Title: Evidence of a causal and modifiable relationship between kidney function and circulating trimethylamine *N*-oxide with implications for heart and kidney disorders.

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

Affiliations

[1]Section of Biomolecular Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK

[2]Section of Genomic & Environmental Medicine, National Heart & Lung Institute, Imperial College London, London, UK

[3]Sorbonne Université, INSERM, Nutrition and obesities; systemic approaches (NutriOmics), Paris, France

[4]Assistance Publique Hôpitaux de Paris, Pitie-Salpêtrière Hospital, Paris France

[5]Medical Department III - Endocrinology, Nephrology, Rheumatology, University of Leipzig Medical Center, Leipzig, Germany

[6]The Wallenberg Laboratory, Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[7]Department of Surgery and Cancer, Imperial College London, London W12 0HS, United Kingdom

[8]Structural and Computational Biology, European Molecular Biology Laboratory, Heidelberg, Germany

[9]Experimental and Clinical Research Center, a cooperation of Charité-Universitätsmedizin and the Max-Delbrück Center, Berlin, Germany

[10]Max Delbrück Center for Molecular Medicine (MDC), Berlin, Germany

[11]Charité University Hospital, Berlin, Germany

[12]DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, Berlin, Germany

[13]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[14]Danone Research, Palaiseau France

[15]Laboratory of Molecular Bacteriology, Department of Microbiology and Immunology, Rega Institute, KU Leuven, Leuven, Belgium

[16]Center for Microbiology, VIB, Leuven, Belgium

[17]Sorbonne Université, IRD, Unité de Modélisation Mathématique et Informatique des Systèmes Complexes, UMMISCO, F-93143, Bondy, France

[18]Diabetic Kidney Disease Centre, Renal Unit, Barts Health National Health Service Trust, The Royal London Hospital, London, UK

[19]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

[20]European Genomics Institute for Diabetes, EGENODIA, INSERM U1283, CNRS UMR8199, Institut Pasteur de Lille, Lille University Hospital, University of Lille, Lille, France

[21]Metagenopolis, INRA, AgroParisTech, Université Paris-Saclay, Paris, France

[22]Department of Bioscience, School of Science and Technology, Nottingham Trent University, Nottingham, UK

[23]Institute of Cardiometabolism and Nutrition, ICAN, INSERM, 1166, Paris, France

[24]Department of Cardiology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

[25]Université de Paris Cité, INSERM UMR 1124, Paris, France

[26]Department of Medicine, Bornholms Hospital, Rønne, Denmark

[27]Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

[28]Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany.

[29]Yonsei Frontier Lab (YFL), Yonsei University, Seoul 03722, South Korea.

[230]Center for Clinical Metabolic Research, Gentofte University Hospital, Copenhagen, Denmark

[31]Department of Clinical and Movement Neurosciences, University College London, London NW3 2PF, UK

[32]McGill Genome Centre, McGill University and Genome Québec Innovation Center, Montréal, Quebec, Canada.

***Joint first authors.**

[#]Joint Senior authors.

[+]Corresponding authors

68 **Correspondence to** Petros Andrikopoulos (p.andrikopoulos04@imperial.ac.uk), Karine Clément

69 (karine.clement@inserm.fr) or Marc-Emmanuel Dumas (m.dumas@imperial.ac.uk)

70

Abstract

Objectives: The host-microbiota co-metabolite trimethylamine *N*-oxide (TMAO) is linked to increased thrombotic and cardiovascular risks. Here we, sought to i) characterize which host variables contribute to fasting serum TMAO levels in real-life settings ii) identify potential actionable therapeutic means related to circulating TMAO.

Design: We applied "explainable" machine learning, univariate-, multivariate- and mediation analyses of fasting plasma TMAO concentration and a multitude of bioclinical phenotypes in 1,741 adult Europeans of the MetaCardis study. We expanded and validated our epidemiological findings in mechanistic studies in human renal fibroblasts and a murine model of kidney fibrosis following TMAO exposure.

Results: Next to age, kidney function was the primary variable predicting circulating TMAO in MetaCardis, with microbiota composition and diet playing minor, albeit significant roles. Mediation analysis revealed a causal relationship between TMAO and kidney function decline that strengthened at more severe stages of cardiometabolic disease. We corroborated our findings in preclinical models where TMAO exposure augmented conversion of human renal fibroblasts into myofibroblasts and increased kidney scarring *in vivo*. Mechanistically, TMAO aggravated kidney fibrosis due to ERK1/2 hyperactivation synergistically with TGF-β1 signaling. Consistent with our findings, patients receiving next-generation glucose-lowering drugs with reno-protective properties, had significantly lower circulating TMAO when compared to propensity-score matched control individuals.

Conclusion: After age, kidney function is the major determinant of fasting circulating TMAO in adults. Our findings of lower TMAO levels in individuals medicated with reno-protective anti-diabetic drugs suggests a clinically actionable intervention for decreasing TMAO-associated excess cardiovascular risk that merits urgent investigation in human trials.

Data availability statement

Raw shotgun sequencing data that support the findings of this study have been deposited in the European Nucleotide Archive with accession codes PRJEB37249, PRJEB38742, PRJEB41311 and PRJEB46098. Serum NMR and urine NMR metabolome data have been uploaded to Metabolights with accession number MTBLS3429; serum GC-MS and isotopically quantified serum metabolites (UPLC–MS/MS) are available from MassIVE with accession numbers MSV000088042 and MSV000088043, respectively.

Introduction

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[1]. 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[1,2].

Trimethylamine *N*-oxide (TMAO) is the phase-one liver *N*-oxide of trimethylamine (TMA). TMA is a product of the microbial[3,4,5] (predominately Firmicutes) metabolism of phosphatidylcholine[6,7], choline[7] and *L*-carnitine[8,9,10], components of the high-fat, high-red meat western diet. TMA is taken up from the gut via the hepatic portal vein and *N*-oxidised into TMAO by host flavin mono-oxygenase 3. 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[6,7,8,11]. TMAO is therefore proposed to mediate the higher cardiovascular risk associated with high red meat and fat intake[12]. Fish, the consumption of which is associated with reduced incidence of cardiovascular disease[12], is also a rich source of TMAO[13]. 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[14]. Concurrent with diet and microbiota composition, TMAO plasma levels also reflect age[15], sex[16], kidney function[17,18,19] and chronic diseases[5,20]. 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 in real-life settings 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[21], 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 multi-center MetaCardis study. Moreover, by taking advantage of the unique cross-sectional MetaCardis design, we 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 (**Suppl.Fig.1**).

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. Interestingly, patients with T2D in the cohort prescribed new-generation anti-diabetics (GLP-1 Receptor Agonists[22]; GLP-1RAs) with evidenced reno-protective effects[23] had lower serum circulating TMAO levels when compared to propensity-score matched controls.

Results

In a 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[4,20] (**Suppl.Fig.2A**). 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[23]; N=837) comprising obese/overweight individuals presenting with a range of metabolic syndrome features but not overt T2D or ischaemic heart disease (IHD) (**Suppl.Table.1**).

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

We explored how TMAO correlated with bioclinical 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, pFDR=0.06) and higher fasting concentrations of uric acid (rho=0.114, pFDR=0.09) after adjustment for age, sex, country of recruitment (demographics thereafter) and body mass index (BMI) (**Suppl.Fig.2B, Suppl.Table.2**). Higher TMAO also positively associated with indicators of central adiposity, including BMI (rho=0.107, pFDR=0.09), visceral body fat rating (rho=0.122, pFDR=0.09) and waist circumference (rho=0.119, pFDR=0.09), after demographics adjustment (**Suppl.Fig.2C**, **Suppl.Table.3)**. Moreover, BMIS individuals with hypertension (systolic blood pressure over 140mmHg, or receiving therapy for high blood pressure; Methods) had higher plasma TMAO (*P*=0.01, Mann-Whitney test; **Suppl.Fig.2D**). In line with previous studies[13,14,20], objectively dividing BMIS participants into TMAO clusters using k-means[25] 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 (**Suppl.Fig.2E, Suppl.Table.4)**. Traits of cardiometabolic risk included altered eGFR, elevated liver enzymes and systolic blood pressure (**Suppl.Fig.2F-I**).

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

To better understand which variables (**Suppl.Table.5)** affect circulating TMAO most, we trained extreme gradient-boosted decision-tree models. We used five-fold cross-validation to predict the Explained Variance (EV) of each variable group on plasma TMAO in the left-out BMIS participants after 100 iterations (**Figure 1A, Suppl.Table.6**). 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 (**Suppl.Table.7**). The full model, containing all variable groups, accounted, on average, for 18.4% of TMAO variability. The averaged predicted TMAO values by the full model significantly correlated with measured TMAO values (rho=0.473, $P<2.2x10^{-16}$; **Suppl.Fig.2J, Suppl.Table.8**).

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 (**Figure 1B**). Other variable categories displayed negligible contribution to prediction suggesting considerable information overlap with the metabolomic, metagenomic, biological and dietary datasets (**Suppl.Table.9**).

Using feature attribution analysis (SHapley Additive exPlanations; SHAP[21,26]), 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 (**Figure 1C, Suppl.Table.10**). Besides eGFR, additional variables

191 associated with kidney function including plasma urea[27], the uremic toxin *p*-cresol[28] or kidney
192 function decline consequences: i.e. serum albumin[27] were among those affecting model outcomes
193 the most. This analysis suggests that kidney function is a major determinant of circulating TMAO.
194 The impact of eGFR on model outcomes for BMIS individuals was bimodal with values over 90
195 mL/min/1.73m^2, the clinical cut-off value for normal kidney function in adults[25], predicting reduced
196 plasma TMAO and lower values resulting in increased predicted circulating TMAO (**Figure 1D**).

197 We next trained algorithms predicting TMAO using the 24 variables identified by our SHAP
198 analysis ("top SHAP" model) and compared it to models trained by traditional clinical risk factors[29] or
199 the full model (**Figure 1E**). The "top SHAP" model significantly ($P<2.2x10^{-16}$, Mann-Whitney test)
200 improved predictions when compared to the full model (EV 21% vs. 18%, on average), presumably
201 by removing noise, supporting the importance of the variables identified by the SHAP analysis.

202 **Diet and microbiota composition have a modest impact on fasting serum TMAO levels in the BMIS**
203 **cohort**

204 Our ML models suggested that diet plays a modest role in determining circulating TMAO. To
205 corroborate this, we assessed how consumption of food items rich in TMAO dietary precursors[3]
206 affected circulating TMAO in BMIS participants with a Food Frequency Questionnaire (FFQ; N=763).
207 After correcting for demographics and BMI, with the exception of oily fish (rho=0.125, pFDR=0.03),
208 circulating TMAO neither significantly associated with habitual consumption of red meat, eggs or
209 milk (**Suppl.Table.11)** nor with the estimated dietary intake of the micronutrients choline, carnitine
210 or betaine **(Suppl.Table.12)**, broadly in agreement with a previous study[3] (**Figure 2A)**. Additionally,
211 no significant correlations were found between consumption of these food items and the serum
212 levels of TMAO precursors choline, betaine or γ-butyrobetaine[30]. Conversely, circulating TMAO
213 positively correlated with plasma choline, betaine and γ-butyrobetaine, implying an association
214 between the serum concentrations of TMAO precursors.

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; **Figure 2B**). Consistent with previous reports[3,31] and our ML models, multivariate PERMANOVA analysis uncovered a significant, albeit weak, association between circulating TMAO and microbiota composition (*P*=0.001; R^2 = 0.009; **Figure 2B**) after demographics adjustment. TMAO levels significantly associated with absolute 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 (**Suppl.Table.13**, **Figure 2C**). In agreement with Li and colleagues[3], we did not identify any overlap between bacteria associated with circulating TMAO and higher red meat, milk or eggs consumption in BMIS participants (N=761; **Figure 2C, Suppl.Table.13)**. Contrasting low and high TMAO clusters revealed 215 differentially abundant bacterial species between these two groups (**Figure 2D**, **Suppl.Table.14**).

We further analyzed the impact on circulating TMAO of the only species, an unknown bacterium taxonomically closely related to ruminoccoci (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 (**Figure 2E**). CAG01909 was more abundant and prevalent in the high TMAO cluster when compared to the low (**Suppl.Fig.3A-B**). Finally, BMIS individuals harboring CAG01909 had significantly higher circulating TMAO than those that lacked it (**Suppl.Fig.3C**), the CAG1909 abundance significantly correlating with TMAO concentration (**Suppl.Fig.3D**).

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 (**Figure 3A and Suppl.Fig.4A** 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 (**Suppl.Table.17)**. 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 (**Suppl.Fig.4A, Suppl.Table.18)**, 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 (**Figure 3B and Suppl.Fig.4B** respectively**; Suppl.Table.19, Suppl.Table.20**). In patients with T2D, model outcome was mostly affected by eGFR followed by age and the serum concentrations of the uremic toxin *p*-cresol[28] and *D*-threitol, also indicative of kidney function[32]. In patients with IHD, serum butyryl-carnitine, 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 remodelling[33] were the top variables. As in BMIS, the "top SHAP" models significantly improved predictions (EV 24% *versus* 16.2% for T2D; **Figure 3C** and EV 21.3% *versus* 16.1% for IHD; **Suppl.Fig.4C**), 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 (**Figure 3D**).

