A metabolomic profile is associated with the risk of incident coronary heart disease

Anika A.M. Vaarhorst, MSc,[a,s] Aswin Verhoeven, PhD,[b,s] Claudia M. Weller, MD,[c,s] Stefan Böhringer, MD, PhD,[d] Sibel Göraler, MSc,[b] Axel Meissner, PhD,[b] André M. Deelder, PhD,[b] Peter Henneman, PhD,[c] Anton P.M. Gorgels, MD, PhD,[e] Piet A. van den Brandt, PhD,[f,g] Leo J. Schouten, MD, PhD,[g] Marleen M. van Greevenbroek, PhD,[h] Audrey H.H. Merry, PhD,[f] W.M. Monique Verschuren, PhD,[i] Arn M.J.M. van den Maagdenberg, PhD,[c,j] Ko Willems van Dijk, PhD,[c,k] Aaron Isaacs, PhD,[l] Dorret Boomsma, PhD,[m] Ben A. Oostra, PhD,[l] Cornelia M. van Duijn, PhD,[l] J. Wouter Jukema, MD, PhD,[n,o,p] Jolanda M.A. Boer, PhD,[i] Edith Feskens, PhD,[q] Bastiaan T. Heijmans, PhD,[a] and P. Eline Slagboom, PhD[a,r] ***Leiden, Maastricht, Bilthoven, Rotterdam, Amsterdam, Utrecht, and Wageningen, the Netherlands***

Background Metabolomics, defined as the comprehensive identification and quantification of low-molecular-weight metabolites to be found in a biological sample, has been put forward as a potential tool for classifying individuals according to their risk of coronary heart disease (CHD). Here, we investigated whether a single-point blood measurement of the metabolome is associated with and predictive for the risk of CHD.

Methods and results We obtained proton nuclear magnetic resonance spectra in 79 cases who developed CHD during follow-up (median 8.1 years) and in 565 randomly selected individuals. In these spectra, 100 signals representing 36 metabolites were identified. Applying least absolute shrinkage and selection operator regression, we defined a weighted metabolite score consisting of 13 proton nuclear magnetic resonance signals that optimally predicted CHD. This metabolite score, including signals representing a lipid fraction, glucose, valine, ornithine, glutamate, creatinine, glycoproteins, citrate, and 1.5-anhydrosorbitol, was associated with the incidence of CHD independent of traditional risk factors (TRFs) (hazard ratio 1.50, 95% CI 1.12-2.01). Predictive performance of this metabolite score on its own was moderate (*C*-index 0.75, 95% CI 0.70-0.80), but after adding age and sex, the *C*-index was only modestly lower than that of TRFs (*C*-index 0.81, 95% CI 0.77-0.85 and *C*-index 0.82, 95% CI 0.78-0.87, respectively). The metabolite score was also associated with prevalent CHD independent of TRFs (odds ratio 1.59, 95% CI 1.19-2.13).

Conclusion A metabolite score derived from a single-point metabolome measurement is associated with CHD, and metabolomics may be a promising tool for refining and improving the prediction of CHD. (Am Heart J 2014;168:45-52.e7.)

Over the last 50 years, risk factors that are robustly and independently associated with coronary heart disease (CHD), including lipid levels, blood pressure, lifestyle factors, family history, sex, and age, were identified.[1,2] Based on these traditional risk factors (TRFs), scores have been developed to predict CHD risk for an individual.[1,2] The discriminatory capabilities for these scores, as assessed by the C-index, ranges from 0.71 to 0.84.[1]

From the [a]Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands, [b]Department of Parasitology, Leiden University Medical Center, Leiden, The Netherlands, [c]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands, [d]Department of Medical Statistics, Leiden University Medical Center, Leiden, The Netherlands, [e]Department of Cardiology, Maastricht University Medical Centre, Maastricht, The Netherlands, [f]Department of Epidemiology (CAPHRI School for Public Health and Primary Care), Maastricht University, Maastricht, The Netherlands, [g]Department of Epidemiology (GROW School of Oncology and Developmental Biology), Maastricht University, Maastricht, The Netherlands, [h]Department of Internal Medicine (CARIM School for Cardiovascular diseases), Maastricht University Medical Centre, Maastricht, The Netherlands, [i]National Institute for Public Health and the Environment, Bilthoven, The Netherlands, [j]Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands, [k]Department of Endocrinology, Leiden University Medical Center, Leiden, The Netherlands, [l]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands, [m]Biological Psychology, VU University, Amsterdam, The Netherlands, [n]Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands, [o]The Durrer Center for Cardiogenetic Research, Amsterdam, The Netherlands, [p]Interuniversity Cardiology Institute of the Netherlands (ICIN), Utrecht, The Netherlands, [q]Division of Human Nutrition, Wageningen University and Research Center, Wageningen, The Netherlands, and [r]Netherlands Consortium for Healthy Ageing, Leiden, The Netherlands.

[s]These authors contributed equally to this work.

Submitted May 16, 2013; accepted January 24, 2014.

Reprint requests: Anika Vaarhorst, MSc/Eline Slagboom, PhD, Molecular Epidemiology Section, Leiden University Medical Center, PO Box 9600, Zone S5-P, 2300 RC Leiden, the Netherlands.

E-mails: a.a.m.vaarhorst@lumc.nl, P.Slagboom@lumc.nl

0002-8703

© 2014, Mosby, Inc. All rights reserved.

http://dx.doi.org/10.1016/j.ahj.2014.01.019

Metabolomics refers to the identification and quantification of low-molecular-weight metabolites in a biological sample.[3] Recent technological developments made it possible to generate metabolomic profiles of blood samples consisting of 10s to 100s metabolites in a single measurement.[4] These profiles are considered to be promising tools to efficiently capture the predictive information of TRFs and may potentially contribute to further improvement of primary CHD risk prediction.[3]

Several studies have attempted to use a metabolomic approach to diagnose prevalent CHD[5,6] or to predict incident CHD events in individuals free from cardiovascular disease[7-9] or diagnosed with CHD[10-13] or diabetes.[14] According to some studies a metabolomic approach might improve CHD risk prediction.[8-13] For example, in 1 study, proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy improves the prediction of subclinical atherosclerosis in comparison with conventional lipid testing.[8] Other studies found that a baseline metabolomic profile based on mass spectroscopy was associated with incident cardiovascular events in patients diagnosed with CHD[11] or in patients with suspected CHD.[10,12,13] However, none of these studies investigated if a metabolite profile based on low-molecular-weight molecules identified by ^{1}H-NMR spectroscopy could predict incident CHD, defined as an acute myocardial infarction (MI), unstable angina pectoris (UAP), or dead because of CHD, in individuals free from cardiovascular disease. Therefore, we studied the association of an ^{1}H-NMR based metabolite profile with incident CHD in a prospective case-cohort study. Subsequently, a second study was performed in an independent population to test if the selected metabolites were also relevant to classify prevalent CHD.

Materials and methods

Study populations

Primary study. We conducted a prospective case-cohort study within the Monitoring Project on Chronic Disease Risk Factors 1993 to 1997,[15] 1 of the 2 monitoring studies that were included in the Cardiovascular Registry Maastricht study.[16] In total, 6,459 men and women, between 20 and 59 years old at the moment of inclusion, had given informed consent to retrieve information from the municipal registries and from the general practitioner and specialist. The study complied with the Declaration of Helsinki and was approved by the Medical Ethics Committee of TNO (Dutch Organization for Applied Scientific Research). We excluded participants who were younger than 30 years at baseline (n = 1,301); who had had an acute MI, UAP, a coronary artery bypass graft, or a percutaneous transluminal coronary angioplasty before baseline (n = 69); or were lost to follow-up (n = 15), resulting in an eligible cohort of 5,074 participants.

Subcohort selection. From the eligible cohort, a subcohort of 738 participants was randomly drawn. This took place before cardiologic follow-up. EDTA plasma was unavailable for 92 participants, and ^{1}H-NMR analysis failed in 19 participants. For 62 participants, information on TRFs (ie, total cholesterol [TC], high-density lipoprotein cholesterol [HDL-C], systolic blood pressure [SBP], current smoking, body mass index [BMI], current diabetes status, and a parental history of MI) was incomplete, resulting in a subcohort of 565 participants.

Cardiologic follow-up. The cardiologic follow-up has been described in detail earlier[16] and ended on December 31, 2003, with a median follow-up of 8.1 years (range 0.2-10.9 years). During follow-up, 125 participants developed CHD (acute MI n = 55, UAP n = 51, dead due to CHD n = 19 [*International Classification of Diseases*, *Ninth Revision*, 410-414 and *International Statistical Classification of Diseases*, *10th Revision*, I20-I25]). For 31 patients, EDTA plasma for ^{1}H-NMR analysis was unavailable, and in 1 patient, ^{1}H-NMR analysis failed. For 14 patients, information on TRFs was incomplete, resulting in 79 patients.

