Food & Function

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Food Funct.*, 2025, **16**, 7067

Integrating population-based metabolomics with computational microbiome modelling identifies methanol as a urinary biomarker for protective diet–microbiome–host interactions

Kristin Klier, *[a] Ameneh Mehrjerd,[a] Daniel Fässler,[a] Maximilien Franck,[b] Antoine Weihs,[a,c] Kathrin Budde,[d] Martin Bahls,[e,f] Fabian Frost,[g] Ann-Kristin Henning,[d] Almut Heinken,[h] Henry Völzke,[e,i] Marcus Dörr,[e,i] Matthias Nauck,[d,e] Hans Jörgen Grabe,[a,c] Nele Friedrich†[d,e] and Johannes Hertel†[a,e]

Background: Diet–microbiome interactions are core to human health, in particular through bacterial fibre degradation pathways. However, biomarkers reflective of these interactions are not well described. *Methods*: Using the population-based SHIP-START-0 cohort (n = 4017), we combined metabolome-wide screenings with elastic net machine learning models on 33 food items captured using a food frequency questionnaire (FFQ) and 43 targeted urine nuclear magnetic resonance (NMR) metabolites, identifying methanol as a marker of plant-derived food items. We utilised the independent SHIP-START-0 cohort for the replication of food–metabolite associations. Moreover, constraint-based microbiome community modelling using the Human Microbiome data (n = 149) was performed to predict and analyse the contribution of the microbiome to the human methanol pools through bacterial fibre degradation. Finally, we employed prospective survival analysis in the SHIP-START-0 cohort, testing urinary methanol on its predictive value for mortality. *Results*: Among 21 metabolites associated with 17 dietary FFQ variables after correction for multiple testing, urinary methanol emerged as the top hit for a range of plant-derived food items. In line with this, constraint-based community modelling demonstrated that gut microbiomes can produce methanol *via* pectin degradation with the genera *Bacteroides* (68.9%) and *Faecalibacterium* (20.6%) being primarily responsible. Moreover, microbial methanol production capacity was a marker of high microbiome diversity. Finally, prospective survival analysis in SHIP-START-0 revealed that higher urinary methanol is associated with lower all-cause mortality in fully adjusted Cox regressions. *Conclusion*: Integrating population-based metabolomics and computational microbiome modelling identified urinary methanol as a promising biomarker for protective diet–microbiome interactions linked to microbial pectin degradation.

Received 12th February 2025,
Accepted 29th July 2025

DOI: 10.1039/d5fo00761e

rsc.li/food-function

[a]*Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Ellernholzstraße 1-2, 17489 Greifswald, Germany. E-mail: kristin.klier@med.uni-greifswald.de*
[b]*Research Center on Aging, Faculty of Medicine and Health Sciences, University of Sherbrooke, Sherbrooke, QC, Canada*
[c]*German Center for Neurodegenerative Diseases (DZNE), Greifswald, Germany*
[d]*Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany*
[e]*German Centre for Cardiovascular Research (DZHK), Partner Site Greifswald, Greifswald, Germany*
[f]*Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany*
[g]*Department of Medicine A, University Medicine Greifswald, Greifswald, Germany*
[h]*UMRS Inserm1256 NGERE (Nutrition-Genetics-Environmental Risks), Institute of Medical Research (PôleBMS) – University of Lorraine, Vandoeuvre-les-Nancy, France*
[i]*Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany*
†These authors contributed equally.

1. Introduction

Dietary fibres derived from plant-based foods, including whole grains, vegetables, or fruits,[1] have been established to be essential components of a healthy diet.[2,3] Their microbial fermentation products include short-chain fatty acids, which provide energy to colonocytes and have anti-inflammatory and anti-tumour effects.[4] Other degradation products include alkylresorcinols, which display neuroprotective, muscle-protective, and metabolism-positive effects.[5] Reflecting the importance of fibre degradation products, accumulating evidence suggests that a fibre-rich diet reduces the risk of non-communicable diseases, including cardiometabolic diseases and cancer.[1,6]

However, the complex diet–host–microbiome interplay in fibre degradation complicates the interpretation of metabolite–

diet associations. For example, dietary fibre intake has been associated with 2,6-dihydroxybenzoic acid (2,6-DHBA), indolepropionic acid, linolenoyl carnitine, 2-aminophenol, 3,4-DHBA, and proline betaine,[7] but underlying microbial pathways remain elusive. Recent progress in gut microbiome modelling allows for the computational and quantitative description of microbial fibre degradation[8–10] *via* constraint-based reconstruction and analysis approaches. This approach is based on personalised microbiome community models that have been shown to be predictive for host metabolomics traits.[8,10,11] So far, however, COBRA community modelling has not been employed for the interpretation and contextualisation of metabolome–food associations derived from large population studies.

Here, we combine metabolome-wide association studies and COBRA modelling with *in silico* fibre supplementation experiments, identifying methanol as a marker of diet–microbiome interactions linked to pectin degradation. We utilised the SHIP-START-0 cohort (n = 4017)[12] for discovery metabolome-wide association studies and the independent SHIP-TREND-0 cohort (n = 992)[12] for replication and applied COBRA community modelling[13] to samples from the Human Microbiome Project (n = 149). Using data from the SHIP project, we further investigated the association between urinary methanol and health-promoting lifestyle habits and determined whether urinary methanol could be predictive for mortality rates (all causes, cancer, cardiovascular diseases (CVDs)).

2. Methods

2.1 Study population SHIP-START-0 and SHIP-TREND-0

For our analysis, we used the population-based SHIP-study, conducted in north-eastern Germany.[12] Its primary objective was to examine the prevalence and incidence of common risk factors, subclinical disorders, and clinical diseases. The initial data acquisition (SHIP-START-0, 1997–2001, n = 4307) was derived from local registries. Subsequently, a second, independent baseline cohort was selected from the same geographical region (SHIP-TREND-0, 2008–2012, n = 4420), with no participant overlap from the SHIP-START-0 cohort. Further details of the measurements performed can be found elsewhere.[12]

Non-fasting targeted urine nuclear magnetic resonance (NMR) metabolome data were available for n = 4068 individuals of SHIP-START-0. To assess the frequency of food intake and its association with NMR metabolite data, we excluded (1) subjects taking antibiotics (n = 35) and (2) pregnant participants (n = 16). In total, we included n = 4017 SHIP-START-0 individuals.

In the replication study SHIP-TREND-0 (n = 4420), n = 996 individuals with targeted urine NMR measurements were available. However, this urine NMR measurements were conducted exclusively on a subset of fasting participants without self-reported diabetes. As a result, the utilised SHIP-TREND-0 subsample is a predominantly healthy cohort. Excluding (1) pregnant individuals (n = 0) and (2) those on antibiotics (n = 4) we obtained an analysis sample of n = 992 participants (SI1).

2.2 Covariate measurements

In the baseline SHIP examinations, behaviour, socioeconomic data, medical history, and sociodemographic factors were obtained from a computer-assisted interview. Furthermore, extensive medical examinations were performed, including measurements of waist circumference (considered as an indicator of abdominal fat), body height, weight, and blood pressure. The smoking variable was categorised into current smoking and non-smoking. Physical activity in leisure time was defined as either no activity (neither in summer nor in winter) or steady activity (>1 h per week in both summer and winter). Between the discovery and replication studies, covariate definitions differed for alcohol consumption (SHIP-START-0: intake during the last 7 days, SHIP-TREND-0: intake during the last 30 days) and sleeping problems (SHIP-START-0: 5 categorical options, SHIP-TREND-0: 3 categorical options).

Participants were asked to bring their medication prescriptions or package receipts for all medications they had taken in the past seven days. Each medication was recorded and categorised according to the Anatomical Therapeutic Chemical Classification (ATC Index, 2007). For biomarker measurements, blood and urinary samples were collected and either analysed directly or stored at −80 °C. The details of the procedures have been described elsewhere.[14] The assays for analysing the blood and metabolic markers were all conducted by skilled technical personnel following the manufacturer's recommendations. Concentrations of glycated haemoglobin (HbA1c) were measured by high-performance liquid chromatography (Bio-RadDiamat, Munich, Germany) and triglycerides (tg) were determined photometrically (Hitachi 704, Roche, Mannheim, Germany). Urine creatinine concentrations were determined using the Jaffé-method (Hitachi717, Roche Diagnostic, Mannheim, Germany).

Information on the vital status was obtained from population registers at annual intervals. Participants were censored in the event of death or lack of follow-up. The follow-up length was defined as the number of months between the baseline examination and censoring. A request for death certificates (coded by a certified nosologist according to the International Classification of Diseases, 10th version) was made to the local health authority of the residence of death.

Dietary intake in the discovery study SHIP-START-0 was captured by a face-to-face interview using an previously validated food frequency questionnaire (FFQ)[15] including 33 food items. In contrast, the FFQ in SHIP-TREND-0 measured dietary behaviour with a reduced number of 16 food categories. Within both cohorts, food intake was rated on an ordinal scale with 6 options (1: daily or almost daily, 2: several times a week, 3: about once a week, 4: several times a month, 5: once a month or less often, and 6: never or almost never).

2.3 Urinary metabolite quantification using targeted NMR measurements

In SHIP-START-0, spontaneous urine samples were collected from non-fasting individuals. Conversely, SHIP-TREND-0 participants were fasting prior to biosample collection. Details on the NMR measurements are provided in SI2.

In SHIP-START-0, 59 metabolites with their concentrations in millimoles per litre (mM) were identified. In SHIP-TREND-0, urinary methanol measured by NMR was exclusively examined as the primary outcome from the discovery study.

2.4 Outlier detection and data normalisation

Metabolites with over 50% missing quantification in the NMR spectra were excluded from the statistical analysis. As a result, 43 of 59 targeted NMR metabolites were selected for subsequent analysis in SHIP-START-0. Outliers, defined by the 4 standard deviation rule were calculated for each metabolite and excluded from the analysis. For all urinary metabolite concentrations, log-transformations were performed. To compensate for metabolite-specific dilution-concentration relations, we applied a regression-based approach using probabilistic quotient normalisation(PQN) with restricted cubic splines(RCS) using 4 knots.[16]

2.5 Statistical analysis

Descriptive statistics were presented as means and standard deviations for metric variables and proportions for categorical variables. All analyses and graphs were conducted using R (version 4.2.3). *P*-Values were calculated two-sided and multiple testing was corrected using the false discovery rate (FDR) with a threshold of <0.05.