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 fit 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 (**Suppl.Fig.4D**), 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* abundance and circulating TMAO (**Figure 3B**), whilst no taxa strongly influenced TMAO predictions in the IHD sub-cohort (**Suppl.Fig.4B**). *R. timonensis* inversely associated with TMAO concentration (rho=-0.140, *P*=0.0009; **Suppl.Fig.5A**), T2D individuals with detectable *R. timonensis* had significantly lower TMAO levels (**Suppl.Fig.5B**) and *R. timonensis* was depleted in T2D compared to BMIS participants (**Suppl.Fig.5C**), in line with circulating TMAO (**Suppl.Fig.2A**). Conversely, in T2D and IHD CAG01909 abundance or presence, unlike BMIS, did not associate with higher circulating TMAO (**Suppl.Fig.5D-5F**). Further investigation revealed that CAG01909 abundance was inversely associated with metformin intake (rho=-0.141, pFDR=0.042; **Suppl.Fig.5G, Suppl.Table.21**), in accordance with the well-documented effect of drugs on the gut microbiome[34]. 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[35]. 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%, $P<2.2x10^{-16}$ and 7.1%, $P<2.2x10^{-16}$; **Figure 3E**). 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[19] and with studies in animal models of chronic kidney disease (CKD) where TMAO diet-supplementation increased kidney injury[36,37] 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-β1 signaling

Based on TMAO's detrimental association with kidney damage of diverse aetiologies[17,18] similarly to renal fibrosis[38], we investigated the impact of TMAO on human primary renal fibroblasts (HRFs). In HRFs, unlike platelets[39], TMAO stimulation resulted in rapid $[Ca^{2+}]_i$ increase (**Figure 4A**). In endothelial cells the ERK1/2 pathway is activated in a Ca^{2+}-dependent manner[40,41] and ERK1/2 activation exacerbates renal fibrosis[42]. Therefore, we investigated whether a similar pathway operates in HRFs. TMAO challenge increased phospho-ERK1/2 levels in a time- (**Figure 4B**) and dose-dependent (**Figure 4C**, **Suppl.Fig6A**) manner, at concentrations (10-100μM) relevant to human disease (up to 150μM in patients with CKD)[17]. TMAO-induced ERK1/2 activation was inhibited by a

MEK inhibitor, suggesting the activation occurs downstream of Ras-Raf-MEK[42]. Additionally, ERK1/2 activation in response to TMAO was suppressed when intracellular Ca^{2+} was chelated (**Suppl.Fig.6B**) or extracellular Ca^{2+} was removed (**Suppl.Fig.6C**). Moreover, Ca^{2+} influx was sufficient to activate ERK1/2 in HRFs (**Suppl.Fig.6D**); indicating that TMAO-induced 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-β1[42]. Unlike short-term application (up-to 30min), TMAO stimulation for 48h had minimal effect on ERK1/2 or SMAD3 phosphorylation (**Figure 4D, Suppl.Fig.6E-F**) or on the expression of the myofibroblast marker αSMA (**Figure 4E, Suppl.Fig.6G**). Conversely, TMAO dose-dependently augmented ERK1/2 pathway activation and myofibroblast trans-differentiation, when co-administered with TGF-β1 in comparison to TGF-β1 or TMAO alone, without affecting SMAD3 phosphorylation (**Figure 4D-E, Suppl.Fig.6E-F**).

TMAO increases renal fibrosis: evidence from an intervention in 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 co-morbidities[43] that could be also impacted by TMAO. Mice were fed a choline (1% w/w)- or TMAO (0.12% w/w)-supplemented diet for 6 weeks[18]. Subsequently, the ureter of one kidney was ligated whilst the other kidney remained unobstructed (**Suppl.Fig.7A**). In the unobstructed (control) kidneys αSMA staining was not affected by diet. As expected, five days of UUO resulted in a significant increase in renal αSMA staining of the injured kidney. However, unlike the unobstructed kidneys, TMAO or choline diet supplementation resulted in significantly more myofibroblast expansion (**Figure 4F-G**). Western blotting of kidney lysates corroborated augmented αSMA and vimentin (another marker of myofibroblasts[42]) expression in injured kidneys of mice receiving TMAO or choline diets (**Figure 4H-J**). Similar to αSMA, collagen deposition or macrophage infiltration, another indicator of fibrotic kidney damage[38], were not affected by the TMAO or choline diets in the unobstructed kidneys (**Suppl.Fig.7B-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 signalling[38,42] (**Suppl.Fig.7F**).

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-β1-mediated SMAD3 pathway causing a secondary hit in our disease model.

Glucagon-Like Peptide-1 Receptor analogues (GLP-1RAs) intake associates with lower serum TMAO concentration in MetaCardis T2D participants.

Given the strong bi-directional 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 kidney function, using prescribed medication as input. SHAP analysis revealed that anti-hypertensive and anti-cholesterol treatments had a negative impact on predicted eGFR, probably reflecting more advanced disease (**Figure 5A**). Conversely, anti-diabetic treatments had a positive effect, with GLP-1RAs being the drugs with the biggest positive impact on predicted eGFR (**Figure 5A, Suppl.Table.22**). In accordance with their documented reno-protective effect even in T2D patients with no preexisting nephropathy[23], patients with T2D receiving GLP-1RAs had significantly lower circulating TMAO (**Figure 5B**) than individuals matched for age, sex, hypertension and disease group (**Suppl.Fig.8A-E, Suppl.Table.23**). 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-β1 signaling. Consistent with our findings, use of reno-protective drugs, GLP-1RAs[23], was associated with lower circulating TMAO in MetaCardis participants with T2D (**Figure 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[3,31]. We replicated associations of Firmicutes[3,4,5] with circulating TMAO and uncovered novel associations between an unknown bacterium (CAG01909) and *R. timonensis,* a marker of less diverse diet in autistic children[44]. 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[3], we did not find any significant association between habitual consumption of red meat and fasting serum levels of TMAO. With a few notable exceptions[3], 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[8,9,10]. 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%)[45], 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 by Li and colleagues[3], suggests that in real-life settings where individuals habitually consume meat (75 to 233 g/day in European adults[46]), this contributes minimally to fasting circulating TMAO variability, possibly limiting the isolated effect of dietary manipulation on TMAO levels in real-life, 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 as disease progressed 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[29]. Accumulation of butyryl-carnitine has been associated with abnormal mitochondrial lipid β-oxidation in T2D[47], 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[48].

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 pre-clinical work uncovered that TMAO primed renal fibroblasts for conversion to myofibroblast, the primary collagen-producing cells in the kidney[38] and contributed to renal fibrosis, a hallmark of kidney damage irrespective of the underlying cause that significantly contributes to eGFR decline[38]. However, TMAO insult alone was not sufficient to convert renal fibroblasts to myofibroblasts requiring synergy with a pre-existing pathological state, *i.e.*, pro-fibrotic signaling ("second-hit") in the form of TGF-β1 stimulation, the most prominent pro-fibrotic cytokine[38]. TGF-β1 expression is increased in ligated UOO mouse kidneys[49] or kidneys of patients with diabetic nephropathy[50] and circulating TGF-β1 is elevated in conditions that are risk factors for CKD including hypertension, dyslipidemia and T2D[51]. Therefore, our results suggest that TMAO accelerates eGFR decline in concert with existing pro-fibrotic signaling (i.e. TGF-β1; "second-hit"), consistent with observations in animal models of CKD[36,37] and humans[18,19].

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/1.73m2[52] and circulating TMAO levels in healthy adults were not indicative of future atherosclerotic burden[53].

TMAO is an independent risk factor of cardiovascular morbidity and mortality in patients with established IHD or CKD[6,17,18] with accumulating evidence suggesting direct causal effects[40]. 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-β1 signaling (i.e hypertension and T2D[51]). Indeed, we observed that patients with T2D prescribed GLP-1RAs, drugs with documented reno-protective and beneficial cardiovascular effects[22,23], had significantly lower fasting plasma TMAO levels than propensity-score matched controls.

Conclusion

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 observed reno-protection in large pharmaceutical trials[23].

Our work, conceptually advances understanding of how TMAO levels (and therefore associated cardiovascular risk) are regulated in humans with a wide range of Cardiometabolic disease burden in real-life settings. Additionally, we uncover a direct mechanistic link between TMAO and renal fibrosis in conjunction with existing co-morbidities (known to elevate TGF-β1 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.

Author contribution

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., 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. Bacterial cell count measurement: 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. Bioinformatics and statistical analyses: P.A., S.K.F., G.F, S.V.S., R.A., E.L.C., L.P.C., E.P. and E.B. Dietary FQQ analyses: S.A., and B.H. Modelling 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., KC 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.

Sources of funding

This work was supported by European Union's Seventh Framework Program for research, technological development and demonstration under grant agreement HEALTH-F4-2012-305312 (METACARDIS). Assistance Publique-Hôpitaux de Paris (AP-HP) is the promoter of the 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 (Jacobaus prize). S.K.F. received support from Deutsche Forschungsgesellschaft SFB1365 ("RENOPROTECTION") and SFB1470: "HFpEF". P.A. is an ISSF Fellow supported by the Wellcome Trust and Imperial College London (Grant No. PSN104). K.Ch. is an ISSF Fellow supported by the Wellcome Trust and Imperial College London (Grant No. PSM279). M.-E.D. is supported by the NIHR Imperial Biomedical Research Centre, GutsUK, Diabetes UK and by grants from the French National

462 Research Agency (ANR-10-LABX-46, European Genomics Institute for Diabetes), from the National
463 Center for Diabetes Precision Medicine – PreciDIAB, which is jointly supported by the French
464 National Agency for Research (ANR-18-IBHU-0001), by the European Union (FEDER), by the Hauts-
465 de-France Regional Council (Agreement 20001891/NP0025517) and by the European Metropolis of
466 Lille (MEL, Agreement 2019_ESR_11) and by Isite ULNE (R-002-20-TALENT-DUMAS), also jointly
467 funded by ANR (ANR16-IDEX-0004-ULNE), the Hauts-de-France Regional Council (20002845) and by
468 the European Metropolis of Lille (MEL). This research was conducted within the context of the
469 CNRS–Imperial International Research Project METABO-LIC. The Novo Nordisk Foundation Center for
470 Basic Metabolic Research is an independent research institution at the University of Copenhagen
471 partially funded by an unrestricted donation from the Novo Nordisk Foundation.

472 **Acknowledgements**

473 We thank the subjects for their participation in the MetaCardis study and particularly patient
474 associations (Alliance du Coeur and CNAO) for their input and interface, as well as Dr Dominique
475 Bonnefont-Rousselot (Department of Metabolic Biochemistry, Pitié-Salpêtrière hospital) for the
476 analysis of plasma lipid profiles. We thank the nurses, technicians, clinical research assistants and
477 data managers from the Clinical investigation platform at the Institute of Cardiometabolism and
478 Nutrition for patient investigations, at the CRNH (Centre de recherche en Nutrition Humaine CRNH-
479 Ile de France) and, the Clinical Investigation Center (CIC) from Pitié-Salpêtrière Hospital for
480 investigation of healthy controls. Quanta Medical provided regulatory oversight of the clinical study
481 and contributed to the processing and management of electronic data. Similarly, we are indebted to
482 the MetaCardis consortium (http://www.metacardis.net/) collaborators for contributions at multiple
483 levels since the consortium start in 2012. A full list of collaborators is given in the Supplementary
484 Section.

485 **Competing Interests.** KC is a consultant for Danone Research, Ysopia and CONFO therapeutics for
486 work not associated with this study. KC held a collaborative research contract with Danone Research
487 in the context of MetaCardis project. FB is a shareholder of Implexion pharma AB. MB received

488 lecture and/or consultancy fees from AstraZeneca, Boehringer-Ingelheim, Lilly, Novo Nordisk,
489 Novartis and Sanofi.

