



Published in final edited form as:

*Arterioscler Thromb Vasc Biol.* 2019 July ; 39(7): 1475–1482. doi:10.1161/ATVBAHA.118.312236.

## A metabolomic pattern predicts incident coronary heart disease: findings from the Atherosclerosis Risk in Communities study

Zhe Wang, PhD<sup>1</sup>, Cong Zhu, MPH<sup>1</sup>, Vijay Nambi<sup>2,3</sup>, Alanna C. Morrison, PhD<sup>1</sup>, Aaron R. Folsom, MD<sup>4</sup>, Christie M. Ballantyne, MD<sup>3,5</sup>, Eric Boerwinkle, PhD<sup>1,6</sup>, Bing Yu, PhD<sup>1</sup>

<sup>1</sup>Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX

<sup>2</sup>Michael E DeBakey Veterans Affairs Hospital, Houston, Texas, USA

<sup>3</sup>Division of Atherosclerosis and Vascular Medicine, Department of Medicine, Baylor College of Medicine, Houston, TX

<sup>4</sup>Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN

<sup>5</sup>Center for Cardiovascular Disease Prevention, Methodist DeBakey Heart Center, Houston, TX

<sup>6</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX

### Abstract

**Objective:** Alterations in the serum metabolome may be detectable in at-risk individuals prior to the onset of coronary heart disease (CHD). Identifying metabolomic signatures associated with CHD may provide insight into disease etiology and prevention.

**Approach and Results:** Metabolomic profiling (chromatography-mass spectrometry) was performed in 2,232 African Americans and 1,366 European Americans from the Atherosclerosis Risk in Communities (ARIC) study. We applied Cox regression with least absolute shrinkage and selection operator (LASSO) to select metabolites associated with incident CHD. A metabolite risk score (MRS) was constructed to evaluate whether the MRS predicted CHD risk beyond traditional risk factors (TRFs). After 30 years of follow-up, we observed 633 incident CHD cases. Thirty-two metabolites were selected by LASSO to be associated with CHD, and 19 of the 32 showed significant individual associations with CHD, including a sugar substitute, erythritol. Theophylline (HR [95% CI] = 1.16 [1.09- 1.25]) and gamma linolenic acid (HR [95% CI] = 0.89 [0.81- 0.97]) showed the greatest positive and negative associations with CHD, respectively. A 1 SD greater standardized MRS was associated with a 1.37 fold higher risk of CHD (HR [95% CI] = 1.37 [1.27- 1.47]). Adding the MRS to the TRFs significantly improved model predictive performance (30-year risk prediction: C-statistic [95% CI] = 0.016 [0.008- 0.024], continuous net reclassification index [95% CI] = 0.522 [0.480- 0.556], integrated discrimination index [95% CI] = 0.038 [0.019- 0.065]).

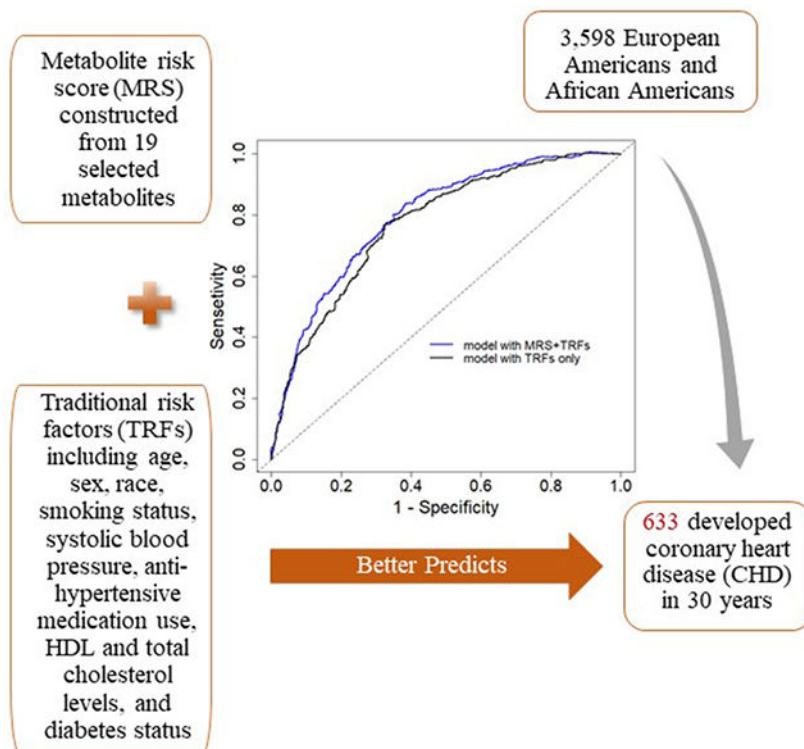
---

**Corresponding author:** Bing Yu, PhD, School of Public Health, RAS E641, University of Texas Health Science Center at Houston, Houston, TX 77030, Telephone: 713-500-9285, bing.yu@uth.tmc.edu.

Disclosure  
None

**Conclusion:** We identified 19 metabolites from known and novel metabolic pathways that collectively improved CHD risk prediction.

### Graphical Abstract



### Keywords

metabolome; biomarker; coronary heart disease; risk prediction; Cardiovascular Disease; Biomarkers; Epidemiology; Risk Factors; Primary Prevention

### Introduction

Despite a decline in coronary heart disease (CHD) incidence in most countries, CHD remains the leading cause of death globally<sup>1,2</sup>. Identifying individuals at increased risk of CHD can improve prospects for delayed onset and improved treatment. Over the last decades, a number of traditional risk factors (TRFs) have been discovered for CHD<sup>3-6</sup>, and those risk factors act synergistically. Multiple studies have demonstrated that integration of novel biomarkers, such as a genetic risk score, cardiac troponin-T or coronary artery calcium, can enhance CHD risk assessment<sup>7</sup>. However, such studies have identified few novel and tractable biological pathways. Other technologies, such as metabolomics, may help to identify novel biomarkers and pathways involved in CHD onset and progression.

The metabolomic approach systematically evaluates small-molecule metabolites in biologic samples that reflect the state of the system or whole organism provide additional insights into disease pathology<sup>8-10</sup>. Studies have successfully identified novel signatures of CHD

risk using metabolomic approaches, and the potential value of such signatures beyond the TRFs has been demonstrated<sup>11–18</sup>. However, these studies were carried out predominantly in individuals of European-ancestry and the spectrum of metabolites was limited.

We hypothesized that a collection of metabolites is associated with CHD risk independent of established risk factors, and adding a combination of CHD related metabolites to TRFs provides improved risk prediction. Hence, we conducted a metabolome-wide analysis of CHD risk in 3,598 African and European Americans from the Atherosclerosis Risk in Communities (ARIC) study to identify metabolomic signatures that better identify individuals at-risk.

## Methods

Because of the sensitive nature of the data collected for this study, requests to access the dataset from qualified researchers trained in human subject confidentiality protocols may be sent to the Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) (<https://biolincc.nhlbi.nih.gov>) and the database of Genotypes and Phenotypes (dbGaP) (Study Accession: phs000090.v1.p1).

