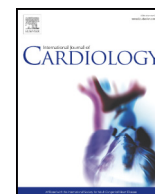




Contents lists available at ScienceDirect

International Journal of Cardiology

journal homepage: www.elsevier.com/locate/ijcard

A plasma lipid signature predicts incident coronary artery disease

Filip Ottosson^{a,*,1}, Payam Emami Khoonsari^{a,b}, Mathias J. Gerl^{c,a}, Kai Simons^c, Olle Melander^a, Céline Fernandez^a^a Department of Clinical Sciences, Lund University, Malmö, Sweden^b Department of Biochemistry and Biophysics, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Stockholm University, Box 1031, SE-17121 Solna, Sweden^c Lipotype GmbH, Dresden, Germany

ARTICLE INFO

Article history:

Received 19 November 2020

Received in revised form 18 January 2021

Accepted 25 January 2021

Available online 3 February 2021

Keywords:

Lipidome

Lipidomics

Coronary artery disease

ABSTRACT

Background: Dyslipidemia is a hallmark of cardiovascular disease but is characterized by crude measurements of triglycerides, HDL- and LDL cholesterol. Lipidomics enables more detailed measurements of plasma lipids, which may help improve risk stratification and understand the pathophysiology of cardiovascular disease.

Methods: Lipidomics was used to measure 184 lipids in plasma samples from the Malmö Diet and Cancer – Cardiovascular Cohort ($N = 3865$), taken at baseline examination. During an average follow-up time of 20.3 years, 536 participants developed coronary artery disease (CAD). Least absolute shrinkage and selection operator (LASSO) were applied to Cox proportional hazards models in order to identify plasma lipids that predict CAD.

Results: Eight plasma lipids improved prediction of future CAD on top of traditional cardiovascular risk factors. Principal component analysis of CAD-associated lipids revealed one principal component (PC2) that was associated with risk of future CAD (HR per SD increment = 1.46, $C \cdot I = 1.35$ – 1.48 , $P < 0.001$). The risk increase for being in the highest quartile of PC2 (HR = 2.33, $P < 0.001$) was higher than being in the top quartile of systolic blood pressure. Addition of PC2 to traditional risk factors achieved an improvement (2%) in the area under the ROC-curve for CAD events occurring within 10 ($P = 0.03$), 15 ($P = 0.003$) and 20 ($P = 0.001$) years of follow-up respectively.

Conclusions: A lipid pattern improve CAD prediction above traditional risk factors, highlighting that conventional lipid-measures insufficiently describe dyslipidemia that is present years before CAD. Identifying this hidden dyslipidemia may help motivate lifestyle and pharmacological interventions early enough to reach a substantial reduction in absolute risk.

© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cardiovascular disease (CVD) is the leading cause of death and a major health concern worldwide [1]. Dyslipidemia is one of the hallmarks of CVD [2], but is characterized by crude measurements of plasma HDL- and LDL-cholesterol and triglycerides. Refinement of these measurements, into a more detailed description of the circulating lipids, could help improve risk stratification and understand pathophysiology [3]. Lipidomics is a powerful discipline where hundreds of different lipids species can be measured in blood plasma [4]. Several lipidomics studies have demonstrated its potential in smaller case-control- and population-based cohorts, by identifying different lipid species to prospectively associate with new-onset CVD [5–7]. Larger studies have also shown that individual lipid species can predict the risk of

cardiovascular death in patients with overt CVD [8]. A large-scale prospective study that investigated a panel of both polar and lipid metabolites in relation to CVD risk found that five phospholipids were the strongest predictors of future CVD [9]. This finding calls for large-scale lipidomics investigations, using prospective cohorts with long follow-up times, relating the plasma lipidome to CVD risk. We recently showed the potential of this approach when lipidomic based lipid measurements were superior at predicting the risk of future type 2 diabetes (T2D) compared to clinically measured levels of HDL, LDL and TG [10]. Similarly, identifying plasma lipid species that predict new-onset coronary artery disease (CAD) could help to refine our view of CAD-related dyslipidemia in order to find molecular targets for pharmacological interventions and to improve risk stratification.

In the present study, we measure 184 different lipid species in participants from the MDC-CC ($N = 3865$), where 536 individuals develop CAD within an average follow-up time of 20.3 years. We attempt to construct a predictive lipid risk score using feature selection least absolute shrinkage and selection operator (LASSO) in Cox proportional hazards models.

* Corresponding author at: Jan Waldenströms gata 35, 214 28, Malmö, Sweden.

E-mail address: Filip.ottosson@med.lu.se (F. Ottosson).

¹ “This author takes responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation”

2. Methods

2.1. Study participants

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 [11]. At baseline, all the MDC-CC participants underwent medical history, physical examination, and laboratory and lifestyle assessment. Of the 5405 overnight-fasted participants, citrate plasma samples were available in 4067 subjects for analysis of the lipidome. Participants with prevalent CAD or with missing values for traditional risk factors were excluded from the analysis, resulting in a cohort of 3856 participants. The ethics committee of Lund University approved the study protocols for MDC-CC (DNR 2009/633), and all participants provided written informed consent.