Main Figures

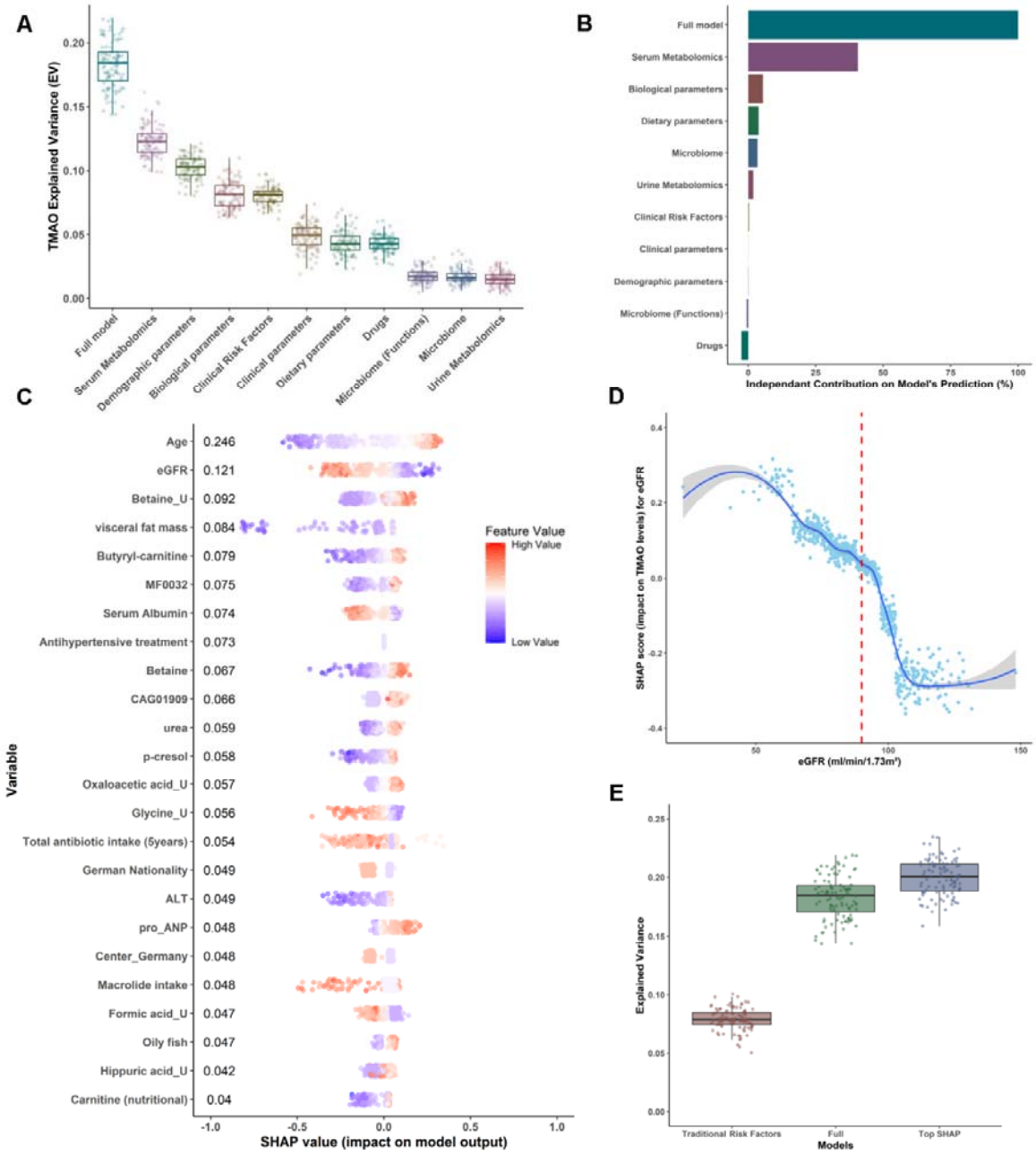
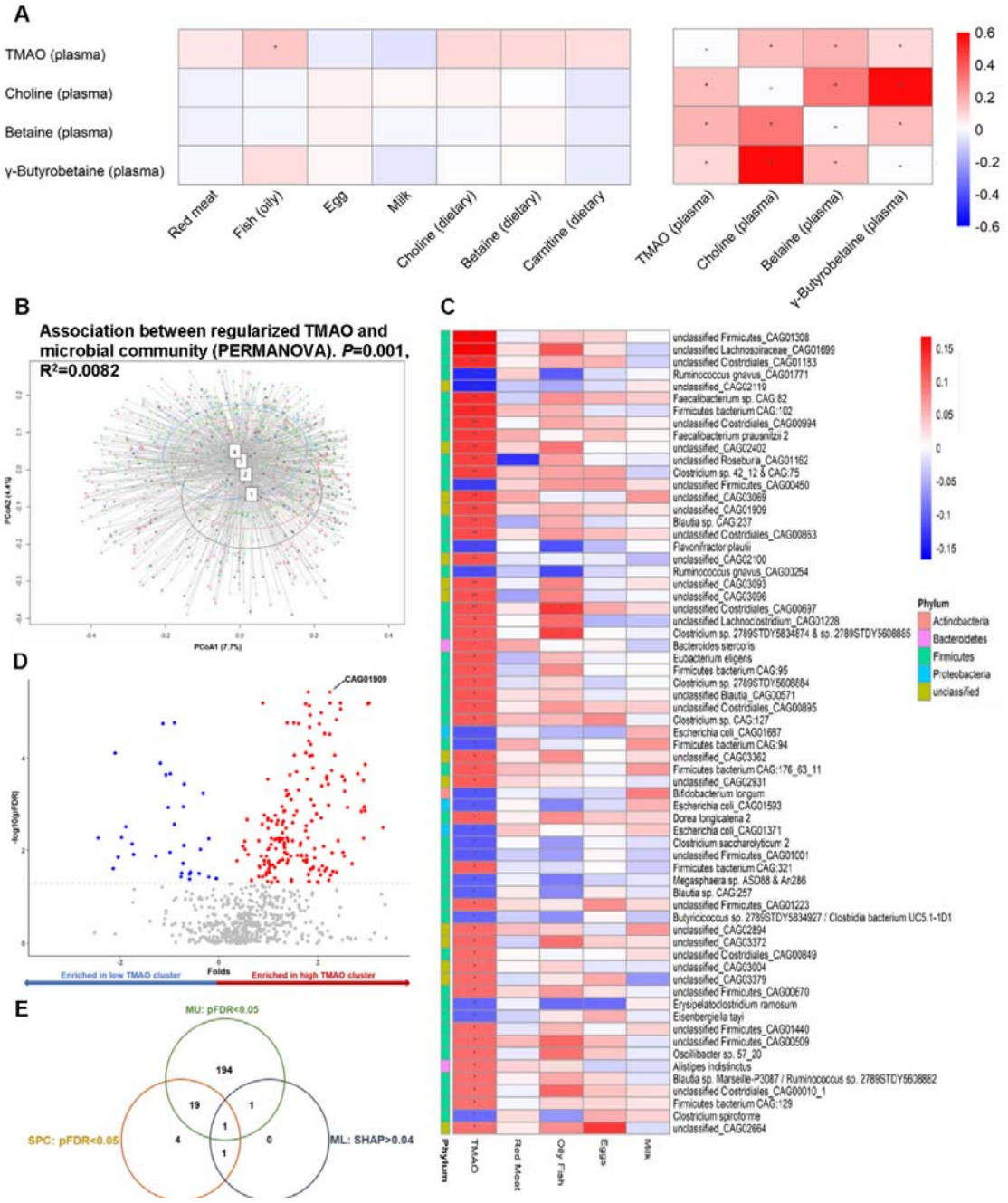


Figure 1. Age and parameters associated with kidney function are the main drivers of circulating TMAO in BMIS MetaCardis participants. (**A**) Coefficients of determination (Explained Variance; EV) of predicted circulating TMAO levels determined by xgboost algorithms after 5-fold cross-validation in the left-out group (**Suppl.Table.6** for N numbers and optimized xgboost parameters per group), trained exclusively on variables from each feature category (**Suppl.Table.5** for a list of variables

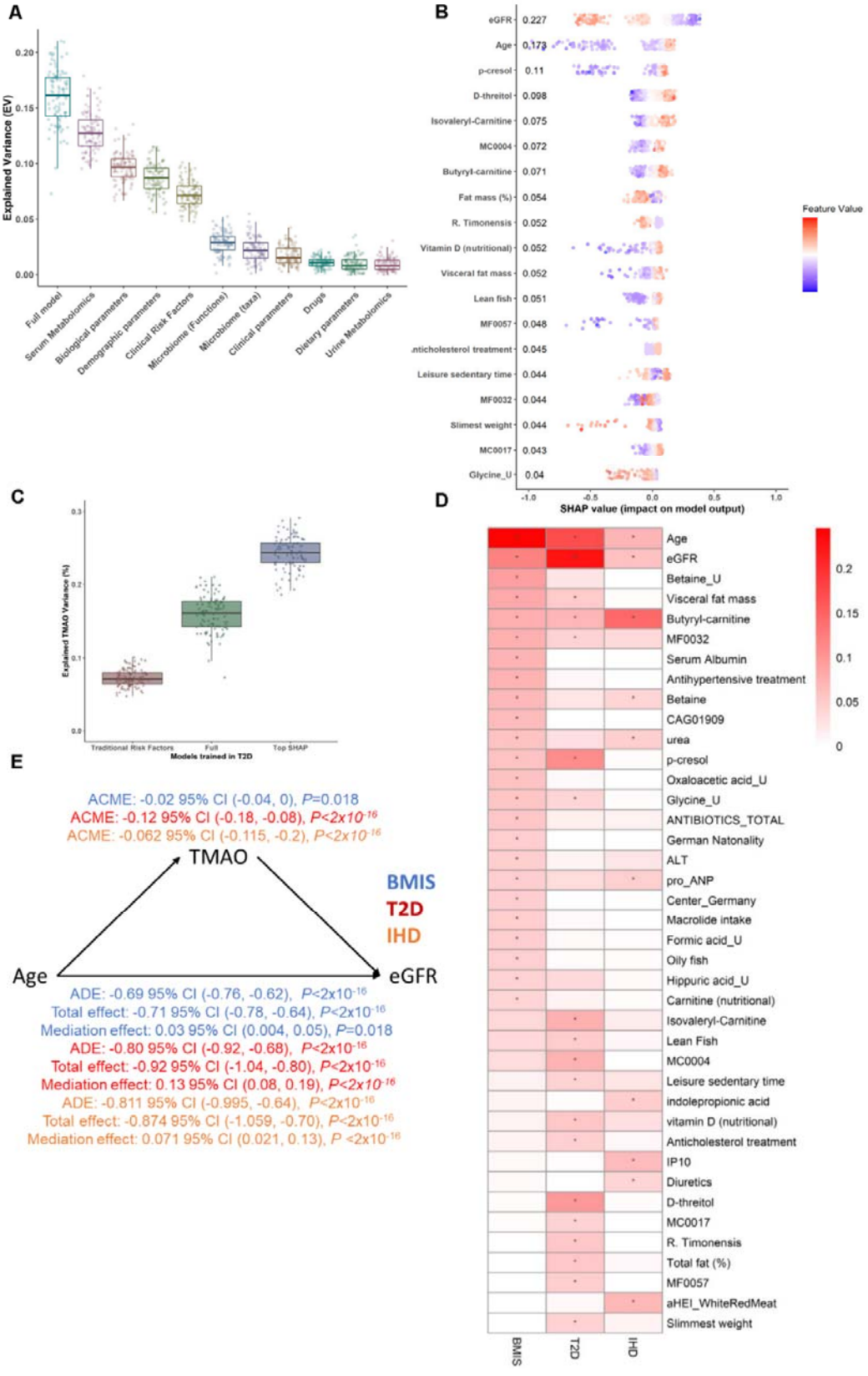
included in each group), or the full model (containing all variables), after 100 iterations (**Suppl.Table.7**). (**B**) Averaged independent predictive contribution of each feature category to full model predictions of plasma TMAO, trained as in (**A**), calculated as the average reduction of the coefficient of determination achieved in relation to the full model (set to 100%) after removing all variables belonging to each feature group after 100 iterations (**Suppl.Table.8**). (**C**) Swarm plots of impact on model output (SHAP values; **Suppl.Table.10**) 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**) Dependance plot of eGFR values (*x-axis*) *versus* their impact on model outcome (*y-axis*, SHAP values) calculated for each individual in the BMIS cohort (N=837) from algorithms trained exclusively on bioclinical variables **(Suppl.Table.5**), vertical red line indicates 90mL/min/1.73m^2. The curve was drawn using locally weighted scatterplot smoothing (LOWESS). (**E**) Boxplots depicting Explained Variance (EV; R^2) of circulating TMAO for BMIS participants computed from algorithms trained on clinical risk factors[29], the full model containing all variables or all 24 variables contributing more than 4% of regularized TMAO standard deviation to model predictions, as determined by SHAP analysis, after 100 iterations.



516

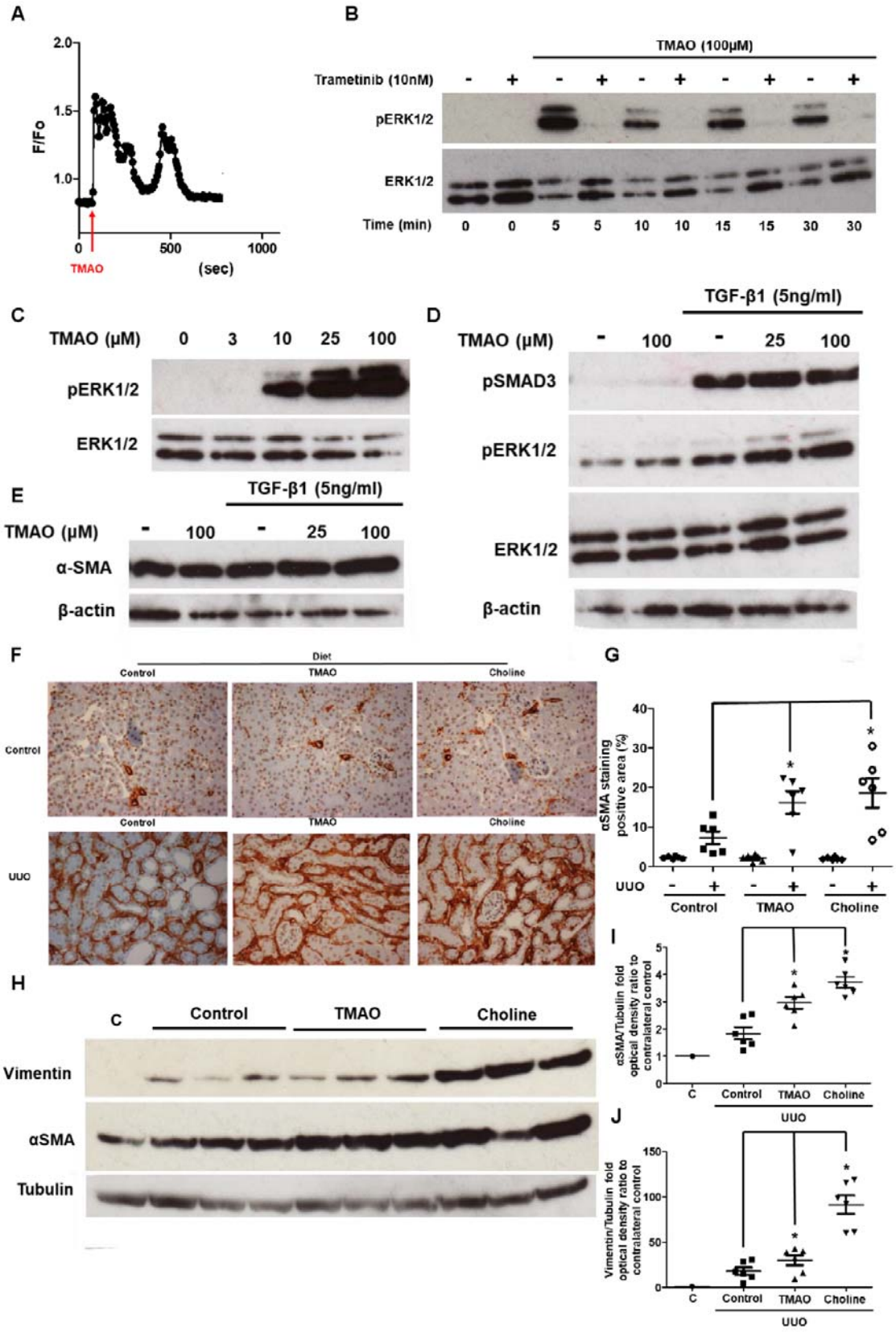
517 **Figure 2. Modest impact of diet and microbiota composition on circulating TMAO in BMIS**
518 **MetaCardis subjects.** (**A**) Associations between circulating TMAO and its precursors with habitual
519 consumption of food items rich in TMAO precursors (N=763; left panel) or with its precursors
520 themselves (right panel) (**Suppl.Table.11, Suppl.Table.12**). (**B**) Principal Coordinates analysis of Bray-
521 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; **Suppl.Table.13**). (**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, pFDR<0.05; **Suppl.Table.14**). (**E**) Venn diagram summarizing the overlap between taxa associating with circulating TMAO according to our three complimentary analyses (SPC: Spearman correlations; ML: Machine learning and feature attribution analysis; MU: Mann-Whitney U test between high and low TMAO clusters). For all *pFDR<0.05, **pFDR<0. 0.01.