## Study Population

The ARIC study is a prospective cohort study of 15,792 individuals aged 45–64 years from four U.S. communities (Forsyth County, NC; Jackson, MS; suburbs of Minneapolis, MN; and Washington County, MD) originally sampled between 1987 and 1989. A detailed description of the ARIC study design was published elsewhere<sup>19</sup>. Metabolomic profiles of baseline stored serum samples were measured in 1,997 African Americans in 2010 (phase 1), and 2,152 African and European Americans in 2014 (phase 2). Included in this analysis were 3,598 participants who did not have prevalent CHD or impaired renal function (estimated glomerular filtration rate (eGFR) < 60 mL/min/1.73 m<sup>2</sup>) at the baseline examination, with CHD follow-up and complete covariate information. The ARIC study was approved by the institutional review boards at each site, and written informed consent was obtained from all study participants.

## Assessment of incident CHD and Covariates

In ARIC, CHD incidence was ascertained by reviewing death certificate and hospital discharge records, and contacting participants annually, identifying hospitalizations and deaths during the prior year<sup>20</sup>. Incident CHD was defined as definite fatal CHD, definite or probable myocardial infarction, silent myocardial infarction between examinations as determined by electrocardiography, or coronary revascularization<sup>19</sup> that occurred on or before December 31<sup>st</sup> 2016. Covariates used in the analyses were measured at the baseline examination or home interview. Blood pressure was measured by trained technicians following a standard protocol<sup>21</sup>. Plasma total cholesterol levels were measured by enzymatic methods<sup>22, 23</sup>. High-density lipoprotein (HDL) cholesterol was measured after precipitation of the plasma with MgCl<sub>2</sub> and dextran sulfate by the method of Warnick et al<sup>24</sup>. Self-reported smoking status (categorized as current smoker, former smoker and never smoker) and antihypertensive medication use during the past 2 weeks were obtained by

questionnaire. Prevalent CHD was defined as a self-report of previous myocardial infarction or coronary reperfusion procedure, or electrocardiography (ECG) evidence of a previous myocardial infarction. Diabetes was defined as a fasting serum glucose level  $\geq 126$  mg/dL or non-fasting serum glucose level  $\geq 200$  mg/dL, a self-reported physician diagnosis of diabetes, or the use of hypoglycemic medication. eGFR was calculated by using the Chronic Kidney Disease Epidemiology Collaboration equation<sup>25</sup>. Carotid intima-media thickness (C-IMT) was measured using the ultrasound procedure that has been previously described<sup>26–28</sup>. Briefly, the C-IMT was assessed in three 1cm segments in the distal common carotid, the carotid artery bifurcation and the proximal internal carotid arteries. We used the mean of the mean measurements for both right and left sides of these segments at the baseline examination for analyses. Plaque presence or absence was adjudicated by trained readers; plaque was adjudicated to be present if two of three of the following criteria were abnormal: wall thickness, wall shape, and wall texture<sup>27, 29</sup>.

### Assessment of the Serum Metabolome

Metabolomic profiling was completed in two phases, 2010 and 2014, using serum samples which had been stored at  $-80^{\circ}\text{C}$  since collection at the baseline examination in 1987–1989. A total of 384 metabolites were detected and semi-quantified by Metabolon Inc. (Durham, USA) using untargeted gas chromatography-mass spectrometry and liquid chromatography-mass spectrometry (GC-MS and LC-MS)<sup>30, 31</sup>. 139 metabolites were excluded because: 1) their values were missing/below-the-detection-limit in more than 25% of the samples in either phase; or 2) their Pearson correlation coefficients were  $< 0.3$  between 97 pairs of duplicate samples measured during both phases. After this assessment, 245 metabolites, including amino acids, lipids, nucleotides, peptides, carbohydrates, cofactors and vitamins, xenobiotics, and energy related metabolites were included in the present study.

### Statistical Analysis

Within each phase, metabolite levels were winsorized at the 1<sup>st</sup> and 99<sup>th</sup> percentile, and missing/below-the-detection-limit values were imputed with the lowest detected value of that metabolite. Prior to the analyses, metabolites were standardized (mean = 0 and SD = 1) within each phase.

To select a subset of the most informative metabolites that were associated with incident CHD, Cox proportional hazard regression models with least absolute shrinkage and selection operator (LASSO) procedures were applied in the combined phase 1 and 2 samples. We fit LASSO with 10-fold cross validation incorporating 245 metabolites and CHD TRFs defined according to ARIC CHD risk score, including age, sex, race, smoking status, systolic blood pressure (SBP), anti-hypertensive medication use, HDL and total cholesterol levels, and diabetes status<sup>4</sup>, as well as phase and study center. The penalty parameter was chosen as the  $\lambda$  that gave the minimum mean error as determined by 10-fold cross validation. To explore the individual effect of each LASSO selected metabolite on incident CHD, we performed analyses using a Cox model adjusting for TRFs, phase and study centers. The false discovery rate of 5% or less was used to define significance for associations between individual metabolites and CHD.

Metabolites selected by the LASSO procedure were further reduced to those that had significant individual associations with incident CHD, and this subset of metabolites were used to construct a metabolomics risk score (MRS). A continuous MRS was constructed using the sum of selected metabolite levels taking into account the direction of effect. The continuous MRS was standardized (mean = 0 and SD = 1) for further analysis. Quartiles of the continuous MRS were analyzed to examine any potential non-linear associations. Cox regression models were used to calculate whether the continuous MRS or MRS quartiles were associated with incident CHD adjusting for the same covariates described above. The proportional hazards assumption was tested in each Cox model and no violation of the assumption was observed. To explore potential sex- and race- specific effects, we performed secondary analyses testing the association between continuous MRS and incident CHD stratified by sex and race separately. Additionally, we tested the association between the MRS and baseline C-IMT and plaque measurement. The significance threshold was defined as  $p < 0.05$  for the MRS analysis.

To investigate whether the MRS improved risk reclassification and prediction, we computed statistical measures of discrimination, including area under the receiver operating characteristic curve (AUC), the continuous net reclassification index (NRI), and the integrated discrimination index (IDI)<sup>32, 33</sup>. The C-statistic and differences in C-statistic were calculated using the method for censored survival data proposed by Uno et al.<sup>34</sup>. Additionally, to confirm the internal validity of the MRS and ensure a nearly unbiased estimate of our MRS model prediction performance, we performed 10-fold cross-validations with 100 replicates and reported the average cross-validation C-statistics<sup>34, 35</sup>. All statistical analyses were carried out using R version 3.4 (R Foundation for Statistical Computing, Vienna, Austria).