Determination of TRFs. At baseline, participants filled in a questionnaire on medical history (including self-reported diabetes), parental history of MI (defined as no parents with MI, 1 parent with MI, or both parents with MI), and lifestyle factors (including smoking). During a medical examination, information on SBP and BMI was collected, and nonfasting EDTA blood samples were taken.[17,18] The blood was centrifuged for 10 minutes at 1000 rpm at 4°C, and EDTA plasma aliquots were stored at −80°C in tubes of 0.5 mL for future analysis or at −20°C for cholesterol determinations. High-density lipoprotein cholesterol and TC levels were determined in the plasma samples stored at −20°C using a Cholesterol oxidase/4-aminophenazone (CHOD-PAP) method.[19]

Secondary study. The Erasmus Rucphen Family study is a population-based study in a genetically isolated community in the Southwest of the Netherlands and includes 3,465 individuals, who are living descendants of 22 couples that had at least 6 children baptized in the community church between 1850 and 1900. Details are described elsewhere.[20] The study was approved by the Medical Ethical Committee of the Erasmus Medical Center in Rotterdam, and all participants gave written informed consent.

Determination of TRFs

All participants filled in questionnaires about lifestyle, personal, and family medical history. During personal interviews performed by study physicians, information on lifestyle factors, medication use, and personal and family medical history was collected. Physical examinations were performed, including measurements of SBP and BMI.[21] In addition, fasted blood samples were taken. An electrocardiogram was performed and scored by an experienced cardiologist. Plasma concentrations of

HDL-C and TC were determined according to standard procedures.[22] *Diabetes* was defined as the use of blood glucose-lowering medication and/or fasting glucose levels of ≥7 mmol/L.

Diagnosis of CHD in the secondary study. Participants were classified as CHD cases if they indicated during the interview or in the questionnaire that they had experienced an MI or underwent a coronary revascularization procedure, they reported angina symptoms in the interview, and/or showed signs of MI on electrocardiogram.

From 2,919 participants, fasting serum samples were available. Good-quality ^{1}H-NMR spectra from 2,415 participants were obtained. For 2,327 of these participants, data on CHD diagnosis were available, and 170 were classified as having CHD.

^{1}H-NMR metabolite profiling

The stored EDTA plasma and serum samples were thawed at 4°C and were mixed by inverting the tubes 10 times. Next, samples (300 μL) were mixed with 300 μL sodium 3-[trimethylsilyl] d_4-propionate buffer (see online Appendix Supplementary material) and transferred into 5-mm nuclear magnetic resonance tubes and kept at 6°C while queued for acquisition. Two-dimensional J-resolved and Carr-Purcell-Meiboom-Gill spectra were acquired on a 600-MHz Bruker Avance II spectrometer (Bruker BioSpin, Karlsruhe, Germany), operating at a sample temperature of 310 K. For details on acquisition, processing, quality control, scaling, and calibration of the ^{1}H-NMR spectra, see online Appendix Supplementary methods.

Using the procedure described in the online Appendix Supplementary material, 100 signals were detected and quantified in the ^{1}H-NMR spectra of every individual in the primary study. For 76 signals, metabolites were assigned (see online Appendix Supplementary methods and online Appendix Supplementary Table I for all ^{1}H-NMR signals and their assigned metabolites). These signals represented 36 different compounds (ie, after subtracting the signals representing free EDTA). Signals representing calcium-EDTA and magnesium-EDTA complexes may give an indication of the levels of calcium and magnesium ions, respectively. Using the same procedure, 68 of 100 signals detected in the primary study were detected in the secondary study. For 54 of these 68 signals, metabolites could be assigned, representing 28 different compounds.

Statistical analysis

In the primary study, Cox regression, adjusted for delayed entry, and according to the method of Prentice to adjust for the case-cohort design was performed to see whether baseline characteristics were associated with incident CHD.[23] Age in years was used as the time-scale variable.

Before analysis, the ^{1}H-NMR signals were transformed to Z-scores. We selected a subset of the most informative signals for CHD prediction, using least absolute shrinkage and selection operator (LASSO) regression,[24] and performed 10-fold cross-validation to determine the tuning parameter.[24,25] This set was further reduced to signals that could be detected in both studies. The linear predictor of the Cox model was used as a weighted metabolite score (sum of regression coefficients multiplied by the corresponding covariate values).

Cox regression, according to the method of Prentice to adjust for the case-cohort design, was used to calculate whether this metabolite score was associated with incident CHD before and after adjusting for TRFs.[23]

To investigate whether the metabolite score improved risk discrimination, Harrell's concordance index (*C*-index),[26] the net reclassification index (NRI), and the integrative discrimination index (IDI) were calculated.[27] For the NRI, the following risk categories were applied: 0% to <5%, 5% to <10%, 10% to <20%, ≥20%. Next, using analysis of variance, we investigated to what extent TRFs can explain the variance in the metabolite score in the subcohort. Stata/SE version 11.2 was used to calculate the *C*-index. R-package PredictABEL version 1.2-1 was used to calculate the IDI. For all other analyses, R version 2.14.1 was used.

In the secondary study, raw ^{1}H-NMR signal data were adjusted for kinship by linear regression in GenABEL.[28] The residuals for all signals were transformed into Z-scores. Logistic regression was performed to assess the association of TRFs with prevalent CHD. To assess whether the metabolite signals selected by LASSO regression in the primary study were also relevant for the identification of prevalent CHD cases, we put these signals into one logistic regression model to determine their individual regression coefficients. Next, we used logistic regression analysis to test the association of this metabolite score, thus with weights based on the secondary study, with prevalent CHD. Finally, we tested to what extent TRFs can explain the variance in the metabolite score using analysis of variance. For these analyses, PASW statistics version 18 (SPSS-IBM, Chicago, IL) was used.

This study was funded by a grant from the Netherlands Heart Foundation (2006B195), the Netherlands Consortium for Healthy Ageing (05060810) and by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO 184.021.007).

The Monitoring Project on Chronic Disease Risk Factors was financially supported by the Ministry of Public Health, Welfare and Sports of the Netherlands.

The Erasmus Rucphen Family study was funded by the Centre for Medical Systems Biology in the framework of the Netherlands Genomics Initiatives and by the European Network of Genomic and Genetic Epidemiology consortium.

Table I. Traditional risk factors and their association with incident CHD in the primary study

	Cases (n = 79)	Subcohort (n = 565)*	HR (95% CI)†	HR (95% CI)‡
Age (y)	51.9 ± 6.1	44.8 ± 8.3	–	–
TC (mmol/L)	5.9 ± 1.0	5.3 ± 1.0	1.30 (1.03-1.64)	1.21 (0.93-1.58)
HDL-C (mmol/L)	1.1 ± 0.3	1.3 ± 0.4	0.12 (0.05-0.30)	0.34 (0.12-0.94)
SBP (mm Hg)	134.5 ± 16.8	121.5 ± 15.1	1.04 (1.02-1.05)	1.03 (1.01-1.05)
BMI (kg/m^2)	27.6 ± 4.6	25.4 ± 3.8	1.10 (1.03-1.17)	1.03 (0.95-1.11)
Men	79.6% (63)	44.6% (252)	4.95 (2.72-9.02)	3.30 (1.65-6.57)
Current smoking	51.9% (41)	38.8% (219)	2.16 (1.29-3.62)	1.84 (1.02-3.32)
Diabetes	3.8% (3)	0.5% (3)	3.99 (0.70-22.84)	3.34 (0.41-27.43)
Parental history of MI	50.6% (40)	40.2% (227)	1.32 (0.88-1.98)	1.24 (0.78-1.98)

Data are expressed as mean ± SD or % (n).
*Including 10 cases.
†Univariate HR was calculated per unit increase for age, TC, HDL-C, SBP, BMI, and for the categorical traits. Age in years was used as the time-scale variable.
‡All variables were added into 1 multivariable Cox proportional hazards model, except for age, which was used as the time-scale variable.

The funding sources had no involvement in the design and conduct of the study or with the collection, management, analysis, or interpretation of the data and reporting or the decision to submit the manuscript for publication.

Results

Primary study

The TRFs were associated with incident CHD. When all TRFs were entered into a Cox proportional hazards model with age in years as the time-scale variable, HDL-C, SBP, sex, and current smoking remained independently associated with incident CHD (Table I). See online Appendix Supplementary Table II for the baseline characteristics before excluding individuals with missing data. For the association of the individual ^{1}H-NMR signals with incident CHD, see online Appendix Supplementary Table I.

We determined 16 ^{1}H-NMR signals as the best prediction subset using LASSO regression, of which 13 were available in the secondary study (Figure 1). These 16 ^{1}H-NMR signals represent creatinine, serine, glucose, 1,5-anhydrosorbitol, trimethylamine N-oxide (TMAO), ornithine, citrate, glutamate, glycoproteins, an unsaturated lipid structure, valine, and 5 nonannotated signals located at 3.924, 3.145, 2.412, 1.391, and 0.988 ppm. From the 13 signals present in both the primary and secondary study, a weighted metabolite score was constructed using the corresponding coefficients (Figure 1). This metabolite score was normally distributed in cases and subcohort members (online Appendix Supplementary Figure) and associated with incident CHD (hazard ratio [HR] per SD 1.91, 95% CI 1.50-2.44). After adjusting for TRFs, this metabolite score remained associated with incident CHD (HR/SD 1.50, 95% CI 1.12-2.01). For the results of the metabolite score based on the 16 signals, see online Appendix Supplementary Table III.

