;121:1402–1411
10. Suvitaival T, Bondia-Pons I, Yetukuri L, et al. Lipidome as a predictive tool in progression to type 2 diabetes in Finnish men. Metabolism 2018;78:1–12
11. Lu L, Koulman A, Petry CJ, et al. An unbiased lipidomics approach identifies early second trimester lipids predictive of maternal glycemic traits and gestational diabetes mellitus. Diabetes Care 2016;39:2232–2239
12. Lappas M, Mundra PA, Wong G, et al. The prediction of type 2 diabetes in women with previous gestational diabetes mellitus using lipidomics. Diabetologia 2015;58:1436–1442
13. Bylesjo M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. J Chemometr 2006;20:341–351
14. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). J Chemometr 2002;16:119–128
15. Rosvall M, Janzon L, Berglund G, Engström G, Hedblad B. Incident coronary events and case fatality in relation to common carotid intima-media thickness. J Intern Med 2005;257:430–437
16. Fernandez C, Rysä J, Almgren P, et al. Plasma levels of the proprotein convertase furin and incidence of diabetes and mortality. J Intern Med 2018;284:377–387
17. Hellstrand S, Sonestedt E, Ericson U, et al. Intake levels of dietary long-chain PUFAs modify the association between genetic variation in FADS and LDL-C. J Lipid Res 2012;53:1183–1189
18. Ericson U, Hindy G, Drake I, et al. Dietary and genetic risk scores and incidence of type 2 diabetes. Genes Nutr 2018;13:13
19. Enhörning S, Hedblad B, Nilsson PM, Engström G, Melander O. Copeptin is an independent predictor of diabetic heart disease and death. Am Heart J 2015;169:549–556.e1
20. Surma MA, Herzog R, Vasilj A, et al. An automated shotgun lipidomics platform for high throughput, comprehensive, and quantitative analysis of blood plasma intact lipids. Eur J Lipid Sci Technol 2015;117:1540–1549
21. Herzog R, Schuhmann K, Schwudke D, et al. LipidXplorer: a software for consensual cross-platform lipidomics. PLoS One 2012;7:e29851
22. Wold S, Sjostrom M, Eriksson L. PLS-regression: a basic tool of chemometrics. Chemometr Intell Lab 2001;58:109–130
23. Uno H, Tian L, Cai T, Kohane IS, Wei LJ. A unified inference procedure for a class of measures to assess improvement in risk prediction

systems with survival data. Stat Med 2013;32:2430–2442
24. Dean JM, Lodhi IJ. Structural and functional roles of ether lipids. Protein Cell 2018;9:196–206
25. Braverman NE, Moser AB. Functions of plasmalogen lipids in health and disease. Biochim Biophys Acta 2012;1822:1442–1452
26. Fernandez C, Sandin M, Sampaio JL, et al. Plasma lipid composition and risk of developing cardiovascular disease. PLoS One 2013;8:e71846
27. Schmitz G, Ruebsaamen K. Metabolism and atherogenic disease association of lysophosphatidylcholine. Atherosclerosis 2010;208:10–18
28. Fedorowski A, Burri P, Hulthén L, Melander O. The metabolic syndrome and risk of myocardial infarction in familial hypertension (hypertension heredity in Malmö evaluation study). J Hypertens 2009;27:109–117
29. Frahnow T, Osterhoff MA, Hornemann S, et al. Heritability and responses to high fat diet of plasma lipidomics in a twin study. Sci Rep 2017;7:3750
30. Toledo E, Wang DD, Ruiz-Canela M, et al. Plasma lipidomic profiles and cardiovascular events in a randomized intervention trial with the Mediterranean diet. Am J Clin Nutr 2017;106:973–983
31. Grace MS, Dempsey PC, Sethi P, et al. Breaking up prolonged sitting alters the postprandial plasma lipidomic profile of adults with type 2 diabetes. J Clin Endocrinol Metab 2017;102:1991–1999