The gut microbiota-related metabolite phenylacetylglutamine associates with increased risk of incident coronary artery disease

Filip Ottosson\textsuperscript{a}, Louise Brunkwall\textsuperscript{a}, Einar Smith\textsuperscript{a}, Marju Orho-Melander\textsuperscript{a}, Peter M. Nilsson\textsuperscript{a}, Céline Fernandez\textsuperscript{a}, and Olle Melander\textsuperscript{a,b}

Objective: The gut microbiota is increasingly being implicated in cardiovascular health. Metabolites produced by bacteria have been suggested to be mediators in the bacterial action on cardiovascular health. We aimed to identify gut microbiota-related plasma metabolites and test whether these metabolites associate with future risk of coronary artery disease (CAD).

Methods: Nontargeted metabolomics was performed using liquid chromatography-mass spectrometry in order to measure 1446 metabolite features in the Malmö Offspring Study (MOS) (N = 776). The gut microbiota was characterized using 16S rRNA sequencing. Gut bacteria-related metabolites were measured in two independent prospective cohorts, the Malmö Diet and Cancer – Cardiovascular Cohort (MDC-CC) (N = 3361) and the Malmö Preventive Project (MPP) (N = 880), in order to investigate the associations between gut bacteria-related metabolites and risk of CAD.

Results: In MOS, 33 metabolite features were significantly (P < 4.8e-7) correlated with at least one operational taxonomic unit. Phenylacetylglutamine (PAG) was associated with an increased risk of future CAD, using inverse variance weighted meta-analysis of age and sex-adjusted logistic regression models in MDC-CC and MPP. PAG remained significantly associated with CAD (OR = 1.17, 95% CI = 1.06–1.29, P = 1.9e-3) after adjustments for cardiovascular risk factors.

Conclusion: The levels of 33 plasma metabolites were correlated with the gut microbiota. Out of these, PAG was associated with an increased risk of future CAD independently of other cardiovascular risk factors. Our results highlight a link between the gut microbiota and CAD risk and should encourage further studies testing if modification of PAG levels inhibits development of CAD.

Keywords: coronary artery disease, gut microbiome, metabolism, microbiota, phenylacetylglutamine

Abbreviations: CAD, coronary artery disease; CKD, chronic kidney disease; CMD, cardiometabolic disease; CRP, C-reactive protein; CV, coefficient of variation; CVD, cardiovascular disease; eGFR, estimated glomerular filtration rate; FRS, Framingham risk score; GBR, gut bacteria-related; ICD, International Classification of Diseases; MDC-CC, The Malmö Diet and Cancer – Cardiovascular Cohort; MOS, The Malmö Offspring Study; MPP, The Malmö Preventive Project; OR, odds ratio; OTU, operational taxonomic units; PAG, phenylacetylglutamine; PWV, pulse wave velocity; QIIME, quantitative insight into microbial ecology; SCAAR, Swedish Coronary Angiography and Angioplasty Registry; T2DM, type 2 diabetes; TMAO, trimethylamine-N-oxide

INTRODUCTION

The human microbiome, the microbial genes colonizing the human body, outnumbers the size of the human genome with several folds. In particular, the microbes residing in the human gut, the gut microbiota [1], is increasingly being implicated in human health [2], including modulating the risk of type 2 diabetes (T2DM) [3] and cardiovascular disease (CVD) [3]. As coronary artery disease (CAD) is incompletely explained by conventional cardiovascular risk factors [4], knowledge of the molecular influence of the gut microbiota on cardiovascular health could help by introducing new therapeutic and predictive avenues for CVD.