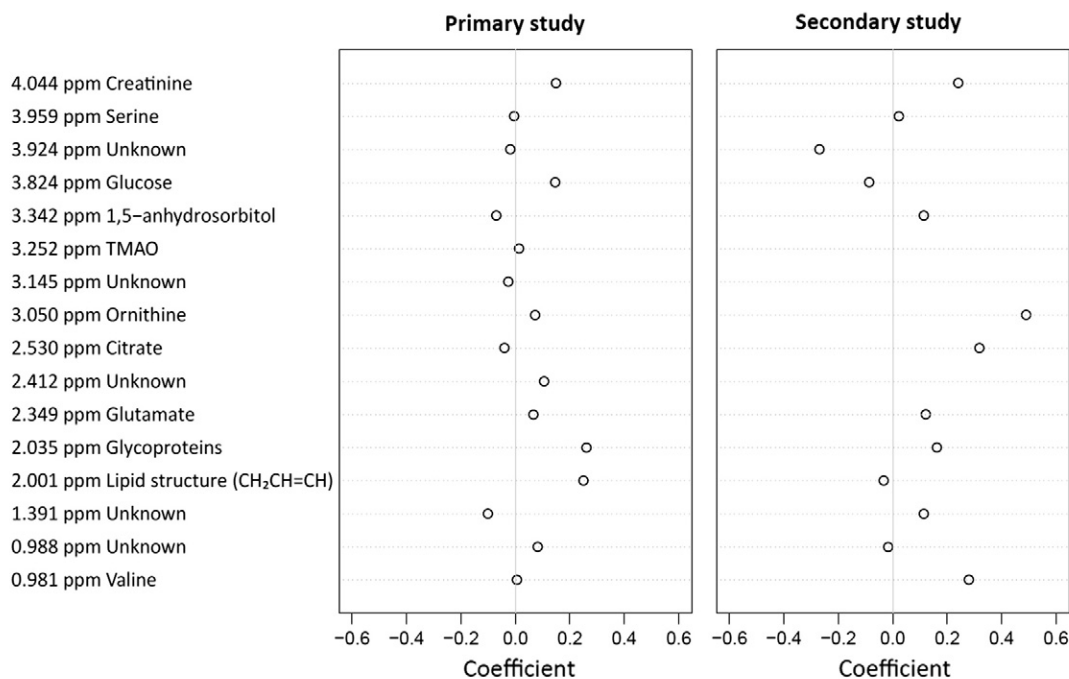
The metabolite score had a *C*-index of 0.75 (95% CI 0.70-0.80). Adding age and sex to the metabolite score, resulted in a *C*-index of 0.81 (95% CI 0.77-0.85), which is similar to a *C*-index when only TRFs are included (*C*-index 0.82, 95% CI 0.78-0.87, P = .327). When the metabolite score was added to a model containing all TRFs, the *C*-index increased from 0.82 to 0.84, which was nonsignificant (P = .107). Both the improvement in the NRI (NRI_{total} = 0.038, P = .209) and the IDI (0.012, P = .091) were nonsignificant (online Appendix Supplementary Tables IV and V). Inspecting *C*-indices for individual TRFs and evaluating improvement of adding the metabolite score indicated that the metabolite score improved the *C*-indices of all individual TRFs (online Appendix Supplementary Table VI).

We tested to what extent TRFs explain the variance in the metabolite score in the subcohort of the primary study. High-density lipoprotein cholesterol, sex, BMI, TC, SBP, age, and diabetes explained respective 16.4%, 11.7%, 10.0%, 7.9%, 5.6%, 4.1%, and 2.3% of the variance in the metabolite score. Current smoking and parental history of MI all explained <1% of the variance in the metabolite score (Figure 2). When all TRFs were combined, 32.6% of the variance in the metabolite score was explained.

Secondary study

To test if the 13 metabolite signals selected in the primary study were relevant for the identification of prevalent CHD cases, we investigated 170 CHD cases and 2,157 controls for which equivalent metabolomics profiles were obtained. In this nonprospective cohort, combining all TRFs in 1 logistic regression model resulted in only age, sex, and parental history of MI to be independently associated with prevalent CHD (Table II). See online Appendix Supplementary Table VII for the baseline characteristics before excluding individuals with missing data.

Figure 1



The subset of signals selected using 10-fold cross-validated LASSO regression and their coefficients in the primary study (left panel) and the secondary study (right panel).

The metabolite score, based on the 13 best predicting signals in the case-cohort study, but with weights based on the secondary study (Figure 2), was associated with prevalent CHD before (odd ratio [OR] 2.72, $P < .001$) and after adjusting for TRFs (OR 1.59, P = .002). After excluding statin users (n = 299), similar results were obtained (online Appendix Supplementary Table VIII).

The proportion of variance in the metabolite score explained by age, SBP, and diabetes was higher for the secondary study than for primary study, whereas the variables TC, HDL-C, and BMI explained a lower proportion of the variance (Figure 2). The proportion of variance explained by sex, current smoking, and parental history of MI was comparable for both studies. With all TRFs combined, 27.3% of the variance of the metabolite score could be explained, which is comparable with that of the primary study.

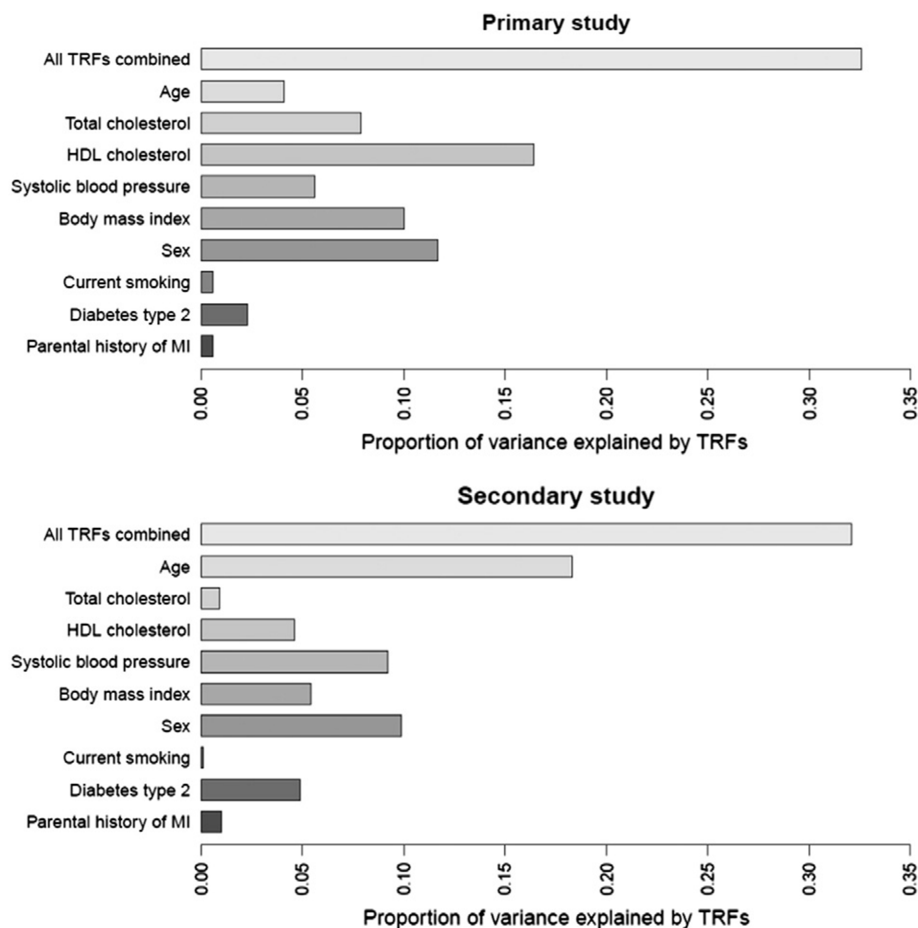
Discussion

A metabolite score, based on ^{1}H-NMR spectroscopy, is significantly associated with incident CHD independent of TRFs. When combined with age and sex, this score was as predictive for incident CHD as all TRFs combined. A score based on the same ^{1}H-NMR signals was also associated with prevalent CHD, independent of TRFs.

The observation that the metabolite score could not improve CHD risk prediction beyond TRFs in individuals free from CHD is in line with a previous study published by El Harchaoui et al.[7] However, Würtz et al[8] found that metabolites measured by ^{1}H-NMR spectroscopy improved risk stratification for subclinical atherosclerosis in comparison with conventional lipids. In this study, both ^{1}H-NMR determined lipoproteins and low-molecular-weight metabolites were included.[8] We only included information on low-molecular-weight metabolites, whereas in the study by El Harchaoui et al, only information on lipoproteins was included. Perhaps the combination of both lipoproteins and low-molecular-weight metabolites results in the optimal prediction of CHD.