2.5.1 Metabolic signatures of dietary intake. Fully adjusted multiple linear regression analyses were conducted to associate dietary intake (predictor of interest) with the metabolite profile (response variable). Testing 33 food frequency categories on 43 urinary metabolites resulted in 43 × 33 regressions, utilising heteroscedastic robust standard errors (HRSE). Nonlinearities were modelled using RCS with four knots using the default setting of the R-package "rms".[17] Additionally, three continuous beverage variables (coffee, decaffeinated coffee, and tea in cups per day) were analyzed separately but analogously. The model's explication ensued:

$$\begin{aligned}\text{Metabolite concentration} = {} & \beta_0 + \beta_1(\text{food/beverage item}) \\ & + \beta_2(\text{rcs}(\text{PQN}, 4)) \\ & + \beta_3(\text{rcs}(\text{age}, 4) \times \text{sex}) \\ & + \beta_4(\text{rcs}(\text{waist circumference}, 4)) \\ & + \beta_5(\text{rcs}(\text{eGFR}, 4)) \\ & + \beta_6(\text{pH-value}) \\ & + \beta_7(\text{physical inactivity}) \\ & + \beta_8(\text{smoking status}) \\ & + \beta_9(\text{alcohol intake} \times \text{sex}) \\ & + \beta_{10}(\log(\text{GGT})) \\ & + \beta_{11}(\text{hypertension}) \\ & + \beta_{12}(\text{education years}) \\ & + \beta_{13}(\text{sleeping problems}) \\ & + \beta_{14}(\text{tg}) + \beta_{15}(\text{white blood cells}) \\ & + \beta_{16}(\text{red blood cells}) \\ & + \beta_{17}(\text{total-hdl-cholesterol ratio}) \\ & + \beta_{18}(\text{presence of diabetes}) \\ & + \beta_{19}(\text{time since last-meal}) + e\end{aligned} \quad (1)$$

To assess the statistical significance of the categorical variable food item, a global Wald-test was performed.

Furthermore, the direction of the association between food intake frequency and urinary metabolites was assessed by repeating the analysis to compare the frequent and rare food-item categories: 1 ("every day or almost every day") and 2 ("several times a week") *versus* 5 ("about once a month or fewer") and 6 ("never or almost never").

For external validation, the analyses were replicated within SHIP-TREND-0 focusing on methanol as the top result. The multiple linear regression models were performed in an analogous way to previous analyses with 16 dietary food categories available in SHIP-TREND-0. Here, food item frequencies (ranging from 1: daily to 6: never) selected by fewer than 10 individuals were reclassified (fresh fruit intake with a frequency of 6 were assigned to 5, and fried potatoes, pasta and rice intake with a frequency of 1 were reclassified to 2).

2.5.2 Predicting urinary methanol *via* FFQ using machine learning. We utilised elastic nets to determine the prediction of urinary methanol concentrations through the dietary intake frequency. Elastic nets integrate the sparsity-inducing Lasso penalty and the coefficient-shrinking L2 regularization from Ridge regression.[18] The idea is to improve the reduction of overfitting in addition to the reduction of dimensionality. The objective function with elastic net regularisation is as follows:

$$\arg\min_{\beta} \frac{1}{2n}\sum_{i=1}^{n}(y_i - \tilde{y}_i) + \gamma\left(\alpha\sum_{j=1}^{n}\left|\beta_j\right| + (1-\alpha)\sum_{j=1}^{n}\left(\beta_j{}^2\right)\right) \quad (2)$$

Here, n represents the number of predictors, y_i is the observed target value for the i-th data point, and $\tilde{y}_i$ is the predicted target value for the i-th data point, based on the linear regression model. The α is the mixing parameter that determines the combination of L1 and L2 regularisation in the elastic net. In addition, γ is the regularization parameter, controlling the strength of regularisation, and β_i represents the coefficient for the j-th feature. As the outcome variable, the residuals of the urinary methanol concentrations were used after regression out the dilution *via* regression-based normalisation.[16] Dummy variables were generated for the food items, with the lowest intake frequency considered as the reference category. Dietary intake categories with fewer than 5% individuals selecting a specific food frequency were omitted from the analysis. We used k-Nearest Neighbour imputation[19] to deal with missing data in the FFQ data. For assessing model fit, 10-fold internal cross-validation was utilised. Finally, *r*-squared measures, mean absolute errors and root mean squared errors were used to assess the model's performance.

2.5.3 Urinary methanol and lifestyle factors. Next, the potential associations of urinary methanol with basic and behavioural covariates, physiological parameters, and clinical phenotypes were analysed in the SHIP-START-0 and SHIP-TREND-0 cohort. We performed multiple linear regressions, incorporating HRSE, RCS and interaction terms with an identical model setup as in previous regressions (Table S1). In addition, the explained variance (incremental R^2) was calculated for the different variables to determine the

explained variance of urinary methanol concentration related to the investigated factors.

2.5.4 Constraint-based modelling. We obtained relative abundances of 149 samples from the Human Microbiome Project,[20] previously mapped to genome-scale reconstructions.[21] Initially, we quantified species onto the reference set of 7302 microbial metabolic reconstructions of AGORA2.[8] Subsequently, microbial community modelling was performed using the Microbiome Modelling Toolbox.[22,23] The mode of operation is explained in more detail in ref. 10, 22 and 23. Briefly, community models were generated by combining genome-scale reconstructions from AGORA2 that were present in the abundance table for each sample. The community biomass reaction was then parametrized by weighting by the biomass reactions of microbes present by its corresponding abundance. In each simulation, we calculated maximum community net secretion fluxes for each metabolite that is present in the lumen compartment for each sample, calculating the maximum net secretion for 791 metabolites in total, including methanol. One simulation was done by only using the Average European Diet constraints that does not neither include an uptake rate of pectin or xylan and is included in the COBRA toolbox.[13] Then, we conducted four simulations utilizing the Average European Diet constraints of the virtual metabolic human database,[24] where we additionally gradually incremented the diet constraints (*i.e.* the maximal uptake rate) of pectin and xylan for each model respectively. This approach was based on the premise, that an average apples weights approximately 200 g, contains about 1% pectin and possesses a molecular weight of roughly 100 kilodaltons (kDa). This translates to a dietary constraint expressed as:

$$\begin{aligned} &2\,\text{g per } 100\,\text{kDa} \times 1\,\text{per person per day} \\ &= 0.2\,\text{mmol per person per day.} \end{aligned} \tag{3}$$

We gradually stacked up the diet constraints of pectin to the Average European Diet by 0.2 mmol per person per day at each step, up to 0.8 mmol per person per day.

We conducted a comparative analysis of our findings by choosing xylan, which served as a control polysaccharide. To facilitate this comparison, we scaled the uptake rate of xylan to an initial uptake rate equivalent to 0.2 mmol per person per day of pectin by the amount of carbon atoms:

$$\begin{aligned} &\frac{\#\,\text{C-atoms of pectin}}{\#\,\text{C-atoms of xylan}} \times 0.2\,\text{mmol per person per day} \\ &= \frac{2535}{2640} \times 0.2\,\text{mmol per person per day} \\ &\approx 0.192\,\text{mmol per person per day.} \end{aligned} \tag{4}$$

As done with pectin, we gradually increased the diet constraint of xylan from 0.192 mmol per person per day up to 0.768 mmol per person per day.

To calculate the individual maximum secretion potential of methanol of each microbe in a COBRA community model, we utilized the predictMicrobeContributions function of the COBRAtoolbox, where we used the Average European Diet and added the before applied maximum pectin constraint. With this function, instead of maximizing the combined net secretion reaction of the community model, each internal exchange reaction of each microbe present into the lumen compartment gets maximised. The average direct production effect of a species on methanol secretion was calculated as the product of the mean abundance and the regression slope of the species methanol production against the species abundance. The total effect of a species on methanol secretion was defined as the product of the mean abundance with the regression slope of the community methanol production against the species abundance. The ecological effect is then defined by the difference between direct and total effect. For details, see Hertel *et al.*[9] All simulations were performed in MATLAB (Mathworks, Inc.) version R2021a, with IBM CPLEX(IBM) as the linear programming solver. The simulations were carried out using the COBRA Toolbox[13] and the Microbiome Modelling Toolbox.[22,23] Using a linear regression, we additionally tested the association between methanol and hippuric acid, a recently reported urinary marker of microbiome diversity.[14]

2.5.5 Urinary methanol and mortality rates. Finally, we examined methanol as a predictive biomarker in prospective survival analysis. By using the follow-up data of the SHIP-START-0 cohort, a total of 18 ($3 \times 3 \times 2$) Cox proportional hazard models were calculated, involving for each mortality rate of all causes, cancer and CVD the calculation of the *p*-values for the linear effect, non-linear effect (RCS), and global effect associated with urinary methanol under different covariate adjustments. The global effect assessed the collective significance of the RCS transformation of the methanol concentrations, while the nonlinearity test determines if the introduction of the non-linear terms in the RCS significantly improves the model fit in comparison with linear modelling.

Two different covariate adjustment models were conducted. The first analysis model involved:

$$\begin{aligned} &\text{Mortality rate of all causes/cancer/CVD} \\ &= \beta_0 + \beta_1(\text{urinary methanol concentration}) \\ &+ \beta_2(\text{rcs}(\text{PQN}, 4)) + \beta_3(\text{pH-value}) \\ &+ \beta_4(\text{rcs}(\text{eGFR}, 4)) + \beta_4(\text{rcs}(\text{age}, 4) \times \text{sex}) \\ &+ \beta_5(\text{waist circumference}) + \beta_6(\text{myocardial infarction}) \\ &+ \beta_7(\text{hypertension}) + \beta_8(\text{tg}) + \beta_9(\text{white blood cells}) \\ &+ \beta_{10}(\text{total-hdl-cholesterol ratio}) \\ &+ \beta_{11}(\text{presence of diabetes}) + e \end{aligned} \tag{5}$$

In the second Cox regression, additionally health-related behaviours were included as covariates, to evaluate the remaining predictive effect of methanol on mortality cases:

$$\begin{aligned} &\text{Mortality rate of all causes/cancer/CVD} \\ &= [\text{first analysis model}] + \beta_{11}(\text{physical inactivity}) \\ &+ \beta_{12}(\text{education years}) \\ &+ \beta_{13}(\text{alcohol intake during last 7 days} \times \text{sex}) \\ &+ \beta_{14}(\text{smoking status}) + \beta_{15}(\text{sleep problems}) + e \end{aligned} \tag{6}$$

The results were visualised by computing Kaplan–Meier curves for tertiles of the regression-normalised urinary metha-

nol concentration. Schoenfeld residuals were examined to evaluate the proportional hazard assumption.