Several studies have shown that the circulating levels of metabolites produced by bacteria, such as trimethylamine-N-oxide (TMAO) [5], secondary bile acids [6] and branched-chain amino acids [7], are partly regulated by the gut microbiota. This is particularly interesting as metabolomic studies have revealed disturbed levels of plasma metabolites that precede CVDs [8–10] but many of the mechanisms behind these alterations are unknown.
Some of these alterations could be influenced by the gut microbiota and their metabolites. Towards this direction, a well-known example is TMAO, whose precursor trimethylamine is being produced by choline-degrading bacteria in the gut. TMAO has been shown to promote atherosclerosis and associated with future risk of CAD, forming a link between the gut microbiota and cardiovascular health [5]. Although such studies have helped pinpointing molecular mediators in the connection between the gut microbiota and CVD, comprehensive nontargeted metabolomic studies are needed to find novel mediators of the gut microbiota action on cardiovascular health. In that light, we performed nontargeted metabolomics and 16S rRNA sequencing in 776 participants from the Malmö Offspring Study (MOS), in order to identify plasma metabolites that associate with gut microbes. The gut microbiota-related plasma metabolites were subsequently investigated in two independent Swedish prospective cohorts, constituting over 4000 participants and 800 incident cases of CAD, in order to study whether their levels associated with the risk of future CAD.

METHODS

Study samples

The Malmö diet and Cancer study is a population-based prospective cohort, where 28,098 individuals were enrolled between 1991 and 1996 in Malmö, Sweden [11]. From this cohort, 5405 participants were randomly selected to participate in the Malmö Diet and Cancer—Cardiovascular Cohort (MDC-CC), designed to study the epidemiology of carotid artery disease [12]. Among the 5405 participants who came fasted, citrate plasma was obtained from 3799 participants for analysis. The 3799 participants included are compared with participants with missing plasma samples in Table S1, http://links.lww.com/HJH/B411. Small, but statistically significant differences pointing towards more unfavorable cardiometabolic parameters in excluded versus included participants were seen. Participants with history of CAD (N=88) or diabetes (N=376) were excluded from the analyses. In the remaining population (N=3361), 423 participants developed CAD within an average follow-up time of 19.0 years.

The Malmö Preventive Project (MPP) is a population-based prospective cohort of 33,346 individuals, enrolled between 1974 and 1992. Between 2002 and 2006, 18,240 individuals (65–80 years old) were re-examined [13] for cardiometabolic risk factors and overnight fasting EDTA plasma was collected and anthropometric measurements were performed. Fecal samples were donated by 776 participants and frozen at −80 °C and were subjected to 16S rRNA sequencing. Nontargeted metabolomics was performed on EDTA plasma samples. The ethics committee of Lund University approved the study protocols for MOS (DNR 2012/594), MPP and MDC-CC (DNR 2009/633), and all participants provided written informed consent.

Endpoint definitions and biochemical measurements

CAD was defined as coronary revascularization, fatal or nonfatal myocardial infarction, or death attributable to ischemic heart disease. The study participants were followed for incident CAD through record-linkage using the Swedish personal identification number with the previously validated Swedish Hospital Discharge Register, the Swedish Cause of Death Register and the Swedish Coronary Angiography and Angioplasty Registry (SCAAR) [14]. International Classification of Diseases (ICD) codes and details about biochemical measurements are found in Material S1.

Liquid chromatography-mass spectrometry

Profiling of plasma metabolites was performed using a UPLC-QTOF-MS System (Agilent Technologies 1290 LC, 6550 MS, Santa Clara, California, USA) and Toronto Research Chemicals (Toronto, Canada). One quality control (QC) sample was thawed and extracted by addition of six volumes of extraction solution. The extraction solution consisted of 80:20 methanol/water containing stable isotope-labeled internal standards, purchased from Cambridge Isotope Laboratories (Andover, Massachusetts, USA) and Toronto Research Chemicals (Toronto, Canada). Detailed information about internal standards is found in Table S2, http://links.lww.com/HJH/B411. Extracted samples were separated on an Acquity UPLC BEH Amide column (1.7 μm, 2.1 × 100 mm; Waters Corporation, Milford, Massachusetts, USA). One quality control (QC) sample was injected every eight analytical sample in order to monitor and estimate the analytical precision, as measured by the coefficient of variation (CV) (Figure S1, http://links.lww.com/HJH/B411).

