



## **Evidence of a causal and modifiable relationship between kidney function and circulating trimethylamine N-oxide**

Downloaded from: <https://research.chalmers.se>, 2025-07-01 04:40 UTC

Citation for the original published paper (version of record):

Andrikopoulos, P., Aron-Wisnewsy, J., Chakaroun, R. et al (2023). Evidence of a causal and modifiable relationship between kidney function and circulating trimethylamine N-oxide. *Nature Communications*, 14(1).  
<http://dx.doi.org/10.1038/s41467-023-39824-4>


N.B. When citing this work, cite the original published paper.

# Evidence of a causal and modifiable relationship between kidney function and circulating trimethylamine *N*-oxide

Received: 2 November 2022

Accepted: 30 June 2023

Published online: 20 September 2023

 Check for updates

A list of authors and their affiliations appears at the end of the paper

The host-microbiota co-metabolite trimethylamine *N*-oxide (TMAO) is linked to increased cardiovascular risk but how its circulating levels are regulated remains unclear. We applied “explainable” machine learning, univariate, multivariate and mediation analyses of fasting plasma TMAO concentration and a multitude of phenotypes in 1,741 adult Europeans of the MetaCardis study. Here we show that next to age, kidney function is the primary variable predicting circulating TMAO, with microbiota composition and diet playing minor, albeit significant, roles. Mediation analysis suggests a causal relationship between TMAO and kidney function that we corroborate in preclinical models where TMAO exposure increases kidney scarring. Consistent with our findings, patients receiving glucose-lowering drugs with reno-protective properties have significantly lower circulating TMAO when compared to propensity-score matched control individuals. Our analyses uncover a bidirectional relationship between kidney function and TMAO that can potentially be modified by reno-protective anti-diabetic drugs and suggest a clinically actionable intervention for decreasing TMAO-associated excess cardiovascular risk.

Over the past two decades, the central role of the commensal gut microbiota in pathologies such as atherosclerosis and type-2 diabetes (T2D) has gained prominence<sup>1</sup>. The microbiota can influence host pathophysiology by producing molecules that directly alter metabolism and/or modulate cellular signaling either locally in the gut or systemically via the circulation<sup>1,2</sup>.

Trimethylamine *N*-oxide (TMAO) is the phase-one liver *N*-oxide of trimethylamine (TMA). TMA is a product of the microbial<sup>3–6</sup> (predominately Firmicutes) metabolism of phosphatidylcholine<sup>7,8</sup>, choline<sup>8</sup>, and *L*-carnitine<sup>9–11</sup>, components of the high-fat, high red meat western diet. TMA is taken up from the gut via the hepatic portal vein and *N*-oxidized into TMAO by host flavin mono-oxygenase<sup>3,12</sup>. High circulating levels of TMAO have been linked to increased thrombotic and cardiovascular risk in animal and human studies, even after adjustment for known cardiovascular risk factors<sup>7–9,13</sup>. TMAO is therefore proposed to mediate the higher cardiovascular risk associated with high red meat and fat intake<sup>14</sup>. Fish, the consumption of which is

associated with reduced incidence of cardiovascular disease<sup>14</sup>, is also a rich source of TMAO<sup>15</sup>. Dietary TMAO is subject to retro-conversion: i.e., it can undergo microbial reduction to TMA by *Enterobacteriaceae* followed by hepatic conversion back to TMAO<sup>16</sup>. Concurrent with diet and microbiota composition, TMAO plasma levels also reflect age<sup>17</sup>, sex<sup>18</sup>, kidney function<sup>19–21</sup>, and chronic diseases<sup>7,22</sup>. To date, the relative contribution of each of these factors to circulating TMAO levels and, therefore, elevated cardiovascular risk remains unclear. Understanding how serum TMAO levels are regulated could uncover host TMAO mechanistic targets and identify modifiable and actionable therapeutic factors to lower circulating TMAO levels.

Here, by using a data-driven “explainable” machine learning (ML) strategy<sup>23</sup>, multivariate and univariate analyses of epidemiological data and mechanistic studies in cultured cells and rodents, we sought to objectively identify variables influencing serum TMAO levels in participants of the European multicenter MetaCardis study. Moreover, by taking advantage of the unique cross-sectional MetaCardis design, we

✉ e-mail: [p.andrikopoulos04@imperial.ac.uk](mailto:p.andrikopoulos04@imperial.ac.uk); [karine.clement@inserm.fr](mailto:karine.clement@inserm.fr); [m.dumas@imperial.ac.uk](mailto:m.dumas@imperial.ac.uk)

queried how variables influencing circulating TMAO manifest at different stages of cardiometabolic disease. Capitalizing on the ML analysis, we aimed to identify (i) novel host TMAO-related mechanistic targets and (ii) a rationale for future interventions that could reduce circulating TMAO levels and thereby decrease associated excess cardiovascular risk.

In epidemiological studies, we observed that kidney function is the main modifiable factor consistently regulating fasting serum TMAO levels, and our preclinical studies align with the suggestion that elevated circulating TMAO adversely affects kidney function, by increasing kidney fibrotic injury. Further supporting the strong interplay between kidney function and fasting circulating TMAO, patients with T2D in the cohort prescribed new-generation anti-diabetics (GLP-1 Receptor Agonists<sup>24</sup>; GLP-1RAs) with evidenced reno-protective effects<sup>25</sup> had lower serum circulating TMAO levels when compared to propensity-score matched controls (Fig. 1).

## Results

In the first analysis, we investigated how TMAO levels change depending on disease classification in the MetaCardis population. We confirmed, circulating TMAO significantly increased with cardiometabolic disease severity, in line with previous reports<sup>7,22</sup> (Supplementary Fig. 1A). To identify determinants of circulating TMAO manifesting at prodromal stages of cardiometabolic disease, we focused on the subset of the MetaCardis cohort termed MetaCardis Body Mass Index Spectrum subset (BMIS<sup>26</sup>;  $N=837$ ) comprising obese/overweight individuals presenting with a range of metabolic syndrome features but not overt T2D or ischemic heart disease (IHD) (Supplementary Table 1).

### Fasting serum concentration of TMAO is associated with worse cardiometabolic profiles in BMIS MetaCardis population

We explored how TMAO correlated with biochemical variables related to cardiometabolic health in BMIS individuals. In this group, circulating TMAO was associated with reduced values of estimated glomerular filtration rate (eGFR, Spearman  $\rho = -0.124$ ,  $p\text{FDR} = 0.06$ ) and higher fasting concentrations of uric acid ( $\rho = 0.114$ ,  $p\text{FDR} = 0.09$ ) after adjustment for age, sex, country of recruitment (demographics thereafter) and body mass index (BMI) (Supplementary Fig. 1B). Higher TMAO also positively associated with indicators of central adiposity, including BMI ( $\rho = 0.107$ ,  $p\text{FDR} = 0.09$ ), visceral body fat rating ( $\rho = 0.122$ ,  $p\text{FDR} = 0.09$ ) and waist circumference ( $\rho = 0.119$ ,  $p\text{FDR} = 0.09$ ), after demographics adjustment (Supplementary Fig. 1C). Moreover, BMIS individuals with hypertension (systolic blood pressure over 140 mmHg, or receiving therapy for high blood pressure; “Methods”) had higher plasma TMAO ( $P = 0.01$ , Mann–Whitney test; Supplementary Fig. 1D). In line with previous studies<sup>7,20</sup>, objectively dividing BMIS participants into TMAO clusters using k-means<sup>27</sup> revealed that those in the cluster with the highest TMAO levels had consistently worse cardiometabolic traits and were significantly older when compared to those in the cluster with lowest TMAO levels (Supplementary Fig. 1E and Supplementary Table 2). Traits of cardiometabolic risk included altered eGFR, elevated liver enzymes, and systolic blood pressure (Supplementary Fig. 1F–I).

### Age and altered kidney function variables are the main drivers of circulating TMAO in BMIS

To better understand which variables (Supplemental Data 1) affect circulating TMAO most, we trained extreme gradient-boosted decision-tree models. We used fivefold cross-validation to predict the Explained Variance (EV) of each variable group on plasma TMAO in the left-out BMIS participants after 100 iterations (Fig. 2A). Microbiota composition alone performed poorly (EV 2%) whilst diet, another purported major contributor to TMAO production, explained less than 5% of TMAO variance. Serum metabolomics, excluding TMAO and its

precursor TMA, was the best predictor (EV 12%) with demographics second, explaining 10% of circulating TMAO variance. <sup>1</sup>H-NMR urine metabolomics, excluding TMA and dimethylamine, was the worst predictor explaining 1.5% of TMAO variance and correcting for urinary creatinine, computed by <sup>1</sup>H-NMR (“Methods”), did not improve predictions explaining 1% of TMAO variance on average (Supplementary Fig. 2A). The full model, containing all variable groups, accounted, on average, for 18.4% of TMAO variability. The average predicted TMAO values by the full model significantly correlated with measured TMAO values ( $\rho = 0.473$ ,  $P < 2.2 \times 10^{-16}$ ; Supplementary Fig. 2B).

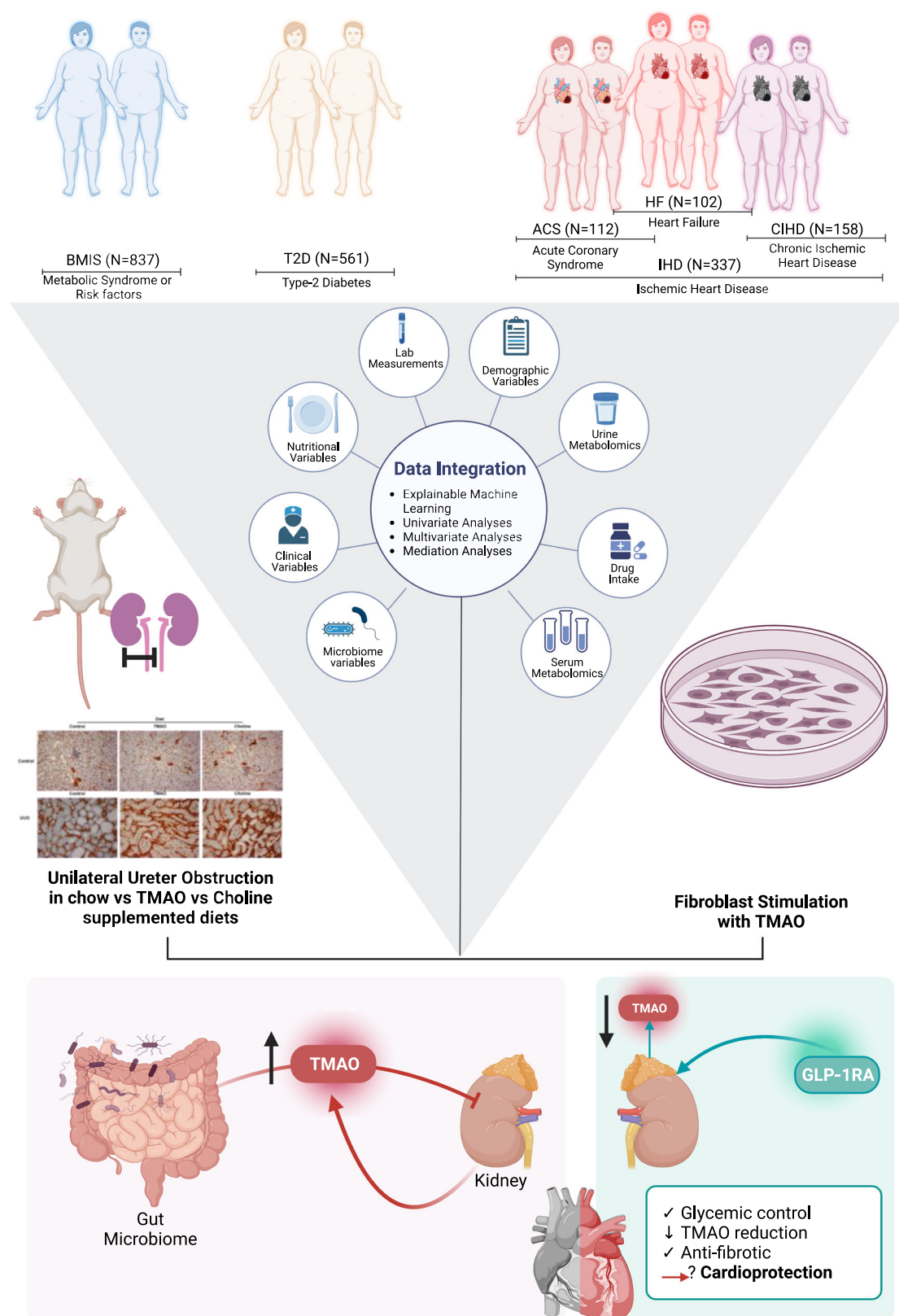
We next assessed the independent contribution of each variable group to the predictive power of the full model by training algorithms as above but removing one feature group at a time for 100 iterations. Almost 40% of the explainable variance of the full model (set to 100%) was contributed by serum metabolomic variables, with biological, dietary and microbiota taxonomic variables adding 5.5%, 4%, and 3.4%, respectively, independently explained variance (Fig. 2B). Other variable categories displayed negligible contribution to prediction suggesting considerable information overlap with the metabolomic, metagenomic, biological and dietary datasets.

Using feature attribution analysis (SHapley Additive exPlanations; SHAP<sup>23,28</sup>), we assessed how individual variables drive TMAO-predicting models. SHAP analysis identified 24 variables that contributed more than 4% of the regularized TMAO standard deviation (SD) to model outcomes. Of those, age affected predictions the most, followed by eGFR, urinary betaine, percentage of visceral body fat and serum butyryl-carnitine (Fig. 2C). Besides eGFR, additional variables associated with kidney function including plasma urea<sup>29</sup>, the uremic toxin *p*-cresol<sup>30</sup> or markers of kidney function decline (i.e., serum albumin<sup>29</sup>), were among those affecting model outcomes the most. This analysis suggests that kidney function is a major determinant of circulating TMAO. The impact of eGFR on model outcomes for BMIS individuals was bimodal with values over 90 mL/min/1.73 m<sup>2</sup>, the clinical cut-off value for normal kidney function in adults<sup>25</sup>, predicting reduced plasma TMAO and lower values resulting in increased predicted circulating TMAO (Fig. 2D).

We next trained algorithms predicting TMAO using the 24 variables identified by our SHAP analysis (“top SHAP” model) and compared it to models trained by traditional clinical risk factors<sup>31</sup> or the full model (Fig. 2E). The “top SHAP” model significantly ( $P < 2.2 \times 10^{-16}$ , Mann–Whitney test) improved predictions when compared to the full model (EV 21% vs. 18%, on average), presumably by removing noise, supporting the importance of the variables identified by the SHAP analysis.