2.2. Endpoint definition and biochemical measurements

CAD was defined as coronary revascularization, fatal or non-fatal myocardial infarction, or death attributable to ischemic heart disease. The study subjects were followed for incident CAD through record-linkage using the Swedish personal identification number with the previously validated Swedish Hospital Discharge Register, the Swedish Cause of Death Register and the Swedish Coronary Angiography and Angioplasty Registry (SCAAR) [12]. International Classification of Diseases (ICD) codes and details about biochemical measurements are found in Material S1.

2.3. Lipidomics analysis

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 intra-run reproducibility. All batches were measured within 4 weeks.

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 [13].

Post-processing spectra were analyzed with an in house– developed lipid identification software based on LipidXplorer [14]. Data post processing, normalization and batch correction was performed as described previously [10]. Lipids that were present in at least 70% of the analytical samples were included in the subsequent statistical analysis, resulting in a total of 184 lipids. Lipid identifiers of the SwissLipids database [15] (<http://www.swisslipids.org>) are provided in Table S1.

2.4. Statistical analysis

Missing lipid values (6.2%) were imputed using random forest-based imputation. All lipid values were then log10-transformed, mean-centered and unit-variance scaled. Least absolute shrinkage and selection operator (LASSO) regularization (10-fold cross-validation) was used to select lipids that improve prediction of incident CAD in Cox proportional hazard models. Eight traditional risk factors, age, sex, BMI, HDL cholesterol, LDL cholesterol, systolic blood pressure (SBP), smoking, anti-hypertensive treatment and prevalent diabetes, were set to be constantly selected in the model by setting the penalty values to zero.

Dimensionality reduction of CAD-associated lipids was performed using principal component analysis (PCA). The number of components were automatically detected by using the autofit option and each principal component was mean-centered and unit-variance scaled.

Hazard-ratios for each CAD-associated lipid and principal component was calculated by Cox proportional hazards models. The first model was adjusted for age and sex and the second model was additionally adjusted for BMI, HDL cholesterol, LDL cholesterol, SBP, anti-hypertensive treatment, prevalent diabetes and smoking.

To assess and compare the contribution of each principal component to risk prediction of future CAD relative to the traditional risk factors, we constructed a base model (i.e. a model based on sex and age) followed by construction of 18 other models each based on the model and one of the risk factors or principal components. The concordance index of each model were calculated and compared to assess the relative contribution of each factor to the prediction.

The performance in risk prediction of using lipidomic measurements, compared to a basic model including all traditional risk factors, was assessed by comparing the area under the ROC curves (AUC) for three different censoring time points (prediction horizons), 10 years, 15 years and 20 years. Statistical difference in the AUC was assessed using Wald tests. Finally, the inter-quartile risk of future CAD was estimated for SBP and the second principal component, using Cox proportional hazards models, adjusted for all traditional risk factors.

All statistical analyses were performed in R 3.6.2. PCA performed in the *ropis* [16] package (v1.18.8), the imputation in the *missForest* [17] package (v1.4) and LASSO in the *glmnet* package (v3.0.2) interfaced through *mlr* package [18] (v2.17.1).

3. Results

Plasma lipidomics was performed in the MDC-CC ($N = 3329$) in order to measure 184 lipid species. The baseline characteristics of the MDC-CC can be found in Table 1. All participants were free from CAD at baseline examination but 536 participants developed CAD during an average follow-up time of 20.3 years. All investigated traditional CAD risk factors, age, sex, BMI, LDL cholesterol, HDL cholesterol, systolic blood pressure, smoking, prevalence of diabetes and anti-hypertensive treatment, were different between participants who developed CAD ($N = 536$) before 2016 and those who remained free from CAD ($N = 2793$).

The feature selection algorithm LASSO was used to select lipid species that were associated with incident CAD, independently of traditional CAD risk factors, using Cox regression models. The Cox regression models were adjusted for age, sex, BMI, systolic blood pressure, anti-hypertensive treatment, smoking, the prevalence of diabetes and fasting levels of HDL and LDL cholesterol. In total, eight lipid species, phosphatidylcholine (PC) 15:0;0_18:2;0, PC 17:0;0_20:3;0, PC 16:0;0_20:1;0, PC 16:2;0_18:0;0, sphingomyelin (SM) 34:1;2, diacylglycerol (DAG) 18:1;0_18:3;0, phosphatidylinositol (PI) 16:0;0_20:4;0 and sterol ester (CE) 18:0;0, contributed to improved prediction of incident CAD on top of the traditional CAD risk factors.