534

Figure 3. 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 (**Suppl.Table.17**) determined by boosted decision trees (**Suppl.Table.15** for N numbers and optimized xgboost parameters per variable group), trained exclusively on variables from each variable category (**Suppl.Table.5**, 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; **Suppl.Table.19**) 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[31], 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).



557

Figure 4. TMAO promotes myofibroblast differentiation and exacerbates renal fibrotic injury. (**A**) Representative ratiometric traces (340/380nm) from Human Renal Fibroblasts (HRFs) loaded with the Ca^{2+} indicator Fura-2 and stimulated with 100μM TMAO. (**B**) Serum-starved HRFs were pre-incubated with the MEK inhibitor trametinib (10nM) for 30min prior to stimulation with 100μM 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 30min and stimulated with TGF-β1 (5nM) or vehicle for 24h. Phospho-ERK1/2, phospho-SMAD3 (**D**) and alpha-smooth muscle actin (αSMA) (**E**) levels were probed with Western blot. (**F**) Immunostaining with αSMA of kidney sections (20x magnification) from obstructed (UUO; 5days 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 6weeks prior to surgery, as indicated. N=6 per group. (**G**) Quantification of positive αSMA staining as (%) of stained area/field of view averaged from 5 images per animal. (**H**) Western blot of whole-kidney lysates for αSMA and vimentin expression. Membranes were subsequently stripped and re-probed for tubulin, as loading control. A representative photomicrograph from n=2 Western blots with n=2–6 animals in each group is shown. OD of the (**I**) αSMA 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±SEM of data from n=4–6 animals per group. *P<0.05 versus the UUO control.

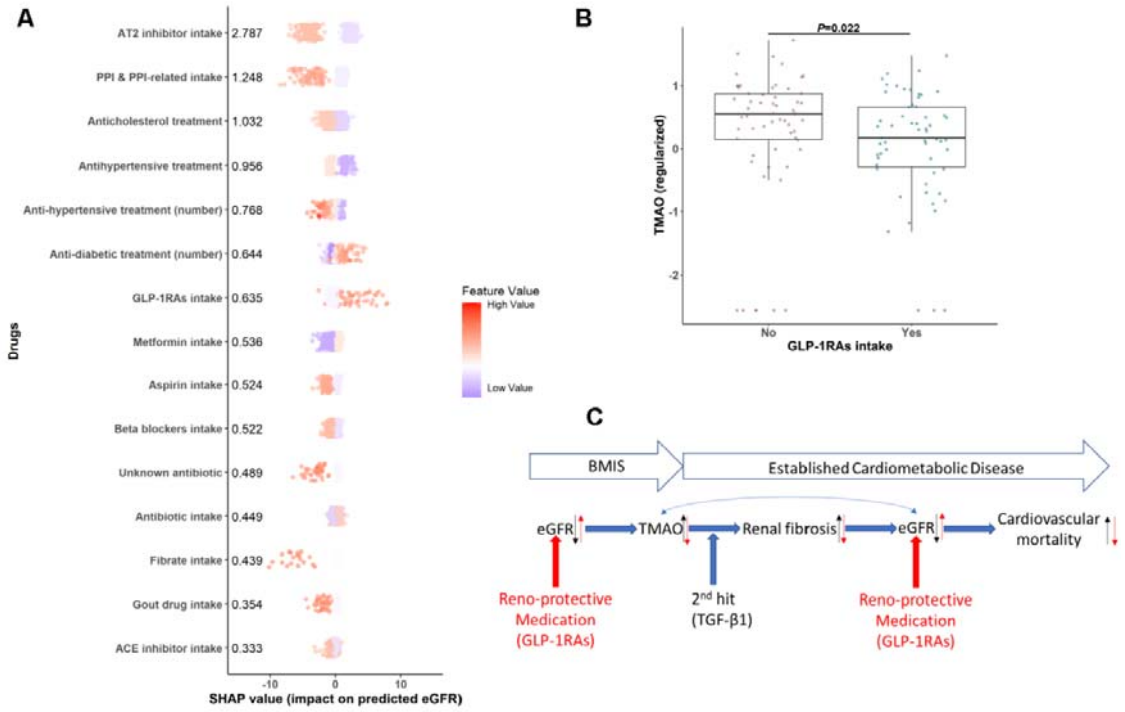
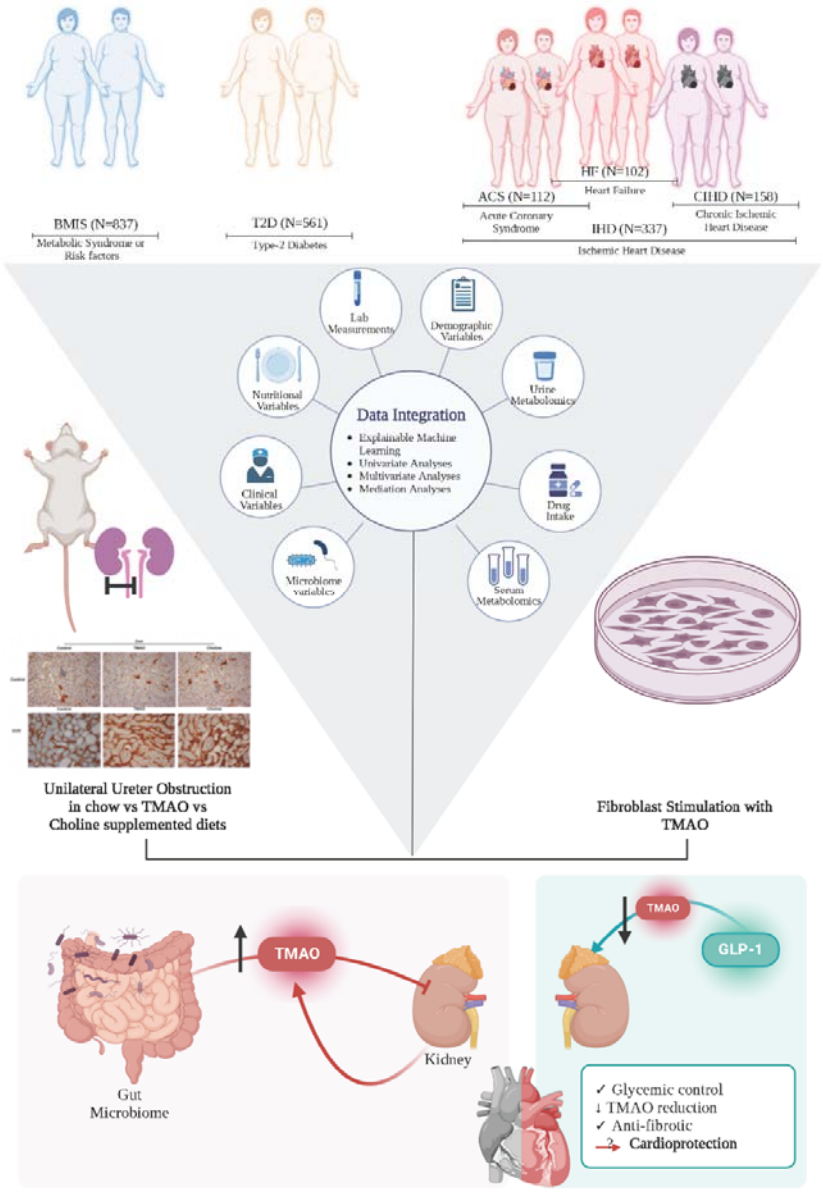


Figure 5. Reno-protective medication is associated with reduced circulating TMAO in MetaCardis participants with T2D. (**A**) Swarm plots of impact on model eGFR predictions (SHAP values; **Suppl.Table.22**) for MetaCardis T2D individuals (N=561) for the top 15 drugs, as determined by xgboost algorithms 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, BMI, disease group and hypertension status (N=59) (**Suppl.Table.23** for group characteristics). *P*-value determined by Mann-Whitney U test. (**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

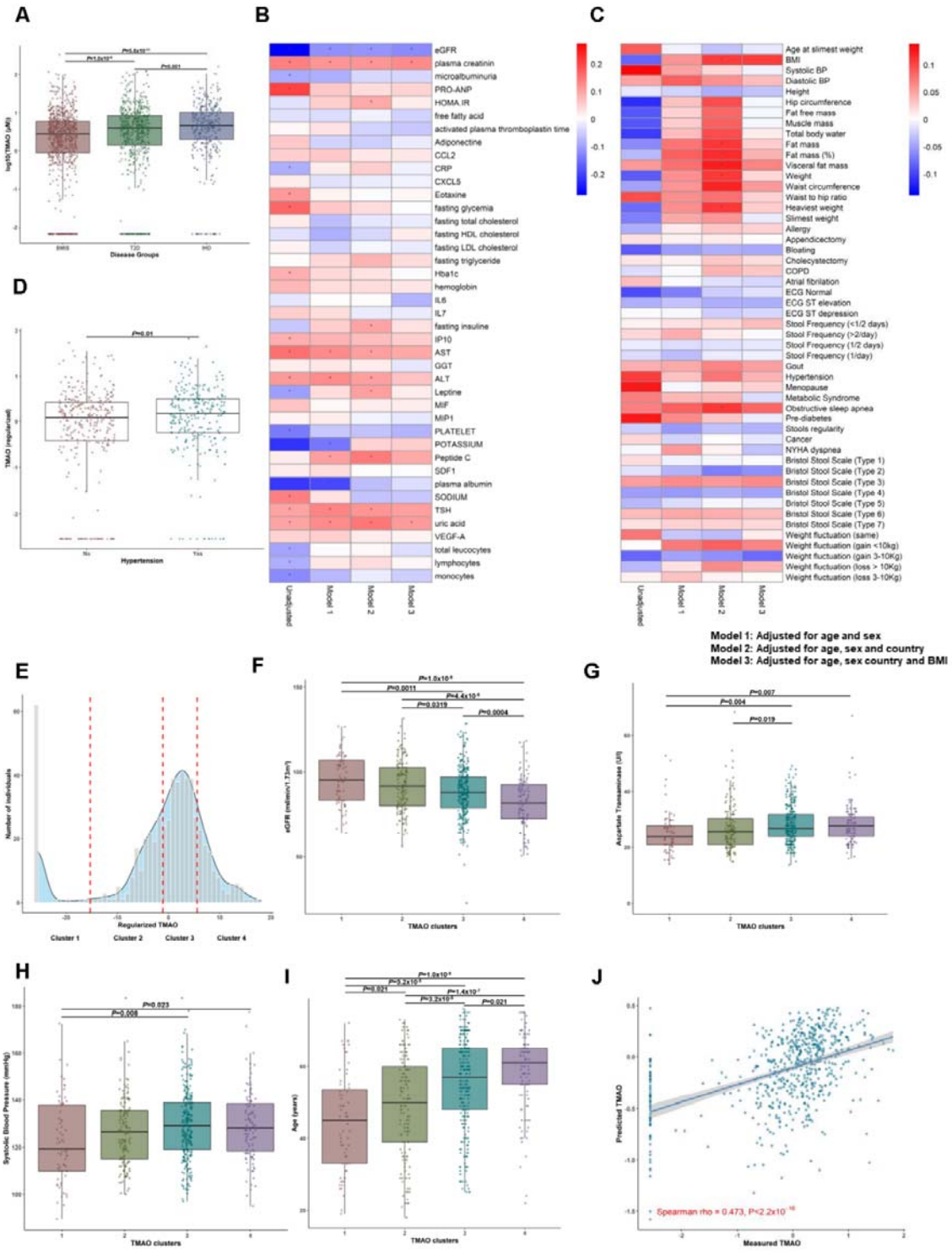
pathophysiology (two-hit" model) further negatively impacting renal clearance. Accordingly, medication with reno-protective properties (red arrows), such as GLP-1RAs, reduce circulating TMAO levels thereby potentially moderating its adverse effect on kidney function.