## Results

During 30 years of follow-up, we identified 633 (17.6%) incident CHD cases among 3,598 African and European American ARIC participants (Table 1). The CHD incidence rate was lower in European Americans as compared to African Americans. In general, ARIC European Americans showed a more favorable baseline CHD risk profile, including less current smokers, lower SBP levels, and less prevalent diabetes, except for lipid profiles, which were similar between African and European Americans. The C-IMT and plaque presence measurements did not differ between the two groups. Baseline characteristics by incidence CHD cases and non-cases were summarized in Supplemental Table I.

Thirty-two metabolites were identified by LASSO as the best subset predictive of incident CHD. We examined each of the selected metabolites for its association with CHD risk, and 19 out of 32 metabolites had significant individual associations, including 8 amino acids, 3 lipids, 2 nucleotides, 1 peptide, 1 carbohydrates, and 4 xenobiotics (Supplemental Table II). The median hazard ratio (HR) per SD was 1.11 across the 19 metabolites; the 25<sup>th</sup> percentile and 75<sup>th</sup> percentile HR per SD was 1.08 and 1.14. Theophylline (HR [95% confidence interval (CI)] = 1.16 [1.09- 1.25]) and linolenate [ $\alpha$  or  $\gamma$ ; (18:3n3 or 6)] (HR [95% CI] = 0.89 [0.81- 0.97]) had the highest and lowest HRs for CHD, respectively (Figure 1).

Fourteen out of 19 selected metabolites were not correlated ( $|r| < 0.3$ ), while five other metabolites showed moderate inter-correlations (Supplemental Figure I).

The MRS constructed from the 19 metabolites was approximately normally distributed (Supplemental Figure II). Baseline participants' characteristics by deciles of the MRS were summarized in Supplemental Table III. A 1 SD increase in the standardized MRS was associated with a 1.37 fold greater risk of incident CHD adjusting for TRFs (HR [95% CI] = 1.37 [1.27-1.47], Table 2). There was a graded association between the MRS quartiles and risk of incident CHD events ( $p$  for linear trend =  $1.82 \times 10^{-13}$ ) (Table 2), where participants with the highest quartile of MRS had more than 2 fold higher risk of developing CHD compared to participants in the lowest quartile (Q4 vs. Q1 HR [95% CI] = 2.21 [1.72-2.84]). Additional analyses adjusting for BMI and total energy intake did not change the association between MRS and CHD risk. Sex-stratified and race-stratified analyses did not suggest different continuous MRS associations with CHD among males and females, or among African- and European- Americans (Supplemental Table IV). The MRS we created for CHD risk was significantly associated with greater intima-media thickness (log transformed C-IMT measurement, coefficient = 0.007,  $p = 0.03$ ). But the association between the MRS and plaque presence was modest (Odds Ratio = 0.07,  $p = 0.07$ ).

Adding the continuous MRS to the TRFs model significantly improved model discrimination performance. The C-statistic [95% CI] increased from 0.724 [0.704-0.744] to 0.740 [0.720-0.759] for 30-year CHD risk prediction (C-statistic [95% CI] = 0.016 [0.008-0.024], Figure 2). Addition of the MRS also showed significant incremental benefit in the analysis of the continuous NRI and IDI (continuous NRI [95% CI] = 0.522 [0.480-0.556],  $p < 0.001$ , and IDI [95% CI] = 0.038 [0.019-0.065],  $p = 0.002$ ). Adding MRS quartiles to the TRFs model similarly improved model performance. The C-statistic [95% CI] for 30-year CHD risk prediction was 0.014 [0.005- 0.022]), and the continuous NRI [95% CI] = 0.475 [0.271-0.634] and IDI [95% CI] = 0.048 [0.014- 0.085].

Using 10-fold cross validation with 100 replicates, we computed cross-validated C-statistics for the TRFs only model (corrected C-statistic: 0.718) and the model with continuous MRS and the TRFs (corrected C-statistic: 0.735). Cross validation did not change the conclusion that adding the continuous MRS to the TRFs model improved discrimination (cross validation corrected C-statistic vs. one time estimated C-statistic: 0.017 vs. 0.016).

## Discussion

In this population based prospective study of African and European Americans, we identified a set of 19 metabolites, including amino acids, lipids, peptides, carbohydrates, nucleotide and xenobiotics, that were collectively associated with CHD risk during thirty years of follow-up. A MRS constructed from the 19 metabolites were positively association with C-IMT, a well-described surrogate marker for cardiovascular disease<sup>36</sup>, and showed a graded association with CHD risk that were consistent across race and sex. Adding the MRS to the TRFs model significantly improved 30-year incident CHD risk prediction and discrimination. The predictive ability of the MRS was internally validated. Our findings

indicate that a metabolite panel can be used to improve CHD risk prediction, compared with TRFs alone.

Previous metabolomic studies have identified multiple metabolites that are associated with CHD risk<sup>11–18, 37–41</sup>, some of which were corroborated in our study. For example, metabolites involved in amino acid metabolism, including asparagine, dimethylglycine, and short chain acylcarnitines such as 2-methylbutyrylcarnitine (C5), were associated positively with CHD occurrence in this and other studies<sup>37, 38, 40, 41</sup>. We observed that  $\alpha$ -linoleic acid was negatively associated with CHD, which is consistent with previous metabolomic study findings, and in line with previous observations that polyunsaturated fatty acids were inversely associated with cardiovascular disease in contrast to saturated and monounsaturated fatty acids<sup>42, 43</sup>. Other metabolites such as 13-HODE + 9-HODE (oxidized derivatives of linoleic acid), uridines, and mannose also have been previously reported to be associated with CHD<sup>14, 16, 38, 39, 44, 45</sup>. Some metabolites reported previously were not included in our MRS, such as TMAO (14, 15) which was not in our analyzed metabolomic panel. Other CHD metabolites, for examples, polyunsaturated fatty acids, were not selected probably because of high correlations with other LASSO selected metabolites, as LASSO tends to select only one variable from a group of highly correlated variables<sup>46</sup>.

Twelve of our 19 selected metabolites have not been previously reported to be associated with CHD (Supplemental Table II). Among these, four metabolites have been shown previously to be related to one or more CHD risk factors. For example, higher urinary excretion levels of 4-vinylphenol sulfate, a styrene metabolite, has been reported to associate with smoking<sup>47</sup>, which can be explained as styrene is one of many chemicals found in cigarettes. DNA methylation studies have previously reported several CpG loci that associate with serum 4-vinylphenol sulfate levels<sup>48</sup>, which are the same CpG loci associated with tobacco smoking<sup>49</sup>. Greater N-acetylthreonine levels was identified to be a biomarker for the progression of renal dysfunction<sup>50</sup>, and N-acetyl-1-methylhistidine, a metabolite involved in histidine metabolism, has been associated positively with incident chronic kidney disease<sup>51</sup>. Three metabolites have been associated with heart failure risk: higher levels of N-acetylalanine and p-cresol sulfate<sup>52</sup>, and lower levels of 1-arachidonoylglycerophosphocholine (20:4n6), also known as lysoPC (20:4)<sup>53</sup>. In addition, p-cresol sulfate levels have been reported to be higher among those elderlies as compared to younger age group<sup>54</sup>, and we observed the same pattern at baseline (data not shown). Such observations suggest that some of the metabolites we selected might partially reflect aging process, a major contributor to CHD risk. Theophylline, which we found associated with the highest CHD HR, is a drug used to treat asthma and chronic obstructive pulmonary disease (COPD), and is well known for having cardiotoxic side effects<sup>55, 56</sup>. We conducted sensitivity analyses by excluding COPD cases, defined by Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria<sup>57</sup>, or individuals with self-reported ever asthma. The association between theophylline and CHD risk in the fully adjusted model remained unchanged (data not shown), although the association still may reflect residual confounding effects of smoking.