Table 1

Baseline characteristics of the participants from the Malmö Diet and Cancer – Cardiovascular Cohort (MDC-CC) ($N = 3865$). Differences between traditional cardiovascular risk factors between controls and incident CAD were calculated using two-sample *t*-tests. Standard deviations for all continuous risk factors are given in parentheses.

Trait	Controls ($N = 3329$)	Incident CAD ($N = 536$)	<i>P</i>
Age (years)	57.3 (± 6.0)	59.3 (± 5.7)	<0.001
Sex (% female)	62.7	41.2	<0.001
BMI (kg/m^2)	25.5 (± 3.8)	26.3 (± 4.1)	<0.001
LDL cholesterol (mmol/l)	4.14 (± 1.0)	4.39 (± 1.0)	<0.001
HDL cholesterol (mmol/l)	1.42 (± 0.4)	1.29 (± 0.4)	<0.001
Systolic blood pressure (mmHG)	141 (± 19)	148 (± 19)	<0.001
Current Smokers (%)	26.0	34.0	<0.001
Diabetes (%)	6.8	14.7	<0.001
Antihypertensive treatment (%)	14.5	22.2	<0.001
Lipid-lowering treatment (%)	1.5	3.5	<0.001

Next, we aimed to create a lipid score based on the eight CAD-associated lipids species that can predict future CAD. Using PCA, the variation in eight lipids was reduced to three principal components (PC) that explained 34%, 16% and 12% respectively of the total variation of the 8 CAD-associated lipids. To evaluate how much the lipidomics features contribute to CAD prediction compared to traditional CAD risk factors, all lipidomic features and traditional CAD risk factors were iteratively added to a basic Cox regression model with only age and sex as covariates. Systolic blood pressure was the risk factor that achieved the largest improvement in c-index (2.0%) compared to the basic model, followed by prevalent diabetes (1.3%), smoking (1.0%) and HDL cholesterol (1.0%) (Fig. 1). Among lipidomic features, PC2 improved the prediction the most (0.9%), followed by PC15:0;0_18:2;0 (0.6%).

The association between all lipidomic features (the three lipid PCs and the eight CAD associated lipids) and incident CAD was assessed using Cox regression models, adjusted for all CAD risk factors. PC2 was the most strongly associated lipidomic feature (HR per SD increment = 1.46, C-I = 1.35–1.48, $P < 0.001$). Among the individual lipids, three species, SM 34:1;2, PC 16:0;0_20:1;0 and PC O 16:2;0_18:0;0 were associated with increased risk and four lipids species, PC 15:0;0_18:2;0, DAG 18:1;0_18:3;0, PI 16:0;0_20:4;0 and CE 18:0;0 were associated with a decreased risk of incident CAD (Fig. 2). Additional adjustment for lipid-lowering treatment, triglycerides and total cholesterol did not influence the associations (Table S2). Being in the top quartile of PC2 was associated with a more than doubled risk of developing CAD compared to the lowest quartile (HR = 2.33, C-I = 1.81–3.00, $P < 0.001$). In comparison, this risk increase was slightly higher than what was