Only metabolite features that were measured in more than 80% of the participants in MOS were selected for statistical analysis (N=1446) in order to ensure high repeatability of studied metabolite features. Gut bacteria-related (GBR) metabolites were measured in MDC-CC and MPP and included in statistical analysis if CV less than 5% in the QC samples and present in more than 80% of the participants. Detailed information about the analytical procedures is found in Supplementary Material, S2, http://links.lww.com/HJH/B411 and the study workflow is presented in Figure S2, http://links.lww.com/HJH/B411.
Microbial metabolite phenylacetylglutamine and CVD

In MDC-CC and MPP, associations between 21 GBR metabolites and incident CAD were investigated using logistic regression models. Model 1 was adjusted for age and sex, whereas model 2 was additionally adjusted for SBP, smoking status, antihypertensive treatment, BMI and fasting levels of HDL and LDL cholesterol, triglycerides and glucose. Due to skewed distributions, metabolite data was log2-transformed prior to logistic regression analyses. Inverse-variance weighted meta-analyses (fixed intercepts) of logistic regression estimates from MDC-CC and MPP were performed for both model 1 and model 2. The significance threshold was set to P = 2.5e-3 (0.05/20 metabolites). Definition of the Framingham risk score (FRS) is found in Supplementary Material S1, http://links.lww.com/HJH/B411 and Tables S7–S12, http://links.lww.com/HJH/B411. All statistical analyses were performed using R 3.6.0. Meta-analyses were performed in meta and partial Spearman’s correlations using ppcor.

RESULTS

Using nontargeted metabolomics, a total of 1446 metabolite features were measured in MOS (N = 776). The cohort characteristics can be found in Table 1. Among measured metabolite features, 33 significantly (P < 4.8e-7) correlated with at least one out of 72 characterized gut bacterial OTUs, using partial Spearman’s correlation tests adjusted for age and sex (Fig. 1). In total, we identified 64 significant correlations between plasma metabolite levels and gut microbes. Metabolites that correlated with at least one OTU were considered to be GBR metabolites and included nine identified, six putatively annotated and eighteen unknown metabolites. A total of 20 OTUs, 4 phyla and 16 genera, were significantly correlated with at least one metabolite feature (Fig. 1 and Table S3–S4, http://links.lww.com/HJH/B412).

The two metabolites, Phenylacetylglutamine (PAG) and the unidentified metabolite X174, showed the highest number of significantly correlated (N = 7) OTUs each. Levels of PAG were significantly correlated with the bacterial phylum Ruminococcaceae with four genera in the Lachnospiraceae family and with an unknown genus in the family Christensenellaceae and in the order SHA-98 (Table 2). A majority Metabolic annotation was performed for gut bacteria-related metabolite features by matching spectra and retention times against an in-house metabolite library and publicly available databases. Annotations followed the Metabolomics Standard Initiative guidelines [16]. Definitions of annotation confidence levels are given in Supplementary Material S3, Table S13, http://links.lww.com/HJH/B411 and Figure S3, http://links.lww.com/HJH/B411.

16S rRNA sequencing

Faecal samples were collected in plastic tubes at home and stored in home freezer until they were brought to the clinic where they were stored at ~80 °C. The microbial DNA was extracted using the QIAamp column Stool Kit (300+2bp) and the V1–V3 region of the 16S rRNA gene was amplified and sequenced on a HiSeq Illumina at the GATC Biotech (Constance, Germany). The fastq files were aligned by FLASH[17] and binned together to operational taxonomical units (OTUs) using QIIME (Quantitative Insight Into Microbial Ecology 1.9.1) [18]. The sequences were matched with the reference database Greengene (13.8). Microbes that only occurred once or twice in the dataset and low-abundant bacteria (≤0.01%) were removed, leaving 64 bacteria characterized at genus level and belonging to eight microbial phyla, for further analysis. Square brackets indicate the Greengenes database notation for proposed taxonomy. Absolute counts were normalized with cumulative sum scaling (R package Metagenomvis).

Statistical analyses

In MOS, analyzes of correlations between plasma metabolite levels and gut bacteria were performed using partial Spearman’s correlation tests, adjusted for age and sex. In total, 104 112 (72 OTUs * 1446 metabolites), independent tests were performed, resulting in a Bonferroni-adjusted significance threshold of P = 4.8e-7 (0.05/104 112). Metabolites that were significantly correlated with at least one OTU were considered as GBR metabolite and were considered in further analyses. Correlations between 35 GBR metabolite levels and seven traditional cardiovascular risk factors were analyzed using partial Spearman’s correlation tests, adjusted for age and sex.