The metabolite score represents the metabolites valine, ornithine, glucose, 1.5-anhydrosorbitol, creatinine, an unsaturated lipid structure, glutamate, glycoproteins, citrate, and TMAO, of which TMAO was not available in the secondary study. Most of these metabolites have been associated with CHD or CHD risk factors before.[11,13,29-31] Valine has been associated with metabolic risk factors,[30] insulin resistance,[32] incident type 2 diabetes,[31] and future cardiovascular events.[11] Ornithine is produced by splitting of urea from arginine, resulting in a lower bioavailability of arginine. Arginine is necessary to produce nitric oxide, which is essential for a normal

Figure 2



The proportion of variance in the metabolite score explained by TRFs in the primary study (upper panel) and the secondary study (lower panel).

endothelial function.[29] This pathway has been linked to CHD[10] and CHD mortality.[29] The presence of glucose and 1,5-anhydrosorbitol, a short-term marker for glycemic control,[33] could indicate that our metabolite score marks individuals at higher risk for developing diabetes or insulin resistance and thereby CHD. Low creatinine levels are a marker for high HDL-C and low low-density lipoprotein cholesterol levels.[34] Thus, the presence of creatinine and an unsaturated lipid structure could indicate that our metabolite risk score is a marker for an unfavorable lipid profile. This is confirmed by the proportion of variance explained by HDL-C and TC levels, 7.9% and 16.4%. In the secondary study, these explained variances are only 0.9% and 4.6%, but this discrepancy might be caused by statin treatment, resulting in lower cholesterol levels for the cases compared with controls. A secondary explanation for the incorporation of creatinine in the metabolite score is that elevated creatinine levels may indicate kidney dysfunction, which is associated with cardiovascular disease.[35] Increased TMAO levels have been associated with cardiovascular risk before.[13,36] Thus, it seems that the LASSO procedure selected relevant metabolites that have been associated with CHD and CHD risk factors before.

Several issues have to be resolved before it can be concluded if a metabolomic approach is useful for CHD risk prediction. First, the known, quantifiable serum metabolome consists of 4,229 metabolites,[37] of which only 36 (0.9%) were included in this study. Other studies that use ^{1}H-NMR spectroscopy also incorporated lipoproteins.[11,12] Therefore, we hope that we can achieve better in follow-up studies when incorporating H-NMR–determined lipoproteins in addition to low-molecular-weight metabolites in our analysis. Moreover, other metabolomic platforms should be also be included.[37] Second, the 16 signals provided by our study should be measured in large prospective cohorts for replication and to determine universally applicable weights. The current

Table II. Traditional risk factors and their association with prevalent CHD in the secondary study

	Cases (n = 170)	Controls (n = 2157)	OR (95% CI)*	OR (95% CI)†
Age (y)	60.9 ± 11.7	47.7 ± 14.0	1.08 (1.07-1.10)	1.08 (1.05-1.10)
TC (mmol/L)	5.0 ± 1.1	5.6 ± 1.1	0.57 (0.49-0.67)	0.47 (0.37-0.59)
HDL-C (mmol/L)	1.2 ± 0.3	1.3 ± 0.4	0.30 (0.18-0.49)	1.25 (0.60-2.61)
SBP (mm Hg)	148.2 ± 22.9	139.1 ± 19.7	1.02 (1.02-1.03)	1.00 (0.99-1.01)
BMI (kg/m^2)	28.1 ± 4.4	26.8 ± 4.6	1.06 (1.03-1.09)	1.01 (0.96-1.07)
Men	61.2% (104)	42.5% (917)	2.40 (1.79-3.22)	1,93 (1.18-3.14)
Current smoking	36.5% (62)	39.1% (844)	0.90 (0.66-1.24)	1.18 (0.75-1.88)
Diabetes	12.9% (22)	4.3% (93)	3.44 (2.15-5.49)	1.25 (0.56-2.79)
Parental history of MI	31.2% (53)	19.9% (430)	2.03 (1.53-2.71)	1.60 (1.16-2.21)

Data are expressed as mean ± SD or % (n).
* Univariate OR was calculated per unit increase for age, total cholesterol, HDL-C, SBP, BMI, and for the categorical traits.
† All the variables were added into 1 multivariable logistic regression model.

study is too small for that purpose. Third, we had nonfasting samples in the primary study and fasted samples in the secondary study. However, we still found that the ^{1}H-NMR signals selected in the primary nonfasted study were also associated with prevalent CHD in the fasted secondary study. This indicates that we have selected ^{1}H-NMR signals that are robust whether fasted or nonfasted samples are used. Fourth, constructing robust prediction models constitutes a statistical challenge, especially in a high-dimensional setting. In our case, model selection by LASSO regression resulted in predictor selection that eliminated high correlations among predictors. This can lead to reduced transferability of prediction models as correlation structures of predictors can vary between studies. A wide variety of penalized regression models are available (eg, elastic net, ridge regression), and further research is needed to select the appropriate methods for the application at hand.

Conclusion

A metabolite score derived from a single point metabolome measurement is associated with the risk of CHD independent of TRFs but does not improve risk prediction beyond TRFs. On the other hand, LASSO regression resulted in the selection of relevant metabolites, suggesting that more comprehensive metabolomic methods may be promising tools to further improve upon CHD disease understanding and risk stratification.

Acknowledgements

We thank the MORGEN-project steering committee consisting of (not mentioning coauthors of this article) H.B. Bueno de Mesquita, H.A. Smit, and J.C. Seidell. Furthermore, we thank the epidemiologists and field workers of the Municipal Health Services in Maastricht for their contribution to baseline data collection and those involved in logistic management (A. Jansen and J. Steenbrink) and data management (A. Blokstra, A. van Kessel, P. Steinberger, E. den Hoedt, I. Toxopeus, J. van der Laan). We thank A. Hilton, V. Visser, P. Erkens, S. Philippens, J. Bremen, and B. Bleijlevens for assistance in clinical data collection. Statistics Netherlands is acknowledged for providing data on causes of death. We thank Joris Deelen, Dennis Kremer, Ruud van der Breggen, Yingchang Lu, Mariken Tijhuis, and Cecile Povel for assistance with practical work.

We thank the participants from the Genetic Research in Isolated Populations, Erasmus Rucphen Family studies, as well as the general practitioner and other clinicians who made this work possible.

References

1. Conroy R. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J 2003;24(11): 987-1003.
2. Wilson PW, D'Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. Circulation 1998;97(18): 1837-47.
3. Ellis DI, Dunn WB, Griffin JL, et al. Metabolic fingerprinting as a diagnostic tool. Pharmacogenomics 2007;8(9):1243-66.
4. Beckonert O, Keun HC, Ebbels TM, et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nat Protoc 2007;2(11): 2692-703.
5. Kirschenlohr HL, Griffin JL, Clarke SC, et al. Proton NMR analysis of plasma is a weak predictor of coronary artery disease. Nat Med 2006;12(6):705-10.
6. Brindle JT, Antti H, Holmes E, et al. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nat Med 2002;8(12):1439-44.
7. El Harchaoui K, van der Steeg WA, Stroes ES, et al. Value of low-density lipoprotein particle number and size as predictors of coronary artery disease in apparently healthy men and women: the EPIC-Norfolk Prospective Population Study. J Am Coll Cardiol 2007;49(5):547-53.
8. Wurtz P, Raiko JR, Magnussen CG, et al. High-throughput quantification of circulating metabolites improves prediction of subclinical atherosclerosis. Eur Heart J 2012;33(18):2307-16.