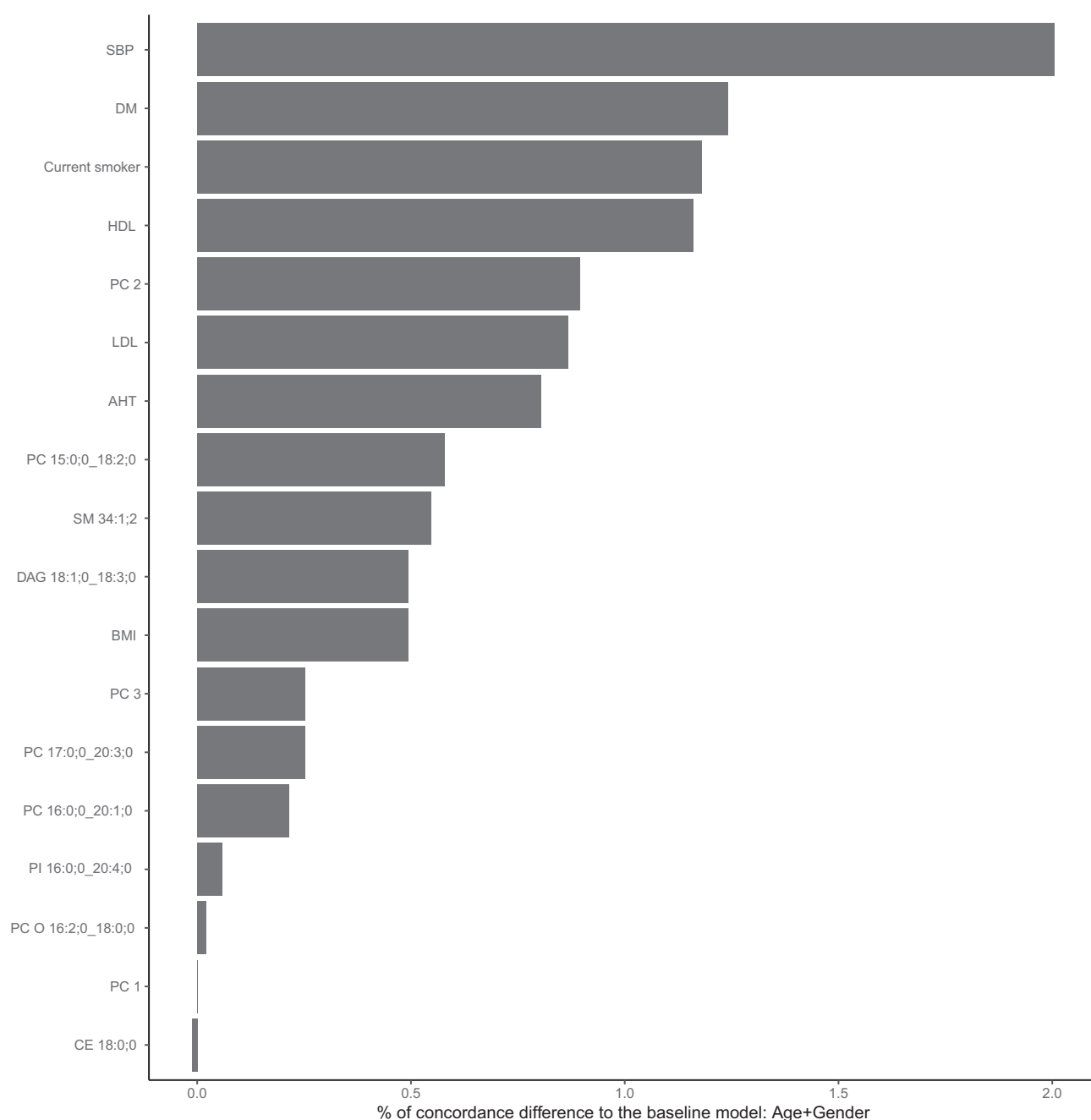


Fig. 1. Improvements (%) in concordance-index (c-index) when iteratively adding each variable to a baseline Cox proportional hazards model including age and sex as covariates.