598 **Supplemental Figures**



599

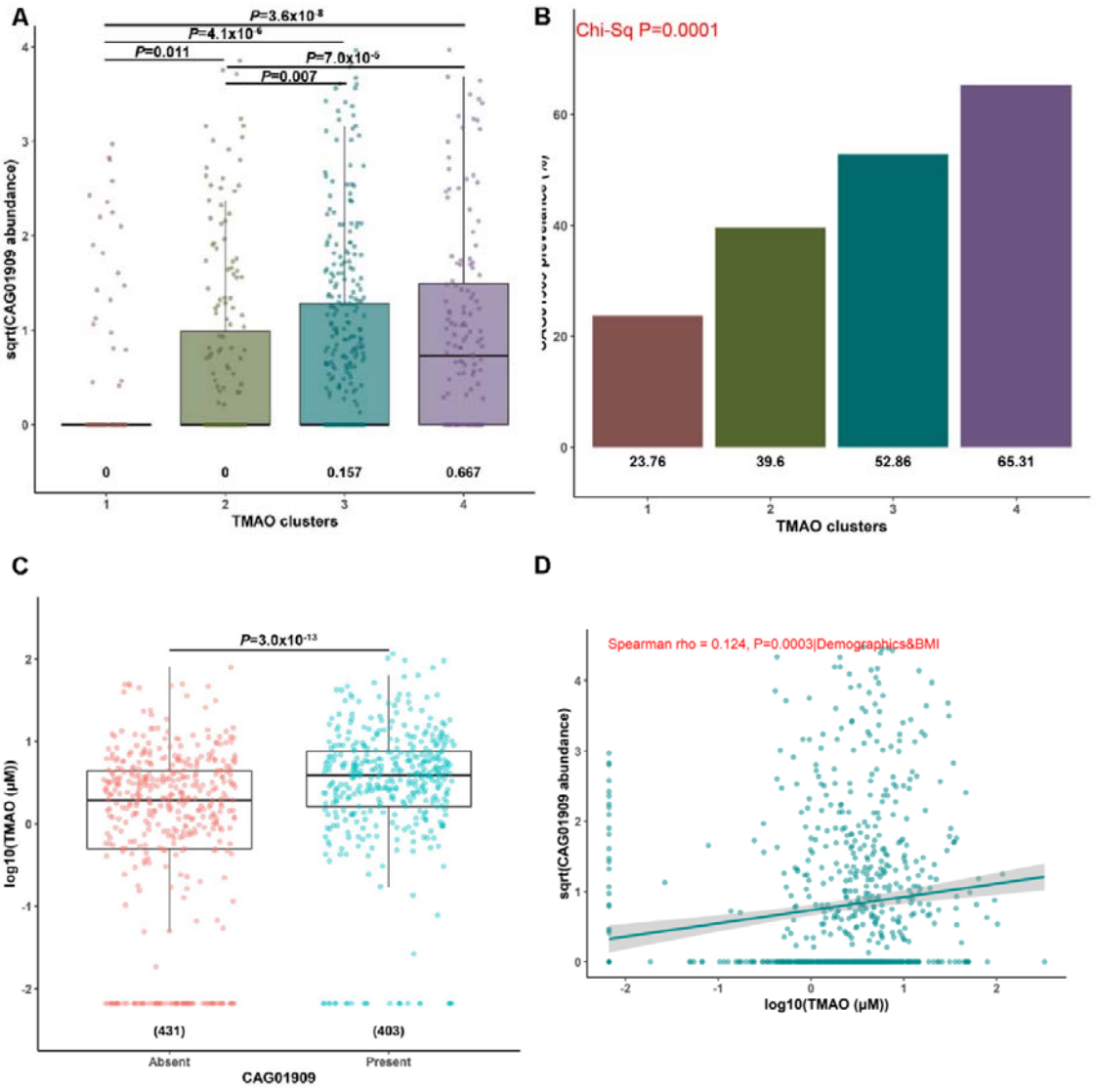
600 **Supplemental Figure 1. Overview of study design and aims.** Here we used an integrated approach
601 comprising Machine Learning (ML), multivariate, univariate and mediation analyses to objectively
602 characterize which host parameters contribute to plasma TMAO levels in the multi-center European
603 MetaCardis study. Capitalizing on the bioinformatics analysis we aimed to i) uncover novel host-
604 TMAO mechanistic targets and ii) identify actionable approaches (*e.g.* precision nutrition and/or
605 pharmacology interventions) to reduce systemic TMAO levels.



606

607 **Supplemental Figure 2. TMAO is associated with worse cardiometabolic profiles in study**
608 **participants with metabolic syndrome-related risk factors or pathologies but not over**
609 **cardiometabolic disease (BMIS subgroup N=581).** **(A)** Boxplot illustrating differences in fasting

circulating TMAO levels in the three MetaCardis patient groups (**Suppl.Table.1** for sub-cohort characteristics). *P* determined by Mann-Whitney U test. **(B)** Spearman correlations between circulating TMAO levels and bioclinical variables unadjusted, adjusted for age and sex (Model 1), age sex and country of recruitment (Model 2) or age, sex country of recruitment and BMI (Model 3), *pFDR<0.1. (**Suppl.Table.2**). (**C**) Spearman correlations between plasma TMAO and clinical variables, corrected as in (**A**), (**Suppl.Table.3**). (**D**) Comparison of circulating TMAO levels between BMIS participants classed as hypertensives, *P* determined by Mann-Whitney U test. (**E**) Density plot illustrating the distribution of regularized circulating TMAO levels in BMIS (N=582). The population was split into four TMAO clusters using the k-means algorithm. Dashed red lines denote the cut-off values for this distribution. Between group comparisons of selected variables for BMIS participants split into clusters according to their circulating TMAO levels for eGFR (**F**; mL/min/1.73m^2), aspartate transaminase (**G**; U/l), systolic blood pressure (**H**; mmHg) and age at the time of recruitment (**I**, years). *P* values were determined with pairwise Mann-Whitney U tests corrected for multiple comparisons with the Benjamini-Hochberg method. (**J**) Predicted (averaged after 100 iterations; y-axis) regularized plasma TMAO of BMIS participants by the BMIS-trained full-model *versus* actual measured regularized TMAO values (*x-axis*) for BMIS individuals (N=582) (**Suppl.Table.8**); insert Spearman rho and *P*-value of predicted *versus* measured values.

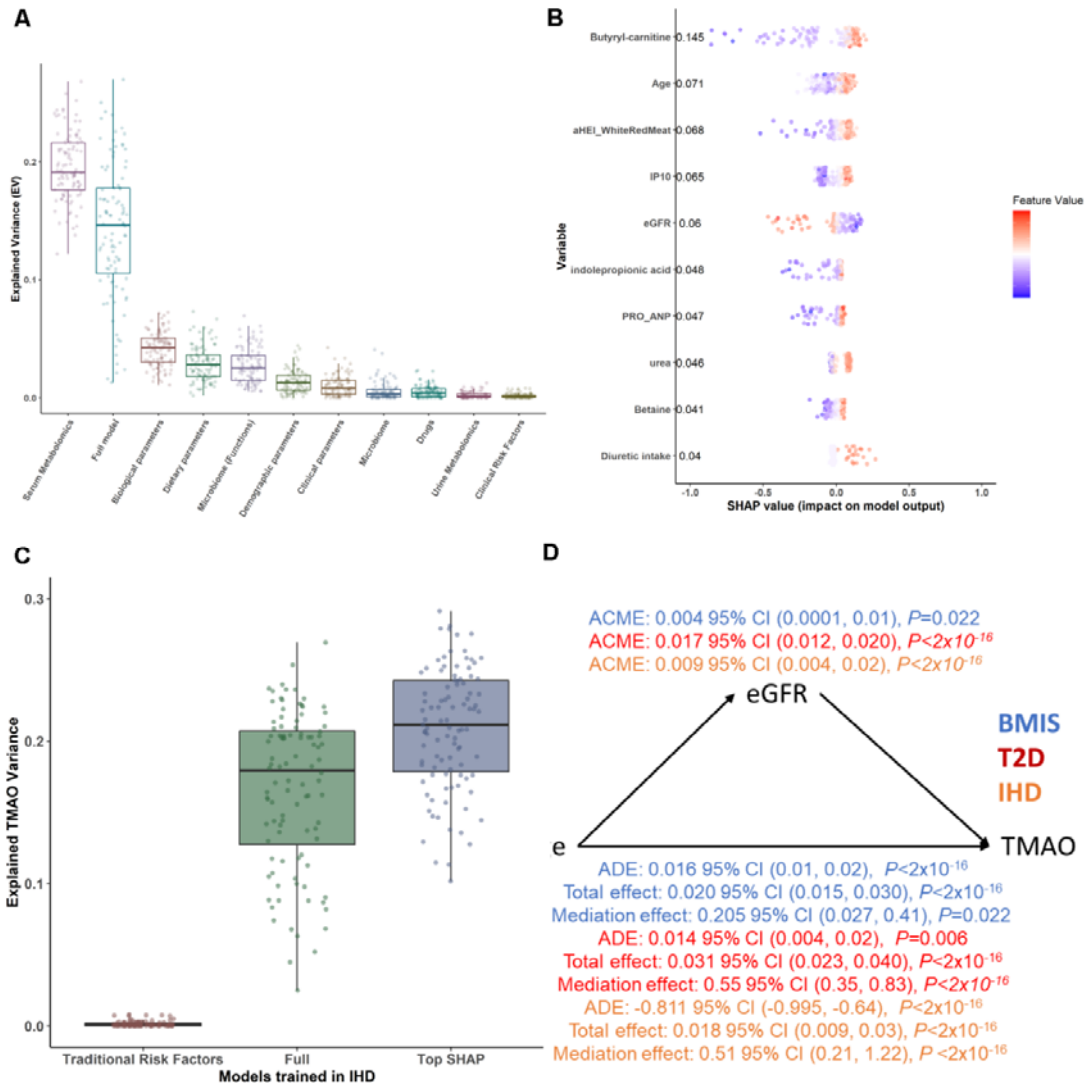


Supplemental Figure 3. CAG01909, associates with higher circulating TMAO in BMIS (N=834) individuals. Comparisons of normalized to bacterial load expression levels (square-root transformed) of (**A**) CAG01909, an unknown bacterium in BMIS participants divided into clusters according to circulating TMAO levels with the k-means algorithm. Median abundance level for each cluster is shown below the corresponding box plot, all pairwise comparisons performed with Mann-Whitney U tests (corrected for multiple testing with the Benjamini-Hochberg method). (**B**) Prevalence ((%); in parentheses under corresponding boxplots) of *CAG01909* in BMIS participants (N=834) split into TMAO clusters as in (A). *P* value determined by Chi-square test. (**C**) BMIS participants were split into

636 those where CAG01909 is present in their gut microbiota (N=401) and those where it is absent
637 (N=433) and circulating log-transformed TMAO levels between the two groups were compared
638 (Mann-Whitney U test). (**D**) Linear-regression-based scatterplot showing correlation between
639 normalized by bacterial load CAG01909 abundance levels (square-root transformed for visualization
640 purposes) and TMAO (log_{10}-transformed). Adjusted (age, sex, country of recruitment and BMI)
641 Spearman rho=0.123 and pFDR=0.032 for CAG01909 (**Suppl.Table.13**).