9. Magnusson M, Lewis GD, Ericson U, et al. A diabetes-predictive amino acid score and future cardiovascular disease. Eur Heart J 2013;34(26):1982-9.
10. Wang Z, Tang WH, Cho L, et al. Targeted metabolomic evaluation of arginine methylation and cardiovascular risks: potential mechanisms beyond nitric oxide synthase inhibition. Arterioscler Thromb Vasc Biol 2009;29(9):1383-91.
11. Shah SH, Bain JR, Muehlbauer MJ, et al. Association of a peripheral blood metabolic profile with coronary artery disease and risk of subsequent cardiovascular events. Circ Cardiovasc Genet 2010;3(2): 207-14.
12. Shah SH, Sun JL, Stevens RD, et al. Baseline metabolomic profiles predict cardiovascular events in patients at risk for coronary artery disease. Am Heart J 2012;163(5):844-50. [e1].
13. Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 2011;472(7341):57-63.
14. Roussel R, Mentre F, Bouchemal N, et al. NMR-based prediction of cardiovascular risk in diabetes. Nat Med 2007;13(4):399-400.
15. Picavet HS, Schouten JS. Physical load in daily life and low back problems in the general population—the MORGEN study. Prev Med 2000;31(5):506-12.
16. Merry AH, Boer JM, Schouten LJ, et al. Validity of coronary heart diseases and heart failure based on hospital discharge and mortality data in the Netherlands using the cardiovascular registry Maastricht cohort study. Eur J Epidemiol 2009;24(5):237-47.
17. Lean MEJ, Han TS, Seidell JC. Impairment of health and quality of life in people with large waist circumference. Lancet 1998;351(9106):853-6.
18. Van Leer EM, Seidell JC, Kromhout D. Dietary calcium, potassium, magnesium and blood pressure in the Netherlands. Int J Epidemiol 1995;24(6):1117-23.
19. Kattermann R, Jaworek D, Moller G, et al. Multicentre study of a new enzymatic method of cholesterol determination. J Clin Chem Clin Biochem 1984;22(3):245-51.
20. Aulchenko YS, Heutink P, Mackay I, et al. Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet 2004;12(7):527-34.
21. Sayed-Tabatabaei FA, van Rijn MJ, Schut AF, et al. Heritability of the function and structure of the arterial wall: findings of the Erasmus Rucphen Family (ERF) study. Stroke 2005;36(11):2351-6.
22. van Gent CM, van der Voort HA, de Bruyn AM, et al. Cholesterol determinations. A comparative study of methods with special reference to enzymatic procedures. Clin Chim Acta 1977;75(2): 243-51.
23. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika 1986;73(1):1-11.
24. Goeman JJ. L1 penalized estimation in the Cox proportional hazards model. Biom J 2010;52(1):70-84.
25. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997;16(4):385-95.
26. Harrell Jr FE, Califf RM, Pryor DB, et al. Evaluating the yield of medical tests. JAMA 1982;247(18):2543-6.
27. Pencina MJ, D'Agostino Sr RB, D'Agostino Jr RB, et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27(2):157-72. [discussion 207–12].
28. Aulchenko YS, Ripke S, Isaacs A, et al. GenABEL: an R library for genome-wide association analysis. Bioinformatics 2007;23(10): 1294-6.
29. Sourij H, Meinitzer A, Pilz S, et al. Arginine bioavailability ratios are associated with cardiovascular mortality in patients referred to coronary angiography. Atherosclerosis 2011;218(1):220-5.
30. Cheng S, Rhee EP, Larson MG, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. Circulation 2012;125(18):2222-31.
31. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. Nat Med 2011;17(4):448-53.
32. Wurtz P, Soininen P, Kangas AJ, et al. Branched-chain and aromatic amino acids are predictors of insulin resistance in young adults. Diabetes Care 2013;36(3):648-55.
33. Dungan KM. 1,5-anhydroglucitol (GlycoMark) as a marker of short-term glycemic control and glycemic excursions. Expert Rev Mol Diagn 2008;8(1):9-19.
34. Bernini P, Bertini I, Luchinat C, et al. The cardiovascular risk of healthy individuals studied by NMR metabonomics of plasma samples. J Proteome Res 2011;10(11):4983-92.
35. Go AS, Chertow GM, Fan D, et al. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. N Engl J Med 2004;351(13):1296-305.
36. Tang WH, Wang Z, Levison BS, et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. N Engl J Med 2013;368(17):1575-84.
37. Psychogios N, Hau DD, Peng J, et al. The human serum metabolome. PLoS One 2011;6(2):e16957.

Appendix. A metabolomic profile is associated with the risk of incident coronary heart disease

Table of contents

Supplementary materials and methods 2
Acquisition and processing of ^{1}H-NMR spectra 2
Quality controls, scaling, and calibration of the H-NMR spectra 2
Supplementary Tables 5
Supplementary Table I 5
Supplementary Table II 17
Supplementary Table III 18
Supplementary Table IV 19
Supplementary Table V 20
Supplementary Table VI 21
Supplementary Table VII 23
Supplementary Table VIII 24
Supplementary Figure 25
References 26

Supplementary materials and methods

Acquisition and processing of ^{1}H-NMR spectra

All proton nuclear magnetic resonance (^{1}H-NMR) experiments were acquired on a 600-MHz Bruker Avance II spectrometer (Bruker BioSpin, Karlsruhe, Germany) equipped with a 5-mm triple resonance inverse (TCI) cryogenic probe head with Z-gradient system and automatic tuning and matching. All experiments were recorded at 310K. Temperature calibration was done before each batch of measurements using the method of Findeisen et al.[38] The duration of the $\pi/2$ pulses were automatically calibrated for each individual sample using a homonuclear-gated nutation experiment on the locked and shimmed samples after automatic tuning and matching of the probe head.[39]

The stored EDTA plasma and serum samples were thawed at 4°C and were mixed by inverting the tubes 10 times. Next, samples (300 μL) were mixed with 300 μL 75 mM disodium phosphate buffer in H_2O/D_2O (80/20) with a pH of 7.4 containing 6.15 mM NaN_3 and 4.64 mM sodium 3-[trimethylsilyl] d_4-propionate using a Gilson 215 liquid handler in combination with a Bruker SampleTrack system. Samples were transferred into 5-mm SampleJet NMR tubes in 96 tube racks using a modified Gilson 215 tube filling station and kept at 6°C on a SampleJet sample changer while queued for acquisition.

For water suppression, presaturation of the water resonance with an effective field of γB_1 = 25 Hz was applied during the relaxation delay.[40] J-resolved spectra (JRES)[41] were recorded with a relaxation delay of 2 seconds and a total of 1 scan for each increment in the indirect dimension. A data matrix of 40 × 12,288 data points was collected covering a sweep width of 78 × 10,000 Hz. A sine-shaped window function was applied, and the data were zero-filled to 256 × 16,384 complex data points before Fourier transformation. The resulting data matrix was tilted along the rows by shifting each row (k) by 0.4992 * (128 − k) points and symmetrized about the central horizontal lines to compensate for the skew of the multiplets in the F1 dimension. For T2-filtered [38]H NMR spectra, a standard 1D Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence,[42,43] was used with a relaxation delay of 4 seconds. A pulse train of 128 refocusing pulses with individual spin echo delays of 0.6 ms were applied resulting in a total T2 filtering delay of 78 ms. A total of 73,728 data points covering a spectral width of 12,019 Hz were collected using 16 scans. The Free Induction Decay (FID) was zero-filled to 131,072 complex data points, and an exponential window function was applied with a line broadening factor of 1.0 Hz before Fourier transformation. The spectra were automatically phased and baseline corrected.

Quality control, scaling, and calibration of the ^{1}H-NMR spectra

Further data processing was performed in Matlab (R2009a; The Mathworks, Inc, Natick, MA). The spectra and associated data were converted into Matlab files using in-house code. First, the spectra were combined into 1 file while removing superfluous information. For CPMG, this included dropping the imaginary part of the spectrum, whereas for the JRES spectra, the sum projection along the indirect dimension was taken. Quality control on the set of ^{1}H NMR spectra was carried out by examining a set of spectroscopic parameters such as shim values and intensity of the water signal and subsequently visually inspecting the spectra. The spectra that failed the quality control were not included for further analysis. The spectra were then scaled with respect to the sensitivity of the receiver coil, as determined from the pulse length that was automatically calibrated for each sample.[44] After subtracting a constant value as a simple baseline correction, the spectra were calibrated with respect to the anomeric resonance of α-D-glucose (δ = 5.23 ppm).[45] Because there are small deviations of the peak position in the different ^{1}H-NMR spectra, alignment was performed using the correlation optimized warping algorithm by Tomasi et al.[46] This was performed actively for the CPMG spectra, after which the same warping was applied to the JRES projection. Peaks in the JRES projection were picked by finding the signals that were above the surrounding spectral area by more than the estimated noise level. Peaks in different spectra were grouped according to similarity in peak position and intensity. The intensity and the position give good initial guesses for the deconvolution of the peaks by fitting groups of mixed Gaussian-Lorentzian line shapes to isolated spectral areas using a Simplex optimization algorithm. As the fitting algorithm incidentally fails to converge properly, values further from the median than 3 times the interquartile range are discarded. Using partial least squares regression, the remaining peak intensities

were used to build a linear model that predicts all intensities directly from the nonwarped spectrum, yielding also values for the cases where the peak picking failed or the deconvolution values were discarded, and eliminating the problem of faulty warping.

For 76 signals, metabolites were assigned using information from previously reported plasma/serum metabolites,[43,45,47,48] the Human Metabolome Database[49] and the Pearson correlation coefficients between the signal intensities.