To confirm that tree-based ML models are the most appropriate for our analysis we also built Least Absolute Shrinkage and Selection Operator (LASSO) models to predict circulating TMAO in the left-out group using again fivefold cross-validation with all the available variables (full model) as input in BMIS ( $N=582$ ; “Methods”). LASSO explained on average 14.9% of circulating TMAO variance after 100 iterations (Supplementary Fig. 3A; see source data for lambda and R<sup>2</sup> values of each iteration) as opposed to 18.4% by boosted trees for the full model (Fig. 2A). This analysis supports the appropriateness of tree-based ML models for predicting circulating TMAO in our population.

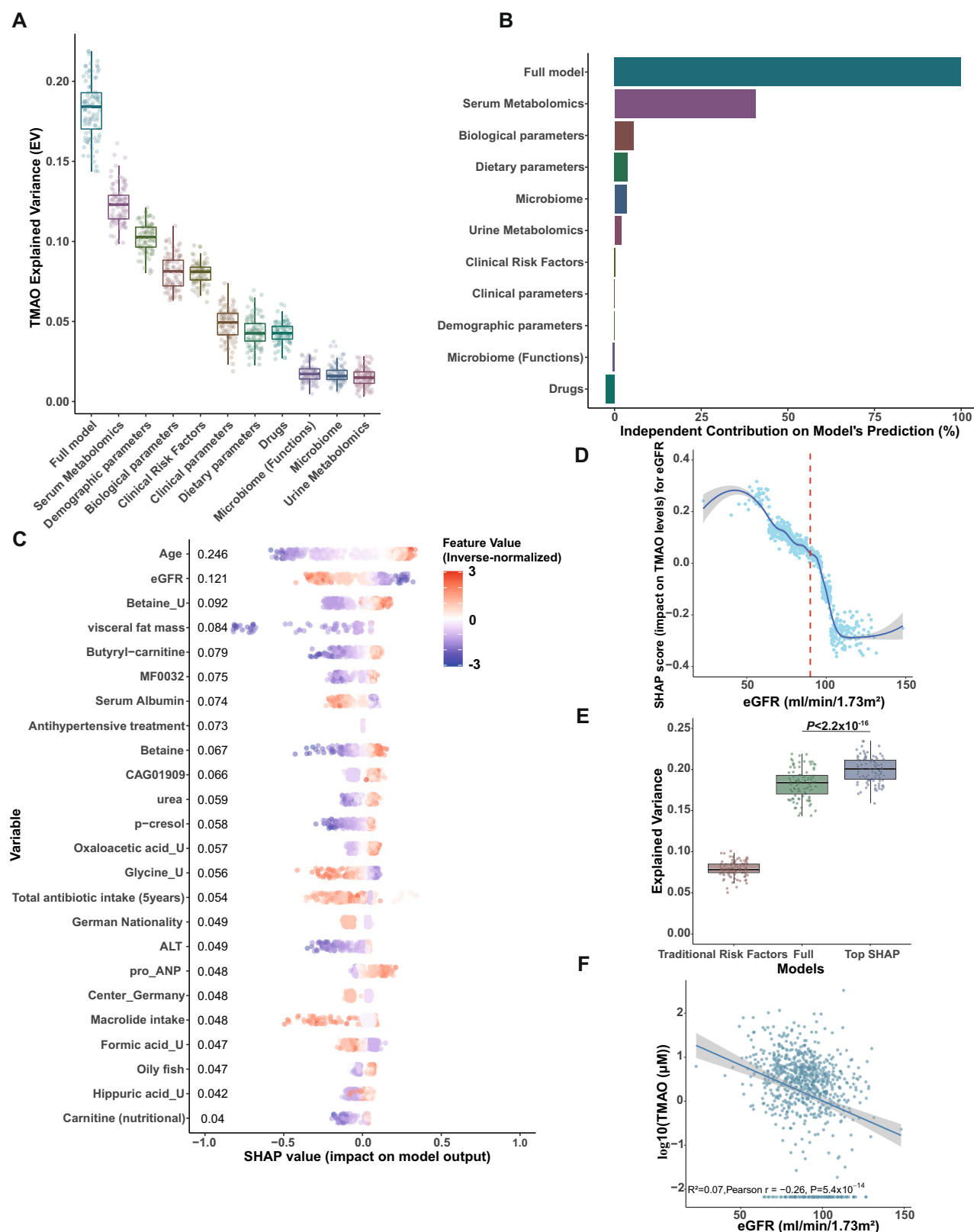
To determine how much of the variances of TMAO and of the other metabolites most strongly associated with its levels in our ML models (Fig. 2C; butyryl-carnitine, betaine, *p*-cresol and betaine\_U, oxaloacetate\_U) is explained by eGFR we built linear-regression models with each metabolite as the dependent variable and eGFR as the independent variable. Kidney function explained 7% of TMAO variance in BMIS (Fig. 2F; Pearson’s  $r = -0.26$ ,  $P = 5.4 \times 10^{-14}$ ,  $N = 837$ ) whilst the explained variance for the other metabolites ranged from 6% to 14% for *p*-cresol and urinary oxaloacetic acid, respectively (Supplementary Fig. 4A–E). To further assess the varying relationship of metabolites with eGFR we computed boosted trees models predicting eGFR with



**Fig. 1 | Overview of study design and main findings.** Here we used an integrated approach comprising Machine Learning (ML), multivariate, univariate, and mediation analyses to objectively characterize host parameters contributing to plasma TMAO levels in the multicenter European MetaCardis study. We observed that kidney function is the main modifiable factor consistently regulating fasting serum TMAO levels (Figs. 2–4) and corroborated our epidemiological findings in

preclinical models where TMAO increased kidney scarring (Fig. 5). Further supporting the strong interplay between kidney function and fasting circulating TMAO, patients with T2D in the cohort prescribed new-generation anti-diabetics (GLP-1 Receptor Agonists; GLP-1RAs) with reno-protective effects had lower serum circulating TMAO levels when compared to propensity-score matched controls (Fig. 6). Created with BioRender.com.





serum metabolomics as the input variable, with a similar methodology to TMAO. Serum metabolomics predicted on average 25% of eGFR variance after 100 iterations. From all the metabolites in our analysis TMAO was the top microbiota-derived compound that affected most strongly eGFR predictions (Supplementary Fig. 4F) in line with its reported exclusive glomerular secretion<sup>32</sup>. Collectively this analysis

suggests that in our population, TMAO is strongly interlinked with eGFR.

Several reports have previously highlighted the inverse correlation between TMAO and kidney function mostly in patients with Chronic Kidney Disease (CKD)<sup>10,19,20,32–34</sup>, but there is limited evidence for the predominance of this relationship in the non-clinical range. The

**Fig. 2 | Age and parameters associated with kidney function are the main drivers of circulating TMAO in BMIS MetaCardis participants.** **A** Coefficients of determination (Explained Variance; EV) of predicted circulating TMAO levels determined by xgboost algorithms after fivefold cross-validation in the left-out group (Supplementary Table 3 for  $n$  numbers and xgboost parameters), trained exclusively on variables from each feature category (Supplemental Data 1 for variables included in each group), or the full model (all variables), after 100 iterations. **B** Averaged independent predictive contribution of each feature category to full model predictions of plasma TMAO, trained as in **(A)**, calculated as the average reduction of EV achieved in relation to the full model (set to 100%) after removing all variables belonging to each feature group after 100 iterations. **C** Swarm plots of impact on model output (SHAP values) for each BMIS individual with complete phenotypic data ( $N=582$ ) for all variables contributing more than 4% to model predictions of regularized TMAO standard deviation, as determined by xgboost algorithms trained on each feature category. Mean absolute SHAP values from all BMIS participants ( $N=582$ ) are shown (in descending order) next to each variable. Individual dots, representing each participant, are colored by the inverse-

normalized value of the corresponding variable. U denotes urinary metabolites. **D** Dependence plot of eGFR values ( $x$  axis) versus their impact on model outcome ( $y$  axis) calculated for each individual in BMIS ( $N=837$ ) from algorithms trained on bioclinical variables, vertical red line indicates 90 mL/min/1.73 m<sup>2</sup>. The curve was drawn using locally weighted scatterplot smoothing (LOWESS) and the shaded area indicates 95% confidence interval (CI). **E** Boxplots depicting EV ( $R^2$ ) of circulating TMAO for BMIS participants computed from algorithms trained on clinical risk factors<sup>29</sup>, the full model or all 24 variables contributing more than 4% of regularized TMAO standard deviation to predictions, as determined by SHAP analysis, after 100 iterations. Significance was determined by the two-sided Mann–Whitney test. **F** Linear-regression-based scatterplot showing correlation between serum TMAO (log-transformed) and estimated Glomerular Filtration Rate (eGFR, mL/min/1.73 m<sup>2</sup>). Insert: unadjusted Pearson's  $r$ ,  $P$  value and explained variance ( $R^2$ ). Shaded area indicates 95% CI. **A**, **E** Center lines denote medians, box limits indicate the 25th and 75th percentiles, whiskers extend to the minimal and maximal values. Source data are provided as a Source Data file.

novelty of the present study includes ranking of a multitude of factors contributing to circulating TMAO levels and identification of kidney function as the top modifiable factor in a non-CKD population across the cardiometabolic disease spectrum.

### Diet and microbiota composition have a modest impact on fasting serum TMAO levels in BMIS individuals

Our ML models suggested that diet plays a modest role in determining circulating TMAO. To corroborate this, we assessed how consumption of food items rich in TMAO dietary precursors<sup>3</sup> affected circulating TMAO in BMIS participants with a Food Frequency Questionnaire<sup>35</sup> (FFQ;  $N=763$ ). After correcting for demographics and BMI, with the exception of oily fish ( $\rho=0.125$ ,  $pFDR=0.03$ ), circulating TMAO was neither significantly associated with habitual consumption of red meat, eggs or milk nor with the estimated dietary intake of the micronutrients choline, carnitine or betaine, broadly in agreement with a previous study<sup>7</sup> (Fig. 3A). In addition, no significant correlations were found between consumption of these food items and the serum levels of TMAO precursors choline, betaine or  $\gamma$ -butyrobetaine<sup>36</sup>. Conversely, circulating TMAO positively correlated with plasma choline, betaine and  $\gamma$ -butyrobetaine, implying an association between the serum concentrations of TMAO precursors. Our findings that diet and particularly meat consumption does not associate with increased TMAO levels in our population do not contradict a number of well-designed human interventional trials<sup>9–11</sup> that have established a clear link between meat intake or  $L$ -carnitine supplementation and TMAO circulating levels. Instead, collectively, our analyses suggest that in non-interventional settings where most individuals consume meat daily (75–233 g/day for European adults<sup>37</sup>), clearance by the kidney and not dietary intake of TMAO precursors is the major determinant of fasting circulating TMAO levels and therefore of the excess cardiovascular risk associated with elevated TMAO (vegetarians and strict vegans aside, who are infrequent in the MetaCardis cohort with only 65/1741 participants reporting no red meat consumption).

To identify putative microbial taxa influencing serum plasma TMAO, we performed multivariate and univariate analyses. Principal coordinates analysis of Bray–Curtis dissimilarity matrices at the species level for BMIS individuals stratified into TMAO clusters revealed a significant difference in microbiota composition between clusters ( $P=0.033$ ; Fig. 3B). Consistent with previous reports<sup>3,38</sup> and our ML models, multivariate PERMANOVA analysis uncovered a significant, albeit weak, association between circulating TMAO and microbiota composition ( $P=0.001$ ;  $R^2=0.008$ , Fig. 3B) after demographics adjustment. TMAO levels significantly associated with the quantitative abundance of 65 bacterial species (corrected for bacterial load) in BMIS individuals after correcting for demographics and BMI, primarily (44/65) of the Firmicutes phylum (Fig. 3C). In agreement with Li et al.<sup>3</sup>,

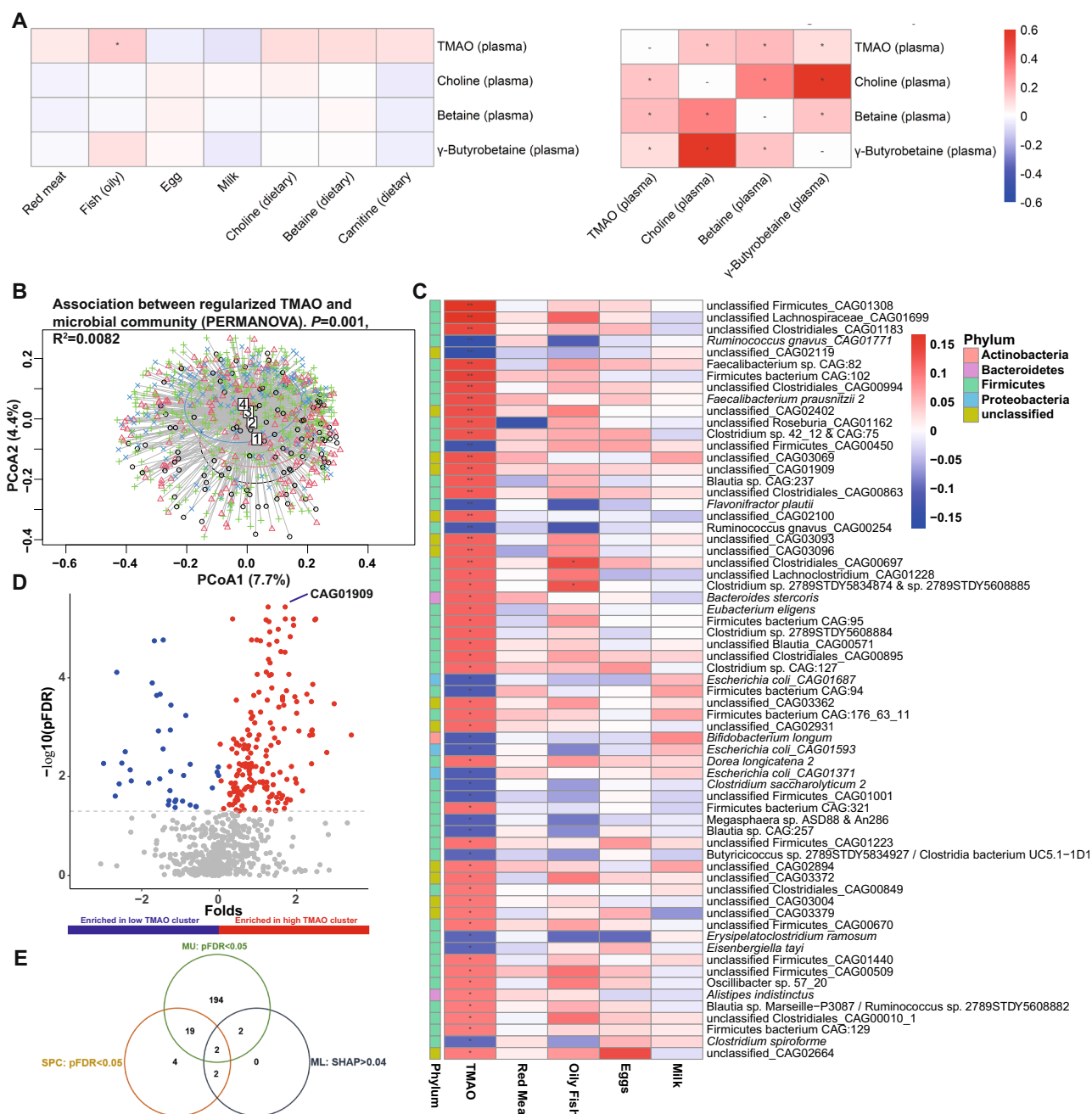
we did not identify any overlap between bacteria associated with circulating TMAO and higher red meat, milk or egg consumption in BMIS participants ( $N=761$ ; Fig. 3C). Contrasting low and high TMAO clusters revealed 215 differentially abundant bacterial species between these two groups (Fig. 3D; Mann–Whitney test).