642

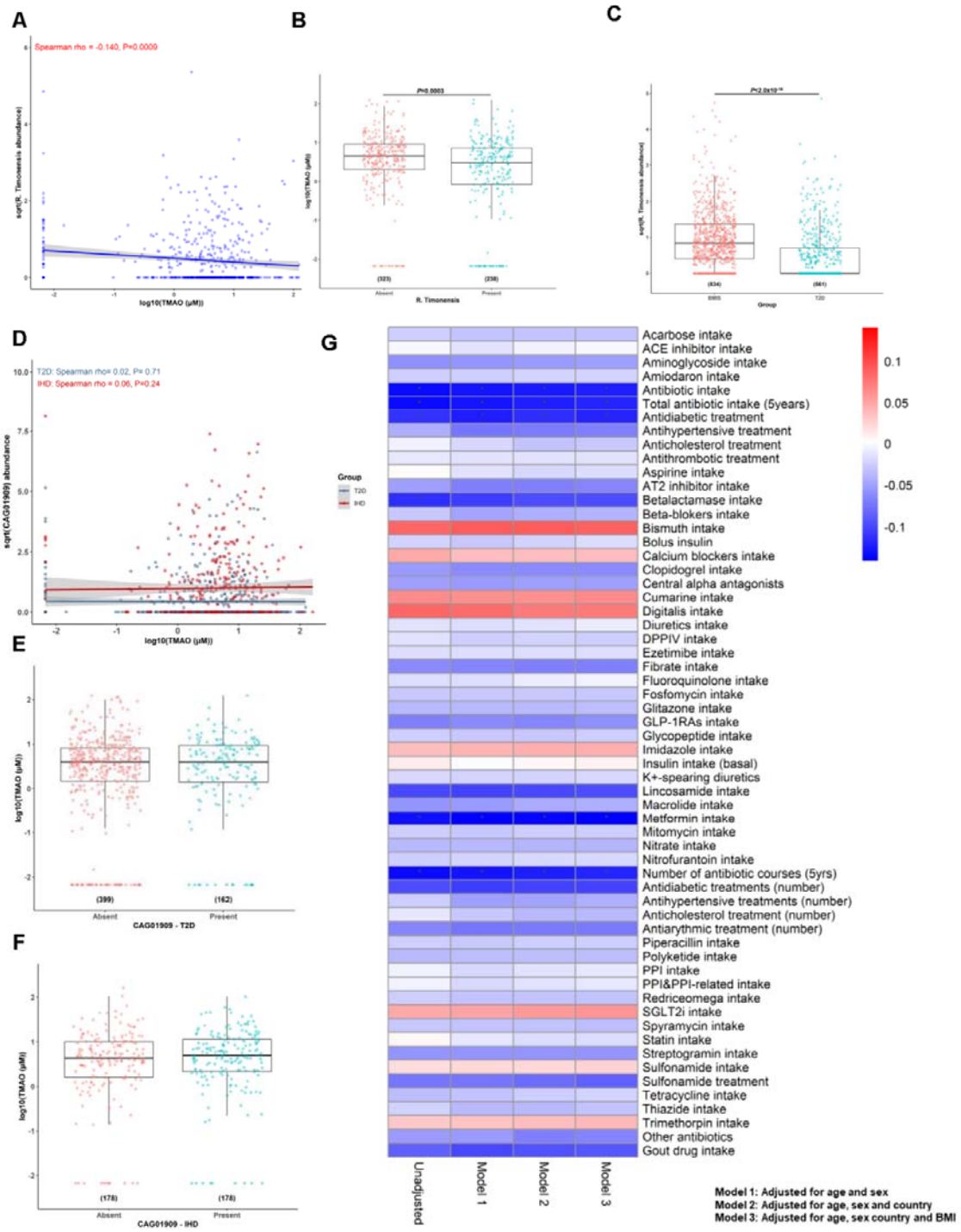
643



644

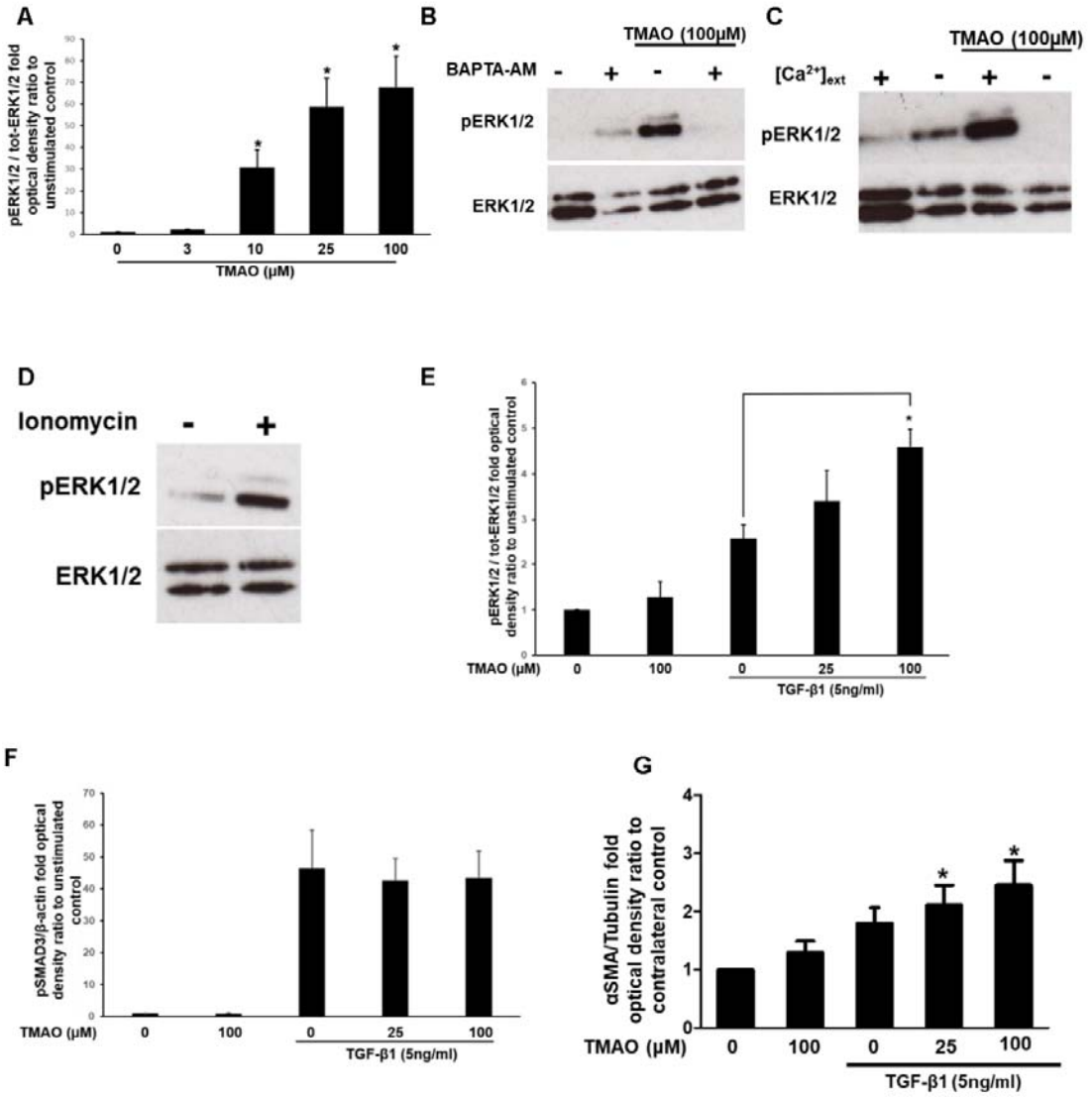
645 **Supplemental Figure 4. Boosted decision tree algorithms predict circulating TMAO in IHD**
646 **MetaCardis participants. (A)** Explained Variance (EV) of predicted serum TMAO levels
647 (**Suppl.Table.18**) determined by boosted decision trees (methods), trained exclusively on variables
648 from each feature category (**Suppl.Table.5** for a list of variables included in each group), or the full
649 model (containing all variables), after 100 iterations (**Suppl.Table.16** for N numbers and optimized
650 xgboost parameters per variable group) in IHD MetaCardis individuals. (**B**) Swarm plots of SHAP
651 values (impact on model outcome; *x-axis*) for each IHD MetaCardis participant with complete
652 phenotypic data (N=221), for all variables contributing to model predictions more than 4% of

regularized TMAO standard deviation, computed from xgboost algorithms trained on each feature category. Numbers denote mean absolute SHAP values from all IHD participants (in descending order) next to their corresponding variable (**Suppl.Table.20**). Dots, representing IHD individuals, are colored by the inverse-normalized value of their corresponding variable. (**C**) Boxplots depicting Explained Variance (EV; R^2) of circulating TMAO in IHD individuals computed by algorithms trained on clinical risk factors[29], the full model containing all variables or all the variables contributing more than 4% of regularized TMAO standard deviation to IHD model predictions, as determined by SHAP analysis, after 100 iterations. (**D**) Mediation analysis computing the direct effect of eGFR on TMAO increase with age in BMIS (blue), T2D (red) or IHD (orange) MetaCardis participants. ADE: Average direct effect (of Age on TMAO); ACME: average mediation effect (of eGFR on TMAO); Total effect: (cumulative effect of age and eGFR on TMAO (ADE + ACME)); Mediation effect: (% of the effect of age on circulating TMAO attributed to eGFR).



Supplemental Figure 5. In MetaCardis T2D participants *R. timonensis* inversely associates with circulating TMAO whilst BMIS-identified CAG01909 does not in overt disease. (**A**) Linear-regression-based scatterplot showing associations between normalized by bacterial load *R. timonensis* abundance levels (square-root transformed for visualization purposes) and TMAO (log_{10}-

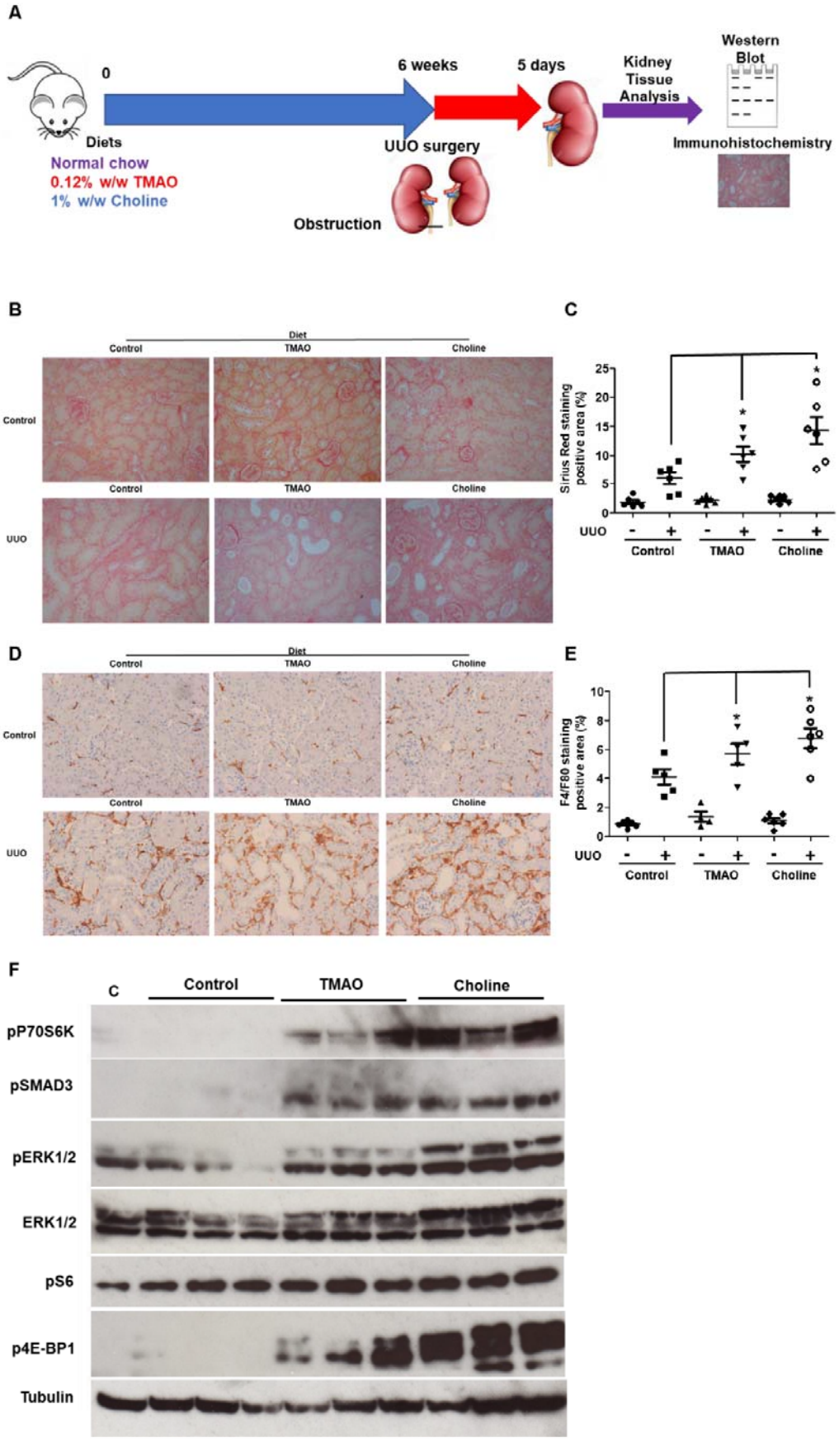
transformed). Unadjusted spearman rho=-0.140 and P=0.009. (**B**) Comparison of log-transformed circulating TMAO between T2D individuals with detectable (present) and undetectable (absent) fecal *R. timonensis* (Mann-Whitney test). Number of individuals in each group shown in parenthesis. (**C**) Boxplot illustrating differences in *R. timonensis* fecal abundance in the BMIS and T2D MetaCardis patient groups (group sizes in parentheses), *P* determined by Mann-Whitney U test. (**D**) Linear-regression-based scatterplot showing correlation between normalized by bacterial load CAG01909 expression levels (square-root transformed for visualization purposes) and TMAO (log_{10}-transformed) in T2D and IHD MetaCardis subjects. Comparison of log-transformed circulating TMAO between T2D (**E**) and IHD (**F**) MetaCardis subjects with detectable (present) or undetectable (absent) fecal CAG01909 (Mann-Whitney test). Number of individuals in each group shown in parenthesis. (**G**) Heatmap illustrating Spearman correlations between bacterial load-normalized abundance of CAG01909 and intake of medication in T2D MetaCardis individuals (N=561) unadjusted, adjusted for age and sex (Model 1), age sex and country of recruitment (Model 2) or age, sex country of recruitment and BMI (Model 3), *pFDR<0.1. (**Suppl.Table.21**).



684

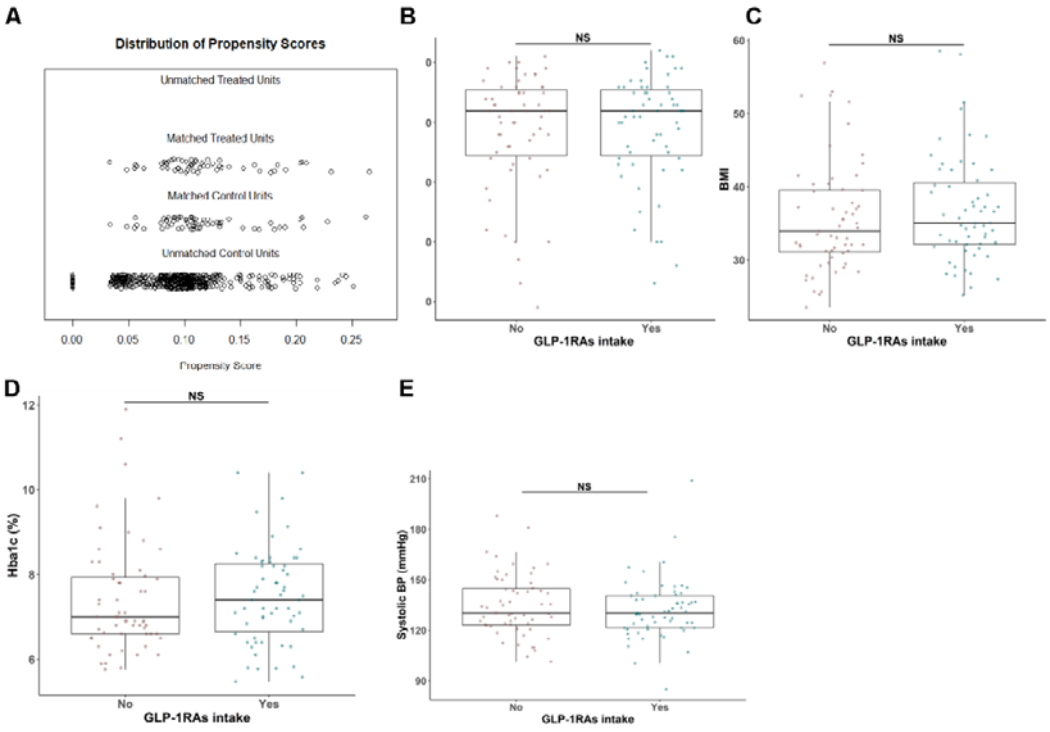
685 **Supplemental Figure 6. TMAO activates ERK1/2 via a Ca^{2+}-sensitive pathway in human renal**
686 **fibroblasts. (A)** Optical density (OD) of pERK1/2 levels normalized against total ERK1/2 for
687 experiments represented in **Figure 4C**. The normalized density of the control unstimulated samples
688 was arbitrarily set to 1. (**B**) Serum-starved Human Renal Fibroblasts (HRFs) were loaded with the Ca^{2+}
689 chelator BAPTA-AM (20μM; 20min) and then stimulated with TMAO (100μM) for 10min. ERK1/2
690 activation was probed by Western blot and membranes were stripped and re-probed for total
691 ERK1/2, a representative image of N=3 independent experiments is shown. (**C**) HRFs were serum-
692 starved in complete physiological medium for 45min, the medium was aspirated and cells were

693 incubated for further 15min in physiological medium with Ca^{2+} omitted from the buffer.
694 Subsequently, cells were stimulated with 100µM TMAO for 10min and ERK1/2 activation was probed
695 as in (**B**). (**D**) Serum-starved HRFs were stimulated with the Ca^{2+} ionophore ionomycin (100µM) and
696 phospho-ERK1/2 levels were probed as in (B). OD of pERK1/2 (**E**) or pSMAD3 (**F**) normalized by total
697 ERK1/2 or β-actin respectively for experiments represented in **Figure 4D**. The normalized density of
698 the control unstimulated samples was arbitrarily set to 1. (**G**) OD of αSMA normalized by β-actin
699 levels for experiments represented in **Figure 4E**. For all N=3, error bars represent ±SEM.