Supplementary Table I. An overview of all peaks identified in the 2-dimensional J-resolved ^{1}H-NMR spectrum and their association with incident CHD in the primary prospective case-cohort study

		Cases		Subcohort		Association with CHD	
Chemical shift	**Assignment**	**Mean (SD)**	**Min: Max**	**Mean (SD)**	**Min: Max**	**HR (95% CI)**	***P***
0.873 ppm*	Lipids (CH3)†	0.45 (1.48)	−0.93: 6.67	−0.05 (0.94)	−0.97: 6.67	1.40 (1.13-1.74)	.002
0.929 ppm*	Isoleucine	0.27 (0.99)	−1.60: 2.89	−0.04 (1.00)	−2.02: 5.08	1.14 (0.90-1.45)	.282
0.947 ppm*	Leucine	0.30 (1.05)	−1.67: 3.83	−0.04 (1.00)	−1.72: 4.98	1.23 (0.98-1.55)	.074
0.952 ppm*	Unknown	0.01 (1.24)	−3.14: 3.00	0.00 (0.97)	−4.29: 3.21	0.99 (0.77-1.26)	.921
0.957 ppm*	Leucine	0.24 (1.00)	−1.78: 2.79	−0.04 (0.99)	−2.10: 4.07	1.17 (0.92-1.50)	.207
0.965 ppm	Unknown	0.07 (1.13)	−3.56: 2.45	−0.02 (1.00)	−3.56: 4.76	0.97 (0.75-1.24)	.780
0.973 ppm*	Unknown	0.03 (1.00)	−2.35: 2.93	−0.01 (1.00)	−3.13: 3.77	0.93 (0.71-1.20)	.568
0.981 ppm*	Valine	0.33 (0.98)	−1.65: 3.26	−0.05 (0.99)	−2.37: 4.43	1.23 (0.97-1.57)	.090
0.988 ppm*	Unknown	0.24 (0.87)	−1.88: 2.69	−0.03 (1.01)	−7.36: 4.19	1.22 (0.96-1.55)	.102
1.000 ppm*	Isoleucine	0.27 (0.96)	−1.79: 2.48	−0.03 (1.00)	−2.07: 4.31	1.17 (0.92-1.49)	.214
1.032 ppm*	Valine	0.29 (1.02)	−1.88: 3.26	−0.04 (0.99)	−2.35: 4.36	1.18 (0.92-1.51)	.193
1.054 ppm	Unknown	0.26 (1.38)	−1.94: 5.02	−0.04 (0.93)	−1.66: 3.99	1.21 (0.95-1.55)	.131
1.063 ppm*	Ketoisovalerate	0.12 (1.27)	−1.89: 7.24	−0.01 (0.96)	−1.90: 4.89	1.09 (0.84-1.41)	.509
1.163 ppm*	Unknown	0.14 (1.77)	−0.18: 15.48	−0.02 (0.83)	−0.21: 14.04	1.11 (0.94-1.31)	.235
1.174 ppm*	Ethanol	0.14 (1.63)	−0.13: 14.20	−0.02 (0.87)	−0.13: 14.52	1.10 (0.93-1.29)	.280
1.191 ppm*	3-Hydroxybutyrate	0.08 (0.96)	−0.57: 6.77	−0.02 (1.00)	−0.58: 9.75	1.10 (0.90-1.34)	.362
1.212 ppm*	Unknown	0.15 (1.44)	−1.33: 10.34	−0.02 (0.92)	−2.03: 5.48	1.17 (0.91-1.52)	.228
1.264 ppm*	Lipids (CH2)†	0.42 (1.39)	−0.91: 5.79	−0.05 (0.95)	−0.94: 6.23	1.36 (1.10-1.69)	.005
1.319 ppm*	Lactate	0.20 (1.06)	−1.51: 4.43	−0.02 (0.99)	−1.90: 4.41	1.22 (0.95-1.56)	.120
1.391 ppm*	Unknown	−0.14 (1.02)	−1.99: 2.69	0.02 (1.00)	−2.60: 4.35	0.78 (0.60-1.00)	.053
1.407 ppm*	Unknown	0.01 (1.00)	−1.78: 2.19	0.00 (1.00)	−2.53: 3.24	1.02 (0.79-1.32)	.904
1.425 ppm*	Unknown	0.17 (1.06)	−1.93: 2.78	−0.02 (1.00)	−1.98: 3.51	1.18 (0.92-1.52)	.198
1.471 ppm*	Alanine	0.22 (0.93)	−1.72: 2.54	−0.03 (1.00)	−2.35: 3.45	1.13 (0.88-1.45)	.344
1.706 ppm*	Unknown	0.24 (1.20)	−1.77: 4.62	−0.03 (0.98)	−2.67: 4.62	1.21 (0.92-1.58)	.172
1.908 ppm*	Acetate	0.13 (0.88)	−0.77: 3.84	−0.02 (1.01)	−0.81: 16.58	1.04 (0.87-1.24)	.656
2.001 ppm*	Lipids (CH*2CH=CH)†	0.49 (1.72)	−1.01: 8.67	−0.05 (0.91)	−1.19: 8.67	1.37 (1.13-1.67)	.001
2.035 ppm*	Glycoproteins	0.52 (1.18)	−1.51: 3.85	−0.06 (0.96)	−2.05: 4.88	1.32 (1.03-1.69)	.028
2.066 ppm*	glycoproteins	0.32 (1.12)	−1.59: 4.33	−0.04 (0.99)	−2.51: 4.33	1.20 (0.93-1.55)	.159
2.099 ppm*	Glutamine/glutamate	−0.01 (0.88)	−1.85: 2.77	−0.01 (1.01)	−2.43: 3.58	0.97 (0.76-1.22)	.773
2.125 ppm*	Glutamine/glutamate	0.00 (1.05)	−2.41: 3.66	0.00 (1.01)	−2.48: 3.66	1.03 (0.81-1.32)	.791
2.221 ppm*	Lipids (CH2CO)†	0.40 (1.55)	−0.76: 7.24	−0.05 (0.93)	−0.84: 7.24	1.35 (1.10-1.66)	.004
2.260 ppm*	Valine	0.11 (1.15)	−2.10: 2.71	−0.02 (0.98)	−2.11: 4.41	1.01 (0.78-1.30)	.954
2.301 ppm*	3-Hydroxybutyrate	0.01 (0.63)	−0.55: 2.71	0.00 (1.03)	−0.64: 11.41	1.04 (0.82-1.30)	.766
2.349 ppm*	Glutamate	0.50 (1.21)	−0.91: 4.24	−0.06 (0.96)	−1.68: 4.24	1.38 (1.09-1.75)	.007
2.364 ppm*	Pyruvate	−0.19 (0.99)	−1.47: 3.13	0.02 (1.00)	−1.66: 3.91	0.96 (0.73-1.26)	.768
2.393 ppm*	3-Hydroxybutyrate	0.16 (0.86)	−0.94: 3.82	−0.02 (1.01)	−0.96: 10.08	1.13 (0.92-1.39)	.242
2.412 ppm	Unknown	0.32 (0.99)	−2.37: 2.67	−0.04 (0.99)	−2.64: 3.11	1.22 (0.96-1.55)	.110
2.430 ppm*	Glutamine	−0.22 (1.04)	−1.69: 2.10	0.02 (0.99)	−1.79: 3.30	0.81 (0.63-1.05)	.107
2.530 ppm*	Citrate	−0.10 (0.81)	−1.52: 2.13	0.01 (1.02)	−2.56: 4.43	0.77 (0.59-0.99)	.045
2.645 ppm*	Citrate	−0.19 (0.76)	−1.81: 1.84	0.02 (1.02)	−2.53: 4.56	0.73 (0.56-0.95)	.017
2.690 ppm	Mg-EDTA	0.11 (1.03)	−1.47: 2.79	−0.01 (0.99)	−1.80: 3.88	1.05 (0.82-1.33)	.710
2.737 ppm	Sarcosine	0.44 (1.57)	−1.45: 9.23	−0.05 (0.96)	−1.45: 9.23	1.32 (1.08-1.61)	.007
2.801 ppm	Aspartate	−0.04 (1.00)	−1.70: 4.01	0.01 (1.00)	−2.36: 4.75	0.84 (0.65-1.09)	.191
2.858 ppm	Asparagine	0.13 (0.98)	−1.50: 5.23	−0.02 (1.00)	−2.95: 10.18	1.18 (0.94-1.47)	.148
2.914 ppm*	Dimethylglycine	0.17 (1.04)	−2.07: 2.53	−0.02 (0.99)	−2.42: 3.57	1.02 (0.79-1.32)	.875
3.024 ppm*	Lysine	−0.08 (0.83)	−2.04: 1.85	0.01 (1.02)	−2.24: 3.89	0.81 (0.63-1.06)	.119
3.034 ppm*	Creatine + creatinine	0.17 (1.18)	−1.72: 6.82	−0.03 (0.97)	−2.05: 4.07	0.96 (0.75-1.23)	.754
3.050 ppm*	Ornithine	0.24 (1.11)	−2.03: 2.55	−0.03 (0.99)	−2.49: 3.42	1.07 (0.82-1.4)	.614
3.090 ppm	Ca-EDTA	0.07 (1.13)	−2.09: 4.19	0.00 (1.01)	−2.25: 4.19	0.95 (0.75-1.22)	.700

Supplementary Table I (*continued*)