We further analyzed the impact on circulating TMAO of the only species, an unknown bacterium taxonomically closely related to *Ruminococci* (CAG01909), that contributed at least 4% of the TMAO SD in our prediction models and was significant ( $pFDR<0.05$ ) for both correlation and differential abundance analyses (Fig. 3E). CAG01909 was more abundant and prevalent in the high TMAO cluster when compared to the low (Supplementary Fig. 5A, B). Finally, BMIS individuals harboring CAG01909 had significantly higher circulating TMAO than those that lacked it (Supplementary Fig. 5C), the CAG01909 abundance significantly correlating with TMAO concentration (Supplementary Fig. 5D).

This analysis corroborated the ML models suggesting a small, albeit significant, contribution of microbiota variations on circulating TMAO and identified a bacterium newly associated with higher serum TMAO in BMIS.

### Signatures predicting circulating TMAO shift in different disease groups

To identify putative common variables driving circulating TMAO levels in different disease groups, we trained ML algorithms for T2D ( $N=561$ ) and IHD sub-cohorts ( $N=356$ ). Similar to BMIS, microbiota composition and dietary variables alone performed poorly, explaining on average 2% and 1% of TMAO variance, respectively, in T2D and IHD individuals (Fig. 4A and Supplementary Fig. 6A, respectively). For the T2D cohort, serum metabolomics was again the best predictor (EV 12.8%) followed by bioclinical (EV 9.6%) and demographic variables (EV 8.7%) with the full model accounting for 16.2% of TMAO variance. For IHD, individual feature categories—except for serum metabolomics (EV 19.5%)—performed poorly, with the full model accounting for 16.1% of circulating TMAO variance (Supplementary Fig. 6A), possibly reflecting reduced power due to smaller sample size and/or IHD group heterogeneity. Feature attribution analysis revealed 19 variables in the T2D cohort and 10 variables in the IHD cohort that contributed more than 4% of regularized TMAO SD to model outcome (Fig. 4B and Supplementary Fig. 6B, respectively). In patients with T2D, model outcome was mostly affected by eGFR followed by age and the serum concentrations of the uremic toxin  $p$ -cresol<sup>30</sup> and  $D$ -threitol, also indicative of kidney function<sup>39</sup>. In patients with IHD, serum butyrylcarnitine, followed by age, alternative healthy eating score (aHEI) related to the ratio of white to red meat intake and the levels of the proinflammatory cytokine IP10, a marker of adverse cardiac remodeling<sup>40</sup> were the top variables. As in BMIS, the “top SHAP”



**Fig. 3 | Modest impact of diet and microbiota composition on circulating TMAO in BMIS MetaCardis subjects.** **A** Associations between circulating TMAO and its precursors with habitual consumption of food items rich in TMAO precursors ( $N = 763$ ; left panel) or with its precursors themselves (right panel). **B** Principal coordinates analysis of Bray–Curtis dissimilarity matrices of participants ( $N = 834$ ) stratified in TMAO clusters by the k-means algorithm (1 the lowest, 4 the highest) at the species level (input; 699 species present in at least 20% of the BMIS population). Insert: PERMANOVA (999 iterations) of taxonomic Bray–Curtis dissimilarity matrices association with regularized TMAO levels with age, sex, and country of recruitment as covariates. **C** Overlap of microbiome taxa significantly associated with circulating TMAO (Spearman partial correlations adjusted for age,

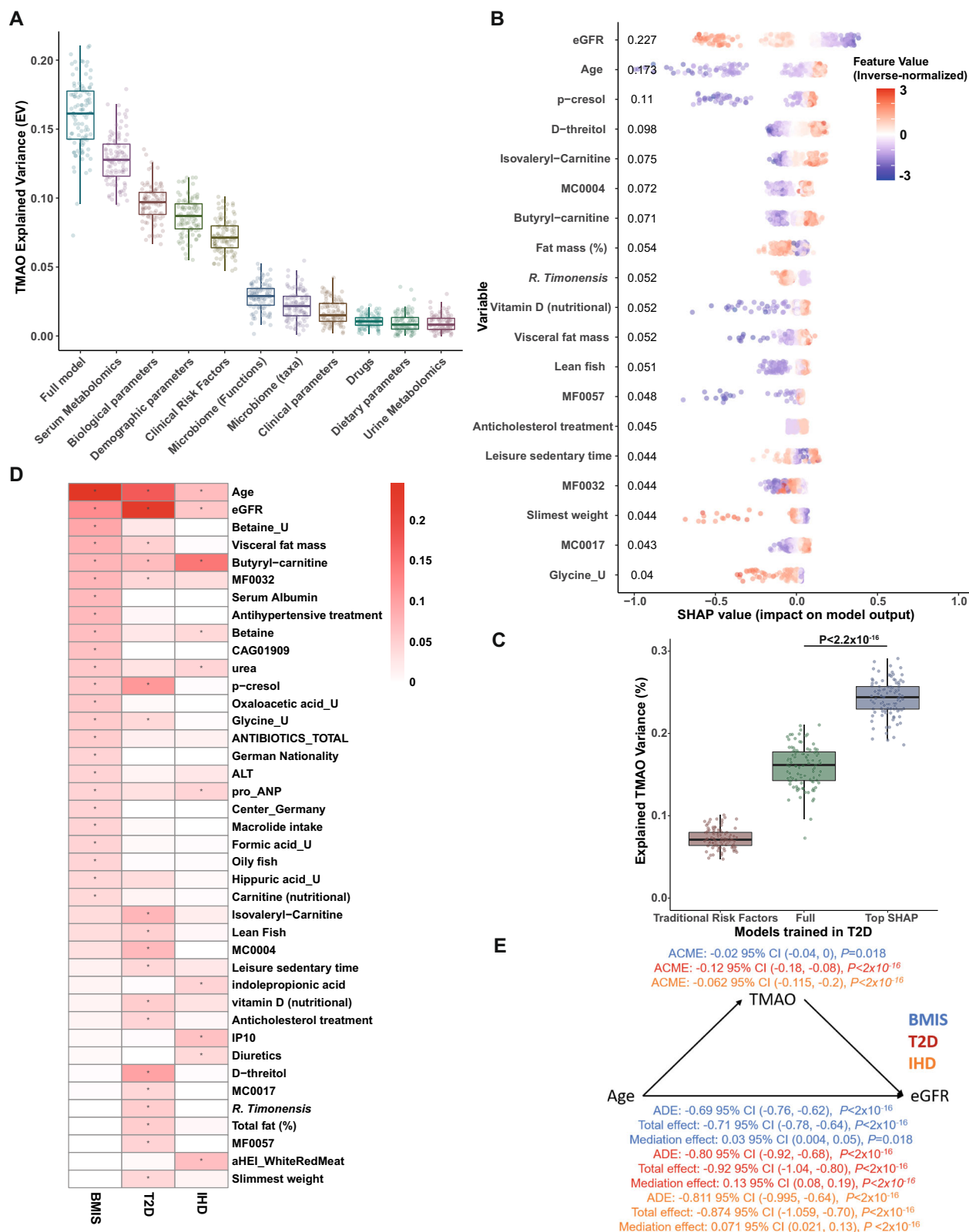
sex, country of recruitment and BMI) and the consumption of food items rich in TMAO precursors in BMIS participants ( $N = 763$ ). **D** Volcano plot of differential bacterial species abundances between BMIS participants in the lowest ( $N = 101$ ) and highest ( $N = 147$ ) TMAO clusters (blue; taxa significantly depleted, red; taxa significantly enriched in the high TMAO cluster respectively, two-sided Mann–Whitney  $U$  test,  $pFDR < 0.05$ ). **E** Venn diagram summarizing the overlap between taxa associating with circulating TMAO according to our three complementary analyses (SPC Spearman correlations, ML machine learning and feature attribution analysis; MU: two-sided Mann–Whitney  $U$  test between high and low TMAO clusters). For all  $*pFDR < 0.05$ ,  $**pFDR < 0.01$ . Source data are provided as a Source Data file.

models significantly improved predictions (EV 24% versus 16.2% for T2D; Fig. 4C and EV 21.3% versus 16.1% for IHD; Supplementary Fig. 6C) when compared to models trained with clinical risk factors or the full model.

Eight variables contributing more than 4% of TMAO SD were shared between BMIS and T2D, including, eGFR, age, and  $p$ -cresol

whilst there were six common features between BMIS and IHD (including eGFR, urea, age and butyryl-carnitine). Only three variables (eGFR, age and butyryl-carnitine) strongly contributed as predictors across all three disease groups (Fig. 4D).

We thus identified age and kidney function as the prominent variables influencing fasting circulating TMAO levels. To further



substantiate this relationship, we evaluated whether data fitted a model where eGFR causally mediates the increase of TMAO concentration with age. Under this model, kidney function achieved significance as mediating ~20% of the positive association between age and circulating TMAO levels in BMIS with its impact further strengthening in the T2D (55%) and IHD (51%) disease groups (Supplementary Fig. 6D), thus corroborating the ML analysis.

### Microbiota composition modestly affects plasma TMAO levels in T2D

In patients with T2D, ML and SHAP analyses revealed an inverse association between fecal *Romboutsia timonensis* quantitative abundance (corrected for bacterial load) and circulating TMAO (Fig. 4B), whilst no taxa strongly influenced TMAO predictions in the IHD sub-cohort (Supplementary Fig. 6B). *R. timonensis* inversely associated



**Fig. 4 | Signatures predicting circulating TMAO shift in different disease groups and TMAO causally mediates eGFR decline with age. A** Explained Variance (EV) of predicted serum TMAO levels determined by boosted decision trees (Supplementary Table 4 for *N* numbers and optimized xgboost parameters per variable group), trained exclusively on variables from each variable category (Supplementary Data 1 for a list of variables included in each group), or the full model (containing all variables), after 100 iterations in T2D MetaCardis patients. **B** Swarm plots of SHAP values (impact on model outcome) for each T2D MetaCardis participant with complete phenotypic data (*N* = 387); represented by individual dots, for all variables contributing more than 4% to model predictions of regularized TMAO standard deviation, computed from xgboost algorithms trained on each feature category. Numbers denote mean absolute SHAP values from all T2D participants (in descending order) next to their corresponding variable. Dots are colored by the inverse-normalized value of their corresponding variable. **C** Boxplots depicting Explained Variance (EV;  $R^2$ ) of circulating TMAO in T2D individuals calculated by

algorithms trained on clinical risk factors<sup>31</sup>, the full model containing all variables or all the variables contributing more than 4% of regularized TMAO standard deviation to T2D model predictions, as determined by SHAP analysis, after 100 iterations. **D** Heatmap depicting all the variables contributing at least 4% of regularized TMAO standard deviation in model predictions as determined by SHAP analysis in at least one of the MetaCardis disease groups. \*Mean absolute SHAP value > 0.04. **E** Mediation analysis (see “Methods”) computing the direct effect of TMAO on eGFR decline with age in BMIS (blue), T2D (red) or IHD (orange) MetaCardis participants. ADE: Average direct effect (of age on eGFR); ACME: average causal mediated effect (of TMAO on eGFR); Total effect: (cumulative effect of age and TMAO on eGFR (ADE + ACME)); Mediation effect: (% of the effect of age on eGFR attributed to TMAO). **A, C** Center lines denote medians, box limits indicate the 25th and 75th percentiles, whiskers extend to the minimal and maximal values. Source data are provided as a Source Data file.

with TMAO concentration ( $\rho = -0.140$ ,  $P = 0.0009$ ; Supplementary Fig. 7A), T2D individuals with detectable *R. timonensis* had significantly lower TMAO levels (Supplementary Fig. 7B) and *R. timonensis* was depleted in T2D compared to BMIS participants (Supplementary Fig. 7C), in line with circulating TMAO (Fig. 1A). Conversely, in T2D and IHD CAG01909 abundance or presence, unlike BMIS, did not associate with higher circulating TMAO (Supplementary Figs. 7D–5F). Further investigation revealed that CAG01909 abundance was inversely associated with metformin intake ( $\rho = -0.141$ ,  $\text{pFDR} = 0.042$ ; Supplementary Fig. 7G), in accordance with the well-documented effect of drugs on the gut microbiome<sup>41,42</sup>. Irrespective of TMAO associations with individual species, similarly to BMIS multivariate PERMANOVA analysis uncovered a significant, albeit weak, association between circulating TMAO and microbiota composition ( $P = 0.001$ ;  $R^2 = 0.005$ ) after adjustment for demographics.

Collectively, these analyses suggest that the composition of the gut microbiota may only modestly influence circulating levels of TMAO in patients with T2D, and in subjects with IHD or T2D abundance of the CAG01909 taxon is not linked with serum TMAO.

### Inference analysis suggests that TMAO may causally mediate the decline of eGFR with age

ML analysis identified age and kidney function as the most prominent variables influencing TMAO. Therefore, we tested the inverse relationship: i.e., whether TMAO may mediate eGFR decline with age<sup>43</sup>. In BMIS, TMAO modestly, albeit significantly, modulated the inverse relationship between age and eGFR (mediation effect 3%,  $P = 0.018$ ), whilst in T2D and IHD the impact of TMAO on kidney function decline with age strengthened (mediation effect 13%,  $P < 2.2 \times 10^{-16}$  and 7.1%,  $P < 2.2 \times 10^{-16}$ ; Fig. 4E). Mediation analysis thus indicated that TMAO, far from being a bystander, directly and adversely affects kidney function. Our finding is in agreement with a prospective study reporting that baseline TMAO positively associated with rates of eGFR decline<sup>21</sup> and with studies in animal models of CKD where TMAO diet supplementation increased kidney injury<sup>44,45</sup> and reduced eGFR.

Interestingly, mediation analysis showed that TMAO's adverse effect on kidney function strengthened at more severe stages of cardiometabolic disease, implying TMAO synergy with existing pathology (thereby providing “2-hit model”). Accordingly, we next employed preclinical models of kidney injury to gain mechanistic insights into the potential interplay between TMAO and kidney function and the putative molecular nature of the 2nd hit.