700

Supplemental Figure 7. TMAO or choline diet exacerbates collagen deposition, macrophage infiltration and activate pro-fibrotic signaling in kidneys of mice that underwent Unilateral Ureter Obstruction (UUO) surgery. (**A**) Outline of the animal experiments. (**B**) Sirius red staining of kidney sections (20x magnification) from obstructed (UUO; 5days 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 6weeks prior to surgery, as indicated. N=6 per group. (**C**) Quantification of collagen deposition staining as (%) of positive Sirius red area/field of view averaged from 5 images per animal. (**D**) Immunostaining of kidney sections (20x magnification) from UUO or control kidneys as in (**B**) with the macrophage marker F4/F80. (**E**) Quantification of macrophage infiltration as % of F4/F80 staining/ field of view of images from (**D**); N=6 animals/group. (**F**) Western blot for profibrotic signaling pathways activation from UUO kidneys from animals treated as in (**A**). C: contralateral kidney from a control-diet fed animal; TMAO: 0.12% w/w TMAO diet; Choline: 1% w/w Choline diet.

716 

717 **Supplemental Figure 8. Propensity-score matching of MetaCardis patients with T2D. (A)** Propensity
718 scores of MetaCardis participants with T2D taking GLP-1 Receptor agonists (GLP-1RAs); N=59) and
719 matched controls (N=59) using nearest neighbors matching with age, sex, disease severity group and
720 hypertension status as covariates (**Suppl.Table.23**). Comparisons of age (**B**), BMI (**C**), glycated
721 hemoglobin (**D**) and systolic blood pressure (**E**), between MetaCardis patients with T2D diabetics
722 prescribed GLP-1RAs (N=59) and matched controls (as in A; N=59). Mann-Whitney U tests were used
723 to determine statistical significance for B-E. NS: non-significant (P>0.05).

724

Supplemental Tables

Supplemental Table 1. Characteristics of MetaCardis sub-cohorts (BMIS, T2D, IHD).

Supplemental Table 2. Spearman correlations between circulating TMAO and biological variables in BMIS.

Supplemental Table 3. Spearman correlations between circulating TMAO and clinical variables in BMIS.

Supplemental Table 4. BMIS TMAO cluster characteristics.

Supplemental Table 5. List of all the variables per Group variable category.

Supplemental Table 6. Model hyperparameters and N numbers for each variable group in BMIS.

Supplemental Table 7. Explained Variance (EV) of models trained with each variable group in BMIS for each iteration.

Supplemental Table 8. Predicted *versus* measured circulating TMAO in BMIS subjects.

Supplemental Table 9. Contributions to predictions of the full model by each variable group.

Supplemental Table 10. Average SHAP values for all variables for each BMIS subject.

Supplemental Table 11. Spearman correlations between circulating TMAO and habitual intake of 37 food categories in BMIS.

Supplemental Table 12. Spearman correlations between circulating TMAO and calculated habitual intake of micronutrients in BMIS.

Supplemental Table 13. Spearman correlations between circulating TMAO and bacterial species abundance in BMIS.

Supplemental Table 14. Differential taxon abundance between high *versus* low TMAO clusters in BMIS.

Supplemental Table 15. Model hyperparameters and N numbers for each variable group in T2D.

Supplemental Table 16. Model hyperparameters and N numbers for each variable group in IHD.

Supplemental Table 17. Explained Variance (EV) of models trained with each variable group in T2D for each iteration.

Supplemental Table 18. Explained Variance (EV) of models trained with each variable group in IHD for each iteration.

Supplemental Table 19. Average impact on TMAO predictions (SHAP values) for all variables for each T2D subject.

Supplemental Table 20. Average impact on TMAO predictions (SHAP values) for all variables for each IHD subject.

Supplemental Table 21. Spearman correlations between CAG01909 abundance and intake of drugs in T2D.

Supplemental Table 22. Average impact on predicted eGFR (SHAP values) for prescribed medication for MetaCardis T2D subjects.

Supplemental Table 23. Characteristics of MetaCardis participants prescribed GLP-1Ras and matched controls.

Supplemental Table 24. Materials.

References

1. Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol* 2021;**19**:55-71. doi: 10.1038/s41579-020-0433-9 pmid:https://pubmed.ncbi.nlm.nih.gov/32887946

2. Agus A, Clément K, Sokol H. Gut microbiota-derived metabolites as central regulators in metabolic disorders. Gut 2021;70:1174-1182. doi: 10.1136/gutjnl-2020-323071 pmid:https://pubmed.ncbi.nlm.nih.gov/33272977

3. Li J, Li Y, Ivey KL, et al. Interplay between diet and gut microbiome, and circulating concentrations of trimethylamine N-oxide: findings from a longitudinal cohort of US men. Gut 2022;71:724-733..doi: 10.1136/gutjnl-2020-322473 pmid:https://pubmed.ncbi.nlm.nih.gov/33926968

4. Romano KA, Vivas EI, Amador-Noguez D, et al. Intestinal microbiota composition modulates choline bioavailability from diet and accumulation of the proatherogenic metabolite trimethylamine-N-oxide. mBio 2015;6:e02481.doi: 10.1128/mBio.02481-14 pmid:https://pubmed.ncbi.nlm.nih.gov/25784704/

5. Rath S, Heidrich B, Pieper DH, et al. Uncovering the trimethylamine-producing bacteria of the human gut microbiota. Microbiome 2017;5:54.doi: 10.1186/s40168-017-0271-9.

6. Wang Z, Klipfell E, Bennett BJ, *et al*. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 2011;**472**:57-63.doi:10.1038/nature09922 pmid:https://pubmed.ncbi.nlm.nih.gov/21475195

7. Tang WH, Wang Z, Levison BS, et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. N Engl J Med 2013;368:1575-84.doi:10.1056/NEJMoa1109400 pmid:https://pubmed.ncbi.nlm.nih.gov/23614584

8. Koeth RA, Wang Z, Levison BS, *et al*. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med* 2013;19:576-85.doi: 10.1038/nm.3145 pmid:https://pubmed.ncbi.nlm.nih.gov/23563705

796 9. Wang Z, Bergeron N, Levison BS, et al. Impact of chronic dietary red meat, white meat, or
797 non-meat protein on trimethylamine N-oxide metabolism and renal excretion in healthy men and
798 women. Eur Heart J 2019;40:583-594.doi: 10.1093/eurheartj/ehy799
799 pmid:https://pubmed.ncbi.nlm.nih.gov/30535398
800 10. Wu WK, Chen CC, Liu PY, et al. Identification of TMAO-producer phenotype and host–diet–
801 gut dysbiosis by carnitine challenge test in human and germ-free mice. Gut 2019;68:1439–1449.doi:
802 10.1136/gutjnl-2018-317155 pmid:https://pubmed.ncbi.nlm.nih.gov/30377191
803 11. Li XS, Obeid S, Klingenberg R, *et al*. Gut microbiota-dependent trimethylamine N-oxide in
804 acute coronary syndromes: a prognostic marker for incident cardiovascular events beyond
805 traditional risk factors. *Eur Heart J* 2017;**38**:814-824.doi: 10.1093/eurheartj/ehw582
806 pmid:https://pubmed.ncbi.nlm.nih.gov/28077467
807 12. Key TJ, Appleby PN, Bradbury KE, *et al*. Consumption of Meat, Fish, Dairy Products, and Eggs
808 and Risk of Ischemic Heart Disease. *Circulation* 2019;**139**:2835-2845.
809 doi:10.1161/CIRCULATIONAHA.118.038813 pmid:https://pubmed.ncbi.nlm.nih.gov/31006335/
810 13. Gibson R, Lau CH, Loo RL, *et al*. The association of fish consumption and its urinary
811 metabolites with cardiovascular risk factors: the International Study of Macro-/Micronutrients and
812 Blood Pressure (INTERMAP). *Am J Clin Nutr* 2020;**111**:280-290.doi:10.1093/ajcn/nqz293
813 pmid:https://pubmed.ncbi.nlm.nih.gov/31782492
814 14. Hoyles L, Jiménez-Pranteda ML, Chilloux J, *et al*. Metabolic retroconversion of
815 trimethylamine N-oxide and the gut microbiota. Microbiome 2018;6:73.doi: 10.1186/s40168-018-
816 0461-0 pmid:https://pubmed.ncbi.nlm.nih.gov/29678198
817 15. Brunt VE, Gioscia-Ryan RA, Casso AB, *et al*. Trimethylamine-N-Oxide Promotes Age-Related
818 Vascular Oxidative Stress and Endothelial Dysfunction in Mice and Healthy Humans. *Hypertension*
819 2020;**76**:101-112.doi: 10.1161/HYPERTENSIONAHA.120.14759

16. Bennett BJ, Vallim TQ, Wang Z, *et al*. Trimethylamine-N-oxide, a metabolite associated with atherosclerosis, exhibits complex genetic and dietary regulation. *Cell Metab* 2013;**17**:49-60.doi: 10.1016/j.cmet.2012.12.011 pmid:https://pubmed.ncbi.nlm.nih.gov/23312283

17. Stubbs JR, House JA, Ocque AJ, *et al*. Serum Trimethylamine-N-Oxide is Elevated in CKD and Correlates with Coronary Atherosclerosis Burden. *J Am Soc Nephrol* 2016;**27**:305-13. doi:10.1681/ASN.2014111063 pmid:https://pubmed.ncbi.nlm.nih.gov/26229137

18. Tang WHW, Wang Z, Kennedy DJ, *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* 2015;**116**:448-55.doi: 10.1161/CIRCRESAHA.116.305360 pmid:https://pubmed.ncbi.nlm.nih.gov/25599331

19. Manor O, Zubair N, Conomos MP, *et al*. A Multi-omic Association Study of Trimethylamine N-Oxide. Cell Rep 2018;**24**:935-946.doi: 10.1016/j.celrep.2018.06.096 pmid:https://pubmed.ncbi.nlm.nih.gov/30044989

20. Zhuang R, Ge X, Han L, et al. Gut microbe-generated metabolite trimethylamine N-oxide and the risk of diabetes: A systematic review and dose-response meta-analysis. Obes Rev 2019;20:883-894. doi: 10.1111/obr.12843 pmid:https://pubmed.ncbi.nlm.nih.gov/30868721

21. Bar N, Korem T, Weissbrod O, *et al*. A reference map of potential determinants for the human serum metabolome. *Nature* 2020;**588**:135-140.doi:10.1038/s41586-020-2896-2 pmid:https://pubmed.ncbi.nlm.nih.gov/33177712

22. Wilding JPH. Batterham RL, Calanna S, *et al*. Once-Weekly Semaglutide in Adults with Overweight or Obesity. *N Engl J Med* 2021;**384**:989-1002. doi:10.1056/NEJMoa2032183 pmid:https://pubmed.ncbi.nlm.nih.gov/33567185