<table>
<thead>
<tr>
<th>Trait</th>
<th>MDC-CC</th>
<th>MPP</th>
<th>MOS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Whole cohort (N = 3362)</td>
<td>Incident CAD (N = 336)</td>
<td>Controls (N = 496)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>57.5 (±6.0)</td>
<td>57.7 (±5.8)</td>
<td>68.7 (±5.9)</td>
</tr>
<tr>
<td>Sex (% female)</td>
<td>58.8</td>
<td>56.7</td>
<td>37.2</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>25.4 (±3.7)</td>
<td>27.4 (±6.6)</td>
<td>26.5 (±4.2)</td>
</tr>
<tr>
<td>Fasting glucose (mmol/l)</td>
<td>4.9 (±0.4)</td>
<td>5.2 (±0.4)</td>
<td>5.4 (±0.5)</td>
</tr>
<tr>
<td>LDL cholesterol (mmol/l)</td>
<td>4.2 (±1.0)</td>
<td>4.3 (±1.1)</td>
<td>3.7 (±0.9)</td>
</tr>
<tr>
<td>HDL cholesterol (mmol/l)</td>
<td>1.4 (±0.4)</td>
<td>1.3 (±0.3)</td>
<td>1.4 (±0.4)</td>
</tr>
<tr>
<td>Triglycerides (mmol/l)</td>
<td>1.3 (±0.6)</td>
<td>1.5 (±0.7)</td>
<td>1.2 (±0.6)</td>
</tr>
<tr>
<td>Current smokers (%)</td>
<td>27.5</td>
<td>28.7</td>
<td>18.1</td>
</tr>
<tr>
<td>Diabetes (%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Plasma metabolomics in MPP was performed in 880 individuals, 384 coronary artery disease and 496 remained free from disease. In MOS, plasma metabolomics was performed in 776 individuals. In MDC, plasma metabolomics was performed in 336 participants, where 336 developed CAD. Table displays the average of traditional risk factors for cardiometabolic disease in the three groups. Numbers in parenthesis indicate the standard deviations. HDL, high-density lipoprotein; LDL, low-density lipoprotein; MDC-CC, Malmö Diet and Cancer – Cardiovascular Cohort MPP, Malmö Preventive Project; MOS, Malmö Offspring Study.
(67%) of the GBR metabolites were only significantly correlated with one OTU. Among the OTUs, the phylum Firmicutes was correlated with the highest number of metabolites (N=14), followed by the genus Blautia (N=10).

We next investigated whether GBR metabolites were related to cardiometabolic disease (CMD) risk factors in MOS. There were significant correlations between several metabolites and CMD risk factors, in particular with BMI (N=15), smoking status (N=12) and fasting levels of HDL cholesterol (N=10) and triglycerides (N=15) (Fig. 2). GBR metabolite levels were in general less strongly correlated with SBP (N=6) and fasting levels of LDL cholesterol (N=7) and glucose (N=5). All correlations between GBR metabolites and CMD risk factors in MOS are found in Table S5, http://links.lww.com/HJH/B412.

Glutamate and beta-carotene were the GBR metabolites that correlated most strongly positively and beta-carotene the most strongly inversely correlated with BMI, triglycerides and HDL cholesterol. Beta-carotene, PC 22:5/P18:1 and LysoPE 22:0 were correlated with six out of seven CMD risk factors. Seven GBR metabolites, including PAG, were not significantly associated with any CMD risk factor.

In additional metabolomic analyses, GBR metabolites were measured in two independent Swedish prospective cohorts: MDC-CC and MPP. Among 33 GBR metabolite features, 20 were measured in both MDC-CC and MPP. Differences in the participants’ baseline age, sex distribution and cardiometabolic risk factors exist between the three investigated cohorts (Table 1). We next investigated whether levels of GBR metabolites are associated with incident CAD. In primary analyses, age and sex-adjusted logistic regression models were applied in both MDC-CC and MPP and combined using inverse variance weighted meta-analysis to identify GBR metabolites associated with incident CAD. Increased levels of PAG were significantly (P<2.5e-3) associated with an increased risk of CAD in logistic regression models adjusted for age and sex (Fig. 3). Trigonelline and cystine were nominally associated (P<0.05) with an increased and LysoPCO 16:0 with a decreased risk of CAD. All associations between GBR metabolites and incident CAD were considered as significant (*) at P<4.8e-7.