Chemical shift	Assignment	Cases		Subcohort		Association with CHD	
		Mean (SD)	Min: Max	Mean (SD)	Min: Max	HR (95% CI)	*P*
3.117 ppm	Ca-EDTA	0.08 (1.12)	−3.84: 4.04	−0.02 (0.99)	−3.84: 4.55	1.01 (0.79-1.28)	.957
3.145 ppm*	Unknown	0.02 (1.15)	−2.57: 5.57	−0.01 (0.98)	−2.57: 6.04	0.88 (0.66-1.16)	.348
3.200 ppm	EDTA	0.11 (0.84)	−2.88: 2.04	−0.01 (1.03)	−3.39: 9.57	1.16 (0.91-1.48)	.228
3.240 ppm*	Glucose	0.32 (1.49)	−1.38: 9.56	−0.05 (0.90)	−2.24: 7.29	1.2 (0.98-1.47)	.078
3.252 ppm	TMAO	0.32 (0.93)	−2.60: 3.46	−0.04 (1.00)	−3.26: 6.43	1.17 (0.93-1.47)	.174
3.267 ppm*	1,5-Anhydrosorbitol	−0.15 (1.07)	−5.08: 1.98	0.01 (0.99)	−4.19: 3.42	1.01 (0.8-1.28)	.927
3.312 ppm	Unknown	0.20 (0.88)	−1.80: 3.71	−0.02 (1.01)	−2.00: 11.23	1.09 (0.82-1.44)	.555
3.333 ppm*	Proline	0.17 (0.91)	−1.44: 2.67	−0.03 (1.01)	−1.80: 4.80	1.02 (0.8-1.29)	.902
3.342 ppm*	1,5-Anhydrosorbitol	−0.16 (0.92)	−1.96: 2.10	0.01 (1.01)	−1.85: 3.53	0.86 (0.66-1.13)	.287
3.348 ppm	Unknown	−0.02 (0.90)	−1.75: 3.21	0.01 (1.01)	−2.67: 3.57	1 (0.79-1.27)	.978
3.354 ppm*	Unknown	0.11 (1.53)	−1.15: 10.50	−0.02 (0.90)	−1.15: 9.50	1.21 (0.95-1.53)	.124
3.362 ppm	Unknown	0.17 (1.31)	−2.79: 4.62	−0.02 (0.95)	−2.79: 3.72	1.21 (0.93-1.56)	.154
3.401 ppm*	Glucose	0.38 (1.48)	−1.69: 9.93	−0.05 (0.90)	−1.95: 7.82	1.23 (1.01-1.49)	.036
3.458 ppm	Glucose	0.30 (1.47)	−1.53: 9.22	−0.04 (0.91)	−1.71: 6.45	1.26 (1.01-1.57)	.040
3.487 ppm*	Glucose	0.37 (1.43)	−1.64: 9.52	−0.05 (0.91)	−2.03: 7.82	1.22 (1.01-1.47)	.042
3.531 ppm*	Glucose	0.39 (1.48)	−1.64: 9.76	−0.05 (0.90)	−1.83: 7.82	1.23 (1.02-1.49)	.033
3.551 ppm*	Unknown	−0.18 (0.90)	−1.53: 3.77	0.02 (1.01)	−4.31: 4.53	0.81 (0.62-1.05)	.111
3.599 ppm	EDTA	0.12 (0.84)	−2.86: 2.52	−0.01 (1.02)	−3.41: 9.90	1.18 (0.93-1.5)	.175
3.633 ppm	Unknown	0.00 (0.51)	−2.47: 0.65	0.00 (1.05)	−22.4: 2.58	0.97 (0.81-1.16)	.755
3.649 ppm*	Ethanol	0.13 (1.59)	−0.41: 13.79	−0.02 (0.88)	−0.41: 14.69	1.09 (0.92-1.29)	.303
3.712 ppm*	Glucose	0.33 (1.46)	−1.71: 9.30	−0.05 (0.91)	−1.76: 7.51	1.19 (0.98-1.45)	.078
3.721 ppm*	Glucose	0.36 (1.45)	−1.68: 9.67	−0.05 (0.91)	−2.31: 7.73	1.23 (1.01-1.49)	.040
3.735 ppm	3-Phosphoglycerate	0.22 (1.12)	−1.53: 3.37	−0.04 (0.98)	−1.53: 2.95	1.19 (0.94-1.52)	.156
3.747 ppm*	Unknown	0.12 (1.02)	−2.19: 2.28	−0.01 (1.00)	−3.09: 2.77	1.09 (0.85-1.39)	.509
3.759 ppm*	Glucose	0.32 (1.46)	−1.47: 9.83	−0.04 (0.91)	−2.29: 7.89	1.21 (0.99-1.47)	.061
3.765 ppm	3-Phosphoglycerate	0.25 (1.01)	−2.35: 2.15	−0.04 (0.99)	−3.13: 4.02	1.24 (0.97-1.58)	.087
3.772 ppm	3-Phosphoglycerate	0.23 (1.07)	−1.33: 2.90	−0.03 (0.98)	−1.35: 3.08	1.21 (0.95-1.53)	.119
3.779 ppm*	Alanine	−0.01 (0.97)	−2.80: 2.26	0.01 (1.01)	−3.58: 3.92	1 (0.78-1.27)	.977
3.818 ppm*	Unknown	0.04 (1.05)	−1.82: 4.81	0.00 (0.99)	−2.09: 6.76	1 (0.79-1.28)	.983
3.824 ppm*	Glucose	0.39 (1.40)	−2.49: 7.92	−0.05 (0.92)	−1.89: 6.32	1.29 (1.04-1.59)	.019
3.838 ppm*	Glucose	0.37 (1.5)	−1.69: 9.94	−0.05 (0.89)	−2.07: 7.37	1.25 (1.02-1.52)	.031
3.876 ppm*	1,5-Anhydrosorbitol	0.00 (1.23)	−5.59: 2.69	0.00 (0.96)	−5.37: 5.01	1.02 (0.8-1.29)	.888
3.888 ppm	Unknown	0.03 (1.10)	−2.28: 4.80	−0.01 (0.99)	−2.50: 3.68	0.96 (0.75-1.23)	.745
3.893 ppm*	Glucose	0.36 (1.47)	−1.57: 9.91	−0.05 (0.90)	−1.96: 7.83	1.22 (1.01-1.49)	.042
3.924 ppm*	Unknown	0.00 (0.93)	−2.30: 2.62	0.00 (1.01)	−2.64: 3.36	0.78 (0.6-1.01)	.062
3.959 ppm*	Unknown	−0.04 (0.77)	−1.65: 1.42	0.01 (1.02)	−2.11: 9.13	0.93 (0.71-1.21)	.570
3.974 ppm*	1,5-Anhydrosorbitol	−0.10 (1.02)	−2.12: 2.85	0.01 (1.00)	−2.85: 2.98	1.02 (0.78-1.32)	.897
3.981 ppm	Multiple metabolites	−0.16 (0.90)	−1.87: 2.73	0.02 (1.01)	−2.46: 5.23	0.86 (0.66-1.12)	.259
4.044 ppm*	Creatinine	0.38 (1.47)	−1.85: 10.37	−0.05 (0.90)	−2.51: 4.16	1.35 (1.04-1.76)	.027
4.103 ppm*	Lactate	0.20 (1.07)	−1.41: 4.57	−0.02 (0.99)	−1.82: 4.60	1.23 (0.96-1.57)	.100
4.121 ppm*	Proline	0.13 (0.93)	−2.01: 2.20	−0.02 (1.01)	−2.20: 4.39	0.97 (0.76-1.25)	.831
4.167 ppm	Phosphorylcholine	−0.02 (0.87)	−1.60: 2.65	0.00 (1.02)	−2.68: 6.10	0.82 (0.61-1.1)	.179
4.223 ppm	3-Phosphoglycerate	−0.01 (0.96)	−1.02: 4.42	0.00 (1.00)	−1.20: 4.84	0.98 (0.76-1.27)	.894
4.239 ppm*	Unknown	−0.12 (0.88)	−1.62: 2.64	0.02 (1.02)	−2.00: 4.14	0.92 (0.72-1.18)	.519
4.503 ppm*	Glucose	0.26 (1.14)	−1.52: 5.96	−0.03 (0.97)	−2.54: 4.61	1.23 (0.98-1.55)	.074
5.180 ppm*	Unknown	0.29 (1.42)	−2.20: 7.36	−0.03 (0.94)	−2.20: 5.14	1.27 (1.01-1.6)	.046
5.230 ppm*	Glucose	0.38 (1.47)	−1.79: 9.75	−0.05 (0.90)	−1.98: 7.97	1.23 (1.02-1.48)	.035
5.299 ppm*	Lipids (CH=CH)†	0.41 (1.65)	−0.61: 7.82	−0.04 (0.91)	−0.76: 7.82	1.37 (1.12-1.67)	.002
6.890 ppm*	Tyrosine	0.03 (0.90)	−1.74: 2.89	−0.01 (1.02)	−2.38: 5.14	0.85 (0.66-1.1)	.212
7.186 ppm*	Tyrosine	0.05 (1.01)	−1.50: 2.95	−0.01 (1.00)	−2.15: 4.98	0.91 (0.7-1.17)	.444
8.451 ppm*	Formate	0.02 (1.03)	−1.49: 4.45	0.00 (1.00)	−1.30: 9.41	0.9 (0.65-1.23)	.495

Mean is expressed as area under the curve.
* Detected and quantified in the confirmation study.
† The term in parenthesis indicates the structural feature of the lipid measured by ^{1}H-NMR spectroscopy.

Supplementary Table II. Baseline characteristics of the traditional CHD risk factors in the primary study before removing participants with missing data

	Cases (n = 125)		Subcohort (n = 738)*	
	No. missing	**Mean ± SD/n (%)**	**No. missing**	**Mean ± SD/n (%)**
Age (y)	0	50.49 ± 7.11	0	44.57 ± 8.23
TC (mmol/L)	2	5.91 ± 0.99	13	5.33 ± 0.98
HDL-C (mmol/L)	2	1.11 ± 0.28	13	1.33 ± 0.36
SBP (mm Hg)	0	133.18 ± 17.83	1	121.23 ± 14.61
BMI (kg/m^2)	0	27.69 ± 4.45	0	25.48 ± 3.87
Men	0	95 (76.0%)	0	323 (43.8%)
Current smoking	2	69 (56.1%)	5	283 (38.6%)
Diabetes	0	6 (4.8%)	0	5 (0.7%)
Parental history of MI	14	58 (52.3%)	68	264 (39.4%)

* Including 10 cases.