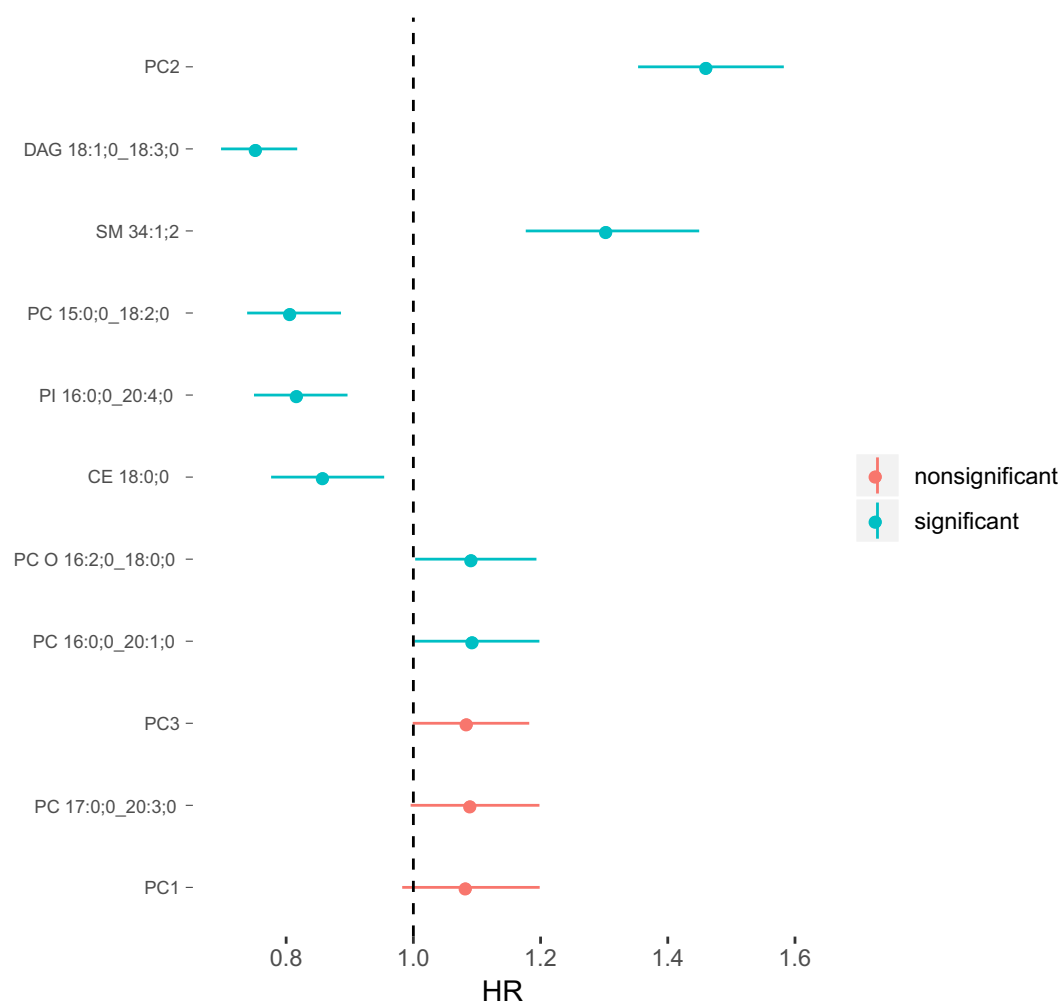


Fig. 2. Associations between lipid measurements and incident CAD. Hazard ratios (HR) are expressed per standard deviation increment of the lipid. Error bars indicate the 95% confidence intervals of the HR. Associations were considered significant at $p < 0.05$. Principal components are denoted as PC1–PC3. DAG = diacylglycerol, SM = Sphingomyelin, PC = Phosphatidylcholine, PI = phosphoinositol, CE = Cholesteryl ester and PC O = Ether phosphatidylcholine.

observed for being in the highest quartile of systolic blood pressure (SBP) (HR = 2.14, C.I. = 1.61–2.86, $P < 0.001$) (Table S3). Neither principal component 1 ($P = 0.11$) nor principal component 3 (0.053) were significantly associated with incident CAD.

Finally, we investigated whether PC2 could improve the prediction of incident CAD, above what is achieved using traditional CAD risk factors. Adding PC2 on top of traditional risk factors achieved a statistically significant improvement in the AUC for incident CAD events occurring within 10 ($P = 0.03$), 15 ($P = 0.003$) and 20 ($P = 0.001$) years respectively (Fig. 3). The best risk prediction (AUC = 0.81) was achieved for participants who developed CAD within 10 years after the examination. However, the improvements in AUC, when adding PC2 on top of traditional risk factors was 2% irrespective of time to disease.

4. Discussion

Here, we show that measurements of lipid species can be used to identify lipidomic patterns that associate with new-onset of CAD, up to 20 years prior to the disease. Importantly, this CAD predictive lipid pattern improves risk prediction compared to traditional risk factors for CAD, including conventionally measured lipids.

We identified eight individual lipid molecules associated with incident CAD. SM 34:1;2 was the lipid species most strongly associated with an increased risk of incident CAD. Sphingomyelins have previously

been suggested to be involved in the development of atherosclerosis [19] and insulin resistance [20,21]. However, previous lipidomic prospective studies have highlighted several sphingomyelin species to be associated with both increased [5] risk of CAD and decreased risk of type T2D [22,23] and heart failure [24]. Recently it was shown that a panel of 32 sphingolipids, including 12 sphingomyelins, could discriminate between patients with CAD and healthy controls [25]. Among these sphingomyelins, SM 34:1;2 was significantly increased in patients with CAD. Despite being associated with increased risk of CAD, SM 34:1;2 has previously been shown to associate strongly with lower body-fat percentage [26]. These seemingly contradictory associations seen for SM 34:1;2 indicate that despite being associated with an increased risk of CAD, its levels are significantly lower in individuals typically more prone to develop T2D. Since several other sphingomyelins have shown inverse associations with risk of T2D, it is intriguing to speculate that specific sphingomyelins, including SM 34:1;2, may have opposing associations with risk of T2DM and cardiovascular disease, as similarly has been shown for the pharmacological lowering of LDL cholesterol [27].

DAG 18:1;0_18:3;0 was the lipid species with the strongest association with a lower risk of future CAD. This finding was surprising since DAGs previously have been related to insulin resistance [28]. Moreover, total levels of DAGs, as well as DAG 18:1;0_18:3;0, has been shown to be increased in patients with stable CAD compared to healthy controls [29].