**TABLE 2.** Associations between operational taxonomic units and phenylacetylglutamine and The Framingham Risk Score in the Malmö Offspring Study (N = 776)

<table>
<thead>
<tr>
<th>Gut bacteria</th>
<th>Phenylacetylglutamine</th>
<th>Framingham risk score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rho</td>
<td>P</td>
</tr>
<tr>
<td>Unknown genus of Christensenellaceae</td>
<td>0.27</td>
<td>1.2e-14</td>
</tr>
<tr>
<td>L.Baunia</td>
<td>−0.25</td>
<td>3.8e-13</td>
</tr>
<tr>
<td>Unknown genus of SHA-98</td>
<td>0.21</td>
<td>1.7e-9</td>
</tr>
<tr>
<td>L.Dorea</td>
<td>−0.20</td>
<td>2.4e-8</td>
</tr>
<tr>
<td>Unknown genus of Lachnospiraceae</td>
<td>−0.20</td>
<td>2.9e-8</td>
</tr>
<tr>
<td>Verrucomicrobia</td>
<td>0.18</td>
<td>2.4e-7</td>
</tr>
<tr>
<td>L.Ruminococcus</td>
<td>−0.18</td>
<td>3.8e-7</td>
</tr>
</tbody>
</table>

Spearman’s correlation coefficients (rho) are partial correlations, adjusted for age and sex. All operational taxonomic units are genera except Verrucomicrobia, which is a phylum.
Microbial metabolite phenylacetylglutamine and CVD

FIGURE 2 Correlations between plasma levels of 33 metabolites and BMI, smoking status and levels of high-density lipoprotein cholesterol and triglycerides in the Malmö Offspring Study (N = 776). Correlation coefficients are partial Spearman’s correlation coefficients, adjusted for age and sex.

CAD in both cohorts separately are presented in Table S5, http://links.lww.com/HJH/B412. In secondary analysis, additionally adjusted for CMD risk factors, the association between PAG and incident CAD remained significant (OR = 1.17, CI = 1.06–1.29, P = 1.9e-3). As PAG was associated with incident CAD independently of traditional risk factors, we hypothesized that the association could be mediated or confounded by decreased renal function, inflammation or arterial stiffness. This hypothesis was based on previous studies showing that PAG levels correlate with renal function [19] and arterial stiffness [20] and that inflammation might be a link between the gut microbiota and cardiovascular disease [3]. PAG was negatively correlated with the estimated glomerular filtration rate (eGFR), in both MPP (rho = −0.27, P < 0.001) and MDC-CC (rho = −0.18, P < 0.001) but the association with incident CAD was only marginally influenced when further adjusting for eGFR (OR = 1.15, CI = 1.04–1.27, P = 6.1e-3). In MDC-CC, it was shown that the association between PAG and incident CAD was not influenced by adjustments for C-reactive protein (CRP) and that there were no significant correlation between PAG and CRP (rho = 0.01, P = 0.65). Finally, PAG was shown not to be correlated with pulse wave velocity (PWV) in MOS (rho = −0.01, P = 0.76). In MOS, PAG was significantly correlated with gut microbiota diversity, as measured with Shannon’s diversity index (rho = 0.18, P = 1.5e-6). Moreover, four out of seven OTUs that were significantly correlated with PAG levels were correlated with FRS (Table 2). *Blautia* and unknown genus of *Lachnospiraceae* and *Ruminococcus* were positively correlated with FRS but inversely correlated with PAG, whereas SHA-98 was inversely correlated with FRS and positively correlated with PAG. As levels of PAG were positively correlated with gut microbe abundances that are more common in individuals with lower traditional risk for cardiovascular disease, we speculate that the association between PAG and incident CAD was different in individuals with low compared with high FRS. However, by stratifying on FRS (above vs. below FRS >12% 10-year risk), it was shown that the associations between PAG and incident CAD were significant in both low-risk (OR = 1.16, CI = 1.00–1.34, P = 0.05) and high-risk (OR = 1.15,
CI = 1.02–1.30,  \( P = 0.02 \) participants from MDC-CC and MPP.