An interesting metabolite associated positively with CHD in our study was erythritol. Erythritol is a sugar substitute widely used in processed foods, especially artificial

sweetened beverages (ASB). Like conventional sugar sweetened beverages, greater ASB intake has been associated with negative health outcomes, including increased risk for CHD<sup>58, 59</sup> and hypertension<sup>60</sup>. Studies of ASB often fail to distinguish the sugar replacement sweetener used in these beverages; therefore, the etiologic connection of sugar substitutes with CHD remains poorly understood. Hootman et al. showed that erythritol can be synthesized from glucose via the pentose-phosphate pathway and associated with adiposity gain<sup>61</sup>, which is in contrast to previously evidence that erythritol cannot be metabolized in humans<sup>62</sup>. The association between erythritol and CHD we observed may be explained by weight gain<sup>63</sup>; however, additional adjustment for body mass index did not change the association we observed between erythritol and CHD risk (data not shown).

The metabolite, 1-arachidonoylglycerophosphocholine (20:4n6) [LysoPC (20:4)], showed a novel negative association with CHD. LysoPC is mainly formed by the enzyme, lecithin:cholesterol acyltransferase (LCAT), which converts cholesterol and phosphatidylcholine to cholesterol esters and lysophosphatidylcholine<sup>64, 65</sup>. It has been reported that lysoPC (20:4) levels are lower in heart failure patients with reduced ejection fraction as compared to controls<sup>53</sup>. lysoPC (20:4) has been suggested to be a potential biomarkers for discriminating CHD patients from controls (lower levels in CHD patients)<sup>66</sup>. Fan et al. (54) provided evidence supporting multiple lysoPCs, for example, lysoPC (16:0), lysoPC (18:1), and lysoPC (20:3), as metabolites with relatively lower concentrations in CHD patients or in more severe patients vs. controls. Our results further document an association of lysoPC with CHD risk.

Arginylphenylalanine, a dipeptide composed of arginine and phenylalanine, was negatively associated with CHD risk. Although no previous study has identified the dipeptide arginylphenylalanine as associated with CHD, its two components, arginine and phenylalanine, have been recognized for their roles in cardiovascular diseases<sup>16, 67-69</sup>. Arginine has substantial cardiovascular benefits, including lowering blood pressure, peripheral vascular resistance and plasma homocysteine, as well as improving endothelial function<sup>68, 69</sup>. Previous metabolomic studies associated higher serum phenylalanine levels with increased cardiovascular risk<sup>16, 67</sup>. The possible pathophysiology of arginylphenylalanine in CHD development is unknown, but we speculate it may be via these two amino acids since it is a short-lived intermediate on the way to amino acids degradation pathways.

To our knowledge, this study is the first to evaluate prospectively whether a comprehensive set of metabolites predicts CHD in a biracial population. Our study has some limitations that are worth noting. We did not replicate our findings in other independent population-based studies. Yet, our use of LASSO and a false discovery rate of 5% strengthened the selection of metabolites. Our MRS included 19 metabolites with significant individual associations with CHD risk. Further investigation of LASSO selected metabolites not included in the MRS is warranted. In addition, LASSO picks the statistically most significant metabolites for CHD prediction, not necessarily the most biologically relevant ones. We assessed the metabolome on samples drawn in 1987-89 to predict CHD over 30 years. ARIC participants' metabolomes and confounding variables may have changed significantly over the 30 years of follow-up. The lack of longitudinal metabolomic data did not allow us to examine the



change of these metabolites in relation to CHD risk. Furthermore, the metabolomic panel we used focused on small molecules, which did not contain complex lipids. Information on complex lipids, such as lipidomic profiles, may provide additional insights for CHD risk 42, 43.

We selected metabolites based on their associations with incidence CHD beyond traditional risk factors, including conventional lipid variables. Our data demonstrated that adding MRS to current conventional risk scores led to improved CHD prediction. Although the strength of association of MRS was similar to conventional lipid parameters used in CHD risk prediction, it is premature to suggest routine assessment of metabolites in clinical practice. Future investigation is warranted to examine long-term metabolic changes in relation to CHD risk prediction, costs and availability of the measurement assays, appropriate therapeutic cutpoints, and clinical evidence of benefits for achieving these cutpoints based on randomized clinical trials.

In conclusion, a MRS constructed from 19 metabolites was associated with the risk of developing CHD and improved CHD 30 years risk prediction beyond TRFs. Our results suggest several new pathways in the etiology of CHD, and highlight the use of metabolomics in CHD risk prediction. Further work is warranted to replicate our novel findings and shed light on potential mechanisms in CHD etiology.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgement

The authors thank the staff and participants of the ARIC study for their important contributions.

Sources of Funding:

The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. (HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700005I, HHSN268201700004I). The metabolomics measurements were supported by the National Genome Research Institute (HG004402). This project also was supported by the American Heart Association (Yu, 17SDG33661228).

## Abbreviations:

<b>TRFs</b>	Traditional risk factors
<b>MRS</b>	Metabolomics risk score
<b>LASSO</b>	Least absolute shrinkage and selection operator
<b>ARIC study</b>	Atherosclerosis Risk in Communities study
<b>CHD</b>	Coronary heart disease