### TMAO increases trans-differentiation of human primary renal fibroblasts into myofibroblasts in conjunction with TGF- $\beta$ 1 signaling

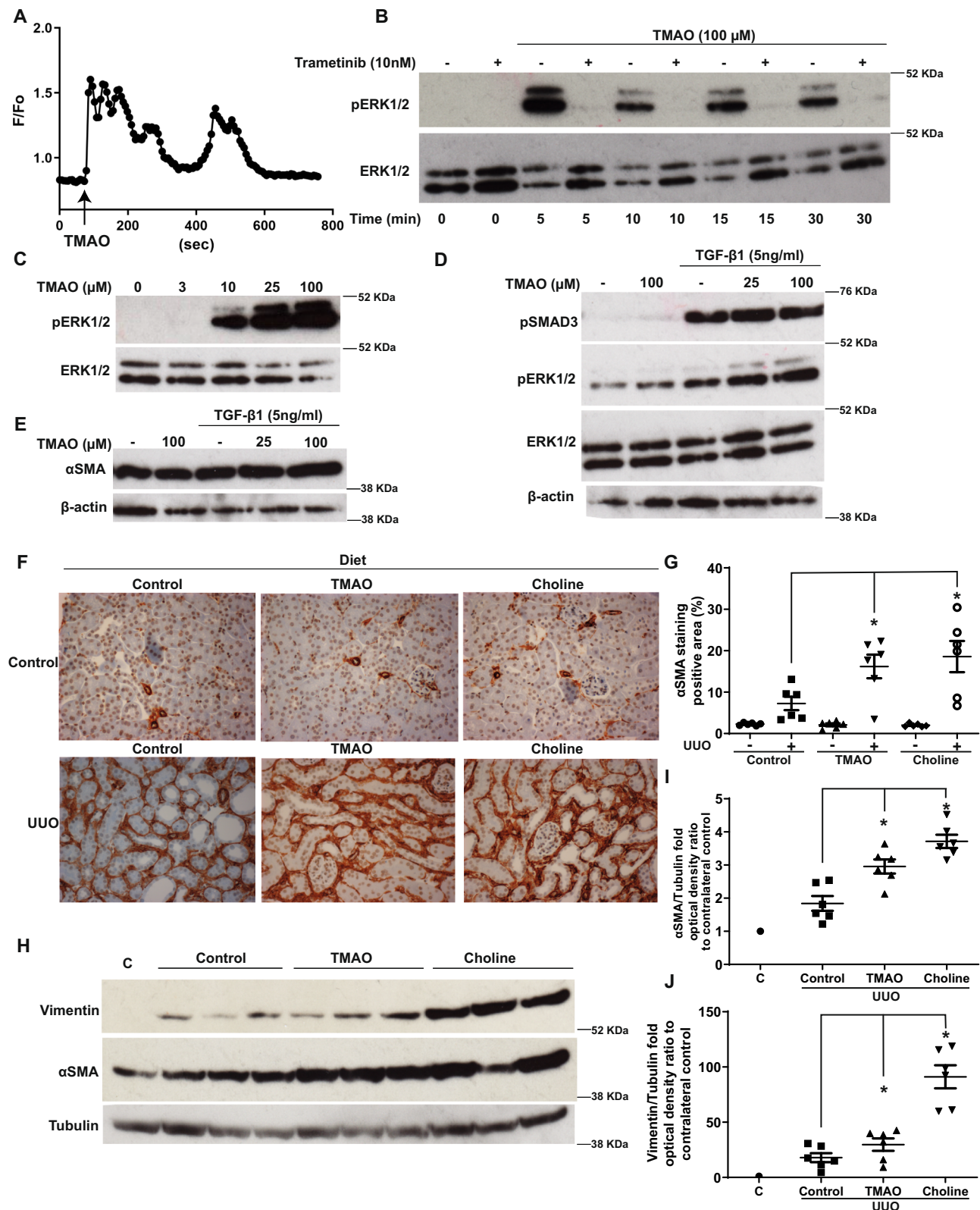
Based on TMAO's detrimental association with kidney damage of diverse aetiologies<sup>19,20</sup> similar to renal fibrosis<sup>46</sup>, we investigated the impact of TMAO on human primary renal fibroblasts (HRFs). In HRFs,

unlike platelets<sup>47</sup>, TMAO stimulation resulted in rapid  $[\text{Ca}^{2+}]_i$  increase (Fig. 5A). In endothelial cells, the ERK1/2 pathway is activated in a  $\text{Ca}^{2+}$ -dependent manner<sup>48,49</sup> and ERK1/2 activation exacerbates renal fibrosis<sup>50</sup>. Therefore, we investigated whether a similar pathway operates in HRFs. TMAO challenge increased phospho-ERK1/2 levels in a time- (Fig. 5B) and dose-dependent (Fig. 5C and Supplementary Fig. 8A) manner, at concentrations (10–100  $\mu\text{M}$ ) relevant to human disease (up to 150  $\mu\text{M}$  in patients with CKD)<sup>19</sup>. TMAO-induced ERK1/2 activation was inhibited by a MEK inhibitor, suggesting the activation occurs downstream of Ras-Raf-MEK<sup>50</sup>. In addition, ERK1/2 activation in response to TMAO was suppressed when intracellular  $\text{Ca}^{2+}$  was chelated (Supplementary Fig. 8B) or extracellular  $\text{Ca}^{2+}$  was removed (Supplementary Fig. 8C). Moreover,  $\text{Ca}^{2+}$  influx was sufficient to activate ERK1/2 in HRFs (Supplementary Fig. 8D), indicating that TMAO-induced  $\text{Ca}^{2+}$  influx was required for ERK1/2 activation. We have reported that ERK1/2 phosphorylation was required for HRF trans-differentiation to myofibroblasts in response to TGF- $\beta$ 1<sup>50</sup>. Unlike short-term application (up to 30 min), TMAO stimulation for 48 h had minimal effect on ERK1/2 or SMAD3 phosphorylation (Fig. 5D and Supplementary Fig. 8E, F) or on the expression of the myofibroblast marker  $\alpha\text{SMA}$  (Fig. 5E and Supplementary Fig. 8G). Conversely, TMAO dose-dependently augmented ERK1/2 pathway activation and myofibroblast trans-differentiation, when co-administered with TGF- $\beta$ 1 in comparison to TGF- $\beta$ 1 or TMAO alone, without affecting SMAD3 phosphorylation (Fig. 5D, E and Supplementary Fig. 8E, F).

### TMAO increases renal fibrosis: evidence from an intervention in male mice

To corroborate our in vitro findings suggesting that TMAO directly increases myofibroblast trans-differentiation and establish in vivo relevance, we performed Unilateral Ureter Obstruction (UUO) surgery in mice, a model assessing renal fibrosis in the absence of other comorbidities<sup>31</sup> that could be also impacted by TMAO. Male mice were fed a choline (1% w/w)- or TMAO (0.12% w/w)-supplemented diet for 6 weeks<sup>20</sup>. Subsequently, the ureter of one kidney was ligated whilst the other kidney remained unobstructed (Supplementary Fig. 9A). In the unobstructed (control) kidneys,  $\alpha\text{SMA}$  staining was not affected by diet. As expected, five days of UUO resulted in a significant increase in renal  $\alpha\text{SMA}$  staining of the injured kidney. However, unlike the unobstructed kidneys, TMAO or choline diet supplementation resulted in significantly more myofibroblast expansion (Fig. 5F, G). Western blotting of kidney lysates corroborated augmented  $\alpha\text{SMA}$  and vimentin (another marker of myofibroblasts<sup>50</sup>) expression in injured kidneys of mice receiving TMAO or choline diets (Fig. 5H–J). Similar to  $\alpha\text{SMA}$ , collagen deposition or macrophage infiltration, another indicator of fibrotic kidney damage<sup>46</sup>, were not affected by the TMAO or choline diets in the unobstructed kidneys (Supplementary Fig. 9B–E). Conversely, collagen and macrophage staining of kidney slides from UUO kidneys of mice consuming TMAO or choline diets were significantly





enhanced compared to controls. Thus, in our experiments in a murine model of kidney fibrosis, TMAO or choline diet supplementation resulted in “hyperactivation” of ERK1/2, mTORC1 and SMAD3 pro-fibrotic signaling<sup>46,50</sup> (Supplementary Fig. 9F).

Collectively, our in vivo and in vitro findings are consistent with TMAO aggravating kidney fibrosis due to ERK1/2 hyperactivation synergistically, probably with activation of the TGF- $\beta$ 1-mediated SMAD3 pathway causing a secondary hit in our disease model.

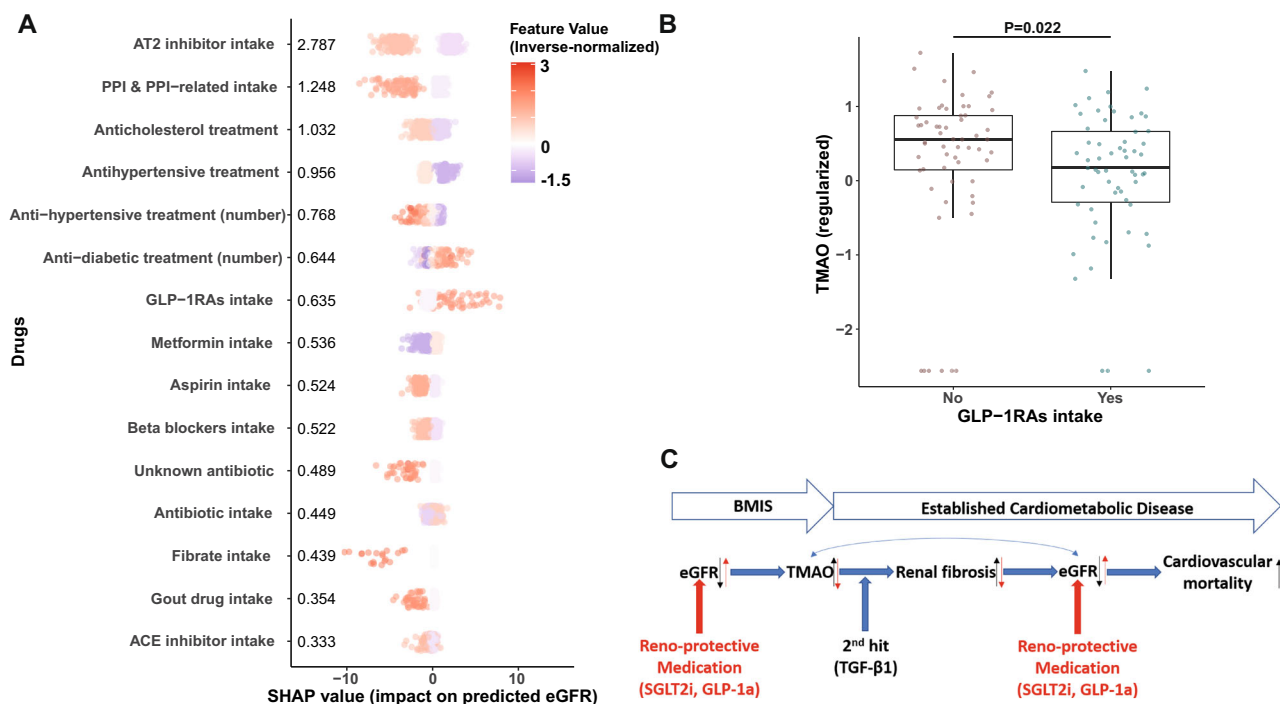
### Glucagon-like peptide-1 receptor analogs (GLP-1RAs) intake associates with lower serum TMAO concentration in MetaCardis T2D participants

Given the strong bidirectional connection we uncovered between serum TMAO and eGFR, we hypothesized that use of reno-protective medication could be linked with lower circulating TMAO. To identify suitable drugs, we trained algorithms to predict eGFR in MetaCardis patients with T2D, where TMAO appears to have the biggest impact on

**Fig. 5 | TMAO promotes myofibroblast differentiation and exacerbates renal fibrotic injury.** **A** Representative ratiometric traces (340/380 nm) from Human Renal Fibroblasts (HRFs) loaded with the  $\text{Ca}^{2+}$  indicator Fura-2 and stimulated with 100  $\mu\text{M}$  TMAO. **B** Serum-starved HRFs were preincubated with the MEK inhibitor trametinib (10 nM) for 30 min prior to stimulation with 100  $\mu\text{M}$  TMAO for the indicated times. Phospho-ERK1/2 levels were probed by Western blot; membranes were stripped and re-probed for total ERK1/2. **C** Serum-starved HRFs were stimulated with the indicated concentrations of TMAO and phospho-ERK1/2 and total ERK1/2 levels were determined as in **(B)**. HRFs in complete medium were preincubated with the indicated concentrations of TMAO for 30 min and stimulated with TGF- $\beta$ 1 (5 nM) or vehicle for 24 h. Phospho-ERK1/2, phospho-SMAD3 (**D**) and alpha-smooth muscle actin ( $\alpha\text{SMA}$ ) (**E**) levels were probed with western blot. B-actin levels for (**D**, **E**) were probed in parallel western blots. For **B**–**E** a representative image from three independent biological repeats is shown. **F** Immunostaining with

$\alpha\text{SMA}$  of kidney sections ( $\times 20$  magnification) from obstructed (UUO; 5 days post-surgery) or contralateral sham-operated (control) kidneys. Animals were fed normal chow (control), a diet containing 0.12% w/w TMAO (TMAO) or 1% choline w/w (Choline) for 6 weeks prior to surgery, as indicated.  $n = 6$  per group.

**G** Quantification of positive  $\alpha\text{SMA}$  staining as (%) of stained area/field of view averaged from 5 images per animal. **H** Western blot of whole-kidney lysates for  $\alpha\text{SMA}$  and vimentin expression. Tubulin, as loading control, was probed in parallel western blots. A representative photomicrograph from  $n = 2$  Western blots with  $n = 1$  animals for control (non-ligated kidneys, chow diet; **C**)  $n = 6$  animals for all other groups is shown. OD of the (**I**)  $\alpha\text{SMA}$  and (**J**) vimentin bands in (**H**) normalized against tubulin. The normalized density of the sham-control samples was arbitrarily set to 1. For all graphs, error bars represent the mean  $\pm$  SEM of data from  $n = 4$ –6 animals per group.  $^*P < 0.05$  versus the UUO control. Source data are provided as a Source Data file.