**DISCUSSION**

We here connect several plasma metabolites with the gut microbiota and show that plasma levels of the GBR metabolite PAG are associated with future risk of CAD in the primary preventive setting independently of traditional cardiovascular risk factors. Moreover, several other GBR metabolites were correlated with cardiovascular risk factors. These metabolites may be important mediators in the relation between the gut microbiota and CVD.

PAG is a well known microbial metabolite [21] that is formed from the conjugation of glutamine and phenylacetate, which almost exclusively originates from bacterial phenylalanine metabolism [22]. Several metabolomic studies have related circulating PAG levels to different families of gut microbes such as Coriobacteriaceae [23], Mogibacteriaceae [23], Bifidobacteriaceae [21], with genera in the Lachnospiraceae, Christensenellaceae and Ruminococcaceae families [19] and with gut microbiome diversity [24].
This study is an important replication of the association between plasma levels of PAG and gut microbiome diversity, as PAG was positively correlated with Shannon’s diversity index. Moreover, three genera in the Lachnospiraceae family that were inversely correlated with PAG, were themselves positively correlated with the FRS. Similarly, an uncharacterized genus in the SHA-98 order was positively correlated with PAG and inversely correlated with the FRS. This seemingly paradoxical relationship indicates that bacteria that are more common in individuals with low FRS could be involved in increasing plasma levels of the CAD-related metabolite PAG. Thus, one could speculate that the associations between PAG and incident CAD should be more pronounced in individuals with low FRS. However, our findings show that the association between PAG and incident CAD is significant in both individuals with high and with low of FRS. Moreover, PAG was neither related to any of the individual CMD risk factors nor to FRS. As PAG displays low or no correlation with traditional cardiovascular risk factors, and the association with future CAD was independent of these, it is tempting to speculate that interventions lowering PAG in the circulation might reduce CAD risk on top of any ‘conventional’ primary preventive interventions (e.g. cholesterol and blood pressure-lowering interventions). Importantly, with the average follow-up time of almost 20 years in MDC-CC, our results indicate that such PAG-lowering strategies could be implemented early to potentially reach a substantial reduction in absolute risk. Our findings further strengthen a recent study showing that PAG associated with increased risk of cardiovascular disease in American prospective cohorts. That study further provided experimental evidence that PAG may increase platelet activation via G-protein-coupled receptors, ultimately resulting in increased thrombosis potential [25]. Taken together, this indicates that clinical studies testing if modification of circulating PAG reduces atherosclerosis and CAD risk are warranted.

PAG is well studied in the context of kidney function, where high plasma levels of PAG have been observed in patients with chronic kidney disease. Despite the fact that PAG excretion is mainly dependent on tubular secretion, circulating levels [19] of PAG strongly correlates with lower eGFR, whereas the renal clearance [26] of PAG correlates with higher eGFR. Interestingly, in a smaller study of patients with chronic kidney disease (CKD) circulating PAG was found to be associated with risk of future CVD [26]. The present study replicates the correlation between PAG and eGFR in two separate cohorts, while also showing that the association between PAG and incident CAD is independent of eGFR. PAG has previously been shown to correlate with increased arterial stiffness in women [20]. Given the association between arterial stiffness and increased risk of cardiovascular disease, it is possible that the association between PAG and incident CAD could be driven by arterial stiffness. This is particularly interesting as a possible link has been found between the gut bacterial diversity and arterial stiffness [27]. However, our finding in MOS, where we found no correlation between PAG and PWV, questions whether arterial stiffness is the link between PAG and increased risk of CAD. Furthermore, it has been suggested that inflammation is a link between the gut microbiota and cardiovascular disease [3]. Our findings indicate that the association between PAG and incident CAD is independent of inflammation, as the association remained significant after adjustment for CRP in MDC-CC.

Our study identifies several other GBR metabolites, such as glutamate, hippurate, urobilin, beta-carotene and LPCO 16:1. Although none of these were associated with incident CAD in this study, several of the GBR metabolites were significantly correlated with cardiovascular risk factors. In particular, glutamate was strongly positively correlated and beta-carotene inversely correlated with cardiovascular risk factors. We and others have previously suggested that plasma levels of glutamate could be influenced by the gut microbiota [28,29] and predict future T2DM and CAD [8]. Interestingly, both beta-carotene and LPCO 16:1 have been associated with lower risk [30,31] and urobilin with an increased risk [32] of T2DM in previous studies. This proposes that the gut bacteria Blautia and an unknown genus in the SHA-98 order, that were correlated with the levels of these metabolites, should be further investigated in relation to T2DM. Hippurate has previously been shown to correlate with several gut microbes and microbiota diversity and was further suggested to be protective against the metabolic syndrome [24].