## References

1. Organization WH. World health statistics 2017: Monitoring health for the sdgs Sustainable Development Goals. Geneva: WHO 2017
2. Benjamin EJ, Virani SS, Callaway CW, et al. Heart disease and stroke statistics-2018 update: A report from the american heart association. *Circulation*. 2018;137:e67–e492 [PubMed: 29386200]
3. Vasan RS. Biomarkers of cardiovascular disease: Molecular basis and practical considerations. *Circulation*. 2006;113:2335–2362 [PubMed: 16702488]
4. Chambless LE, Folsom AR, Sharrett AR, Sorlie P, Couper D, Szklo M, Nieto FJ. Coronary heart disease risk prediction in the atherosclerosis risk in communities (aric) study. *J Clin Epidemiol* 2003;56:880–890 [PubMed: 14505774]
5. Ballantyne CM, Hoogeveen RC, Bang H, Coresh J, Folsom AR, Heiss G, Sharrett AR. Lipoprotein-associated phospholipase a2, high-sensitivity c-reactive protein, and risk for incident coronary heart disease in middle-aged men and women in the atherosclerosis risk in communities (aric) study. *Circulation*. 2004;109:837–842 [PubMed: 14757686]
6. Folsom AR, Aleksic N, Catellier D, Juneja HS, Wu KK. C-reactive protein and incident coronary heart disease in the atherosclerosis risk in communities (aric) study. *Am Heart J* 2002;144:233–238 [PubMed: 12177639]
7. Vasan R Biomarkers of cardiovascular disease: Molecular basis and practical considerations. *Circulation*. 2006;113:2335 [PubMed: 16702488]
8. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Analytical chemistry*. 2009;81:6656–6667 [PubMed: 19624122]
9. Madsen R, Lundstedt T, Trygg J. Chemometrics in metabolomics—a review in human disease diagnosis. *Analytica chimica acta*. 2010;659:23–33 [PubMed: 20103103]
10. Villas-Bôas SG, Mas S, Åkesson M, Smedsgaard J, Nielsen J. Mass spectrometry in metabolome analysis. *Mass spectrometry reviews*. 2005;24:613–646 [PubMed: 15389842]
11. Shah SH, Sun J-L, Stevens RD, Bain JR, Muehlbauer MJ, Pieper KS, Haynes C, Hauser ER, Kraus WE, Granger CB, Newgard CB, Califf RM, Newby LK. Baseline metabolomic profiles predict cardiovascular events in patients at risk for coronary artery disease. *American Heart Journal*. 2012;163:844–850.e841 [PubMed: 22607863]
12. Rizza S, Copetti M, Rossi C, Cianfarani MA, Zucchelli M, Luzi A, Pecchioli C, Porzio O, Di Cola G, Urbani A, Pellegrini F, Federici M. Metabolomics signature improves the prediction of cardiovascular events in elderly subjects. *Atherosclerosis*. 2012;232:260–264 [PubMed: 24468136]
13. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, DuGar B, Feldstein AE, Britt EB, Fu X, Chung Y-M. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*. 2011;472:57–63 [PubMed: 21475195]
14. Vaarhorst AA, Verhoeven A, Weller CM, et al. A metabolomic profile is associated with the risk of incident coronary heart disease. *Am Heart J*. 2014;168:45–52.e47 [PubMed: 24952859]
15. Kume S, Araki S-i, Ono N, et al. Predictive properties of plasma amino acid profile for cardiovascular disease in patients with type 2 diabetes. *PLOS ONE*. 2014;9:e101219 [PubMed: 24971671]
16. Würtz P, Havulinna AS, Soininen P, Tynkkynen T, Prieto-Merino D, Tillin T, Ghorbani A, Artati A, Wang Q, Tiainen M. Metabolite profiling and cardiovascular event risk: A prospective study of three population-based cohorts. *Circulation*. 2015;131:774 [PubMed: 25573147]
17. Paynter NP, Balasubramanian R, Giulianini F, et al. Metabolic predictors of incident coronary heart disease in women. *Circulation*. 2018;137:841–853 [PubMed: 29459470]
18. Ganna A, Salihovic S, Sundstrom J, et al. Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. *PLoS Genet* 2014;10:e1004801 [PubMed: 25502724]
19. ARIC Investigators. The atherosclerosis risk in communities (aric) study: Design and objectives. *American journal of epidemiology*. 1989;129:687–702 [PubMed: 2646917]