Additionally, in a sensitivity analysis, a competing risk regression was used to evaluate the independent association of methanol concentrations with the mortality rates of CVD and cancer, using the analogous setup of the two different covariate adjustments models. Moreover, we performed prospective survival analysis to determine the predictive impact of the food categories on the mortality rates of all causes, cancer and CVD, both with and without accounting for methanol in the two different covariate adjustments models.

3. Results and discussion

3.1 Characteristics of study sample SHIP-START-0 and SHIP-TREND-0

The analysis sample of the discovery study SHIP-START-0 comprised n = 4017 participants (age-range: 20–81, 50.4% female) (Table 1 and Table S2, SI1) with FFQ data (33 items, Table S3) and 43 non-fasting urine NMR metabolites (Table S4). In the independent replication SHIP-TREND-0 cohort, n = 992 individuals (age-range: 20–81, 50.9% female), (Table 1 and Table S5, SI1) with fasting urine NMR metabolome data and FFQ data (16 items, Table S6) were included. As SHIP-TREND-0 NMR urine metabolome data was exclusively available in a subset of participants without self-reported diabetes, participants were substantially healthier in general than those in SHIP-START-0, reflected in a wide range of variables (Table 1).

3.2 Metabolic signatures of dietary intake

First, we conducted exploratory metabolome-wide association analysis using fully adjusted multiple linear regression analyses with HRSE for 43 quantified urinary metabolites and 33 food items in SHIP-START-0, exploring associations across all six FFQ categories, plus three continuous beverages (coffee, decaffeinated coffee, and tea in cups per day) analyzed separately.

Prior to corrections for multiple testing (nominal p-value <0.05), associations between 41 NMR urinary metabolites and 32 food items were identified (Table S7.1). After adjusting for multiple testing (false discovery rate (FDR) < 0.05), 42 associ-

Table 1 Descriptive table of SHIP-START-0 and SHIP-TREND-0

	SHIP-START-0 (n = 4017)	SHIP-TREND-0 (n = 992)	p-Value
Demographics and anthropometrics			
Age [years,(SD)[a], (range)]	50 (16.3), (20–81)	50 (13.7), (20–81)	0.679
Sex [no., (% female)]	2024 (50.39)	554 (55.85)	0.002
Waist circumference [cm, (SD)][a]	89.27 (13.87)	88.06 (12.87)	0.013
Lifestyle factors			
Physical inactivity [no., (% yes)]	2309 (57.70)	261 (26.31)	$<2.2 \times 10^{-16}$
Smoking status [no., (% yes)]	1209 (30.20)	217 (21.92)	1.64×10^{-7}
Alcohol intake [g d^{-1}, (IQR)]	4.97 (0, 17.4)[b]	3.99 (1.22, 10.46)[c]	0.239
Education [years, (range)]	11 (10–13)	13 (11–15)	$<2.2 \times 10^{-16}$
Time since the last meal [h, (IQR)]	3.53 (2.58, 4.58)	—	—
Metabolic and blood markers			
C-reactive protein (CRP) [mg L^{-1}, (IQR)]	1.38 (0.68, 3.15)	1.18 (0.62, 2.5)	0.000101
Glomerular filtration rate (eGFR) [mL per min per 1.73 m^2, (IQR)]	79.3 (69.8, 89.1)	89 (78.5, 102.5)	$<2.2 \times 10^{-16}$
Total hdl cholesterol ratio [(IQR)]	4.03 (3.19, 5.11)	3.75 (3.08, 4.57)	7.07×10^{-10}
LDL-C [mmol L^{-1}, (IQR)]	3.49 (2.75, 4.25)	3.36 (2.76,4)	0.001
HDL-C [mmol L^{-1}, (IQR)]	1.39 (1.14, 1.7)	1.43(1.21,1.7)	0.001
Triglycerides [mmol L^{-1}, (IQR)]	1.48 (1.01, 2.27)	1.22 (0.87,1.73)	$<2.2 \times 10^{-16}$
Red blood cells [%, (IQR)]	4.4 (4.12, 4.7)	4.6 (4.4, 4.9)	$<2.2 \times 10^{-16}$
White blood cells [%, (IQR)]	6.4 (5.4, 7.7)	5.46 (4.68, 6.46)	$<2.2 \times 10^{-16}$
HbA1c [%, (IQR)]	5.3 (4.9, 5.8)	5.2 (4.8, 5.5)	4.44×10^{-15}
Gamma-glutamyltransferase [µmol per sll, (IQR)]	0.34 (0.23, 0.56)	0.48 (0.38, 0.67)	$<2.2 \times 10^{-16}$
Health status			
Prevalent T2D [no., (%)]	441 (11.01)	29 (2.93)	$<2.2 \times 10^{-16}$
Hypertension [no., (%)]	1897 (47.35)	602 (60.81)	3.38×10^{-14}
Metabolic syndrome [no., (%)]	1107 (28.06)	212 (21.44)	2.02×10^{-5}
Myocardial infarction [no., (%)]	135 (3.37)	34 (3.43)	$<2.2 \times 10^{-16}$
Chronic kidney disease [no., (%)][d]	350 (8.75)	154 (15.54)	1.43×10^{-9}
Mortality			
All cause mortality [no., (%; YFU)]	1067 (26.56; 11.5)	—	—
CVD mortality [no., (%; YFU)]	329 (8.53; 10.3)	—	—
Cancer mortality [no.; (%; YFU)]	311 (8.06; 10.1)	—	—

Abbrevations: HDL-C, high density lipoprotein cholesterol; LDL-C, low density lipoprotein cholesterol; HbA1c, glycated haemoglobin; T2D, type 2 diabetes; CVD, cardiovascular disease; SD, standard deviation; IQR, interquartile range; YFU, median year of follow up. [a] Variables summarized with means ± SD. [b] Alcohol intake during the last 7 days. [c] Alcohol intake during the last 30 days. [d] Chronic kidney disease defined as the glomerular filtration rate >60 mL per min per 1.73 m^2. Median with the IQR/range for quantitative variables and the number (percentage) for categorical variables are presented if not stated otherwise. Characteristics were compared between the two groups using t-tests for normal distribution continuous variables, the Wilcox-test for non-normal distribution continuous variables, Fisher exact tests for binary variable and the Chi^2 test for categorical variables.

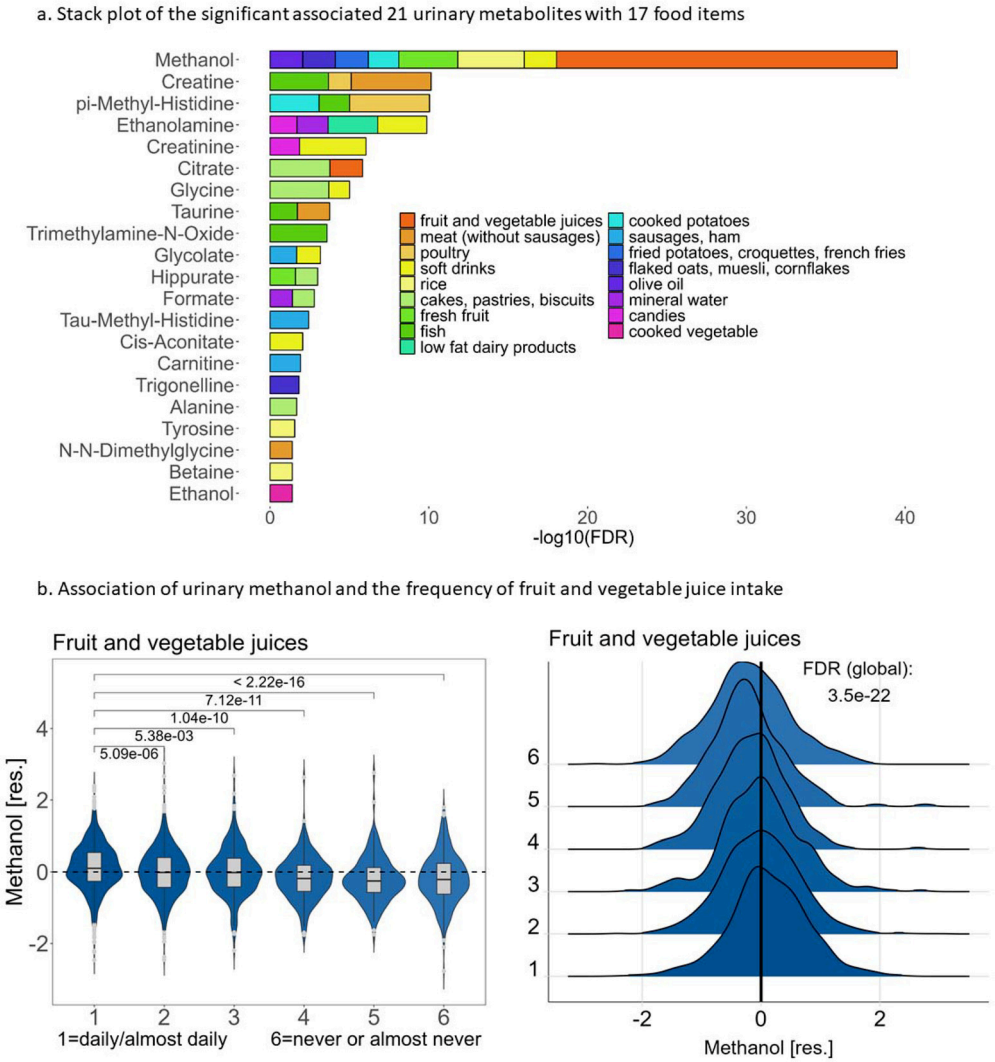

Fig. 1 (a) Stack plot of the significant associated 21 urinary metabolites with 17 food items. The *x*-axis shows the significant negative logarithmised FDR values (−log10(FDR)) of multiple linear regression analysis, whereas the *y*-axis depicts the urinary metabolites. (b) Association of urinary methanol and the frequency of fruit and vegetable juice intake. The figure depicts the top hit association of urinary methanol with the food item "fruit and vegetable juices" using the residuals of the methanol concentration after regressing out the same variables as in the main analysis: PQN (rcs: restricted cubic splines), age (rcs), age−sex interaction term (rcs), sex, waist circumference (rcs), glomerular filtration rate (eGFR) (rcs), pH-value, physical inactivity, smoking status, alcohol intake during the last 7 days, sex−alcohol intake interaction term, gamma-glutamyl transferase (GGT), hypertension, years of education, sleeping problems, triglycerides (tg), white blood cells (wbc), red blood cells (rbc), total-hdl-cholesterol ratio, time since the last meal, and prevalence of diabetes.

ations remained statistically significant among 21 urinary metabolites and 17 food items (Fig. 1a). Urinary methanol was the top hit, being associated particularly with plant-derived food items, such as fruit and vegetable juices (FDR = 3.44×10^{-22}, Fig. 1a and b), rice (FDR = 6.46×10^{-5}), and fresh fruits (FDR = 1.90×10^{-4}). In the continuous-variable analysis, methanol and coffee intake showed a significant inverse association (FDR = 5.58×10^{-22}; b = −0.06, 95%-CI: (−0.07, −0.05), Table S7.2), possibly confounded by higher coffee consumption among smokers (SI4).