**Limitations**

This study has several limitations. The observational nature of the study design makes us unable to prove causal links between gut bacteria, PAG and coronary artery disease. The microbiota assessment was performed using 16S rRNA sequencing, which did not allow us to study the bacteria at the species level. Although collection of fecal samples is the only convenient method to study gut microbiota composition in large cohort material, the microbiota composition of fecal samples may not fully represent the entire gut. Moreover, lack of stool consistency assessments makes us unable to show that correlations between metabolites and gut microbes were independent of stool consistency. Finally, measurement of CRP and PWV were not available in all the three investigated cohorts, preventing us from verifying that the association between PAG and incident CAD was independent of inflammation and arterial stiffness.

In conclusion, we here describe multiple connections between the gut microbiota, the plasma metabolome and cardiovascular risk in one of the largest studies performed involving untargeted metabolomics and prospective endpoint evaluation. Plasma levels of PAG, which is strongly related to gut microbiota composition, independently predicts CAD in the primary preventive setting. Our results highlight a link between the gut microbiota and CAD risk and should encourage further studies testing if modification of PAG levels inhibits development of CAD.

**ACKNOWLEDGEMENTS**

We thank all the participants in the Malmö Offspring Study, the Malmö Preventive Project and the Malmö Diet and Cancer – Cardiovascular Cohort. Additionally, we thank data manager Anders Dahlin for extensive quality control of MOS data, and Johan Hultman for working with setting up the microbiota pipeline.

**Microbial metabolite phenylacetylglutamine and CVD**

This study was supported by: The Swedish Research Council [2433]; the Swedish Heart and Lung Foundation; the Swedish Medical Research Council; the Swedish Society for Medical Research; The Greta and Johan Kock Foundation; Formas; The Swedish Cancer Society; The European Union’s Seventh Framework Programme (FP7/2007-2013) for research, technological development and demonstration under grant Agreement Number 278142 (CVD-ARC); The Swedish Heart Lung Foundation; The Ake Wiberg Foundation; the Swedish Heart Lung Foundation; The Pehr and Sonya Wallenberg Foundation; The Dunér Foundation; The Hjalmar Stelt Foundation; The Nils Ericsson Foundation; the Stockholm County Council; the Region of Skåne and the University of Lund; The Magnus Bergvall Foundation; the Regeringskansliet; and the Swedish National Board of Health and Welfare. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
Funding: The Malmö Offspring Study was supported by the Swedish Research Council, Region Skåne University Hospital, the Heart and Lung Foundation, and the European Foundation for Study of Diabetes (EFSF 2015/338). F.O. was supported by Ernhold Lundström Research Foundation. O.M. was supported by research grants from the Knut and Alice Wallenberg Foundation, Göran Gustafsson Foundation, the Swedish Heart- and Lung Foundation, the Swedish Research Council, the European Research Council ERC-ADG-2019-885003, the Novo Nordisk Foundation, Region Skåne, Skåne University Hospital and Lund University. L.B. was supported by Albert Pålsson Research Foundation. C.F. was supported by the Albert Pålsson Research Foundation, the Crafoord Research Foundation, the Ernhold Lundström Research Foundation, the Royal Physiographic Society of Lund and the Ake Wiberg Foundation. M.O.M. was supported by European Research Council ERC-CoG-2014-640921 (for M.O.-M.), the Novo Nordisk Foundation (NNF18OC0034386), the Swedish Research Council, the Swedish Heart and Lung Foundation and the Swedish Diabetes Foundation. The study was additionally supported by the Swedish Research Council (for Strategic Research Area Exodiab, Dnr 2009–1039), the Swedish Research Council for Strategic Research (Dnr IRC15-0067) and the Swedish Research Council (Linnaeus grant, Dnr 349-2006-237).

Conflicts of interest
There are no conflicts of interest.

REFERENCES