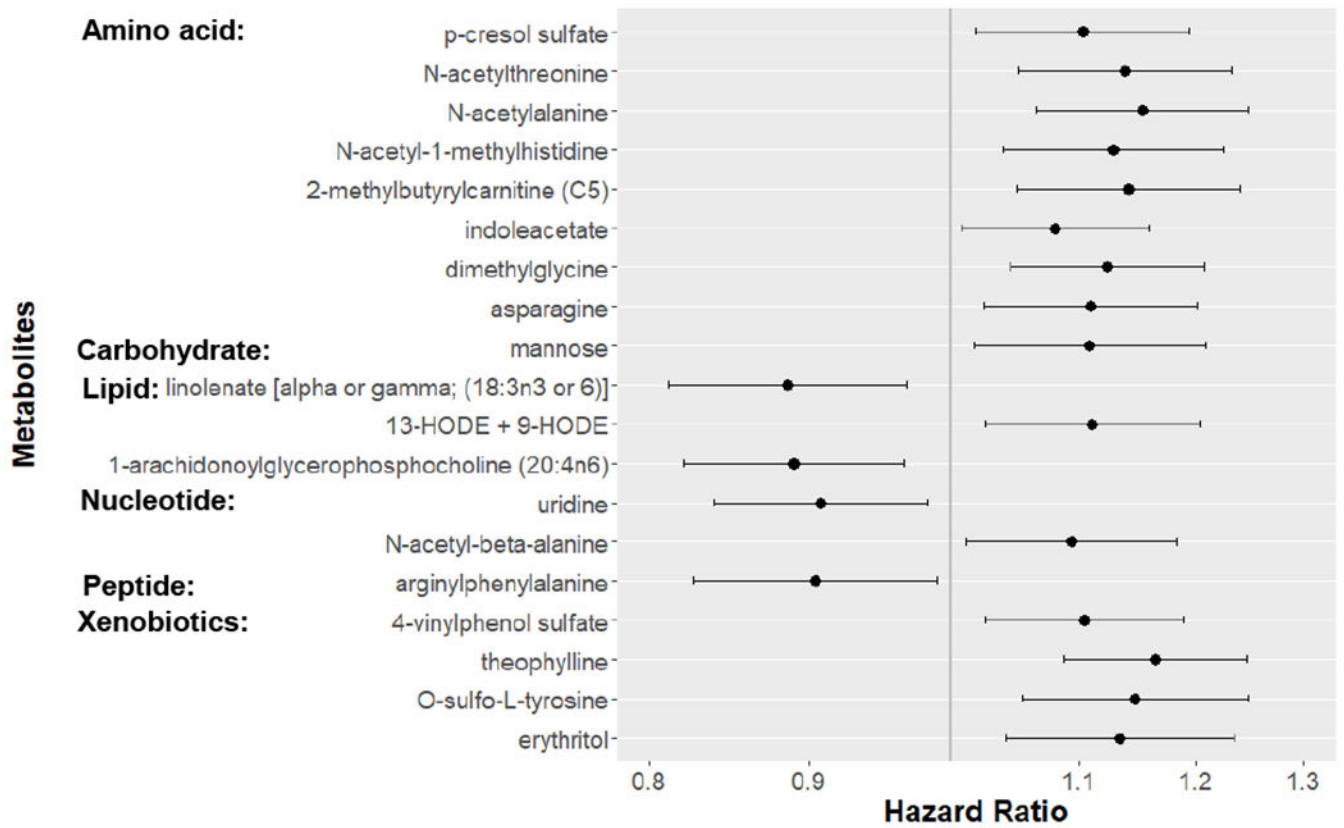
20. White AD, Folsom AR, Chambless LE, Sharret AR, Yang K, Conwill D, Higgins M, Williams OD, Tyroler HA, The AI. Community surveillance of coronary heart disease in the atherosclerosis risk in communities (aric) study: Methods and initial two years' experience. *Journal of Clinical Epidemiology*. 1996;49:223–233 [PubMed: 8606324]
21. Sharrett AR, Hubbard LD, Cooper LS, Sorlie PD, Brothers RJ, Nieto FJ, Pinsky JL, Klein R. Retinal arteriolar diameters and elevated blood pressure: The atherosclerosis risk in communities study. *Am J Epidemiol* 1999;150:263–270 [PubMed: 10430230]
22. Siedel J, Hägele EO, Ziegenhorn J, Wahlefeld AW. Reagent for the enzymatic determination of serum total cholesterol with improved lipolytic efficiency. *Clinical Chemistry*. 1983;29:1075–1080 [PubMed: 6851096]
23. Nagele U, Hagele EO, Sauer G, Wiedemann E, Lehmann P, Wahlefeld AW, Gruber W. Reagent for the enzymatic determination of serum total triglycerides with improved lipolytic efficiency. *Journal of clinical chemistry and clinical biochemistry. Zeitschrift fur klinische Chemie und klinische Biochemie*. 1984;22:165–174 [PubMed: 6716056]
24. Russell Warnick G, Mayfield C, Benderson J, Chen J-S, Albers JJ. Hdl cholesterol quantitation by phosphotungstate-mg<sup>2+</sup> and by dextran sulfate-mn<sup>2+</sup>-polyethylene glycol precipitation, both with enzymic cholesterol assay compared with the lipid research method. *American journal of clinical pathology*. 1982;78:718–723 [PubMed: 6182791]
25. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T. A new equation to estimate glomerular filtration rate. *Annals of internal medicine*. 2009;150:604–612 [PubMed: 19414839]
26. Li R, Cai J, Tegeler C, Sorlie P, Metcalf PA, Heiss G. Reproducibility of extracranial carotid atherosclerotic lesions assessed by b-mode ultrasound: The atherosclerosis risk in communities study. *Ultrasound in medicine & biology*. 1996;22:791–799 [PubMed: 8923698]
27. Li R, Duncan BB, Metcalf PA, Crouse JR 3rd, Sharrett AR, Tyroler HA, Barnes R, Heiss G. B-mode-detected carotid artery plaque in a general population. *Atherosclerosis risk in communities (aric) study investigators. Stroke*. 1994;25:2377–2383 [PubMed: 7974576]
28. High-resolution b-mode ultrasound scanning methods in the atherosclerosis risk in communities study (aric). The aric study group. *Journal of neuroimaging : official journal of the American Society of Neuroimaging*. 1991;1:68–73 [PubMed: 10149803]
29. Hunt KJ, Sharrett AR, Chambless LE, Folsom AR, Evans GW, Heiss G. Acoustic shadowing on b-mode ultrasound of the carotid artery predicts chd. *Ultrasound in medicine & biology*. 2001;27:357–365 [PubMed: 11369121]
30. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal Chem* 2009;81:6656–6667 [PubMed: 19624122]
31. Tetsuya O, Naoya M, Naohisa T, Tetsuya S, Matthew M, Michael VM, John AR, Kirk DB, Lining G. Untargeted metabolomic profiling as an evaluative tool of fenofibrate-induced toxicology in fischer 344 male rats. *Toxicologic Pathology*. 2009;37:521–535 [PubMed: 19458390]
32. Pencina MJ, Agostino RBD, Agostino RBD, Vasan RS. Evaluating the added predictive ability of a new marker: From area under the roc curve to reclassification and beyond. *Statistics in Medicine*. 2008;27:157–172 [PubMed: 17569110]
33. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 2011;30:11–21 [PubMed: 21204120]
34. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med* 2011;30:1105–1117 [PubMed: 21484848]
35. Taylor JMG, Ankerst DP, Andridge RR. Validation of biomarker-based risk prediction models. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2008;14:5977–5983 [PubMed: 18829476]
36. Nambi V, Chambless L, Folsom AR, He M, Hu Y, Mosley T, Volcik K, Boerwinkle E, Ballantyne CM. Carotid intima-media thickness and presence or absence of plaque improves prediction of

- coronary heart disease risk: The aric (atherosclerosis risk in communities) study. *Journal of the American College of Cardiology*. 2010;55:1600–1607 [PubMed: 20378078]
37. Drake KJ, Sidorov VY, McGuinness OP, Wasserman DH, Wikswo JP. Amino acids as metabolic substrates during cardiac ischemia. *Experimental biology and medicine (Maywood, N.J.)*. 2012;237:10.1258/ebm.2012.012025
  38. Zheng Y, Yu B, Alexander D, Steffen LM, Nettleton JA, Boerwinkle E. Metabolomic patterns and alcohol consumption in african americans in the atherosclerosis risk in communities study. *The American journal of clinical nutrition*. 2014;99:1470–1478 [PubMed: 24760976]
  39. Vangaveti V, Baune BT, Kennedy RL. Hydroxyoctadecadienoic acids: Novel regulators of macrophage differentiation and atherogenesis. *Therapeutic Advances in Endocrinology and Metabolism*. 2010;1:51–60 [PubMed: 23148150]
  40. Svingen GF, Ueland PM, Pedersen EK, Schartum-Hansen H, Seifert R, Ebbing M, Loland KH, Tell GS, Nygard O. Plasma dimethylglycine and risk of incident acute myocardial infarction in patients with stable angina pectoris. *Arterioscler Thromb Vasc Biol* 2013;33:2041–2048 [PubMed: 23723367]
  41. Guasch-Ferré M, Zheng Y, Ruiz-Canela M, et al. Plasma acylcarnitines and risk of cardiovascular disease: Effect of mediterranean diet interventions. *The American journal of clinical nutrition*. 2016;103:1408–1416 [PubMed: 27099249]
  42. Toledo E, Wang DD, Ruiz-Canela M, et al. Plasma lipidomic profiles and cardiovascular events in a randomized intervention trial with the mediterranean diet. *The American journal of clinical nutrition*. 2017;106:973–983 [PubMed: 28814398]
  43. Stegemann C, Pechlaner R, Willeit P, Langley SR, Mangino M, Mayr U, Menni C, Moayyeri A, Santer P, Rungger G, Spector TD, Willeit J, Kiechl S, Mayr M. Lipidomics profiling and risk of cardiovascular disease in the prospective population-based bruneck study. *Circulation*. 2014;129:1821–1831 [PubMed: 24622385]
  44. Fretts AM, Mozaffarian D, Siscovick DS, Sitlani C, Psaty BM, Rimm EB, Song X, McKnight B, Spiegelman D, King IB, Lemaitre RN. Plasma phospholipid and dietary  $\alpha$ -linolenic acid, mortality, chd and stroke: The cardiovascular health study. *The British journal of nutrition*. 2014;112:1206–1213 [PubMed: 25159901]
  45. Mardinoglu A, Stancakova A, Lotta LA, Kuusisto J, Boren J, Bluher M, Wareham NJ, Ferrannini E, Groop PH, Laakso M, Langenberg C, Smith U. Plasma mannose levels are associated with incident type 2 diabetes and cardiovascular disease. *Cell metabolism*. 2017;26:281–283 [PubMed: 28768165]
  46. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2005;67:301–320
  47. Manini P, De Palma G, Andreoli R, Goldoni M, Poli D, Lasagni G, Mutti A. [urinary excretion of 4-vinyl phenol after experimental and occupational exposure to styrene]. *Giornale italiano di medicina del lavoro ed ergonomia*. 2003;25 Suppl:61–62
  48. Petersen A-K, Zeilinger S, Kastenmüller G, et al. Epigenetics meets metabolomics: An epigenome-wide association study with blood serum metabolic traits. *Human molecular genetics*. 2014;23:534–545 [PubMed: 24014485]
  49. Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, Weidinger S, Lattka E, Adamski J, Peters A, Strauch K, Waldenberger M, Illig T. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PloS one*. 2013;8:e63812–e63812 [PubMed: 23691101]
  50. Solini A, Manca ML, Penno G, Pugliese G, Cobb JE, Ferrannini E. Prediction of declining renal function and albuminuria in patients with type 2 diabetes by metabolomics. *The Journal of clinical endocrinology and metabolism*. 2016;101:696–704 [PubMed: 26684276]
  51. Yu B, Zheng Y, Alexander D, Morrison AC, Coresh J, Boerwinkle E. Genetic determinants influencing human serum metabolome among african americans. *PLoS Genet* 2014;10
  52. Zheng Y, Yu B, Alexander D, Manolio TA, Aguilar D, Coresh J, Heiss G, Boerwinkle E, Nettleton JA. Associations between metabolomic compounds and incident heart failure among african americans: The aric study. *American journal of epidemiology*. 2013:kwt004