Other significant associations were observed for plant-derived food-items including citrate with fruit and vegetable juices (FDR = 8.79×10^{-3}), hippurate with fresh fruits (FDR = 2.45×10^{-2}), or betaine with rice (FDR = 3.98×10^{-2}). Significant links were also found for animal-derived products, such as creatine with meat (without sausages) (FDR = 9.16×10^{-6}), trimethylamine-*N*-oxide with fish (FDR = 2.62×10^{-4}) and trigonelline associated with coffee intake in the separate continuous-variable analysis (FDR = 3.97×10^{-68}) (Table S7, Fig. 1a), aligning with findings from previous studies.[25–29] For urinary methanol, being associated with 10 food items, we found no similar findings reported in the existing literature. Next, we conducted analogous regressions to identify the directions of metabolite associations with frequent ("every day or almost every day" joined with "several times a week") *versus* rare consumption ("about once a month or fewer" joined with

"never or almost never") of each food item (full results Table S7). Consistent with the first set of regressions, urinary methanol was positively associated with fruit and vegetable juices (b = 0.29, 95%-CI: (0.23, 0.35), FDR = 7.93×10^{-17}), rice (b = 0.16, 95%-CI: (0.08, 0.25), FDR = 9.83×10^{-3}), flaked oats, muesli, and cornflakes (b = 0.13, 95%-CI: (0.06, 0.20), FDR = 1.04×10^{-2}), and cooked vegetables (b = 0.28, 95%-CI: (0.11, 0.44), FDR = 3.18×10^{-2}) (Table 2a, S7 and S8, Fig. 2a). Conversely, food items such as fried potatoes, croquettes, French fries (b = −0.19, 95%-CI: (−0.28, −0.10), FDR = 6.6×10^{-3}) and soft drinks (b = −0.12, 95%-CI: (−0.18, −0.06), FDR = 4.86×10^{-3}) were found to have inverse associations with urinary methanol levels (Table 2a, S7, S8, and Fig. 2a). This pattern was also visible in nominally significant (p-value < 0.05, FDR > 0.05) methanol associations (*e.g.* positive: salad and fresh fruits; negative: cake and pizza, Fig. 2b), further strengthening the conclusion that urinary methanol is linked to a diet, rich in plant-derived food and potentially a high fibre.

In the SHIP-TREND-0 replication cohort, we focused our analysis on urinary methanol concentrations to validate the initial findings from SHIP-START-0. Significant results were again detected between urinary methanol and plant-derived food items (salad or raw vegetables, FDR = 1.30×10^{-2}; fresh fruits, FDR = 4.90×10^{-2}), as well as food items, including cake, biscuits, and cookies (FDR = 3.70×10^{-2}) (Table 2b and S9). In conclusion, the analysis in both SHIP-cohorts indicates that urinary methanol is positively associated with the consumption of plant-derived food items potentially indicative of a fibre-rich diet.

3.3 Predicting urinary methanol *via* FFQ using machine learning

To determine the extent to which urinary methanol concentrations could be predicted from the FFQ data, we adopted a machine learning approach using elastic net regressions with 10-fold cross-validation (Table S10) after imputation of missing FFQ data *via* k-nearest neighbours. The model predicted normalised urinary methanol concentrations based on FFQ data (Table S11), with the two penalisation parameters alpha = 0.899 and lambda = 0.005 and an out-of-sample r-squared value of 0.10 (SI3), meaning that FFQ data could explain 10% of the variance in urinary methanol levels.

Consistent with the previous results, the largest positive coefficient was attributed to the frequent consumption of fruit and vegetable juices ("daily or almost daily", b = 0.31). Conversely, the most negative coefficient was related to the rare intake of flaked oats, muesli and cornflakes ("never or almost never", b = −0.14). Potential reasons for the low amount of explained variance might include the limited accuracy of the FFQ and other non-diet-related influence factors on urinary methanol levels.

3.4 Urinary methanol and lifestyle factors

Consequently, we explored the associations of urinary methanol with basic, behavioural, physiological, and clinical covariates in SHIP-START-0 and SHIP-TREND-0. While we could not identify associations with clinical covariates, we found strong associations between urinary methanol and lifestyle indicators. Urinary methanol was positively associated with education years, whereas it was negatively associated with physical inactivity, smoking and alcohol intake (Table 3 and Table S12). We validated the methanol associations with education, smoking, and alcohol intake in the replication cohort (Table 3 and Table S13, SI4). However, the explained variance for all these factors together was low (Table S12 and S13). In conclusion, urinary methanol showed associations with health-related behaviours in the SHIP cohorts, under both fasting and non-fasting conditions, but the main factors causing inter-individual variation in methanol levels were not identified.

3.5 Constraint-based modelling

Since methanol is a metabolic by-product of plants,[30] stemming mainly from water-soluble dietary fibre pectin metabolisation, we analysed in a further step, the potential contribution of the microbiome to human methanol pools through fibre degradation. Investigating the reference set of 7302 microbial metabolic reconstructions of AGORA2,[8] we discovered 92 strains belonging to 15 genera (mainly *Bacteroides (39 strains)* and *Bacillus (24 strains)*) that are theoretically able to produce methanol through fibre breakdown specifically *via* the degradation of pectin's. To explore the contribution of these species to microbiome methanol production in actual measured microbiomes, we applied COBRA community modelling to 149 individuals with metagenomics data (Table S14) from the Human Microbiome Project. For each individual, a personalised microbiome community model was built, and the maximum secretion potential of methanol was quantified through flux variability simulations under different diet constraints (Table S15).

Gradually incrementing diet constraints (*i.e.*, the maximal uptake rate) of pectin revealed a continuous rise in methanol production with increasing pectin intake (Fig. 3a). This was not observed for incremental increases of xylan diet constraints, which served as a control fibre contributing the same number of additional carbon atoms to the community (Fig. 3a). Since a wide range of microbes may also produce methanol as a by-product of biotin synthesis,[8] this result indicates that the rise in methanol is not driven by a general increase in the availability in carbon sources. Instead, the results point into the direction that microbiome methanol production can be specifically attributed to pectin availability. Thus, the simulations supported the hypothesis that pectin is a primary source of microbiome methanol production. Indeed, a previous study demonstrated that fecal bacteria are capable of releasing methanol through the degradation of pectin.[31] The microbiome's capacity to produce methanol from pectin, also explains the observed association pattern of food items with urinary methanol in the SHIP-cohort. Interestingly, maximal methanol secretion potentials positively correlated with alpha-diversity, as quantified by Shannon entropy (correlation r = 0.34, p-value = 2.78×10^{-5}, Fig. 3d), suggesting that methanol may serve as a potential biomarker indicative of a

Table 2 Significant urinary methanol associations across food categories

Urinary methanol					
Food frequency category	b^a (95% CI)	p-Value^b (all categories)	p-Value^c (frequent *vs.* rare)	FDR^b (all categories)	FDR^c (frequent *vs.* rare)
a. SHIP-START-0: significant associations of urinary methanol across the 33 food categories					
Fruit and vegetable juices	0.29 (0.23, 0.35)	2.42×10^{-25}	5.59×10^{-20}	3.44×10^{-22}	7.93×10^{-17}
Rice	0.16 (0.08, 0.25)	2.28×10^{-7}	1.39×10^{-4}	6.46×10^{-5}	9.83×10^{-3}
Fresh fruits	0.24 (0.06, 0.42)	9.38×10^{-7}	1.82×10^{-2}	1.90×10^{-4}	2.75×10^{-1}
Fried potatoes, croquettes, and French fries	−0.19 (−0.28, −0.1)	8.87×10^{-5}	8.40×10^{-5}	8.39×10^{-3}	6.60×10^{-3}
Flaked oats, muesli, and cornflakes	0.13 (0.06, 0.2)	1.18×10^{-4}	1.54×10^{-4}	8.79×10^{-3}	1.04×10^{-2}
Olive oil	0.11 (0.05, 0.17)	1.10×10^{-4}	8.84×10^{-5}	8.79×10^{-3}	6.60×10^{-3}
Soft drinks	−0.12 (−0.18, −0.06)	1.38×10^{-4}	5.48×10^{-5}	9.35×10^{-3}	4.86×10^{-3}
Cooked potatoes	−0.18 (−0.43, 0.06)	2.01×10^{-4}	9.68×10^{-2}	1.19×10^{-2}	5.49×10^{-1}
White grain bread, black bread, and crispbread	0.05 (−0.02, 0.12)	1.54×10^{-3}	2.04×10^{-1}	5.08×10^{-2}	6.88×10^{-1}
Low fat dairy products	0.09 (0.04, 0.14)	2.180×10^{-3}	8.21×10^{-4}	6.18×10^{-2}	3.64×10^{-2}
Cakes, pastries, and biscuits	−0.08 (−0.14, −0.01)	5.43×10^{-3}	2.98×10^{-2}	1.13×10^{-1}	3.55×10^{-1}
Fish	0.06 (−0.02, 0.14)	1.20×10^{-2}	1.32×10^{-1}	1.92×10^{-1}	6.98×10^{-1}
Salad or raw vegetables	0.11 (0.04, 0.19)	1.64×10^{-2}	5.68×10^{-3}	2.30×10^{-1}	1.37×10^{-1}
Fried sausage, hamburger, doner kebab, and pizza	−0.18 (−0.3, −0.06)	2.46×10^{-2}	2.35×10^{-3}	2.86×10^{-1}	8.02×10^{-2}
Cooked vegetables	0.28 (0.11, 0.44)	2.71×10^{-2}	6.50×10^{-4}	2.94×10^{-1}	3.18×10^{-2}
Candies	0.06 (0, 0.11)	3.91×10^{-2}	5.64×10^{-2}	3.65×10^{-1}	4.44×10^{-1}
Cheese	0.17 (0.07, 0.28)	6.29×10^{-2}	4.68×10^{-3}	4.42×10^{-1}	1.19×10^{-1}
Butter	−0.07 (−0.12, −0.02)	7.39×10^{-2}	5.43×10^{-3}	4.66×10^{-1}	1.33×10^{-1}