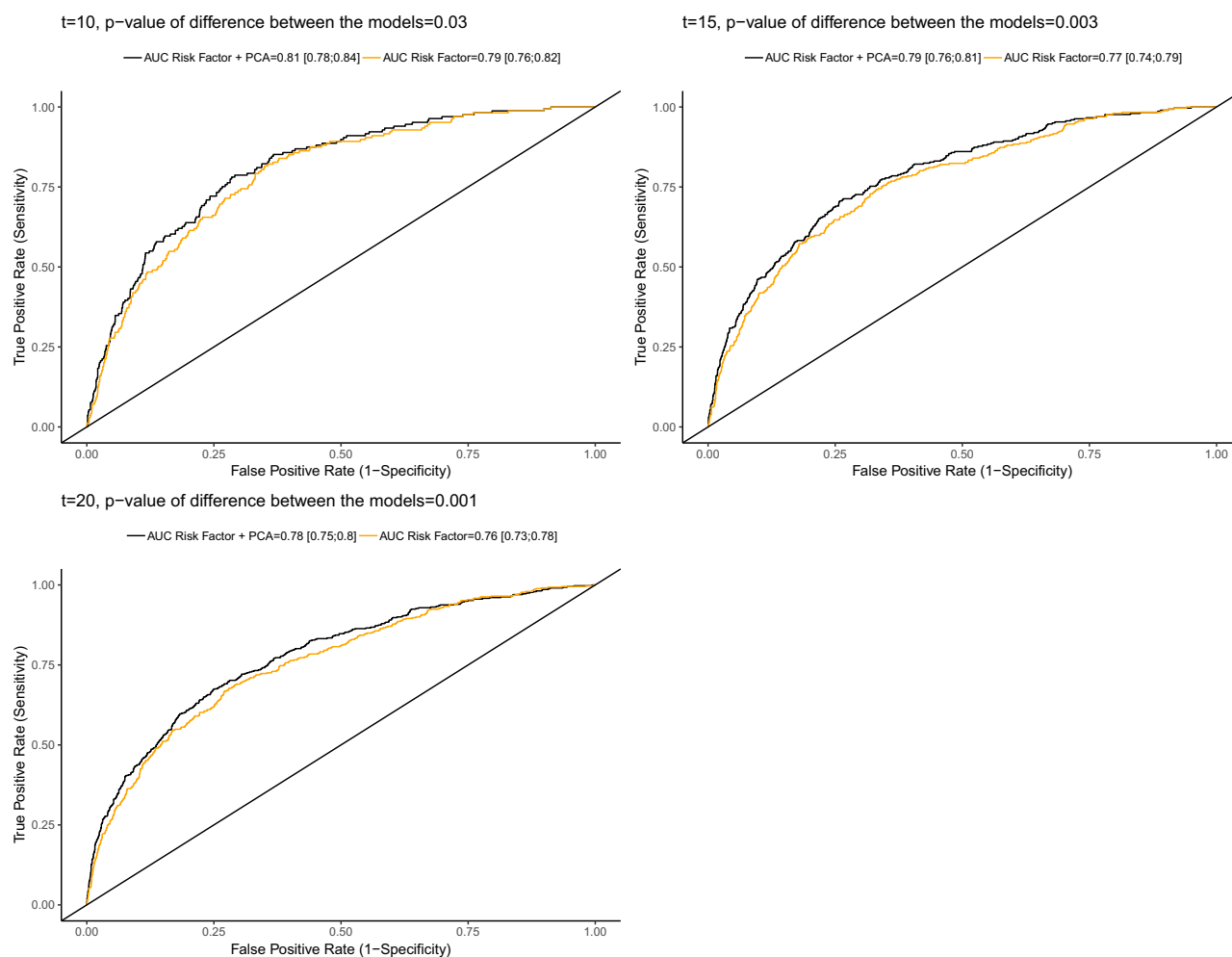


Fig. 3. Receiver operating characteristic (ROC) curves for prediction of future CAD, censored at three different follow-up times (10 years, 15 years and 20 years). The area under the curve (AUC) is calculated for two different models at each time point, where the first model (AUC Risk Factor) includes nine traditional CVD risk factors (age, sex, BMI, HDL cholesterol, LDL cholesterol, systolic blood pressure, smoking, anti-hypertensive treatment and diabetes) and the second model (AUC Risk Factor + PCA) additionally include the CVD-predictive lipidomics principal component (PC2). The 95% confidence interval for each AUC are given in brackets. *P*-values for model differences in AUC at each censoring time are calculated using the Wald test.

There is some support for a protective role of DAG 18:1;0_18:3;0 in cardiometabolic disease, as it was recently shown to be associated with a lower risk of incident T2D [10].

Although only explaining 16% of the variation in these lipids, PC2 was strongly associated with new-onset CAD. The improvements in risk prediction using PC2, on top of age and sex, was only surpassed by systolic blood pressure, diabetes and smoking status and HDL cholesterol. Furthermore, in the fully adjusted prediction model, being in the top quartile of PC2 was associated with a more than doubled risk of incident CAD, which was slightly higher than being in the top quartile of SBP. This is particularly interesting since these associations were adjusted for HDL and LDL cholesterol levels, emphasizing the potential of lipidomic-based risk prediction for CAD.

Adding PC2 to a multivariate prediction model including all traditional risk factors, a significant improvement in the AUC 2% was achieved. This improvement was achieved as long as 20 years prior to disease, indicating that lipidomics measurements could be useful as early markers increased CAD risk.

Several studies have shown the potential of lipidomics in cardiovascular disease. These include studies that find associations between lipid species and CAD [5,6,30], predict cardiovascular mortality in CAD patients [8] and identifying links between genetics, the lipidome and cardiovascular disease risk [31]. Interestingly, a large recent case-control study, comprising over

10,000 participants, found associations between five phosphatidylcholines and new-onset coronary heart disease. However, the five phosphatidylcholines could not improve risk prediction of coronary heart disease, above what was achieved with traditional risk factors [9].