Supplementary Table III. The association between the metabolite score based on 16 ^{1}H-NMR signals and the metabolite score based on the 13 ^{1}H-NMR signals and incident CHD in the primary study

	Univariate analysis*		Adjusted for TRF†	
	HR (95% CI)	***P***	**HR (95% CI)**	***P***
Metabolite score based on 16 ^{1}H-NMR signals	1.93 (1.50-2.48)	<.001	1.58 (1.18-2.12)	.002
Metabolite score based on 13 ^{1}H-NMR signals‡	1.91 (1.50-2.44)	<.001	1.50 (1.12-2.01)	.007

* Univariate HRs were calculated per unit increase for the metabolite scores. Age in years was used as the time-scale variable.
† The TRFs include total cholesterol, HDL-C, SBP, BMI, gender, current smoking, self-reported diabetes, and parental history of MI. Age in years was used as the time-scale variable.
‡ This metabolite score is based on the 13 ^{1}H-NMR signals that were also available in the secondary study.

Supplementary Table IV. Reclassification of participants when the metabolite score was used in addition to a risk score based on traditional risk factors (ie, age, sex, current smoking, TC, HDL-C, BMI, parental history of MI, and self-reported diabetes)

	Model with TRF + metabolite score				
Model with TRF*	**<5%**	**5-<10%**	**10-<20%**	**≥20%**	**Total**
Incident cases					
<5%	39	3	1	0	43
5-<10%	5	13	3	1	22
10-<20%	0	2	4	0	6
≥20%	0	0	1	4	5
Total	44	18	9	5	76
Non-cases					
<5%	515	7	1	0	523
5-<10%	4	15	2	0	21
10-<20%	0	3	7	0	10
≥20%	0	0	0	0	0
Total	519	25	10	0	554

For reclassification, the following risk categories were used: 0% to <5%, 5% to <10%, 10% to <20%, ≥20%.
* The TRFs include age in years, TC, HDL-C, SBP, BMI, gender, current smoking, self-reported diabetes, and parental history of MI.

Supplementary Table V. Improvement in reclassification of incident CHD when using the metabolite score in addition to TRFs (ie, age, sex, current smoking, TC, HDL-C, BMI, parental history of MI, and self-reported diabetes)

	Reclassification				
	P Up	**P Down**	**NRI**	**Z-score**	***P*-value**
Event	0.140	0.096	0.044	0.924	0.178
Non-event	0.018	0.013	-0.005	-1.593	0.944
Total	…	…	0.038	0.809	0.209

For calculating the NRI, the following risk categories were used: 0% to <5%, 5% to <10%, 10% to <20%, ≥20%. Abbreviations: *P Up*, proportion of participants placed into a higher risk category; *P Down*, proportion of participants placed into a lower risk category.

Supplementary Table VI. (A) Comparing the predictive power of the single TRFs, all TRFs combined with and without the metabolite score in the primary case-cohort study using Harrell's *C*-index

	TRFs only	TRF + MS		
	***C*-index (95% CI)**	***C*-index (95% CI)**	**Diff* (SE)**	***P***
Age	0.72 (0.66-0.77)	0.79 (0.74-0.84)	0.08 (0.02)	<.001
TC	0.66 (0.61-0.72)	0.77 (0.72-0.81)	0.10 (0.03)	<.001
HDL-C	0.67 (0.60-0.73)	0.74 (0.68-0.79)	0.07 (0.02)	<.001
SBP	0.71 (0.65-0.77)	0.79 (0.74-0.84)	0.08 (0.02)	<.001
BMI	0.64 (0.58-0.71)	0.74 (0.69-0.80)	0.10 (0.03)	<.001
Male sex	0.66 (0.61-0.71)	0.76 (0.71-0.81)	0.10 (0.02)	<.001
Current smoking	0.57 (0.51-0.62)	0.74 (0.69-0.80)	0.18 (0.03)	<.001
Self-reported diabetes	0.51 (0.49-0.54)	0.76 (0.71-0.81)	0.25 (0.03)	<.001
Parental history of MI	0.55 (0.49-0.60)	0.75 (0.70-0.80)	0.20 (0.03)	<.001
All TRFs combined	0.82 (0.78-0.87)	0.84 (0.80-0.87)	0.01 (0.01)	.107

(B) Comparing the predictive power of the metabolite score combined with and without single TRFs and all TRFs combined in the primary case-cohort study using Harrell's *C*-index.

	***C*-index (95% CI)**	**Diff† (SE)**	***P***
MS only	0.75 (0.74-0.84)	–	–
Age + MS	0.79 (0.74-0.84)	0.038 (0.022)	.090
Total cholesterol + MS	0.77 (0.72-0.81)	0.011 (0.008)	.194
HDL cholesterol +MS	0.74 (0.68-0.79)	0.017 (0.021)	.412
Systolic blood pressure +MS	0.79 (0.74-0.84)	0.035 (0.019)	.064
Body mass index +MS	0.74 (0.69-0.80)	0.011 (0.010)	.308
Male sex +MS	0.76 (0.71-0.81)	0.007 (0.020)	.720
Current smoking +MS	0.74 (0.69-0.80)	0.011 (0.015)	.476
Self-reported diabetes +MS	0.76 (0.71-0.81)	0.002 (0.003)	.453
Parental history of MI +MS	0.75 (0.70-0.80)	0.008 (0.008)	.970
All TRFs combined	0.84 (0.80-0.87)	0.081 (0.023)	<.001

Abbreviations: *MS*, Metabolite score; *Diff*, difference.
* Difference in *C*-indices.
† Difference in *C*-indices between the metabolite score only and the traditional risk factor(s) and the metabolite score combined.

Supplementary Table VII. Baseline characteristics of the traditional CHD risk factors in the secondary study before removing participants with missing data

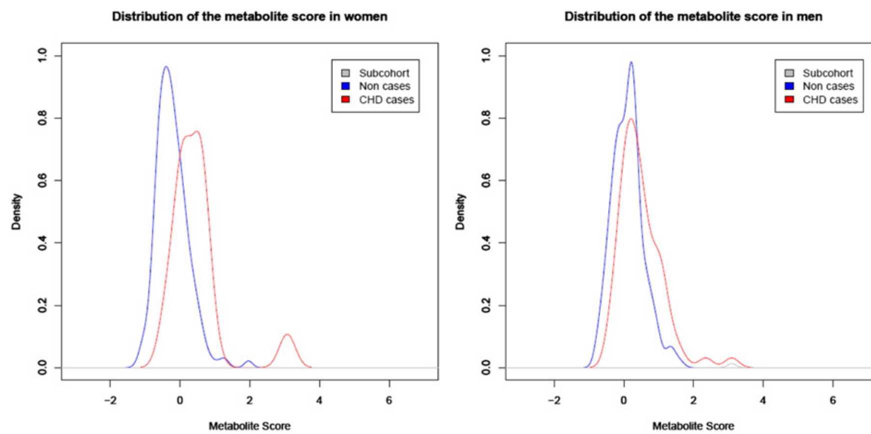
	Cases (n = 174)		Controls (n = 2170)	
	No. missing	**Mean ± SD/n (%)**	**No. missing**	**Mean ± SD/n (%)**
Age (y)	0	60.98 (11.73)	0	47.70 (14.02)
TC (mmol/L)	1	5.00 (1.07)	12	5.60 (1.06)
HDL-C (mmol/L)	1	1.15 (0.32)	12	1.29 (0.37)
SBP (mm Hg)	1	148.31 (22.91)	30	139.18 (19.75)
BMI (kg/m^2)	3	28.14 (4.43)	35	26.79 (4.61)
Men	0	108 (62.1%)	0	924 (42.6%)
Current smoking	1	62 (35.8%)	3	844 (38.9%)
Diabetes	0	25 (14.4%)	0	101 (4.7%)
Parental history of MI	69	54 (31.0%)	729	430 (19.9%)

Supplementary Table VIII. The association between the metabolite score and CHD in the total confirmation study (170 cases and 2,157 controls) and after excluding statin users (n = 299, 77 cases and 1,946 controls left for analysis)

	Crude OR		Adjusted OR	
Metabolite score	**(95% CI)**	***P***	**(95% CI)***	***P***
Total population	2.72 (2.28-3.25)	2.0×10^{-28}	1.59 (1.19-2.13)	.002
Nonstatin users	2.78 (2.16-3.57)	1.2×10^{-15}	2.02 (1.33-3.08)	.001

Odds ratios were calculated per SD increase in metabolite score.
*The metabolite score was adjusted for age, TC, HDL-C, SBP, BMI, sex, smoking, diabetes, and parental history of MI.

Supplementary Figure



Kernel density plots showing the distribution of the metabolite score in the subcohort and in the cases of the primary case-cohort study.