Urinary methanol					
Food items	b^a (95% CI)	p-Value^b (all categories)	p-Value^c (frequent *vs.* rare)	FDR^b (all categories)	FDR^c (frequent *vs.* rare)
b. SHIP-TREND-0: significant association of urinary methanol across the 16 food categories					
Salad or raw vegetables	0.14 (−0.01, 0.3)	7.70×10^{-4}	8.80×10^{-2}	1.20×10^{-2}	4.70×10^{-1}
Cake, biscuits, and cookies	−0.02 (−0.16, 0.12)	4.60×10^{-3}	7.90×10^{-1}	3.70×10^{-2}	8.40×10^{-1}
Fresh fruits	0 (−0.31, 0.3)	9.10×10^{-3}	9.90×10^{-1}	4.80×10^{-2}	9.90×10^{-1}
Sausages and ham	−0.1 (−0.29, 0.09)	3.20×10^{-2}	3.30×10^{-1}	1.30×10^{-1}	6.70×10^{-1}

a. SHIP-START-0: results of multiple linear regressions adjusted for PQN (rcs: restricted cubic splines), age (rcs), age–sex interaction term (rcs), sex, waist circumference (rcs), glomerular filtration rate (eGFR) (rcs), pH-value, physical inactivity, smoking status, alcohol intake during the last 7 days, sex–alcohol intake interaction term, gamma-glutamyl transferase (GGT), hypertension, years of education, sleeping problems, triglycerides (tgs), white blood cells (wbcs), red blood cells (rbcs), total-hdl-cholesterol ratio, time since the last meal, and prevalence of diabetes. b. SHIP-TREND-0: results of multiple linear regressions adjusted for PQN, (rcs), age (rcs), age–sex interaction term (rcs), sex, waist circumference (rcs), eGFR (rcs), pH-value, physical inactivity, smoking status, alcohol intake during the last 30 days, sex–alcohol intake interaction term, GGT, hypertension, years of education, sleeping problems, tgs, wbcs, rbcs, total-hdl-cholesterol ratio, and prevalence of diabetes. [a] b-Values per SD. [b] Overall significant value of the association between urinary metabolites and the categorical variable food items, with all levels considered collectively within the Wald test. [c] Significant values of the associations between methanol and frequent ("every day or almost every day" and "several times a week") *vs.* rare consumption ("about once a month or fewer" and "never or almost never") of food items.

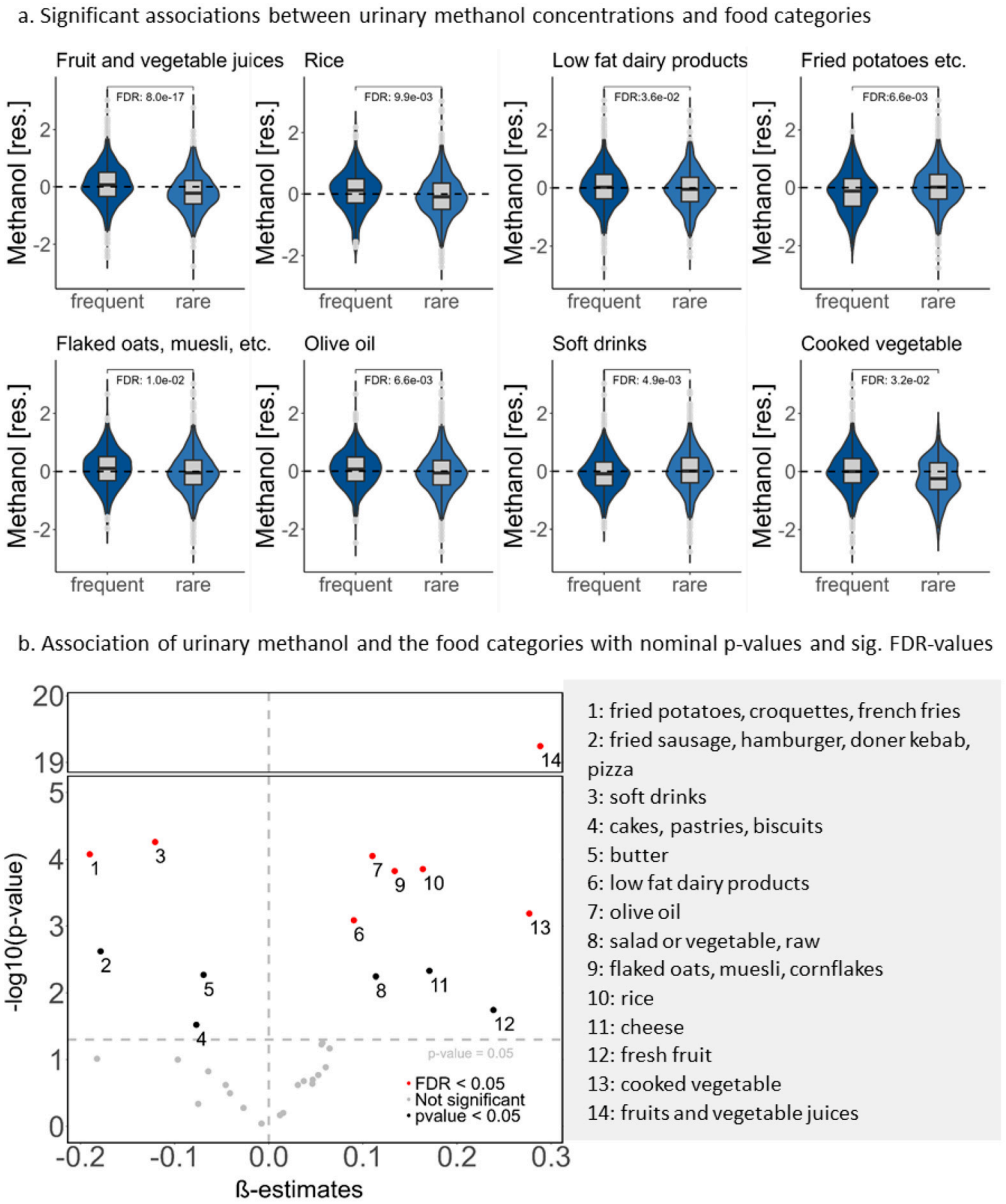

Fig. 2 Urinary methanol concentration linked to the comparison of frequent and rare food item categories. (a) Violin plots depicting the significant correlation between urinary methanol concentrations and food items using residuals of methanol concentration, after regressing out the same variables as in the main analysis: PQN (rcs: restricted cubic splines), age (rcs), age–sex interaction term (rcs), sex, waist circumference (rcs), glomerular filtration rate (eGFR) (rcs), pH-value, physical inactivity, smoking status, alcohol intake during the last 7 days, sex–alcohol intake interaction term, gamma-glutamyl transferase (GGT), hypertension, years of education, sleeping problems, triglycerides (tg), white blood cells (wbc), red blood cells (rbc), total-hdl-cholesterol ratio, time since the last meal, and prevalence of diabetes. (b) Vulcano plot illustrating the association of urinary methanol and the categories of food frequency intake with nominal *p*-values and significant FDR values.

healthy microbiome, which is characterised by high ecological diversity.[32] Furthermore, by testing the association between methanol and hippuric acid, a urinary marker of microbiome diversity, a small but significant inverse association was found (p-value = 0.049, b = −0.03, and 95% CI: (−0.07, −0.00)), with a negligible incremental R^2 of 0.02%, suggesting methanol and hippurate may independently reflect microbiome diversity.

To shed light on the individual microbes responsible for methanol production, we computed the maximum secretion fluxes of each strain present (Table S15). Additionally, using a variation of the analysis routes developed by Hertel *et al.*[9] we calculated the ecological, direct and total contributions of each microbe to the overall community methanol production (Fig. 3c). At the broader genus level, we found that *Bacteroides* (68.9%) and *Faecalibacterium* (20.6%) were together responsible for nearly 90% of the total methanol community production (Fig. 3b). At the species level, *Faecalibacterium prausnitzii* (20.6%) was computed to produce the highest secretion contribution, followed by the species *Bacteroides ovatus* (19.1%) and *Bacteroides stercoris* (18.6%) (Table S15). These

Table 3 Methanol concentration and correlations with lifestyle factors

	SHIP-START-0 (n = 4017)			SHIP-TREND-0 (n = 992)		
	b^g (95% CI)	p-Value	FDR	b^g (95% CI)	p-Value	FDR
Basic covariates[a]						
Sex[a]	—	5.78×10^{-4}	2.43×10^{-3}	—	0.325	0.568
Age[a]	—	8.03×10^{-4}	2.81×10^{-3}	—	0.310	0.568
Waist circumference[b]	—	1.62×10^{-2}	4.25×10^{-2}	—	0.020	0.084
eGFR[c]	—	1.39×10^{-2}	4.17×10^{-2}	—	0.097	0.292
Behaviour covariates[d]						
Smoking	−0.27 (−0.32, −0.22)	8.13×10^{-25}	1.71×10^{-23}	−0.17 (−0.28, −0.07)	0.002	0.016
Alcohol intake	−0.01(−0.02, −0.01)	4.27×10^{-8}	2.99×10^{-7}	−0.02 (−0.03, −0.01)	0.000	0.004
Physical inactivity	−0.095 (−0.05, −0.14)	7.67×10^{-5}	4.03×10^{-4}	−0.08 (−0.17, 0.02)	0.126	0.331
Years of education	0.03 (0.02, 0.04)	2.57×10^{-10}	2.70×10^{-9}	0.03 (0.01, 0.04)	0.002	0.016
Physiological parameters[e]						
tg	0.02 (−0.02, 0.07)	0.257	0.360	−0.04 (−0.13, 0.05)	0.401	0.648
log(crp)	−0.00 (−0.02, 0.02)	0.882	0.882	0.01 (−0.03, 0.06)	0.637	0.787
log(ggt)	−0.01 (−0.04, 0.03)	0.816	0.856	−0.03 (−0.12, 0.06)	0.493	0.701
wbcs	−0.01 (−0.02, 0.00)	0.098	0.171	−0.00 (−0.03, 0.03)	0.951	0.951
rbcs	0.072 (0.01, 0.14)	0.030	0.062	0.09 (−0.03, 0.22)	0.150	0.351
hdl	0.01 (−0.05, 0.07)	0.804	0.856	0.01 (−0.15, 0.13)	0.869	0.912
ldl	−0.01 (−0.04, 0.01)	0.237	0.356	−0.014 (−0.06, 0.04)	0.590	0.774
HbA1c	−0.04 (−0.07, −0.00)	0.025	0.058	−0.11 (−0.19, −0.02)	0.017	0.084
Clinical phenotypes[e]						
Diabetes	0.02 (−0.07, 0.11)	0.645	0.797	−0.03 (−0.24, 0.19)	0.817	0.912
Hypertension	0.05 (−0.01, 0.10)	0.079	0.150	0.01 (−0.09, 0.11)	0.829	0.912
MetS	0.04 (−0.02, 0.09)	0.222	0.356	0.08 (−0.05, 0.20)	0.220	0.461
MI	0.01 (−0.05, 0.07)	0.731	0.853	0.04 (−0.08, 0.16)	0.501	0.701
CKD[f]	0.035 (−0.05, 0.12)	0.414	0.543	0.18 (−0.03, 0.39)	0.090	0.292