To our knowledge, we are the first to show that lipidomic measurements can be used to achieve significant improvements in risk prediction of CAD in a large cohort with long-term follow-up. Importantly, this highlights that conventionally measured lipids are not enough to describe dyslipidemia that is present up to twenty years before CAD. This interpretation is supported by a previous study using a composite endpoint (cardiovascular events) as outcome which also found significant improvement of the C-index [7]. Identifying this hidden dyslipidemia may help motivate lifestyle and pharmacological interventions early enough to reach a substantial reduction in absolute risk. This study has several limitations. We acknowledge that this is an observational study and we thus cannot prove a causal relationship between the plasma lipidome and CAD risk. Moreover, although we were able to adjust for lipid-lowering treatment prescribed at baseline, it is possible that medication prescribed between the baseline examination and last follow-up may influence both lipid levels and CAD risk. Similarly, lipidomic measurements were only available at one time-point, whereas lipid levels could change over time as an effect of life-style changes or pharmacological treatment.

Author contributions

FO drafted the manuscript. CF, FO, and PEK contributed to the data analyses. CF, KS, FO and OM contributed to study concept and design. KS and MJG contributed to acquisition of data. FO, CF, PEK, MJG, KS, and OM provided intellectual contributions to drafting and/or revising the manuscript and approved the final version. CF 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. The manuscript is original research, has not been published and is not submitted to be considered for publication elsewhere, in whole or in part, in any language. There are no potentially overlapping papers in preparation, submission or published.

Sources of Funding

FO is financially supported by Ernhold Lundström Research Foundation, Hulda och E Conrad Mossfelts Stiftelse and the Albert Pahlsson Foundation. OM is supported by Knut and Alice Wallenberg Foundation, Göran Gustafsson Foundation, the Swedish Heart- and Lung Foundation, the Swedish Research Council, the European Research Council ERC-AdG-2019-885003, the Novo Nordisk Foundation, Region Skåne, Skåne University Hospital and Lund University, Swedish Foundation for Strategic Research (Dnr IRC15-0067), AIR Lund (Artificially Intelligent use of Registers at Lund University) research environment (VR; grant no. 2019-00198) and Lund University Infrastructure Grant 2020-2024. PEK is financially supported by the Knut and Alice Wallenberg Foundation as part of the National Bioinformatics Infrastructure Sweden at SciLifeLab. KS is financially supported by the Klaus Tschira Foundation. CF was supported by the Albert Pahlsson Research Foundation, the Crafoord Research Foundation, the Ernhold Lundström Research Foundation, the Royal Physiographic Society of Lund and the Åke Wiberg Foundation.

Disclosures

K.S. is a shareholder 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.

Acknowledgements

The authors would like to thank the staff and participants of the Malmö Diet and Cancer Study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijcard.2021.01.059>.