**Fig. 6 | Reno-protective medication is associated with reduced circulating TMAO in MetaCardis participants with T2D.** **A** Swarm plots of impact on model eGFR predictions (SHAP values) for MetaCardis T2D individuals ( $N = 561$ ) for the top 15 drugs, as determined by xgboost algorithms trained exclusively on prescribed medication. Mean absolute SHAP values from participants with T2D are shown (in descending order) next to each variable. Individual dots, representing each participant, are colored by the inverse-normalized value of the corresponding drug variable. **B** Comparison of circulating regularized TMAO levels between subjects with T2D prescribed GLP-1 receptor agonists (GLP-1RAs;  $N = 59$ ) and non-GLP-1RAs treated subjects with T2D propensity-matched for age, sex, disease group, and hypertension status ( $N = 59$ ) (Supplementary Table 6 for group characteristics).

$P$  value determined by two-sided Mann–Whitney  $U$  test. Center lines denote medians, box limits indicate the 25th and 75th percentiles, whiskers extend to the minimal and maximal values. **C** Summary of the main findings of our study. We demonstrate that eGFR, irrespective of disease stage, is the primary modifiable modulator of circulating TMAO. Far from being a bystander, TMAO significantly accelerates the rate of renal output decline by age, with its effect increasing at advanced stages of disease. TMAO promotes renal fibrosis in conjunction with established pathophysiology (two-hit model) further negatively impacting renal clearance. Accordingly, medication with reno-protective properties (red arrows), such as GLP-1RAs, reduce circulating TMAO levels thereby potentially moderating its adverse effect on kidney function. Source data are provided as a Source Data file.

kidney function, using prescribed medication as input. SHAP analysis revealed that antihypertensive and anti-cholesterol treatments had a negative impact on predicted eGFR, probably reflecting more advanced disease (Fig. 6A). Conversely, anti-diabetic treatments had a positive effect, with GLP-1RAs being the drugs with the biggest positive impact on predicted eGFR (Fig. 6A). In accordance with their documented reno-protective effect even in T2D patients with no preexisting nephropathy<sup>25</sup>, patients with T2D receiving GLP-1RAs had significantly lower circulating TMAO (Fig. 6B) than individuals matched for age, sex, hypertension, and disease group (Supplementary Fig. 8A–E and Supplementary Table 6). This exploratory analysis

suggests that GLP-1RAs may reduce serum TMAO concentration and thereby the associated higher risk of IHD.

## Discussion

With our approach combining epidemiological studies, explorative cellular experiments and murine interventions, we demonstrate that kidney function is the main modifiable factor consistently regulating fasting serum TMAO concentrations and that TMAO adversely impacts on eGFR, at least partially, by increasing kidney scarring synergistically with existing pathophysiology elevating kidney tissue TGF- $\beta$ 1 signaling. Consistent with our findings, use of reno-protective drugs,

GLP-1RAs<sup>25</sup>, was associated with lower circulating TMAO in MetaCardis participants with T2D (Fig. 5C).

Irrespective of cardiometabolic disease severity, gut microbiota composition had a modest ( $R^2 < 0.01$ ), albeit significant, association with serum TMAO levels in the MetaCardis population, in agreement with recent human studies<sup>3,38</sup>. We replicated associations of Firmicutes<sup>3–5</sup> with circulating TMAO and uncovered novel associations between an unknown bacterium (CAG01909) and *R. timonensis*, a marker of less diverse diet in autistic children<sup>52</sup>. However, taxa associated with TMAO diverged between disease modalities. This could at least, partially, be due to medication, suggesting that targeting TMAO production at the species level would be ineffective.

In agreement with a recent study<sup>3</sup>, we did not find any significant association between habitual consumption of red meat and fasting serum levels of TMAO. With a few notable exceptions<sup>3</sup>, the contribution of diet and in particular red meat and *L*-carnitine to TMAO levels, showing an increase following intervention, has predominately been examined in metabolically healthy volunteers<sup>9–11</sup>. In such interventions, often *L*-carnitine has been provided as a dietary supplement, which is poorly absorbed in the small intestine (~12%), as opposed to dietary *L*-carnitine (~71%)<sup>33</sup>, and therefore may be more available for microbial catabolism in the upper gut and large intestine, leading to overestimating its role in TMA and, thereby, TMAO production. Our observations, similar to the report in ref. 3, suggest that in non-interventional settings where individuals habitually consume meat (75 to 233 g/day in European adults<sup>37</sup>), this contributes minimally to fasting circulating TMAO variability, possibly limiting the isolated effect of dietary manipulation on TMAO levels in non-interventional settings, aside strict vegans or vegetarians.

Further, our analyses revealed that the multi-omics signature associated with higher circulating TMAO concentrations and, therefore, elevated cardiovascular risk, shifted from obesity with increased insulin resistance toward overt T2D and IHD with only three variables (age, eGFR, and serum butyryl-carnitine) consistently being strong contributors across models built with the BMIS, T2D and IHD disease groups. This is in agreement with our recent report that the majority of markers of dysmetabolism manifest early during cardiometabolic disease development<sup>31</sup>. Accumulation of butyryl-carnitine has been associated with abnormal mitochondrial lipid  $\beta$ -oxidation in T2D<sup>54</sup>, suggesting a link between TMAO and mitochondrial dysfunction consistent with the finding that TMAO binds to protein kinase R-like endoplasmic reticulum kinase and increases mitochondrial stress<sup>55</sup>.

Far from being a bystander, mediation analysis indicated that TMAO significantly accelerated the rate of kidney function decline by age, with its effect increasing in more severe disease stages. Consistent with this, our preclinical work uncovered that TMAO primed renal fibroblasts for conversion to myofibroblasts, the primary collagen-producing cells in the kidney<sup>46</sup>, and contributed to renal fibrosis, a hallmark of kidney damage irrespective of the underlying cause that significantly contributes to eGFR decline<sup>46</sup>. However, TMAO insult alone was not sufficient to convert renal fibroblasts to myofibroblasts requiring synergy with a preexisting pathological state, i.e., pro-fibrotic signaling (“second-hit”) in the form of TGF- $\beta$ 1 stimulation, the most prominent pro-fibrotic cytokine<sup>46</sup>. TGF- $\beta$ 1 expression is increased in ligated UOO mouse kidneys<sup>56</sup> or kidneys of patients with diabetic nephropathy<sup>57</sup> and circulating TGF- $\beta$ 1 is elevated in conditions that are risk factors for CKD including hypertension, dyslipidemia and T2D<sup>58</sup>. Therefore, our results suggest that TMAO accelerates eGFR decline in concert with existing pro-fibrotic signaling (i.e., TGF- $\beta$ 1; “second-hit”), consistent with observations in animal models of CKD<sup>44,45</sup> and humans<sup>20,21</sup>.

Supporting our assertion that TMAO accelerates cardiovascular disease progression together with existing pathology, at least partially by impacting the kidney, TMAO increased all-cause mortality only in individuals with eGFR <90 mL/min/1.73 m<sup>2</sup> (see ref. 59) and circulating

TMAO levels in healthy adults were not indicative of future atherosclerotic burden<sup>60</sup>.

TMAO is an independent risk factor of cardiovascular morbidity and mortality in patients with established IHD or CKD<sup>18,19</sup> with accumulating evidence suggesting direct causal effects<sup>47</sup>. Our work suggests that reno-protective strategies could potentially lower circulating TMAO and therefore preserve kidney function in individuals with high circulating TMAO in the presence of risk factors known to increase TGF- $\beta$ 1 signaling (i.e., hypertension and T2D<sup>58</sup>). Indeed, we observed that patients with T2D prescribed GLP-1RAs, drugs with documented reno-protective and beneficial cardiovascular effects<sup>24,25</sup>, had significantly lower fasting plasma TMAO levels than propensity-score matched controls. Differences in microbiota composition were predictive of glycemic responses to GLP-1RA intake<sup>61</sup> and further work is required to determine factors influencing GLP-1RA-mediated reno-protection which appear to be independent of improvements in glycemic control<sup>62</sup>.

Collectively, our findings demonstrate that eGFR, irrespective of disease stage, is the primary modifiable modulator of circulating TMAO, which then by promoting renal fibrosis in conjunction with established pathophysiology (“second-hit” model) further negatively impacts kidney function. Accordingly, we observe that intake of GLP-1RAs is associated with lower circulating TMAO levels, thereby potentially moderating the adverse effect of TMAO on kidney function and suggesting a putative mechanism for these drugs that observed renoprotection in large pharmaceutical trials<sup>25</sup>.

Our work conceptually advances the understanding of how TMAO levels (and therefore associated cardiovascular risk) are regulated in humans with a wide range of cardiometabolic disease burden in non-interventional settings. In addition, we uncover a direct mechanistic link between TMAO and renal fibrosis in conjunction with existing comorbidities (known to elevate TGF- $\beta$ 1 signaling) such as hypertension and T2D. Furthermore, our findings suggest that therapeutic modalities preserving kidney function could markedly benefit and reduce cardiovascular risk in individuals with high circulating TMAO in the presence of risk factors (T2D, hypertension, or metabolic syndrome). This merits urgent testing in a longitudinal independent clinical trial.

## Methods

### MetaCardis study design and recruitment

MetaCardis is a cross-sectional study that recruited individuals at increasing stages of dysmetabolism and IHD severity (ranging from metabolically healthy, metabolic syndrome and/or obesity, T2D, IHD), aged 18–75-years old and recruited from Denmark, France and Germany between 2013 and 2015. Patients under care of the participating hospitals meeting the inclusion criteria of the study were invited to enroll. Healthy controls were recruited via public advertisement. Study participants provided written informed consent and the study was undertaken according to Helsinki Declaration-II. Ethical approval was obtained from the Ethics Committee CPP Ile-de France, the Ethical Committees of the Capital Region of Denmark (H-3-2013-145), and Ethics Committee at the Medical Faculty at the University of Leipzig (047-13-28012013). Study design, recruitment and exclusion criteria has been extensively described<sup>26,31,41,63</sup>. The overarching goal of the trial was to investigate the impact of qualitative and quantitative changes in the gut microbiota on the pathogenesis of cardiometabolic diseases (CMDs) and their associated co-morbidities (ClinicalTrials.gov Identifier: NCT02059538). For the present study, patients were subclassified in three groups: BMI-spectrum patients (BMIS<sup>26</sup>;  $N = 837$ ), encompassing MetaCardis participants presenting with metabolic syndrome-related risk factors or conditions (hypertension, as defined by the American Heart Association<sup>64</sup>; obesity, as defined by the World Health Organization<sup>65</sup> and metabolic syndrome, as defined by the International Diabetes Federation<sup>66</sup>) and patients diagnosed with type-2 diabetes (T2D, as defined by the American Diabetes Association<sup>67</sup>;  $N = 561$ ).



or ischemic heart disease (IHD;  $N = 356$ ). The IHD group comprised patients with Acute ( $<15$  days) Coronary Syndrome (ACS;  $N = 106$ ), Chronic IHD (CIHD;  $N = 157$ ) with normal Left Ventricular Ejection Fraction (LVEF) determined by echocardiography and Heart Failure patients (HF;  $N = 93$ , LVEF  $< 45\%$ ). Cardiometabolic disease is used as an umbrella term for all the above cases, and severity of cardiometabolic disease refers, in this manuscript, to the progression from single risk factors such as obesity to overt T2D and cardiac phenotype (ischemic heart disease and heart failure).

### Sample and phenotypic information collection

Biofluid and biomatter collection has been described elsewhere<sup>26,31,41,63</sup>. Briefly, blood samples were collected in the morning after overnight fasting, and fecal samples were collected at home by participants, frozen immediately and transferred to study centers on dry ice within 48 h. All samples were stored at  $-80^{\circ}\text{C}$  until use. Clinical history, medication and phenotypic information were acquired as described<sup>26,31,41,63</sup> with standardized procedures across centers. Participants reported habitual food intake through a customized Food Frequency Questionnaire (FQQ). The participant's responses to the FQQs was validated by three web-based dietary recalls for a subset of study participants ( $N = 324$ ) according to established practices<sup>35</sup>. Biochemical variables were measured in a single center according to standard procedures<sup>31</sup>. Estimated glomerular filtration rate (eGFR) was calculated with the CKD-EPI formula without ethnicity adjustment<sup>68</sup>.

### Metabolic profiling

**<sup>1</sup>H-nuclear magnetic resonance (<sup>1</sup>H-NMR) spectroscopy.** Spectra acquisition, using an Avance spectrometer (Bruker) at 600 MHz; and structural assignments have been extensively described previously<sup>31</sup>. Absolute quantifications were derived using the "In Vitro Diagnostics for research" (IVDr) quantification BI-Quant-UR algorithm (Bruker, <https://www.bruker.com/en/products-and-solutions/mr/nmr-clinical-research-solutions/b-i-quant-ur.html>, v1.1). For some analyses metabolites absolute quantifications were divided by the creatinine concentration, also derived by the IVDr platform from the <sup>1</sup>H-NMR spectra.

**Gas chromatography-coupled mass spectrometry (GC-MS).** Serum samples (100  $\mu\text{l}$ ) were prepared, analyzed and processed as described<sup>31,41</sup>. Briefly, protein was methanol-precipitated, methanol was evaporated to dryness and subsequent to derivatization samples were injected to an Agilent 7890B-5977B Inert Plus GC-MS system. The chromatographic column was an Agilent ZORBAX DB5-MS (30 m  $\times$  250  $\mu\text{m}$   $\times$  0.25  $\mu\text{m}$  + 10 m Duragard). The temperature gradient was 37.5 min long and the mass analyzer was operated in full-scan mode between 50 and 600  $m/z$ . Peaks were annotated with the use of the Fiehn library (Agilent G1676AA Fiehn GC/MS Metabolomics RTL Library, User Guide, Agilent Technologies, [https://www.agilent.com/cs/library/usermanuals/Public/G1676-90001\\_Fiehn.pdf](https://www.agilent.com/cs/library/usermanuals/Public/G1676-90001_Fiehn.pdf)). Metabolic features with low reproducibility or linearity were removed from the dataset, resulting in 102 annotated metabolic features.

**Ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS).** UPLC-MS/MS performed on a Waters Acquity UPLC-Xevo TQ-S UPLC-MS/MS system equipped with an Acquity BEH HILIC (2.1  $\times$  100 mm, 1.7  $\mu\text{m}$ ) chromatographic column was employed to determine TMA, TMAO, choline, betaine,  $\gamma$ -butyrobetaine, betaine aldehyde, butyryl-carnitine, isovaleryl-carnitine, OH-isovaleryl-carnitine, stearoyl-carnitine, oleoyl-carnitine, linoleoyl-carnitine, myristoyl-carnitine, lauroyl-carnitine, and decanoyl-carnitine as described previously<sup>31,41</sup>. TMAO circulating values of all MetaCardis participants were  $\log_{10}$ -transformed, and subsequently the median was subtracted and divided by the standard deviation<sup>23</sup> (SD; regularized TMAO values throughout). MassLynx™ software (Waters corporation, v4.2) was used for data acquisition and analysis.