Abbreviations: eGFR, estimated glomerular filtration rate; tg, triglyceride; crp; c-reactive protein; ggt, gamma-glutamyl transferase; wbcs, white blood cells; rbcs, red blood cells; hdl, high density lipoprotein; ldl, low density lipoprotein; HbA1c, glycated haemoglobin; HDL-C, high density lipoprotein cholesterol; LDL-C, low density lipoprotein cholesterol; MetS, metabolic syndrome; MI, myocardial infarction; CKD, chronic kidney disease; SD, standard deviation; and FDR, false discovery rate. [a] Adjusted for PQN (rcs), pH, and sex–age interaction (rcs). [b] Adjusted for waist circumference (rcs), PQN (rcs), pH, and sex–age interaction (rcs). [c] Adjusted for eGFR (rcs), PQN (rcs), pH, and sex–age interaction (rcs). [d] Adjusted for smoking, alcohol–sex interaction, physical inactivity, education, waist circumference (rcs), eGFR (rcs), sex–age interaction (rcs), PQN (rcs), and pH-value. [e] Variables considered individually with adjustment for smoking, alcohol–sex interaction, physical inactivity, education, waist circumference (rcs), eGFR (rcs), sex–age interaction (rcs), PQN (rcs), and pH-value. [f] Adjustment for CKD, smoking, alcohol–sex interaction, physical inactivity, education, waist circumference (rcs), eGFR (rcs), sex–age interaction (rcs), PQN (rcs), and pH-value. [g] b-Values = per SD.

results align with an earlier study showing that *Faecalibacterium prausnitzii* is capable of degrading various types of pectins.[33]

Further analyses of ecological effects[9] revealed that *Parabacteroides* sp. *D13* exhibited the highest negative ecological effect on community methanol production. Thus, while *Parabacteroides* sp. *D13* itself can produce certain amounts of methanol, community methanol production was indicated to be lower in communities with higher *Parabacteroides* sp. *D13* abundance, which could be explained by competitive effects on other methanol-producing species. In summary, the *in silico* experiments indicate that the microbiome produces methanol from pectin. Communities with greater diversity, which is often viewed as an unspecific protective factor in human health and disease,[32] showed higher methanol secretion potentials with the genera *Bacteroides* and *Faecalibacterium* as primary methanol producers.

3.6 Urinary methanol and mortality rates

Given our previous findings, we further investigated urinary methanol as a predictive biomarker in a prospective survival analysis using the follow-up data of SHIP-START-0. The cohort exhibits an overall mortality rate of 26.6% (1067 deaths), with 30.8% (329 deaths) of the deaths attributed to CVD, and 29.1% (311 deaths) to cancer (Table 1 and Table S16) during a median follow-up of 11.5 person-years (maximum of 21.5 years of follow-up).

Regardless of covariate adjustments, we found strong associations between urinary methanol and cancer, CVD and all-cause mortality in Cox regressions (Fig. 4a). Kaplan–Meier curves for tertiles of regression-normalised urinary methanol concentration (Fig. 4b) revealed higher mortality rates, particularly in the lowest tertile, thereby visually explaining the detection of nonlinearity (Fig. 4b). To address potential overestimation of hazard ratios in cause-specific Cox regressions, we performed competing risk models, in which neither cancer mortality nor cardio vascular disease mortality remained significant (SI5), indicating that the cause-specific findings may need further corroboration. Note that this caveat does not apply to the negative association between urinary methanol levels and all-cause mortality. Noteworthy, methanol stayed predictive for all-cause mortality in models adjusting for methanol-associated food items, providing evidence for a value of urinary methanol as a biomarker beyond FFQ data (Table S17).

The negative association between urinary methanol and mortality deserves explanation since methanol is known to be

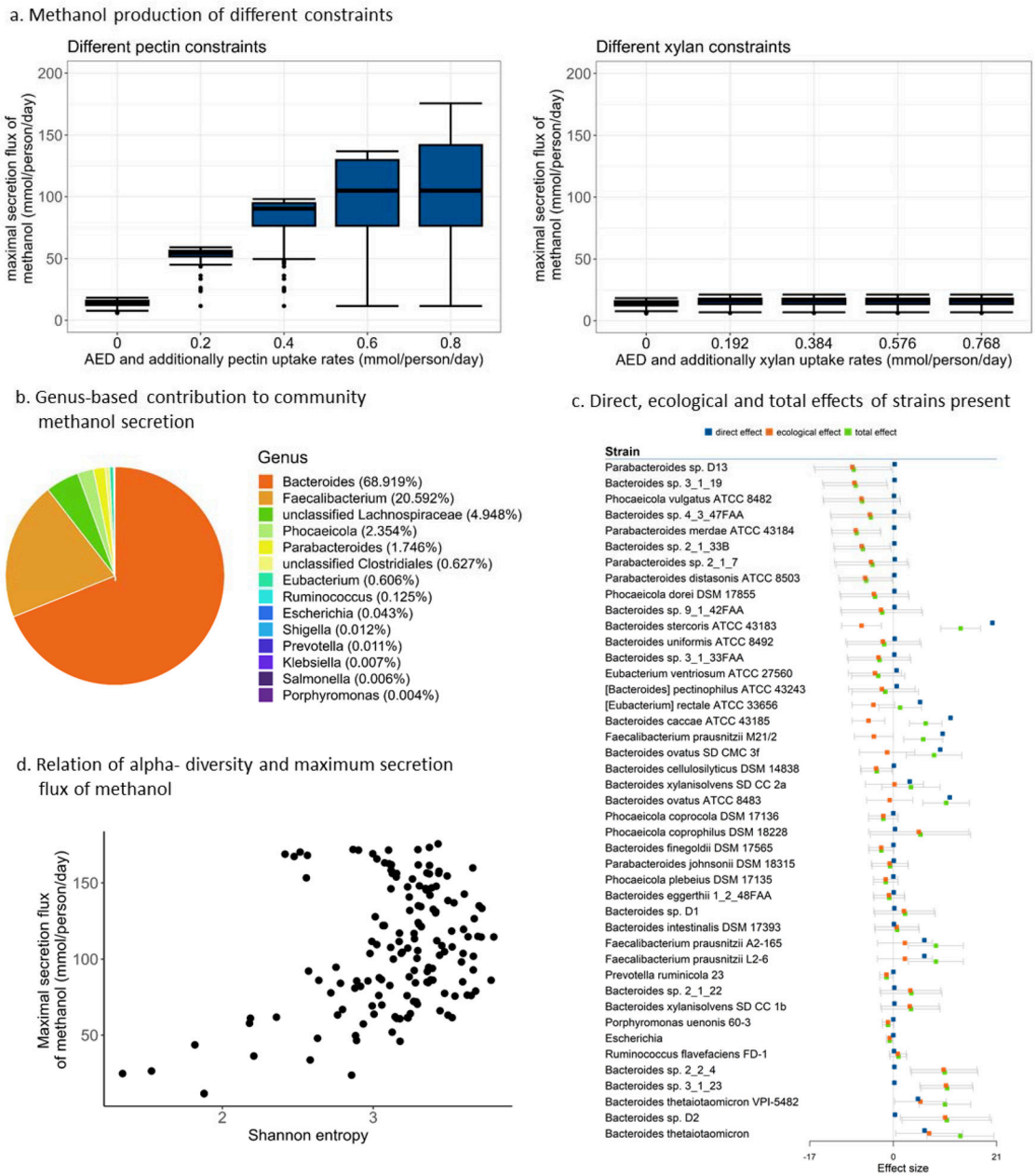

Fig. 3 (a) Maximum secretion fluxes of methanol under an Average European Diet (AED) and under the AED with gradually increasing diet constraints of pectin and xylan. (b) Average maximum individual secretion fluxes of microorganisms summarized at the genus level. (c) Direct, ecological and total effects of the different strains present in the community models calculated based on the individual secretion fluxes. (d) Scatterplot of alpha-diversity measured in Shannon entropy and the maximum secretion flux of methanol (correlation $r = 0.34$, p-value $= 2.78 \times 10^{-5}$).

toxic at high concentrations. Blood methanol levels above 200 mg L^{-1} are associated with adverse effects on the central nervous system, with severe acute toxicity above 500 mg L^{-1} and fatality at levels exceeding 1500 mg L^{-1},[34] known from cases of contamination in alcoholic beverages[35] or occupational exposure.[36] The toxicity of methanol arises from two primary mechanisms. The first one is related to the direct depression of the central nervous system, similarly to ethanol poisoning.[37] The second one involves the conversion of methanol to toxic formaldehyde *via* alcohol dehydrogenase, resulting in cellular hypoxia and several other metabolic disturbances.[37,38] The metabolisation of methanol predominantly occurs in the liver (70–97%), while minor amounts are excreted non-metabolically through urine and lungs.[38]

Nevertheless, methanol is physiologically present in small amounts in humans.[39] Accordingly, the urinary concentration observed in this study can be rated as being in the physiological range. Interestingly, the toxic methanol catabolite formaldehyde has been shown to have positive effects at low doses. In plants, formaldehyde can influence growth and photosyn-

a. Methanol concentration in SHIP-START-0 with cox regression

	*Adjusted Model I**				*Adjusted Model II**			
	HR per SD (95% CI)	p-value linear-effect	p-value non-linearity	p-value global	HR per SD (95% CI)	p-value linear effect	p-value non-linearity	p-value global
Mortality CVD	0.83 (0.71, 0.98)	0.023	0.003	0.006	0.89 (0.76, 1.05)	0.177	0.031	0.07
Mortality Cancer	0.83 (0.71, 0.97)	0.020	0.003	0.009	0.88 (0.75, 1.04)	0.138	0.047	0.104
Mortality all	0.87 (0.80, 0.95)	0.002	3.13E-07	1.29E-06	0.93(0.85, 1.01)	0.087	2.85E-04	0.001