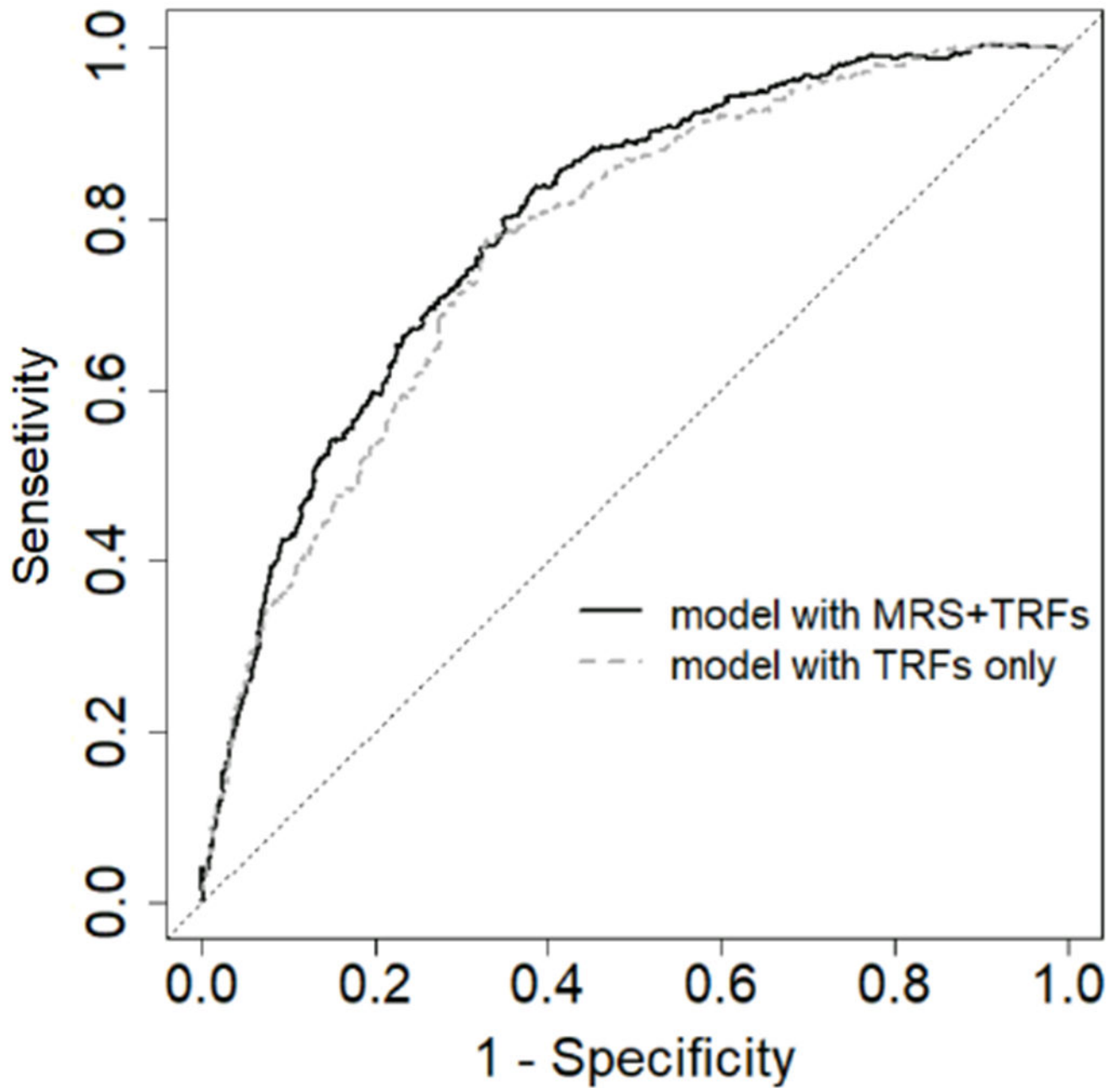
53. Zordoky BN, Sung MM, Ezekowitz J, Mandal R, Han B, Bjorndahl TC, Bouatra S, Anderson T, Oudit GY, Wishart DS, Dyck JRB, Alberta H. Metabolomic fingerprint of heart failure with preserved ejection fraction. *PLOS ONE*. 2015;10:e0124844 [PubMed: 26010610]
54. Wyczalkowska-Tomasik A, Czarkowska-Paczek B, Giebultowicz J, Wroczyński P, Paczek L. Age-dependent increase in serum levels of indoxyl sulphate and p-cresol sulphate is not related to their precursors: Tryptophan and tyrosine. *Geriatrics & Gerontology International*. 2017;17:1022–1026 [PubMed: 27240996]
55. Starakis I, Starakis I, Lekkou A, Blikas A, Labropoulou-Karatza C. Drug-induced cardiotoxicity due to aminophylline treatment: A case report(). *Current Therapeutic Research, Clinical and Experimental*. 2003;64:367–374
56. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. Gold executive summary. *American journal of respiratory and critical care medicine*. 2017;195:557–582 [PubMed: 28128970]
57. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. Nhlbi/who global initiative for chronic obstructive lung disease (gold) workshop summary. *American journal of respiratory and critical care medicine*. 2001;163:1256–1276 [PubMed: 11316667]
58. Fung TT, Malik V, Rexrode KM, Manson JE, Willett WC, Hu FB. Sweetened beverage consumption and risk of coronary heart disease in women. *The American journal of clinical nutrition*. 2009;89:1037–1042 [PubMed: 19211821]
59. de Koning L, Malik VS, Kellogg MD, Rimm EB, Willett WC, Hu FB. Sweetened beverage consumption, incident coronary heart disease, and biomarkers of risk in men. *Circulation*. 2012;125:1735–1741, s1731 [PubMed: 22412070]
60. Cohen L, Curhan G, Forman J. Association of sweetened beverage intake with incident hypertension. *Journal of general internal medicine*. 2012;27:1127–1134 [PubMed: 22539069]
61. Hootman KC, Trezzi JP, Kraemer L, Burwell LS, Dong X, Guertin KA, Jaeger C, Stover PJ, Hiller K, Cassano PA. Erythritol is a pentose-phosphate pathway metabolite and associated with adiposity gain in young adults. *Proceedings of the National Academy of Sciences of the United States of America*. 2017;114:E4233–e4240 [PubMed: 28484010]
62. Hiele M, Ghooys Y, Rutgeerts P, Vantrappen G. Metabolism of erythritol in humans: Comparison with glucose and lactitol. *The British journal of nutrition*. 1993;69:169–176 [PubMed: 8457525]
63. Zheng Y, Manson JE, Yuan C, Liang MH, Grodstein F, Stampfer MJ, Willett WC, Hu FB. Associations of weight gain from early to middle adulthood with major health outcomes later in life. *JAMA*. 2017;318:255–272 [PubMed: 28719691]
64. Kosek AB, Durbin D, Jonas A. Binding affinity and reactivity of lecithin cholesterol acyltransferase with native lipoproteins. *Biochemical and biophysical research communications*. 1999;258:548–551 [PubMed: 10329423]
65. Hirsch-Reinshagen V, Donkin J, Stukas S, Chan J, Wilkinson A, Fan J, Parks JS, Kuivenhoven JA, Lutjohann D, Pritchard H, Wellington CL. Lcat synthesized by primary astrocytes esterifies cholesterol on glia-derived lipoproteins. *Journal of lipid research*. 2009;50:885–893 [PubMed: 19065001]
66. Feng Q, Liu Z, Zhong S, et al. Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease. *Scientific Reports*. 2016;6:22525 [PubMed: 26932197]
67. Tenori L, Hu X, Pantaleo P, Alterini B, Castelli G, Olivotto I, Bertini I, Luchinat C, Gensini GF. Metabolomic fingerprint of heart failure in humans: A nuclear magnetic resonance spectroscopy analysis. *International Journal of Cardiology*. 2013;168:e113–e115 [PubMed: 24007967]
68. Bode-Böger SM, Böger RH, Galland A, Tsikas D, Frölich JC. L-arginine-induced vasodilation in healthy humans: Pharmacokinetic–pharmacodynamic relationship. *British journal of clinical pharmacology*. 1998;46:489–497 [PubMed: 9833603]
69. West SG, Likos-Krick A, Brown P, Mariotti Fo. Oral l-arginine improves hemodynamic responses to stress and reduces plasma homocysteine in hypercholesterolemic men. *The Journal of Nutrition*. 2005;135:212–217 [PubMed: 15671215]