### Metagenomic analysis

Phylogenetic microbiota profiles were built after correction for bacterial load as extensively described<sup>26,31,41,63</sup> using a protocol devised in ref. 69 with modifications. Briefly, total fecal DNA was extracted following the International Human Microbiome Standards (IHMS) guidelines (SOP 07 V2 H) and samples were sequenced using ion-proton technology (ThermoFisher Scientific). Gene abundance profiling was performed using the 9.9 million gene integrated reference catalog of the human microbiome, as described<sup>26,31,41,63</sup>. The following libraries were used to process metagenomic data: METEOR (v.3.2; <https://forgemia.inra.fr/metagenopolis/meteor>), Alientrimmer v0.4.0, Bowtie2 v2.3.4, MetaOMineR (momr, v1.31), Omixer-RPM (v1.0).

### Customized microbial module analysis (GMM)

Manually curated customized module sets focusing on anaerobic bacterial and archaeal fermentation processes relevant to the human gut microbiota were assembled as previously extensively described<sup>26,31,41,63</sup>.

### Statistical analyses

Statistical analysis was undertaken using R (v4.0.3; R Core Team (2020). R: A language and environment for statistical computing. R (v4.0.3) Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>). For comparing two groups, we used Mann-Whitney  $U$  and for multiple group comparisons Kruskal-Wallis tests. Unadjusted Spearman correlations were computed using R, whilst adjusted Spearman correlations with ppcor (v1.1).  $P$  values were corrected for multiple comparisons using the Benjamini-Hochberg method.

### Machine learning (ML) analysis

**Variable groups.** Patient phenotypic variables were separated into ten groups (Supplementary Data 1). Specifically, biological parameters included biochemical and clinical serum laboratory tests, including lipids, glycated hemoglobin, creatinine, and eGFR. Clinical parameters consisted of clinical history, BMI, systolic and diastolic blood pressure, anthropometric variables, and stool frequency and type. Demographic information comprised age, physical activity, educational and income levels, smoking status, ethnicity and country of recruitment. Drug variables included intake of common medication as described<sup>41</sup>, number of antibiotic courses in the last 5 years and number of anti-hypertensive, anti-diabetic and lipid-lowering treatments. Dietary parameters included habitual consumption of 37 food items, daily nutrients intake derived from these food items calculated as in<sup>35</sup>, alternative Healthy Eating<sup>70</sup> (aHEI), Dietary Approaches to Stop Hypertension<sup>71</sup> (DASH) and Dietary Diversity<sup>72</sup> (DDS) scores. Clinical risk factors variables included age, systolic and diastolic blood pressure, glycated hemoglobin levels, fasting cholesterol levels, smoking status and waist circumference as described<sup>31</sup>. Serum metabolomics comprised the absolute or relative levels of 116 circulating metabolites determined by GC-MS or UPLC-MS/MS<sup>31,41</sup>. Urine metabolomics included absolute quantification of 47 urine metabolites from <sup>1</sup>H-NMR spectra calculated with the IVDr algorithm. Microbiota variables included an abundance of 699 bacterial species present in at least 20% of MetaCardis patients corrected for microbial load and the first 10 principal components from a PCA of relative microbial gene abundances<sup>31,41</sup>. Microbiome (Modules) group included abundance of 116 manually curated bacterial modules<sup>31</sup>. In all cases, categorical variables were converted into dummy variables using caret (v6.0.86).

**Boosted decision trees (Xgboost).** We predicted regularized circulating TMAO levels by using gradient-boosting decision trees based on the xgboost algorithm (v1.3.2.1)<sup>73</sup>, co-opting a strategy from ref. 23. Xgboost consistently outperforms other algorithms in Kaggle competitions for tabular data. For each of our 10 variable groups

(Supplementary Table 5) we optimized xgboost models using fivefold cross-validation and two sequential hyperparameter grids searches with early stopping (972 different parameter combinations for each feature group in total) to predict mean-centered and unit variance-scaled (regularized) TMAO levels in the left-out group using root-mean-square error (RMSE) to evaluate model outcomes. After parameter optimization, we predicted circulating TMAO using 5-fold cross-validation for 100 iterations using as input the variables of each group or all variables (full model). For each round, we calculated the coefficient of determination using the rsq function from yardstick (v0.0.7) and the predicted regularized TMAO values. Xgboost models trained with 80% of each patient group participants during cross-validation (five for each round) were saved and used for feature attribution analysis in the left-out group (see below).

To minimize the risk of overfitting, we took four steps:

1. during the model parameter optimization step, training was stopped if predictions were not improved for ten rounds (early\_stopping\_rounds = 10).
2. We incorporated the regularization parameters lambda and gamma in our models, thus making our models more conservative<sup>23</sup>.
3. We introduced randomness by using 0.8–0.9 of the available variables in each variable group (colsample\_bytree=0.8–0.9 depending on parameter optimization; Supplemental Tables 4–6 for specific model parameters) for the training of each tree, as a way to minimize overfitting.
4. all our conclusions are based on ensemble models (the average of 100 independent runs) that in combination with the introduced randomness and the out-of-sample predictions (fivefold cross-validation) makes our models conservative.

Please see also the XGBoost documentation for additional information on model parameters (<https://xgboost.readthedocs.io/en/latest/index.html>).

**SHapley additive exPlanations (SHAP) analysis.** We interpreted our ML models and assigned relative importance to variables influencing circulating TMAO levels by co-opting SHAP values, as expanded for tree-based ML models<sup>28</sup> and recently used to objectively evaluate factors driving metabolite plasma levels in humans<sup>23</sup>. Briefly, for each prediction the SHAP value of a particular variable is the difference in the model output when this variable is included versus when excluded. Variables are added to the model in all possible orderings, and the SHAP value is computed from the average of model outcomes<sup>28</sup>. Variable SHAP values for each individual in the five left-out groups were extracted from corresponding xgboost models trained on the remaining 80% participants for each variable group using SHAP-forxgboost (0.1.0) and averaged over 100 iterations. We assigned relative importance to each variable by computing the mean of the absolute SHAP value for all individuals in each disease group similarly to ref. 23. For swarm plots depicting each individual participants SHAP values, variable values were inverse-normalized using the RNOmni (v1.0.0) package.

**Least absolute shrinkage and selection operator (LASSO).** We used the LASSO algorithm as implemented in the glmnet (v4.1) package to predict regularized TMAO values in BMIS ( $N = 582$ ) participants using all available variables as input (full model). Initially, we used fivefold cross-validation to determine the best regularization parameter lambda (from 100 values ranging from  $10^{-3}$  to  $10^5$ ) for predicting circulating TMAO in the left-out group. We next trained five LASSO regression models in 80% of BMIS participants excluding the corresponding left-out group with the optimal lambda parameter. These models were then used to predict regularized TMAO in the left-out groups and subsequently to calculate the coefficient of determination

using the rsq function from yardstick (v0.0.7). This process was repeated for 100 iterations, similar to the xgboost analysis.

**Clustering.** Clustering was performed using the built-in R k-means function with the Hartigan–Wong algorithm using 25 random sets, as described<sup>27</sup>.

**Multivariate analysis.** To identify microbiota composition differences between individuals split in TMAO clusters we performed multivariate homogeneity analysis of Bray–Curtis dissimilarity matrices at the species level (699 species present in at least 20% of our population) and determined statistical significance with a permutation ANOVA test (999 iterations) using the vegan (v2.5.7) package. Permutation analysis of variance (PERMANOVA) of regularized circulating TMAO versus Bray–Curtis taxonomic dissimilarity matrices with age, sex, and country of recruitment as covariates for 999 iterations were performed with vegan (v2.5.7).

**Mediation analysis.** We performed mediation analysis<sup>74</sup> to assess the putative impact of eGFR on plasma TMAO increase with age and, conversely, the impact of TMAO on eGFR decline with age. We used the Preacher and Hayes bootstrapping method, as implemented in the mediation (v4.5.0) R package<sup>75</sup>, using general linear models with sex and country of recruitment as covariates. Confidence intervals and Bayesian  $P$  values were computed after 999 simulations.

**Propensity-score matching.** MetaCardis participants with T2D were propensity-score matched using the R package MatchIt<sup>76</sup> (v4.1.0) with age, sex, disease group, and hypertension status as covariates using nearest neighbor matching determined by generalized linear models. All covariates were given equal weight.

## Preclinical models

### In vitro experiments

**Cell culture and protein extraction.** Primary human adult kidney fibroblasts from a single donor (DV Biologics, AU009-F) were cultured in 10% FCS low-Glucose DMEM with 1% Pen/Strep antibiotics (Sigma). For acute (up to 30 min) TMAO stimulation, cells (100,000/condition) were serum-starved for 1 h in a physiologic serum-free buffer as previously described<sup>50</sup>. Fibroblasts were preincubated with Tramenitib (10 nM, Selleckchem), BAPTA-AM (20  $\mu$ M, Molecular Probes) or vehicle for 30 min before TMAO (Sigma) stimulation. For longer-term (24 h) stimulation, fibroblasts were in serum-free low-glucose DMEM overnight and subsequently stimulated with TMAO, TGF- $\beta$ 1 (5 ng/ml, R&D), or their combination. At the end of the experiment, fibroblasts were lysed and stored at  $-80^{\circ}\text{C}$  until further use, as described<sup>50</sup>.

**Cytosolic  $[\text{Ca}^{2+}]$  measurements.** Fibroblasts were serum-starved overnight in low-glucose DMEM and subsequently loaded with 5  $\mu$ M fura-2-AM (Molecular Probes) in pH 7.4 Hanks Balanced Salt Solution (HBSS; Sigma) containing  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ . Measurements were obtained on an epifluorescence inverted microscope equipped with a  $\times 20$  fluorite objective. Single-cell intracellular  $\text{Ca}^{2+}$  ( $[\text{Ca}^{2+}]_i$ ) was monitored using excitation at 340 and 380 nm, through a monochromator (Cairn Research). Emitted light was reflected through a 515 nm filter to a QImaging Retiga CCD camera (Cairn Research) and digitized to 12-bit resolution. All imaging data were collected and analyzed using software from Andor.

**Animal procedures.** All animal experiments were conducted in accordance with the United Kingdom Home Office Animals 1986 Scientific Procedures, with local ethical committee approval (Project License 70/8356). Male C57BL/6J mice (Charles River) at 6–8 weeks of age were fed a control, 0.12% w/w TMAO- or 1% w/w choline-containing diets (Teklad Global 18% Protein Rodent Diet (Cat. #2018); 6 animals/



group) for 6 weeks<sup>20</sup>. Food and water were available ad libitum, and mice were held at 12/12 h light/dark cycle at 20–23 °C and 40–60% humidity in standard individually ventilated cages (5–6 animals in each cage). Subsequently, Unilateral Ureter Obstruction (UUO) was established as previously described<sup>50</sup>. Only male mice were used, since this is common practice for this model<sup>50</sup>. Briefly, mice were anesthetized with isoflurane and the abdominal cavity was exposed using midline laparotomy, the right ureter was isolated and tied off 0.5 cm from the pelvis. The left ureter was left unclamped and served as the sham-operated control. At day 5, mice were euthanized by terminal anesthesia with ketamine and cardiac puncture, and subsequently the kidneys were harvested. Both kidneys from each animal (UUO and sham) were cut in half longitudinally. One-half of each kidney was snap-frozen in liquid nitrogen and subsequently stored at –80 °C for Western blotting. The other half was fixed with 10% Formalin (Sigma) for 16 h at 4 °C, transferred to a 70% v/v ethanol solution for a further 24 h and was then paraffin-embedded for immunohistochemistry.

**Western blot.** We used the NuPAGE electrophoresis and buffer system (Invitrogen) for immunoblot analysis of kidney samples or cellular lysates as described<sup>48,49</sup>. Amersham™ Rainbow™ full-range molecular weight (MW) markers (Merck, GERPN800E) were used to determine MW and cut membranes for probing with specific antibodies. Proteins were visualized with ECL Prime (GE Healthcare). Optical densities of bands of interest were determined using ImageJ 1.46r (NIH) and normalized against loading controls. The value of the normalized control sample was arbitrarily set to 1. Membranes were stripped using the Restore Plus reagent (Fisher Scientific), and re-probed with appropriate loading controls (total ERK1/2). In cases where there was overlap in MW between loading controls and proteins of interest, loading controls were run in parallel gels. The following antibodies were used in the present study with dilutions in parentheses: From Cell Signaling Technology, Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) Antibody (1:3000) #9101, p44/42 MAPK (Erk1/2) Antibody (1:5000) #9102, Phospho-Smad3 (Ser423/425) (C25A9) Rabbit mAb (1:1000) #9520, Phospho-p70 S6 Kinase (Thr389) Antibody (1:1000) #9205, Phospho-S6 Ribosomal Protein (Ser235/236) Antibody (1:1000) #2211, Phospho-4E-BP1 (Ser65) Antibody (1:1000) #9451, Vimentin (R28) antibody (1:1000) #3932. From Sigma-Aldrich, monoclonal anti-actin,  $\alpha$ -Smooth Muscle (clone 1A4) A2547 (1:10,000), monoclonal Anti- $\alpha$ -Tubulin antibody (clone DM1A; 1:10,000) T9026, anti- $\beta$ -actin antibody (clone AC-74, 1:10,000) A2228.

**Immunohistochemical staining.** Formalin-fixed kidneys were embedded in paraffin, and 4- $\mu$ m sections were cut by the Barts Cancer Institute Pathology Unit. Staining for Sirius red,  $\alpha$ -Smooth Muscle Actin ( $\alpha$ SMA), and F4/F80 were performed as described<sup>50</sup> with the DISCOVERY XT (Ventana) automated slide processing instrument using the OmniMap reagents (Ventana), according to the manufacturer's recommendations. Images were captured at  $\times 20$  magnification, using a Zeiss AxioPhot microscope with an AxioCam HRc camera. Five kidney cortex images were captured per mouse, and staining was quantified as the percentage of the total area, using ImageJ 1.46r (NIH). The Anti-F4/F80 (Cl:A3-1) antibody (1:50) from Biorad (MCA497) was used.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The Raw shotgun sequencing data that support the findings of this study have been deposited in the European Nucleotide Archive under accession codes “PRJEB37249”, “PRJEB38742”, “PRJEB41311”, and “PRJEB46098”. The Serum NMR and urine NMR metabolome data have been uploaded to Metabolights with accession number “MTBLS3429”.