*I: Adj. for PQN, pH value, eGFR, age, sex, waist circumference, myocardial infarction, hypertension, diabetes, triglycerides, total-HDL-cholesterol ratio, and white blood cells, red blood cells
**II Adj. for PQN, pH value, eGFR, age, sex, waist circumference, myocardial infarction, hypertension, diabetes, triglycerides, total-HDL-cholesterol ratio, white
blood cells, red blood cells, physical activity, education years, alcohol intake during last 7 days and sleep problem

Abbreviation: CVD, cardio vascular diseases; SD, standard deviation

b. Methanol concentration in SHIP-START-0 with cox regression

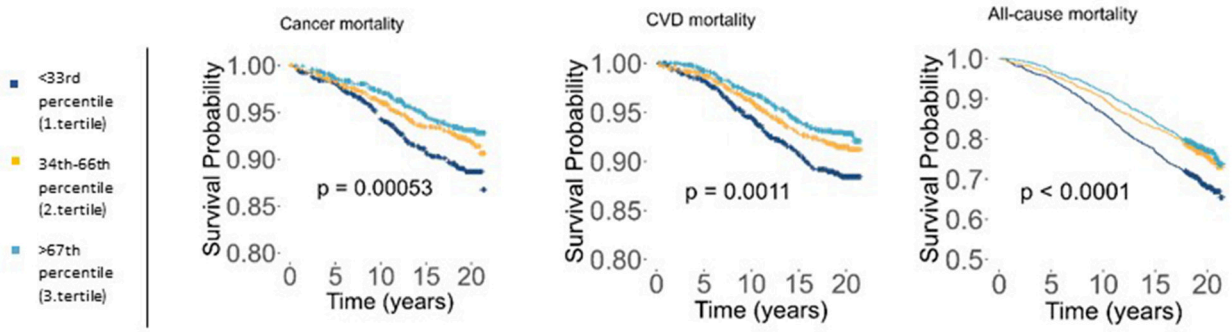

Fig. 4 Association of methanol concentration and mortality rates. Abbreviation: CVD, cardiovascular diseases; SD, standard deviation (a). Methanol concentration in association with mortality rates in SHIP-START-0. (b). Kaplan–Meier survival probability curves as well as the significant values of a log-rank test for the mortality of all causes, cardiovascular diseases and cancer using residuals of methanol concentration, adjusted for the probabilistic quotient normalisation (PQN).

thetic pigments, while in animal cells, it can positively impact cell proliferation and viability.[40,41] Given the hermetic response displayed by the methanol catabolite formaldehyde[40,41] and this study showing a positive link between methanol and longevity, it can be speculated that low doses of methanol are act in a hormetic manner.

Besides the bound form of methanol in pectin, free methanol can be found in plant-based foods (*e.g.* 11–68 mg L^{-1} in fresh squeezed fruit juices),[42] or as a catabolite of aspartame, a synthetic non-nutritive sweetener.[38] It is also present at low levels in most alcoholic beverages, without conferring health risks.[43] Importantly, we found a negative association between urinary methanol concentrations and alcohol intake in the SHIP cohorts, indicating that low concentrations in non-contaminated alcoholic beverages are not a major source of normal human methanol pools. Our study provides support for the contribution of free methanol from dietary sources besides bound methanol from pectin degradation, to human methanol levels, yet their precise contributions remain unclear.[38,39] Pectin, putatively the primary dietary source of methanol, has been shown to enhance the diversity and abundance of beneficial microbial communities.[44] Its microbial degradation improves gastrointestinal immune barrier function through the production of short-chain fatty acids and promotes the adhesion of commensal bacteria while inhibiting the adhesion of pathogens to epithelial cells.[44] This raises the hypothesis that microbial methanol may help modulate gut inflammation, though this remains to be shown. Beyond pectin, methanol as a biomarker for beneficial microbiome–host–diet interaction may also be a marker for a variety of bioactive compounds in foods, such as melanoidins, polyunsaturated fats, inulin, and oligosaccharides, highlighting the

need for further study. For instance, fiber intake enhances SCFA production like butyrate, which may promote methanol generation *via* pathways not yet included in AGORA2.

3.7 Strengths and limitations

Regardless of the robust findings across two independent cohorts and multiple lines of analysis, several limitations have to be discussed. Utilising NMR allowed us to identify and quantify methanol, a metabolite previously missed in food-metabolome analyses. However, the relatively high detection limit and the targeted nature of the applied NMR methodology poses a limitation, restricting the scope of metabolome analysis and thus missing important microbial degradation products such as butyrate. Second, despite the findings related to a potentially high fibre diet–microbiome–host interaction, the FFQ utilised in the SHIP cohort is not specifically tailored to measure fibre intake in a quantitative way. Moreover it does not differentiate between sources from processed and unprocessed foods, such as fruit juices or soft drinks, which may contain fibre-like compounds or additives with differing microbial fermenting and bioactivity, potentially influencing microbial fibre degradation.[45,46] Future studies should explore the relationship between urinary methanol and dietary fibre in greater detail. Additionally, the measurement of dietary habits through the FFQ could be biased by participants' self-reported food-intake frequencies, adding unwanted variability and masking true food–metabolite associations. Given the ordinal nature of the FFQ, quantitative relations between food frequencies and metabolite could not be established, and as such further research is necessary to identify the quantitative relation between fibre intake and urinary methanol. Third, the validity of our findings may be influenced by unmeasured confounding due to the observational design, and the hypothesis of methanol being a marker for pectin intake requires further experimental tests. Fourth, normal fluctuations over time in microbial composition alongside microbial and host metabolic activity causing substantial inter-individual and intra-individual variability, may modulate methanol excretion and thereby affect its predictive value.[47] Finally, the results may not generalise to populations with different diet patterns due to regionally restricted nature of SHIP data, which exclusively originates from north-eastern Germany and shows very low ethnic diversity.[12] However, the integration of mechanistic microbiome community modelling provided additional evidence for a microbiome-mediated contribution to human methanol pools.

4. Conclusion

In conclusion, we have provided consistent evidence that urinary methanol is a valuable biomarker for protective diet–microbiome interactions linked to microbial pectin degradation, as demonstrated in two independent population studies, microbiome modelling of an independent sample with publicly available metagenomics data, and prospective survival analysis.

Author contributions

Conceptualization: K. K. and J. H, data curation: K. K., A. M., D. F., N. F., K. B., and A. H.; formal analysis: K. K., A. M., D. F., A. H., and J. H; funding acquisition: H. V. and J. H.; investigation: K. K., J. H., M. F., A. M., D. F., M. D. and M. B.; methodology: K. B., N. F., A. K. H., F. F., A. H., J. H., and M. N.; project administration: H. V., H. J. G., M. N. and J. H.; resources: K. B., N. F., F. F., H. V., A. K. H. and M. N.; supervision: J. H. and N. F.; validation: K. K., A. M., D. F., M. F., A. W. and J. H.; visualization: K. K., A. M., and D. F.; writing – original draft: K. K.; and writing – review & editing: all authors.

Conflicts of interest

HJG has received travel grants and speakers' honoraria from Indorsia, Neuraxpharm, Servier and Janssen Cilag. The other authors declare that they have no conflicts of interest.

Ethics declarations

The SHIP study was approved by the Ethics Committee at the University Medicine Greifswald, Germany (approval number BB39/08) and conducted in accordance with the principles of the Declaration of Helsinki.

All study participants of the SHIP cohorts provided written informed consent.

Abbreviation

DHBA	Dihydroxybenzoic acid
COBRA	Constraint-based reconstruction and analysis
SHIP	Study of Health in Pomerania
CVD	Cardiovascular diseases
NMR	Nuclear magnetic resonance
HbA1c	Glycated haemoglobin
tg	Triglycerides
FFQ	Food frequency questionnaire
PQN	Probabilistic quotient normalisation
RCS	Restricted cubic splines
HRSE	Heteroscedastic robust standard errors
eGFR	Estimated glomerular filtration rate

Data availability

Due to data protection agreements, SHIP datasets are not publicly available but can be obtained upon request through the SHIP transfer office (**https://transfer.ship-med.uni-greifswald.de/FAIRequest/data-use-intro**). Relative strain abundances

from 149 healthy adult gut microbiome samples from the Human Microbiome Project had been mapped to the nomenclature of AGORA[48] previously.[21] The mapped relative abundances are available at **https://GitHub.com/SysPsyHertel/CodeBase**. All scripts utilised to generate results for this manuscript can be found at **https://GitHub.com/SysPsyHertel/CodeBase**.

Supplementary information containing detailed tables supporting the analyses presented in this manuscript is available at DOI: **https://doi.org/10.1039/d5fo00761e**.

Acknowledgements

SHIP is integrated in the research network Community Medicine of the University Medicine Greifswald, supported by the German Federal State of Mecklenburg-West Pomerania. J. H. received funding from the Deutsche Forschungsgesellschaft (DFG) (grant no. HE9198/1-1). This study was also funded by the DFG–SFB 1597-499552394. The contribution to data collection made by fieldworkers, study physicians, ultrasound technicians, interviewers and laboratory workers is gratefully acknowledged. We are also appreciative of the important support of computer scientists, medical documentarists and administration staff. We also thank former and current members of our Data Safety and Monitoring Committee for valuable comments and advice. Last but not least, we thank all study participants whose personal dedication and commitment have made this project possible.