### Highlights

- Identifying metabolomic signatures may provide insight into disease etiology and prevention.
- Using least absolute shrinkage and selection operator and a false discovery rate of 5%, we selected 19 metabolites that were associated with the risk of developing coronary heart disease, including a sugar substitute, erythritol.
- A metabolite risk score constructed from the 19 selected metabolites improved coronary heart disease 30 years risk prediction beyond traditional risk factors.
- Our results suggest several new pathways in the etiology of coronary heart disease, and highlight the use of metabolomics in CHD risk prediction.



**Figure 1.** Individual metabolite hazard ratios (95% confidence intervals) per SD increment for 19 LASSO metabolites with significant individual associations with coronary heart disease risk.



**Figure 2.** Receiver operating characteristic curves for coronary heart disease 30-year risk prediction, comparing a model with traditional risk factors (TRF) only and a model with TRF plus continuous metabolite risk score (MRS).



**Table 1.**

Baseline characteristics and number of incident coronary heart disease events among participants, Atherosclerosis Risk in Communities study.

	African Americans	European Americans	P-values	Overall
Participants, n	2232	1366		3598
Incident CHD, n (%)	359 (16.1%)	274 (20.1%)	0.003	633 (17.6%)
Age (years)	52.8 ± 5.6	54.2 ± 5.7	<0.001	53.3 ± 5.7
Male, n (%)	801 (35.9%)	602 (44.1%)	<0.001	1403 (40.0%)
SBP (mm Hg)	127.9 ± 21.0	119.1 ± 18.2	<0.001	124.5 ± 20.4
Anti-hypertensive medication use, n (%)	828 (37.1%)	276 (20.2%)	<0.001	1104 (30.7%)
HDL-c (mmol/L)	1.4 ± 0.4	1.3 ± 0.4	<0.001	1.4 ± 0.4
TC (mmol/L)	5.5 ± 1.1	5.6 ± 1.0	0.34	5.5 ± 1.1
Prevalent Diabetes, n (%)	367 (16.4%)	90 (6.6%)	<0.001	457 (12.7%)
Smoking status			<0.001	
Current, n (%)	1070 (47.9%)	556 (40.7%)		1626 (45.3%)
Former, n (%)	526 (23.6%)	463 (33.9%)		989 (27.6%)
Never, n (%)	636 (28.5%)	347 (25.4%)		983 (27.3%)
C-IMT	0.7 (0.1)	0.7 (0.2)	0.9	0.7 (0.1)
Plaque presence (%)	592 (28.1%)	425 (31.6%)	0.03	1017 (29.4%)

Abbreviations: CHD, coronary heart disease; SBP, systolic blood pressure; HDL-c, high-density lipoprotein cholesterol, TC, serum total cholesterol, and C-IMT, carotid intima-media thickness.

Data are expressed as mean ± standard errors for quantitative traits.

**Table 2.**

Hazard ratios relating metabolite risk score with incident coronary heart disease.

Contrast	HR	95% CI	P-value
Continuous MRS (per SD change)	1.37	1.27- 1.47	$<2 \times 10^{-16}$
MRS Quartiles			
Q2 vs. Q1	1.08	0.82- 1.43	0.57
Q3 vs. Q1	1.36	1.04- 1.77	0.02
Q4 vs. Q1	2.21	1.72- 2.84	$6.51 \times 10^{-10}$
Trend			$4.63 \times 10^{-13}$

Abbreviations: HR, hazard ratio; CI, confidence interval; MRS, metabolite risk score; SD, standard deviation. Each model adjusted for age, sex, race, study center, phase, smoking status, systolic blood pressure, anti-hypertensive medication use, diabetes status, total cholesterol, and high-density lipoprotein cholesterol levels.