The Serum GC-MS and isotopically quantified serum metabolites (UPLC-MS/MS) that have been used in this study are deposited in MassIVE with accession numbers “MSV000088042 [<https://doi.org/10.25345/C5CV76>]” and “MSV000088043 [<https://doi.org/10.25345/C58246>]”, respectively. In adherence to EU and national privacy laws, unrestricted access to patient phenotypic data cannot be provided for this study. Interested researchers, wishing to access individual phenotypic data would need to submit argued applications to the relevant National Data Protection Agencies. These are the Danish Data Protection Agency (<https://www.datatilsynet.dk/english>) for phenotypic data from study participants recruited in Denmark, the Federal Commissioner for Data Protection ([https://www.bfdi.bund.de/EN/Home/home\\_node.html](https://www.bfdi.bund.de/EN/Home/home_node.html)) for phenotypic data from study participants recruited in Germany and the Commission Nationale Informatique & Libertés (<https://www.cnil.fr/en/home>) for phenotypic data of study participants recruited in France. Application procedures are given on the outlined websites. If such permission is granted, phenotypic data will be then made available by the corresponding authors within 5 weeks. Source data are provided with this paper.

### Code availability

No custom code or algorithm was used for the analyses conducted in this work. Code and associated phenotypic data to replicate the analyses presented in this work can be obtained by contacting the corresponding authors (please also see the Data availability statement).

### References

1. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* **19**, 55–71 (2021).
2. Agus, A., Clément, K. & Sokol, H. Gut microbiota-derived metabolites as central regulators in metabolic disorders. *Gut* **70**, 1174–1182 (2021).
3. Li, J. et al. Interplay between diet and gut microbiome, and circulating concentrations of trimethylamine N-oxide: findings from a longitudinal cohort of US men. *Gut* **71**, 724–733 (2022).
4. Romano, K. A., Vivas, E. I., Amador-Noguez, D. & Rey, F. E. Intestinal microbiota composition modulates choline bioavailability from diet and accumulation of the proatherogenic metabolite trimethylamine-N-Oxide. *mBio* **6**, e02481–14 (2015).
5. Rath, S., Heidrich, B., Pieper, D. H. & Vital, M. Uncovering the trimethylamine-producing bacteria of the human gut microbiota. *Microbiome* **5**, 54 (2017).
6. Falony, G., Vieira-Silva, S. & Raes, J. Microbiology meets big data: the case of gut microbiota-derived trimethylamine. *Annu. Rev. Microbiol.* **69**, 305–321 (2015).
7. Wang, Z. et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
8. Tang, W. H. W. et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* **368**, 1575–1584 (2013).
9. Koeth, R. A. et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.* **19**, 576–585 (2013).
10. Wang, Z. et al. Impact of chronic dietary red meat, white meat, or non-meat protein on trimethylamine N-oxide metabolism and renal excretion in healthy men and women. *Eur. Heart J.* **40**, 583–594 (2019).
11. Wu, W.-K. et al. Identification of TMAO-producer phenotype and host-diet-gut dysbiosis by carnitine challenge test in human and germ-free mice. *Gut* **68**, 1439–1449 (2019).
12. Treacy, E. Mutations of the flavin-containing monooxygenase gene (FMO3) cause trimethylaminuria, a defect in detoxication. *Hum. Mol. Genet.* **7**, 839–845 (1998).
13. Li, X. S. et al. Gut microbiota-dependent trimethylamine N-oxide in acute coronary syndromes: a prognostic marker for incident

- cardiovascular events beyond traditional risk factors. *Eur. Heart J.* ehw582 <https://doi.org/10.1093/eurheartj/ehw582> (2017).
14. Key, T. J. et al. Consumption of meat, fish, dairy products, and eggs and risk of ischemic heart disease: a prospective study of 7198 incident cases among 409 885 participants in the Pan-European EPIC Cohort. *Circulation* **139**, 2835–2845 (2019).
  15. Gibson, R. et al. The association of fish consumption and its urinary metabolites with cardiovascular risk factors: the International Study of Macro-/Micronutrients and Blood Pressure (INTERMAP). *Am. J. Clin. Nutr.* **111**, 280–290 (2020).
  16. Hoyles, L. et al. Metabolic retroconversion of trimethylamine N-oxide and the gut microbiota. *Microbiome* **6**, 73 (2018).
  17. Brunt, V. E. et al. Trimethylamine-N-oxide promotes age-related vascular oxidative stress and endothelial dysfunction in mice and healthy humans. *Hypertension* **76**, 101–112 (2020).
  18. Bennett, B. J. et al. Trimethylamine-N-oxide, a metabolite associated with atherosclerosis, exhibits complex genetic and dietary regulation. *Cell Metab.* **17**, 49–60 (2013).
  19. Stubbs, J. R. et al. Serum trimethylamine-N-oxide is elevated in CKD and correlates with coronary atherosclerosis burden. *J. Am. Soc. Nephrol.* **27**, 305–313 (2016).
  20. Tang, W. H. W. et al. Gut microbiota-dependent trimethylamine N-oxide (TMAO) pathway contributes to both development of renal insufficiency and mortality risk in chronic kidney disease. *Circ. Res.* **116**, 448–455 (2015).
  21. Manor, O. et al. A multi-omic association study of trimethylamine N-oxide. *Cell Rep.* **24**, 935–946 (2018).
  22. Zhuang, R. et al. Gut microbe-generated metabolite trimethylamine N-oxide and the risk of diabetes: a systematic review and dose-response meta-analysis. *Obes. Rev.* **20**, 883–894 (2019).
  23. Bar, N. et al. A reference map of potential determinants for the human serum metabolome. *Nature* **588**, 135–140 (2020).
  24. Wilding, J. P. H. et al. Once-weekly semaglutide in adults with overweight or obesity. *N. Engl. J. Med.* **384**, 989–1002 (2021).
  25. Shaman, A. M. et al. Effect of the glucagon-like peptide-1 receptor agonists semaglutide and liraglutide on kidney outcomes in patients with type 2 diabetes: pooled analysis of SUSTAIN 6 and LEADER. *Circulation* **145**, 575–585 (2022).
  26. MetaCardis Consortium. et al. Statin therapy is associated with lower prevalence of gut microbiota dysbiosis. *Nature* **581**, 310–315 (2020).
  27. Brial, F. et al. Human and preclinical studies of the host–gut microbiome co-metabolite hippurate as a marker and mediator of metabolic health. *Gut* **70**, 2105–2114 (2021).
  28. Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
  29. Levey, A. S. et al. National kidney foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Ann. Intern. Med.* **139**, 137 (2003).
  30. Vanholder, R., Schepers, E., Pletinck, A., Nagler, E. V. & Glorieux, G. The uremic toxicity of indoxyl sulfate and p-cresyl sulfate: a systematic review. *J. Am. Soc. Nephrol.* **25**, 1897–1907 (2014).
  31. Fromentin, S. et al. Microbiome and metabolome features of the cardiometabolic disease spectrum. *Nat. Med.* **28**, 303–314 (2022).
  32. Pelletier, C. C. et al. Elevation of trimethylamine-N-oxide in chronic kidney disease: contribution of decreased glomerular filtration rate. *Toxins* **11**, 635 (2019).
  33. Zeng, Y. et al. Gut microbiota-derived trimethylamine N-oxide and kidney function: a systematic review and meta-analysis. *Adv. Nutr. Bethesda Md.* **12**, 1286–1304 (2021).
  34. Zhou, Z. et al. Circulating trimethylamine-N-oxide and risk of all-cause and cardiovascular mortality in patients with chronic kidney disease: a systematic review and meta-analysis. *Front. Med.* **9**, 828343 (2022).
  35. Verger, E. O. et al. Dietary assessment in the Metacardis study: development and relative validity of an online food frequency questionnaire. *J. Acad. Nutr. Diet.* **117**, 878–888 (2017).
  36. Koeth, R. A. et al.  $\gamma$ -Butyrobetaine is a proatherogenic intermediate in gut microbial metabolism of L-carnitine to TMAO. *Cell Metab.* **20**, 799–812 (2014).
  37. Cocking, C., Walton, J., Kehoe, L., Cashman, K. D. & Flynn, A. The role of meat in the European diet: current state of knowledge on dietary recommendations, intakes and contribution to energy and nutrient intakes and status. *Nutr. Res. Rev.* **33**, 181–189 (2020).
  38. Ferrell, M. et al. Fecal microbiome composition does not predict diet-induced TMAO production in healthy adults. *J. Am. Heart Assoc.* **10**, e021934 (2021).
  39. Titan, S. M. et al. Metabolites related to eGFR: evaluation of candidate molecules for GFR estimation using untargeted metabolomics. *Clin. Chim. Acta* **489**, 242–248 (2019).
  40. Altara, R. et al. CXCL10 is a circulating inflammatory marker in patients with advanced heart failure: a pilot study. *J. Cardiovasc. Transl. Res.* **9**, 302–314 (2016).
  41. Forslund, S. K. et al. Combinatorial, additive and dose-dependent drug–microbiome associations. *Nature* **600**, 500–505 (2021).
  42. Bryrup, T. et al. Metformin-induced changes of the gut microbiota in healthy young men: results of a non-blinded, one-armed intervention study. *Diabetologia* **62**, 1024–1035 (2019).
  43. Schmitt, R. & Melk, A. Molecular mechanisms of renal aging. *Kidney Int.* **92**, 569–579 (2017).
  44. Gupta, N. et al. Targeted inhibition of gut microbial trimethylamine N-oxide production reduces renal tubulointerstitial fibrosis and functional impairment in a murine model of chronic kidney disease. *Arterioscler. Thromb. Vasc. Biol.* **40**, 1239–1255 (2020).
  45. Fang, Q. et al. Trimethylamine N-oxide exacerbates renal inflammation and fibrosis in rats with diabetic kidney disease. *Front. Physiol.* **12**, 682482 (2021).
  46. Duffield, J. S. Cellular and molecular mechanisms in kidney fibrosis. *J. Clin. Invest.* **124**, 2299–2306 (2014).
  47. Zhu, W. et al. Gut microbial metabolite TMAO enhances platelet hyperreactivity and thrombosis risk. *Cell* **165**, 111–124 (2016).
  48. Andrikopoulos, P., Eccles, S. A. & Yaqoob, M. M. Coupling between the TRPC3 ion channel and the NCX1 transporter contributed to VEGF-induced ERK1/2 activation and angiogenesis in human primary endothelial cells. *Cell. Signal.* **37**, 12–30 (2017).
  49. Andrikopoulos, P. et al. Endothelial angiogenesis and barrier function in response to thrombin require  $\text{Ca}^{2+}$  influx through the  $\text{Na}^{+}/\text{Ca}^{2+}$  exchanger. *J. Biol. Chem.* **290**, 18412–18428 (2015).
  50. Andrikopoulos, P. et al. The MEK inhibitor trametinib ameliorates kidney fibrosis by suppressing ERK1/2 and mTORC1 signaling. *J. Am. Soc. Nephrol.* **30**, 33–49 (2019).
  51. Chevalier, R. L., Forbes, M. S. & Thornhill, B. A. Ureteral obstruction as a model of renal interstitial fibrosis and obstructive nephropathy. *Kidney Int.* **75**, 1145–1152 (2009).
  52. Yap, C. X. et al. Autism-related dietary preferences mediate autism–gut microbiome associations. *Cell* **184**, 5916–5931.e17 (2021).
  53. Rebouche, C. J. Kinetics, pharmacokinetics, and regulation of L-carnitine and acetyl-L-carnitine metabolism. *Ann. N. Y. Acad. Sci.* **1033**, 30–41 (2004).
  54. Zhao, S. et al. The association between acylcarnitine metabolites and cardiovascular disease in Chinese patients with type 2 diabetes mellitus. *Front. Endocrinol.* **11**, 212 (2020).
  55. Chen, S. et al. Trimethylamine N-oxide binds and activates PERK to promote metabolic dysfunction. *Cell Metab.* **30**, 1141–1151.e5 (2019).
  56. Martínez-Klimova, E., Aparicio-Trejo, O. E., Tapia, E. & Pedraza-Chaverri, J. Unilateral ureteral obstruction as a model to investigate fibrosis-attenuating treatments. *Biomolecules* **9**, 141 (2019).

57. Sureshbabu, A., Muhsin, S. A. & Choi, M. E. TGF- $\beta$  signaling in the kidney: profibrotic and protective effects. *Am. J. Physiol. Ren. Physiol.* **310**, F596–F606 (2016).
  58. Suthanthiran, M. et al. Circulating transforming growth factor- $\beta$ 1 levels and the risk for kidney disease in African Americans. *Kidney Int.* **76**, 72–80 (2009).
  59. Gruppen, E. G. et al. TMAO is associated with mortality: impact of modestly impaired renal function. *Sci. Rep.* **7**, 13781 (2017).
  60. Meyer, K. A. et al. Microbiota-dependent metabolite trimethylamine N-oxide and coronary artery calcium in the coronary artery risk development in young adults study (CARDIA). *J. Am. Heart Assoc.* **5**, e003970 (2016).
  61. Tsai, C.-Y. et al. Gut microbial signatures for glycemic responses of GLP-1 receptor agonists in type 2 diabetic patients: a pilot study. *Front. Endocrinol.* **12**, 814770 (2022).
  62. Tuttle, K. R. et al. Dulaglutide versus insulin glargine in patients with type 2 diabetes and moderate-to-severe chronic kidney disease (AWARD-7): a multicentre, open-label, randomised trial. *Lancet Diabetes Endocrinol.* **6**, 605–617 (2018).
  63. Belda, E. et al. Impairment of gut microbial biotin metabolism and host biotin status in severe obesity: effect of biotin and prebiotic supplementation on improved metabolism. *Gut* **71**, 2463–2480 (2022).
  64. Yancy, C. W. et al. 2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. *Circulation* **136**, e137–e161 (2017).
  65. Branca, F., Nikogosian, H. & Lobstein, T. *The Challenge of Obesity in the WHO European Region and the Strategies for Response Summary* (World Health Organization, Regional Office for Europe, 2007).
  66. Alberti, K. G. M., Zimmet, P. & Shaw, J. The metabolic syndrome—a new worldwide definition. *Lancet* **366**, 1059–1062 (2005).
  67. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* **27**, s5–s10 (2004).
  68. Levey, A. S. et al. A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* **150**, 604 (2009).
  69. Vandeputte, D. et al. Quantitative microbiome profiling links gut community variation to microbial load. *Nature* **551**, 507–511 (2017).
  70. The InterAct Consortium. Adherence to predefined dietary patterns and incident type 2 diabetes in European populations: EPIC-InterAct Study. *Diabetologia* **57**, 321–333 (2014).
  71. Tyson, C. C. et al. Short-term effects of the DASH diet in adults with moderate chronic kidney disease: a pilot feeding study. *Clin. Kidney J.* **9**, 592–598 (2016).
  72. Jeurnink, S. M. et al. Variety in vegetable and fruit consumption and the risk of gastric and esophageal cancer in the European prospective investigation into cancer and nutrition. *Int. J. Cancer* **131**, E963–E973 (2012).
  73. Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (ACM, 2016).
  74. MacKinnon, D. P., Fairchild, A. J. & Fritz, M. S. Mediation Analysis. *Annu. Rev. Psychol.* **58**, 593–614 (2007).
  75. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. mediation: R package for causal mediation analysis. *J. Stat. Softw.* **59**, 1–38 (2014).
  76. Ho, D. E., Imai, K., King, G. & Stuart, E. A. MatchIt: nonparametric preprocessing for parametric causal inference. *J. Stat. Softw.* **42**, 1–28 (2011).
- under grant agreement HEALTH-F4-2012-305312 (METACARDIS). Assistance Publique-Hôpitaux de Paris (AP-HP) is the promoter of clinical investigation (MetaCardis). Partial funding supports to K.C. and J.A.-W. were also obtained from Leducq Foundation (TransAtlantic grant), SFN (Société Française de Nutrition), F-CRIN-FORCE network for support, INSERM via ITMO and JPI-Microdiet study and Novo Nordisk foundation (Jacobus prize). S.K.F. received support from Deutsche Forschungsgesellschaft SFB1365 (“RENOPROTECTION”) and SFB1470: “HFpEF”. P.A. and K.Ch. are recipients of a Wellcome ISSF Fellowship (ISSF204834/Z/16/Z). M.-E.D. is supported by the NIHR Imperial Biomedical Research Centre, GutsUK, Diabetes UK and by grants from the French National Research Agency (ANR-10-LABX-46, European Genomics Institute for Diabetes), from the National Center for Diabetes Precision Medicine—PreciDIAB, which is jointly supported by the French National Agency for Research (ANR-18-IBHU-0001), by the European Union (FEDER), by the Hauts-de-France Regional Council (Agreement 20001891/NP0025517) and by the European Metropolis of Lille (MEL, Agreement 2019\_ESR\_11) and by Isite ULNE (R-002-20-TALENT-DUMAS), also jointly funded by ANR (ANR16-IDEX-0004-ULNE), the Hauts-de-France Regional Council (20002845) and by the European Metropolis of Lille (MEL). This research was conducted within the context of the CNRS–Imperial International Research Project METABO-LIC. The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent research institution at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation. We thank the subjects for their participation in the MetaCardis study and particularly patient associations (Alliance du Coeur and CNAO) for their input and interface, as well as Dr Dominique Bonnefont-Rousselot (Department of Metabolic Biochemistry, Pitié-Salpêtrière hospital) for the analysis of plasma lipid profiles. We thank the nurses, technicians, clinical research assistants and data managers from the Clinical investigation platform at the Institute of Cardiometabolism and Nutrition for patient investigations, at the CRNH (Centre de recherche en Nutrition Humaine CRNH-Ile de France) and, the Clinical Investigation Center (CIC) from Pitié-Salpêtrière Hospital for investigation of healthy controls. Quanta Medical provided regulatory oversight of the clinical study and contributed to the processing and management of electronic data.