References

1 W. J. Dahl and M. L. Stewart, Position of the Academy of Nutrition and Dietetics: Health Implications of Dietary Fiber, *J. Acad. Nutr. Diet.*, 2015, **115**(11), 1861–1870.
2 P. Cronin, S. A. Joyce, P. W. O'Toole and E. M. O'Connor, Dietary Fibre Modulates the Gut Microbiota, *Nutrients*, 2021, **13**(5), 1655.
3 V. P. N. Prasadi and I. J. Joye, Dietary Fibre from Whole Grains and Their Benefits on Metabolic Health, *Nutrients*, 2020, **12**(10), 3045.
4 D. Parada Venegas, M. K. De la Fuente, G. Landskron, M. J. González, R. Quera, G. Dijkstra, *et al.*, Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases, *Front. Immunol.*, 2019, **10**, 277.
5 A. A. Zabolotneva, O. P. Shatova, A. A. Sadova, A. V. Shestopalov and S. A. Roumiantsev, An Overview of Alkylresorcinols Biological Properties and Effects, *J. Nutr. Metab.*, 2022, **2022**, 4667607.
6 N. Veronese, M. Solmi, M. G. Caruso, G. Giannelli, A. R. Osella, E. Evangelou, *et al.*, Dietary fiber and health outcomes: an umbrella review of systematic reviews and meta-analyses, *Am. J. Clin. Nutr.*, 2018, **107**(3), 436–444.
7 A. Unión-Caballero, T. Meroño, R. Zamora-Ros, A. L. Rostgaard-Hansen, A. Miñarro, A. Sánchez-Pla, *et al.*, Metabolome biomarkers linking dietary fibre intake with cardiometabolic effects: results from the Danish Diet, Cancer and Health-Next Generations MAX study, *Food Funct.*, 2024, **15**(3), 1643–1654.
8 A. Heinken, J. Hertel, G. Acharya, D. A. Ravcheev, M. Nyga, O. E. Okpala, *et al.*, Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine, *Nat. Biotechnol.*, 2023, **41**(9), 1320–1331.
9 J. Hertel, A. Heinken, F. Martinelli and I. Thiele, Integration of constraint-based modeling with fecal metabolomics reveals large deleterious effects of Fusobacterium spp. on community butyrate production, *Gut Microbes*, 2021, **13**(1), 1–23.
10 J. Hertel, A. Heinken, D. Fässler and I. Thiele, Causal inference on microbiome-metabolome relations in observational host-microbiome data *via* in silico in vivo association pattern analyses, *Cells Rep. Methods*, 2023, 100615.
11 A. Heinken, A. Basile, J. Hertel, C. Thinnes and I. Thiele, Genome-Scale Metabolic Modeling of the Human Microbiome in the Era of Personalized Medicine, *Annu. Rev. Microbiol.*, 2021, **75**, 199–222.
12 H. Völzke, J. Schössow, C. O. Schmidt, C. Jürgens, A. Richter, A. Werner, *et al.*, Cohort Profile Update: The Study of Health in Pomerania (SHIP), *Int. J. Epidemiol.*, 2022, **51**(6), e372–e383.
13 L. Heirendt, S. Arreckx, T. Pfau, S. N. Mendoza, A. Richelle, A. Heinken, *et al.*, Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0, *Nat. Protoc.*, 2019, **14**(3), 639–702.
14 J. Hertel, D. Fässler, A. Heinken, F. U. Weiß, M. Rühlemann, C. Bang, *et al.*, NMR Metabolomics Reveal Urine Markers of Microbiome Diversity and Identify Benzoate Metabolism as a Mediator between High Microbial Alpha Diversity and Metabolic Health, *Metabolites*, 2022, **12**(4), 0308.
15 G. Winkler and A. Döring, Kurzmethoden zur Charakterisierung des Ernährungsmusters: Einsatz und Auswertung eines Food-Frequency-Fragebogens, *Ernähr-Umsch*, 1995, **42**(8), 289–291.
16 J. Hertel, M. Rotter, S. Frenzel, H. U. Zacharias, J. Krumsiek, B. Rathkolb, *et al.*, Dilution correction for dynamically influenced urinary analyte data, *Anal. Chim. Acta*, 2018, **1032**, 18–31.
17 F. Harrell, *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. 2015.
18 P. J. García-Nieto, E. García-Gonzalo and J. P. Paredes-Sánchez, Prediction of the critical temperature of a superconductor by using the WOA/MARS, Ridge, Lasso and Elastic-net machine learning techniques, *Neural. Comput. Appl.*, 2021, **33**(24), 17131–17145.
19 A. Juna, M. Umer, S. Sadiq, H. Karamti, A. Eshmawi, A. Mohamed, *et al.*, Water Quality Prediction Using KNN Imputer and Multilayer Perceptron, *Water*, 2022, **14**, 2592.

20 Human Microbiome Project Consortium, B. A. Methé, K. E. Nelson, M. Pop, H. H. Creasy, M. G. Giglio, C. Huttenhower, D. Gevers, J. F. Petrosino, S. Abubucker, J. H. Badger, *et al.*, A framework for human microbiome research, *Nature*, 2012, **486**(7402), 215–221.
21 A. Heinken, D. A. Ravcheev, F. Baldini, L. Heirendt, R. M. T. Fleming and I. Thiele, Systematic assessment of secondary bile acid metabolism in gut microbes reveals distinct metabolic capabilities in inflammatory bowel disease, *Microbiome*, 2019, **7**(1), 75.
22 F. Baldini, A. Heinken, L. Heirendt, S. Magnusdottir, R. M. T. Fleming and I. Thiele, The Microbiome Modeling Toolbox: from microbial interactions to personalized microbial communities, *Bioinformatics*, 2019, **35**(13), 2332–2334.
23 A. Heinken and I. Thiele, Microbiome Modelling Toolbox 2.0: efficient, tractable modelling of microbiome communities, *Bioinformatics*, 2022, **38**(8), 2367–2378.
24 A. Noronha, J. Modamio, Y. Jarosz, E. Guerard, N. Sompairac, G. Preciat, *et al.*, The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease, *Nucleic Acids Res.*, 2019, **47**(D1), D614–D624.
25 A. Guerra, A. Ticinesi, F. Allegri, A. Nouvenne, B. Prati, S. Pinelli, *et al.*, Insights about urinary hippuric and citric acid as biomarkers of fruit and vegetable intake in patients with kidney stones: The role of age and sex, *Nutrition Burbank*, 2019, **59**, 83–89.
26 C. H. E. Lau, A. P. Siskos, L. Maitre, O. Robinson, T. J. Athersuch, E. J. Want, *et al.*, Determinants of the urinary and serum metabolome in children from six European populations, *BMC Med.*, 2018, **16**(1), 202.
27 M. Lombardo, G. Aulisa, D. Marcon, G. Rizzo, M. G. Tarsisano, L. Di Renzo, *et al.*, Association of Urinary and Plasma Levels of Trimethylamine N-Oxide (TMAO) with Foods, *Nutrients*, 2021, **13**(5), 1426.
28 C. Stella, B. Beckwith-Hall, O. Cloarec, E. Holmes, J. C. Lindon, J. Powell, *et al.*, Susceptibility of human metabolic phenotypes to dietary modulation, *J. Proteome Res.*, 2006, **5**(10), 2780–2788.
29 D. Dobrijević, K. Pastor, N. Nastić, F. Özogul, J. Krulj, B. Kokić, *et al.*, Betaine as a Functional Ingredient: Metabolism, Health-Promoting Attributes, Food Sources, Applications and Analysis Methods, *Molecules*, 2023, **28**(12), 4824.
30 R. Fall and A. A. Benson, Leaf methanol—the simplest natural product from plants, *Trends Plant Sci.*, 1996, **1**(9), 296–301.
31 R. Siragusa, J. Cerda, M. Baig, C. Burgin and F. Robbins, Methanol production from the degradation of pectin by human colonic bacteria, *Am. J. Clin. Nutr.*, 1988, **47**(5), 848–851.
32 J. Lloyd-Price, G. Abu-Ali and C. Huttenhower, The healthy human microbiome, *Genome Med.*, 2016, **8**(1), 51.
33 M. Lopez-Siles, S. H. Duncan, L. J. Garcia-Gil and M. Martinez-Medina, Faecalibacterium prausnitzii: from microbiology to diagnostics and prognostics, *ISME J.*, 2017, **11**(4), 841–852.
34 Committee on toxicity of chemicals in food, consumer products and the environment; COT statement on the effects of chronic dietary exposure to methanol, 2011.
35 S. Abegg, L. Magro, J. van den Broek, S. E. Pratsinis and A. T. Güntner, A pocket-sized device enables detection of methanol adulteration in alcoholic beverages, *Nat. Food*, 2020, **1**(6), 351–354.
36 T. Kawai, T. Yasugi, K. Mizunuma, S. Horiguchi, Y. Hirase, Y. Uchida, *et al.*, Methanol in urine as a biological indicator of occupational exposure to methanol vapor, *Int. Arch. Occup. Environ. Health*, 1991, **63**(5), 311–318.
37 M. Gök, Ö. Tunçyürek, E. Ertekin and Y. Özsunar Dayanır, Metanol İntoksikasyonunda Görüntüleme Bulguları, *Kafkas J. Med. Sci.*, 2017, **7**(2), 158–161.
38 Y. L. Dorokhov, A. V. Shindyapina, E. V. Sheshukova and T. V. Komarova, Metabolic methanol: molecular pathways and physiological roles, *Physiol. Rev.*, 2015, **95**(2), 603–644.
39 Y. L. Dorokhov, T. V. Komarova, I. V. Petrunia, V. S. Kosorukov, R. A. Zinovkin, A. V. Shindyapina, *et al.*, Methanol may function as a cross-kingdom signal, *PLoS One*, 2012, **7**(4), e36122.
40 M. P. Mattson, Hormesis defined, *Ageing Res. Rev.*, 2008, **7**(1), 1–7.
41 E. Agathokleous and E. J. Calabrese, Formaldehyde: Another hormesis-inducing chemical, *Environ. Res.*, 2021, **199**, 111395.
42 C. Y. Hou, Y. S. Lin, Y. T. Wang, C. M. Jiang and M. C. Wu, Effect of storage conditions on methanol content of fruit and vegetable juices, *J. Food Compos. Anal.*, 2008, **21**, 410–415.
43 A. Paine and A. D. Davan, Defining a tolerable concentration of methanol in alcoholic drinks, *Hum. Exp. Toxicol.*, 2001, **20**(11), 563–568.
44 M. Beukema, M. M. Faas and P. de Vos, The effects of different dietary fiber pectin structures on the gastrointestinal immune barrier: impact *via* gut microbiota and direct effects on immune cells, *Exp. Mol. Med.*, 2020, **52**(9), 1364–1376.
45 M. J. Gibney and C. G. Forde, Nutrition research challenges for processed food and health, *Nat. Food*, 2022, **3**(2), 104–109.
46 K. Gerasimidis, K. Bryden, X. Chen, E. Papachristou, A. Verney, M. Roig, *et al.*, The impact of food additives, artificial sweeteners and domestic hygiene products on the human gut microbiome and its fibre fermentation capacity, *Eur. J. Nutr.*, 2020, **59**(7), 3213–3230.
47 L. M. Olsson, F. Boulund, S. Nilsson, M. T. Khan, A. Gummesson, L. Fagerberg, *et al.*, Dynamics of the normal gut microbiota: A longitudinal one-year population study in Sweden, *Cell Host Microbe*, 2022, **30**(5), 726–739.e3.
48 S. Magnúsdóttir, A. Heinken, L. Kutt, D. A. Ravcheev, E. Bauer, A. Noronha, *et al.*, Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota, *Nat. Biotechnol.*, 2017, **35**(1), 81–89.