References

- [1] <https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-cvds>. Accessed.
- [2] K.G. Alberti, P. Zimmet, J. Shaw, Group IDFETFC, The metabolic syndrome—a new worldwide definition, *Lancet*. 366 (9491) (2005) 1059–1062.
- [3] O. Quehenberger, E.A. Dennis, The human plasma lipidome, *N. Engl. J. Med.* 365 (19) (2011) 1812–1823.
- [4] A. Shevchenko, K. Simons, Lipidomics: coming to grips with lipid diversity, *Nat. Rev. Mol. Cell Biol.* 11 (8) (2010) 593–598.
- [5] C. Fernandez, M. Sandin, J.L. Sampaio, et al., Plasma lipid composition and risk of developing cardiovascular disease, *PLoS One* 8 (8) (2013), e71846.
- [6] C. Razquin, L. Liang, E. Toledo, et al., Plasma lipidome patterns associated with cardiovascular risk in the PREDIMED trial: a case-cohort study, *Int. J. Cardiol.* 253 (2018) 126–132.
- [7] C. Stegmann, R. Pechlaner, P. Willeit, et al., Lipidomics profiling and risk of cardiovascular disease in the prospective population-based Bruneck study, *Circulation*. 129 (18) (2014) 1821–1831.
- [8] M. Hilvo, P.J. Meikle, E.R. Pedersen, et al., Development and validation of a ceramide- and phospholipid-based cardiovascular risk estimation score for coronary artery disease patients, *Eur. Heart J.* 41 (3) (2020) 371–380.
- [9] E. Cavus, M. Karakas, F.M. Ojeda, et al., Association of circulating metabolites with risk of coronary heart disease in a European population: results from the biomarkers for cardiovascular risk assessment in Europe (BiomarCaRE) consortium, *JAMA Cardiol.* (2019) 1–10.
- [10] C. Fernandez, M.A. Surma, C. Klose, et al., Plasma lipidome and prediction of type 2 diabetes in the population-based malmo diet and cancer cohort, *Diabetes Care* 43 (2) (2020) 366–373.
- [11] M. Rosvall, L. Janzon, G. Berglund, G. Engstrom, B. Hedblad, Incident coronary events and case fatality in relation to common carotid intima-media thickness, *J. Intern. Med.* 257 (5) (2005) 430–437.
- [12] J.F. Ludvigsson, E. Andersson, A. Ekbom, et al., External review and validation of the Swedish national inpatient register, *BMC Public Health* 11 (2011) 450.
- [13] M.A. Surma, R. Herzog, A. Vasilj, et al., An automated shotgun lipidomics platform for high throughput, comprehensive, and quantitative analysis of blood plasma intact lipids, *Eur. J. Lipid Sci. Technol.* 117 (10) (2015) 1540–1549.
- [14] R. Herzog, K. Schuhmann, D. Schwudke, et al., LipidXplorer: a software for consensual cross-platform lipidomics, *PLoS One* 7 (1) (2012), e29851.
- [15] L. Aimo, R. Liechti, N. Hyka-Nouspikel, et al., The Swiss lipids knowledge base for lipid biology, *Bioinformatics*. 31 (17) (2015) 2860–2866.
- [16] E.A. Thevenot, A. Roux, Y. Xu, E. Ezan, C. Junot, Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses, *J. Proteome Res.* 14 (8) (2015) 3322–3335.
- [17] D.J. Stekhoven, P. Bühlmann, MissForest—non-parametric missing value imputation for mixed-type data, *Bioinformatics*. 28 (1) (2012) 112–118.
- [18] mlr, Machine Learning in R, <http://jmlr.org/papers/v17/15-066.html> Accessed.
- [19] Y. Fan, F. Shi, J. Liu, et al., Selective reduction in the sphingomyelin content of atherogenic lipoproteins inhibits their retention in murine aortas and the subsequent development of atherosclerosis, *Arterioscler. Thromb. Vasc. Biol.* 30 (11) (2010) 2114–2120.
- [20] M. Sugimoto, Y. Shimizu, S. Zhao, et al., Characterization of the role of sphingomyelin synthase 2 in glucose metabolism in whole-body and peripheral tissues in mice, *Biochim Biophys Acta* 1861 (8 Pt A) (2016) 688–702.
- [21] Z. Li, H. Zhang, J. Liu, et al., Reducing plasma membrane sphingomyelin increases insulin sensitivity, *Mol. Cell Biol.* 31 (20) (2011) 4205–4218.
- [22] F. Ottosson, E. Smith, W. Gallo, C. Fernandez, O. Melander, Purine metabolites and carnitine biosynthesis intermediates are biomarkers for incident type 2 diabetes, *J. Clin. Endocrinol. Metab.* 104 (10) (2019) 4921–4930.
- [23] T. Fall, S. Salihovic, S. Brandmaier, 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*. 59 (10) (2016) 2114–2124.
- [24] M. Stenemo, A. Ganna, S. Salihovic, et al., The metabolites urobilin and sphingomyelin (30:1) are associated with incident heart failure in the general population, *ESC Heart Fail.* 6 (4) (2019) 764–773.
- [25] A.M. Poss, J.A. Maschek, J.E. Cox, et al., Machine learning reveals serum sphingolipids as cholesterol-independent biomarkers of coronary artery disease, *J. Clin. Invest.* 130 (3) (2020) 1363–1376.
- [26] M.J. Gerl, C. Klose, M.A. Surma, et al., Machine learning of human plasma lipidomes for obesity estimation in a large population cohort, *PLoS Biol.* 17 (10) (2019), e3000443.
- [27] N. Sattar, D. Preiss, H.M. Murray, et al., Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials, *Lancet*. 375 (9716) (2010) 735–742.
- [28] R.J. Perry, V.T. Samuel, K.F. Petersen, G.I. Shulman, The role of hepatic lipids in hepatic insulin resistance and type 2 diabetes, *Nature*. 510 (7503) (2014) 84–91.
- [29] P.J. Meikle, G. Wong, D. Tsorotes, et al., Plasma lipidomic analysis of stable and unstable coronary artery disease, *Arterioscler. Thromb. Vasc. Biol.* 31 (11) (2011) 2723–2732.
- [30] D.D. Wang, E. Toledo, A. Hruby, et al., Plasma ceramides, Mediterranean diet, and incident cardiovascular disease in the PREDIMED trial (Prevencion con Dieta Mediterranea), *Circulation*. 135 (21) (2017) 2028–2040.
- [31] R. Tabassum, J.T. Ramo, P. Ripatti, et al., Genetic architecture of human plasma lipidome and its link to cardiovascular disease, *Nat. Commun.* 10 (1) (2019) 4329.