## Author contributions

P.A., K.C., and M.E.-D. developed the present project concept and protocol. K.C. (coordinator), O.P., M.S., S.D.E., P.B., J.R., M.-E.D., F.B., and J.N. conceived the overall objectives and study design of the MetaCardis initiative. MetaCardis cohort recruitment, phenotyping and lifestyle: J.A.-W., R.C., T.N., J.-E.S., F.A., L.K., H.V., T.H., G.H., R.I., J.-M.O., and M.B. and supervised by K.C., M.S., and O.P. Data curation: R.C., S.K.F., J.A.-W., and T.N. Quantitative microbiota profiling: S.V.S. and G.F. Serum and urine metabolome profiling: A.M., J.C., M.T.O., and L.H. Biochemical analyses: J.P.G. and C.R. Bioinformatics and statistical analyses: P.A., S.K.F., G.F., S.V.S., R.A., E.L.C., L.P.C., E.P., E.B., D.G., K.Ch., P.F., F.P.-C., M.C., and J.-D.Z. Dietary FFQ analyses: S.A., and B.H. Modeling of phenotypic data: P.A. Cellular in vitro experiments: P.A. Murine in vivo experiments P.A. and J.K. with supervision from M.M.Y. The manuscript was written by P.A., K.C., and M.-E.D. with input from J.A.-W., R.C., S.K.F., M.M.Y., M.S., O.P., and S.D.E. All authors participated in project development, discussion of results and revision of the article and approved the final version for submission.

## Competing interests

K.C. is a consultant for Danone Research, Ysopia, and CONFO therapeutics for work not associated with this study. K.C. held a collaborative research contract with Danone Research in the context of MetaCardis project. F.B. is a shareholder of Implexion Pharma AB. M.B. received lecture and/or consultancy fees from AstraZeneca, Boehringer-Ingelheim, Lilly, Novo Nordisk, Novartis, and Sanofi. The remaining authors declare no competing interests.

## Acknowledgements

This work was supported by European Union’s Seventh Framework Program for research, technological development, and demonstration



## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-39824-4>.

**Correspondence** and requests for materials should be addressed to Petros Andrikopoulos, Karine Clément or Marc-Emmanuel Dumas.

**Peer review information** *Nature Communications* thanks Peirong Lu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Petros Andrikopoulos<sup>1,2,39</sup>✉, Judith Aron-Wisniewsky<sup>3,4,39</sup>, Rima Chakaroun<sup>5,6,39</sup>, Antonis Myridakis<sup>1,7,39</sup>, Sofia K. Forslund<sup>8,9,10,11,12,39</sup>, Trine Nielsen<sup>13</sup>, Solia Adriouch<sup>3</sup>, Bridget Holmes<sup>14</sup>, Julien Chilloux<sup>1</sup>, Sara Vieira-Silva<sup>15,16,17,18</sup>, Gwen Falony<sup>15,16,17,18</sup>, Joe-Elie Salem<sup>19</sup>, Fabrizio Andreelli<sup>3,4</sup>, Eugeni Belda<sup>3,19</sup>, Julius Kieswich<sup>20,21</sup>, Kanta Chechi<sup>1,2</sup>, Francesc Puig-Castellvi<sup>22</sup>, Mickael Chevalier<sup>22</sup>, Emmanuelle Le Chatelier<sup>23</sup>, Michael T. Olanipekun<sup>1</sup>, Lesley Hoyles<sup>24</sup>, Renato Alves<sup>18</sup>, Gerard Helft<sup>4,25</sup>, Richard Isnard<sup>3,4</sup>, Lars Køber<sup>26</sup>, Luis Pedro Coelho<sup>8</sup>, Christine Rouault<sup>3</sup>, Dominique Gauguier<sup>27,28</sup>, Jens Peter Gøtzte<sup>29</sup>, Edi Prifti<sup>3,19</sup>, Philippe Froguel<sup>22,30</sup>, The MetaCardis Consortium\*, Jean-Daniel Zucker<sup>3,19</sup>, Fredrik Bäckhed<sup>6,13</sup>, Henrik Vestergaard<sup>13,31</sup>, Torben Hansen<sup>13</sup>, Jean-Michel Oppert<sup>4</sup>, Matthias Blüher<sup>5</sup>, Jens Nielsen<sup>32</sup>, Jeroen Raes<sup>15,16</sup>, Peer Bork<sup>8,10,33,34</sup>, Muhammad M. Yaqoob<sup>20,21,40</sup>, Michael Stumvoll<sup>18,40</sup>, Oluf Pedersen<sup>13,35,40</sup>, S. Dusko Ehrlich<sup>36,40</sup>, Karine Clément<sup>3,4,40</sup>✉ & Marc-Emmanuel Dumas<sup>1,2,22,28,40</sup>✉

<sup>1</sup>Section of Biomolecular Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK. <sup>2</sup>Section of Genomic & Environmental Medicine, National Heart & Lung Institute, Imperial College London, London, UK. <sup>3</sup>Sorbonne Université, INSERM, Nutrition and obesity; systemic approaches (NutriOmics), Paris, France. <sup>4</sup>Assistance Publique Hôpitaux de Paris, Pitié-Salpêtrière Hospital, Paris, France. <sup>5</sup>Medical Department III—Endocrinology, Nephrology, Rheumatology, University of Leipzig Medical Center, Leipzig, Germany. <sup>6</sup>The Wallenberg Laboratory, Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. <sup>7</sup>Environmental Research Group, MRC Centre for Environment and Health, School of Public Health, Imperial College London, 86 Wood Lane, London W12 0BZ, UK. <sup>8</sup>Structural and Computational Biology, European Molecular Biology Laboratory, Heidelberg, Germany. <sup>9</sup>Experimental and Clinical Research Center, a cooperation of Charité—Universitätsmedizin and the Max-Delbrück Center, Berlin, Germany. <sup>10</sup>Max Delbrück Center for Molecular Medicine (MDC), Berlin, Germany. <sup>11</sup>Charité University Hospital, Berlin, Germany. <sup>12</sup>DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, Berlin, Germany. <sup>13</sup>Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. <sup>14</sup>Danone Research, Palaiseau, France. <sup>15</sup>Laboratory of Molecular Bacteriology, Department of Microbiology and Immunology, Rega Institute, KU Leuven, Leuven, Belgium. <sup>16</sup>Center for Microbiology, VIB, Leuven, Belgium. <sup>17</sup>Institute of Medical Microbiology and Hygiene and Research Center for Immunotherapy (FZI), University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany. <sup>18</sup>Institute of Molecular Biology (IMB), Mainz, Germany. <sup>19</sup>Sorbonne Université, IRD, Unité de Modélisation Mathématique et Informatique des Systèmes Complexes, UMMISCO, F-93143 Bondy, France. <sup>20</sup>Diabetic Kidney Disease Centre, Renal Unit, Barts Health National Health Service Trust, The Royal London Hospital, London, UK. <sup>21</sup>Centre for Translational Medicine and Therapeutics, William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK. <sup>22</sup>European Genomics Institute for Diabetes, EGENODIA, INSERM U1283, CNRS UMR8199, Institut Pasteur de Lille, Lille University Hospital, University of Lille, Lille, France. <sup>23</sup>Metagenopolis, INRA, AgroParisTech, Université Paris-Saclay, Paris, France. <sup>24</sup>Department of Biosciences, School of Science and Technology, Nottingham Trent University, Nottingham, UK. <sup>25</sup>Institute of Cardiometabolism and Nutrition, ICAN, INSERM, 1166 Paris, France. <sup>26</sup>Department of Cardiology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark. <sup>27</sup>INSERM UMR 1124, Université de Paris, 45 rue des Saint-Pères, 75006 Paris, France. <sup>28</sup>McGill Genome Centre, McGill University, 740 Doctor Penfield Avenue, Montreal, QC H3A 0G1, Canada. <sup>29</sup>Department of Clinical Biochemistry, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark. <sup>30</sup>Section of Genetics and Genomics, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, UK. <sup>31</sup>Department of Medicine, Bornholms Hospital, Rønne, Denmark. <sup>32</sup>Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden. <sup>33</sup>Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany. <sup>34</sup>Yonsei Frontier Lab (YFL), Yonsei University, Seoul 03722, South Korea. <sup>35</sup>Center for Clinical Metabolic Research, Gentofte University Hospital, Copenhagen, Denmark. <sup>36</sup>Department of Clinical and Movement Neurosciences, University College London, London NW3 2PF, UK. <sup>39</sup>These authors contributed equally: Petros Andrikopoulos, Judith Aron-Wisniewsky, Rima Chakaroun, Antonis Myridakis, Sofia K. Forslund. <sup>40</sup>These authors jointly supervised this work: Muhammad M. Yaqoob, Michael Stumvoll, Oluf Pedersen, S. Dusko Ehrlich, Karine Clément, Marc-Emmanuel Dumas. \*A list of authors and their affiliations appears at the end of the paper. ✉e-mail: [p.andrikopoulos04@imperial.ac.uk](mailto:p.andrikopoulos04@imperial.ac.uk); [karine.clement@inserm.fr](mailto:karine.clement@inserm.fr); [m.dumas@imperial.ac.uk](mailto:m.dumas@imperial.ac.uk)

## The MetaCardis Consortium

Rohia Alili<sup>3</sup>, Ehm Astrid Andersson Galijatovic<sup>13</sup>, Olivier Barthelemy<sup>4</sup>, Jean-Philippe Bastard<sup>3</sup>, Jean-Paul Batisse<sup>4</sup>, Pierre Bel-Lassen<sup>3</sup>, Magalie Berland<sup>23</sup>, Randa Bittar<sup>4</sup>, Hervé Blottière<sup>23</sup>, Frederic Bosquet<sup>4</sup>, Rachid Boubrit<sup>4</sup>, Olivier Bourron<sup>4</sup>, Mickael Camus<sup>3</sup>, Cecile Ciangura<sup>4</sup>, Jean-Philippe Collet<sup>4</sup>, Arne Dietrich<sup>5</sup>, Morad Djebbar<sup>4</sup>, Angélique Doré<sup>23</sup>, Line Engelbrechtsen<sup>13</sup>, Leopold Fezeu<sup>37</sup>, Sebastien Fromentin<sup>23</sup>, Nicolas Pons<sup>23</sup>, Marianne Graine<sup>3</sup>, Caroline Grünemann<sup>5</sup>, Agnes Hartemann<sup>4</sup>, Bolette Hartmann<sup>13</sup>, Malene Hornbak<sup>13</sup>, Sophie Jaqueminet<sup>4</sup>, Niklas Rye Jørgensen<sup>29</sup>, Hanna Julienne<sup>23</sup>, Johanne Justesen<sup>13</sup>, Judith Kammer<sup>5</sup>, Nikolaj Karup<sup>13</sup>, Ruby Kozłowski<sup>1</sup>, Michael Kuhn<sup>8</sup>, Véronique Lejard<sup>23</sup>, Ivica Letunic<sup>8</sup>, Florence Levenez<sup>23</sup>, Lajos Marko<sup>10</sup>, Laura Martinez-Gili<sup>1</sup>, Robin Massey<sup>23</sup>, Nicolas Maziers<sup>23</sup>, Lucas Moitinho-Silva<sup>8</sup>, Gilles Montalescot<sup>4</sup>, Ana Luisa Neves<sup>1</sup>, Laetitia Pasero Le Pavin<sup>23</sup>, Françoise Pousset<sup>4</sup>, Andrea Rodriguez-Martinez<sup>1</sup>, Sebastien Schmidt<sup>8</sup>, Tatjana Schütz<sup>5</sup>, Lucas Silva<sup>8</sup>, Johanne Silvain<sup>4</sup>, Mathilde Svendstrup<sup>13</sup>, Timothy D. Swartz<sup>3</sup>, Thierry Vanduyvenboden<sup>23</sup>, Eric O. Verger<sup>38</sup> & Stefanie Walther<sup>5</sup>

<sup>37</sup>Nutritional Epidemiology Research Team (EREN), Centre of Research in Epidemiology and Statistics Sorbonne Paris Cité, Inserm (U1153), Inra (U1125), Cnam, Paris 13 University, COMUE Sorbonne Paris Cité, 93017 Bobigny, France. <sup>38</sup>MolSA, Univ Montpellier, CIHEAM-IAMM, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France.