23. Shaman AM, Bain SC, Bakris GL, 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 2022;145:575-585.

845 24. Vieira-Silva S, Falony G, Belda E, *et al*. Statin therapy is associated with lower prevalence of
846 gut microbiota dysbiosis. *Nature* 2020;**581**:310-315.doi:10.1038/s41586-020-2269-x.
847 pmid:https://pubmed.ncbi.nlm.nih.gov/32433607
848 25. Brial F, Chilloux J, Nielsen T, *et al*. Human and preclinical studies of the host-gut microbiome
849 co-metabolite hippurate as a marker and mediator of metabolic health. *Gut* 2021; gutjnl-2020-
850 323314.doi: 10.1136/gutjnl-2020-323314 pmid:https://pubmed.ncbi.nlm.nih.gov/33975870
851 26. Lundberg SM, Erion G, Chen H, et al. From Local Explanations to Global Understanding with
852 Explainable AI for Trees. Nat Mach Intell 2020;2:56-67.doi: 10.1038/s42256-019-0138-9
853 pmid:https://pubmed.ncbi.nlm.nih.gov/32607472
854 27. Levey AS, Coresh J, Balk E, *et al*. National Kidney Foundation practice guidelines for chronic
855 kidney disease: evaluation, classification, and stratification. *Ann Intern Med* 2003;**139**:137-47.
856 doi:10.7326/0003-4819-139-2-200307150-00013 pmid:https://pubmed.ncbi.nlm.nih.gov/12859163
857 28. Vanholder R, Schepers E, Pletinck A, *et al*. The uremic toxicity of indoxyl sulfate and p-cresyl
858 sulfate: a systematic review. *J Am Soc Nephrol* 2014;**25**:1897-90.doi:10.1681/ASN.2013101062
859 pmid:https://pubmed.ncbi.nlm.nih.gov/24812165
860 29. Fromentin S, Forslund SK, Chechi K, *et al*. Microbiome and metabolome features of the
861 cardiometabolic disease spectrum. *Nat Med* 2022;**28**:303-314.doi:10.1038/s41591-022-01688-4.
862 pmid:https://pubmed.ncbi.nlm.nih.gov/35177860
863 30. Koeth RA, Levison BS, Culley MK, *et al*. γ-Butyrobetaine is a proatherogenic intermediate in
864 gut microbial metabolism of L-carnitine to TMAO. *Cell Metab* 2014;**20**:799-812.
865 doi:10.1016/j.cmet.2014.10.006 pmid:https://pubmed.ncbi.nlm.nih.gov/25440057
866 31. Ferrell M, Bazeley P, Wang Z, *et al*. Fecal Microbiome Composition Does Not Predict Diet-
867 Induced TMAO Production in Healthy Adults. *J Am Heart Assoc* 2021;**10**: e021934.
868 doi:10.1161/JAHA.121.021934 pmid: https://pubmed.ncbi.nlm.nih.gov/34713713

869 32. Titan SM, Venturini G, Padilha K, *et al*. Metabolites related to eGFR: Evaluation of candidate
870 molecules for GFR estimation using untargeted metabolomics. *Clin Chim Acta* 2019;**489**:242-248.
871 doi:10.1016/j.cca.2018.08.037 pmid:https://pubmed.ncbi.nlm.nih.gov/30153452
872 33. Altara R, Manca M, Hessel MH, *et al*. CXCL10 Is a Circulating Inflammatory Marker in Patients
873 with Advanced Heart Failure: a Pilot Study. *J Cardiovasc Transl Res* 2016;**9**:302-14.
874 doi:10.1007/s12265-016-9703-3 pmid:https://pubmed.ncbi.nlm.nih.gov/27271043
875 34. Forslund SK, Chakaroun R, Zimmermann-Kogadeeva M, *et al*. Combinatorial, additive and
876 dose-dependent drug-microbiome associations. *Nature* 2021;**600**:500-505.doi:10.1038/s41586-021-
877 04177-9. pmid:https://pubmed.ncbi.nlm.nih.gov/34880489
878 35. Schmitt R, Melk A. Molecular mechanisms of renal aging. *Kidney Int* 2017;**92**:569-
879 579.doi:10.1016/j.kint.2017.02.036 pmid:https://pubmed.ncbi.nlm.nih.gov/28729036
880 36. Gupta N, Buffa JA, Roberts AB, *et al*. Targeted Inhibition of Gut Microbial Trimethylamine N-
881 Oxide Production Reduces Renal Tubulointerstitial Fibrosis and Functional Impairment in a Murine
882 Model of Chronic Kidney Disease. *Arterioscler Thromb Vasc Biol* 2020;**40**:1239-1255.
883 doi:10.1161/ATVBAHA.120.314139 pmid:https://pubmed.ncbi.nlm.nih.gov/32212854
884 37. Fang Q, Zheng B, Liu N, *et al*. Trimethylamine N-Oxide Exacerbates Renal Inflammation and
885 Fibrosis in Rats With Diabetic Kidney Disease. Front Physiol 2021;**12**:682482.
886 doi:10.3389/fphys.2021.682482 pmid:https://pubmed.ncbi.nlm.nih.gov/34220546
887 38. Duffield JS. Cellular and molecular mechanisms in kidney fibrosis. *J Clin Invest*
888 2014;**124**:2299-306.doi:10.1172/JCI72267 pmid:https://pubmed.ncbi.nlm.nih.gov/24892703
889 39. Zhu W, Gregory JC, Org E, *et al*. Gut Microbial Metabolite TMAO Enhances Platelet
890 Hyperreactivity and Thrombosis Risk. *Cell* 2016;**165**:111-124.doi:10.1016/j.cell.2016.02.011
891 pmid:https://pubmed.ncbi.nlm.nih.gov/26972052
892 40. Andrikopoulos P, Eccles SA, Yaqoob MM. Coupling between the TRPC3 ion channel and the
893 NCX1 transporter contributed to VEGF-induced ERK1/2 activation and angiogenesis in human

894 primary endothelial cells. *Cell Signal* 2017;**37**:12-30 doi:10.1016/j.cellsig.2017.05.013
895 pmid:https://pubmed.ncbi.nlm.nih.gov/28535874

896 41. Andrikopoulos P, Kieswich J, Harwood SM, *et al*. Endothelial Angiogenesis and Barrier
897 Function in Response to Thrombin Require Ca2+ Influx through the Na+/Ca2+ Exchanger. *J Biol Chem*
898 2015;**290**:18412-28.doi:10.1074/jbc.M114.628156
899 pmid:https://pubmed.ncbi.nlm.nih.gov/25979335

900 42. Andrikopoulos P, Kieswich J, Pacheco S, *et al*. The MEK Inhibitor Trametinib Ameliorates
901 Kidney Fibrosis by Suppressing ERK1/2 and mTORC1 Signaling. *J Am Soc Nephrol* 2019;**30**:33-49.
902 doi:10.1681/ASN.2018020209 pmid:https://pubmed.ncbi.nlm.nih.gov/30530834

903 43. Chevalier RL, Forbes MS, Thornhill BA. Ureteral obstruction as a model of renal interstitial
904 fibrosis and obstructive nephropathy. *Kidney Int* 2009;**75**:1145-1152.doi:10.1038/ki.2009.86
905 pmid:https://pubmed.ncbi.nlm.nih.gov/19340094

906 44. Yap CX, Henders AK, Alvares GA, *et al*. Autism-related dietary preferences mediate autism-
907 gut microbiome associations. *Cell* 2021;**184**:5916-5931.e17.doi:10.1016/j.cell.2021.10.015
908 pmid:https://pubmed.ncbi.nlm.nih.gov/34767757

909 45. Rebouche CJ. Kinetics, pharmacokinetics, and regulation of L-carnitine and acetyl-L-carnitine
910 metabolism. *Ann N Y Acad* Sci 2004;**1033**:30-41. doi:10.1196/annals.1320.003
911 pmdi:https://pubmed.ncbi.nlm.nih.gov/15591001

912 46. Cocking C, Walton J, Kehoe L, et al. The role of meat in the European diet: current state of
913 knowledge on dietary recommendations, intakes and contribution to energy and nutrient intakes
914 and status. *Nutr Res Rev* 2020;**33**:181-189. doi:10.1017/S0954422419000295
915 pmid:https://pubmed.ncbi.nlm.nih.gov/31918784

916 47. Zhao S, Feng XF, Huang T, *et al*. The Association Between Acylcarnitine Metabolites and
917 Cardiovascular Disease in Chinese Patients With Type 2 Diabetes Mellitus. *Front Endocrinol*
918 *(Lausanne)* 2020;**11**:212. doi:10.3389/fendo.2020.00212
919 pmid:https://pubmed.ncbi.nlm.nih.gov/32431666

48. Chen S, Henderson A, Petriello MC, *et al*. Trimethylamine N-Oxide Binds and Activates PERK to Promote Metabolic Dysfunction. *Cell Metab* 2019;**30**:1141-1151.e5. doi:10.1016/j.cmet.2019.08.021 pmid:https://pubmed.ncbi.nlm.nih.gov/31543404

49. Martinez-Klimova E, Aparicio-Trejo OE, Tapia E, et al. Unilateral Ureteral Obstruction as a Model to Investigate Fibrosis-Attenuating Treatments. *Biomolecules* 2019;**9**:141. doi:10.3390/biom9040141 pmid:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6523883

50. Sureshbabu A, Muhsin SA, Choi ME. TGF-β signaling in the kidney: profibrotic and protective effects. *Am J Physiol Renal Physiol* 2016;**310**:F596–F606. doi:10.1152/ajprenal.00365.2015 pmid:https://pubmed.ncbi.nlm.nih.gov/26739888/

51. Suthanthiran M, Gerber LM, Schwartz JE, *et al*. Circulating transforming growth factor-beta1 levels and the risk for kidney disease in African Americans. *Kidney Int* 2009;**76**:72-80.doi: 10.1038/ki.2009.66 pmid:https://pubmed.ncbi.nlm.nih.gov/19279557/

52. Gruppen EG, Garcia E, Connely MA, *et al*. TMAO is Associated with Mortality: Impact of Modestly Impaired Renal Function. *Sci Rep* 2017;**7**:13781.doi:10.1038/s41598-017-13739-9 pmid:https://pubmed.ncbi.nlm.nih.gov/29061990

53. Meyer KA, Benton TZ, Bennett BJ, *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* 2016;**5**:e003970.doi:10.1161/JAHA.116.003970 pmid:https://pubmed.ncbi.nlm.nih.gov/27792658

Methods

MetaCardis study design and recruitment

Study design, recruitment and exclusion criteria has been extensively described[24,29,34,54]. Patients were subclassified in three groups: BMI-spectrum patients (BMIS[29]; N=837), encompassing MetaCardis participants presenting with metabolic syndrome-related risk factors or conditions (hypertension[55], obesity[56] and metabolic syndrome[57]) and patients diagnosed with type-2 diabetes (T2D[58]; N=561) or ischaemic heart disease (IHD; N=356). The IHD group comprised patients with Acute (<15days) 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%). 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 participants provided written informed consent and the study was undertaken according to Helsinki Declaration-II.

Sample and phenotypic information collection

Biofluid and biomatter collection has been described elsewhere[24,29,34,54]. 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 48h. All samples were stored at -80°C until use. Clinical history, medication and phenotypic information were acquired as described[23,28,34,54] with standardized procedures across centers. Participants reported habitual food intake through a customized Food Frequency Questionnaire (FQQ)[59]. Bioclinical variables were measured in a single center according to standard procedures[28]. Estimated glomerular filtration rate (eGFR) was calculated with the CKD-EPI formula without ethnicity adjustment[61].

Metabolic profiling

964 **^{1}H-Nuclear Magnetic Resonance (^{1}H NMR) spectroscopy**

965 Spectra acquisition, using an Avance spectrometer (Bruker) at 600 MHz; and structural assignments
966 have been extensively described previously[25,29]. Absolute quantifications were derived using the "*In*
967 *Vitro* Diagnostics for research" (IVDr) quantification BI-LISA and BI-Quant algorithms (Bruker).

968 **Gas Chromatography coupled Mass Spectrometry (GC-MS)**

969 Serum samples (100µl) were prepared, analyzed and processed as described[29,35]. Briefly, protein was
970 methanol-precipitated, methanol was evaporated to dryness and subsequent to derivatization
971 samples were injected to an Agilent 7890B-5977B Inert Plus GC-MS system. The chromatographic
972 column was an Agilent ZORBAX DB5- MS (30m x 250µm x 0.25µm + 10m Duragard). The
973 temperature gradient was 37.5min long and the mass analyzer was operated in full scan mode
974 between 50 and600 m/z. Peaks were annotated with the use of the Fiehn library (Agilent G1676AA
975 Fiehn GC/MS Metabolomics RTL Library, User Guide, Agilent Technologies,
976 https://www.agilent.com/cs/library/usermanuals/Public/G1676-90001_Fiehn.pdf). Metabolic
977 features with low reproducibility or linearity were removed from the dataset, resulting in 102
978 annotated metabolic features.

979 **Ultra-Performance Liquid Chromatography-Tandem Mass Spectrometry (UPLC-MS/MS)**

980 UPLC-MS/MS performed on a Waters Acquity UPLC-Xevo TQ-S UPLC-MS/MS system equipped with
981 an Acquity BEH HILIC (2.1×100mm, 1.7µm) chromatographic column was employed to determine
982 TMA, TMAO, choline, betaine, γ-butyrobetaine, betaine aldehyde, butyryl-carnitine, isovaleryl-
983 carnitine, OH-isovaleryl-carnitine, stearoyl-carnitine, oleoyl-carnitine, linoleoyl-carnitine, myristoyl-
984 carnitine, lauroyl-carnitine and decanoyl-carnitine as described previously[28,34].

985 TMAO circulating values of all MetaCardis participants were $\log_{10}$-transformed and subsequently the
986 median was subtracted and divided by the standard deviation[21] (SD; regularized TMAO values
987 thereafter).

Metagenomic analysis.

Phylogenetic microbiota profiles were built after correction for bacterial load as extensively described[24,29,35,54] using a protocol devised by Vandeputte *et al*[61] with modifications. Briefly, total faecal 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[24,29,34,54].

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[24,29,34,54].

Statistical analyses

Statistical analysis was undertaken using R (v4.03)[62]. 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 10 groups (**Suppl.Table.5**). Specifically, biological parameters included biochemical and clinical serum laboratory tests including lipids, glycated haemoglobin, 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[34], number of antibiotic courses in the last 5years and number of antihypertensive, 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[59], alternative Healthy Eating[63] (aHEI), Dietary Approaches to Stop Hypertension[64] (DASH) and Dietary Diversity[65] (DDS) scores. Clinical risk factors variables included age, systolic and diastolic blood pressure, glycated haemoglobin levels, fasting cholesterol levels, smoking status and waist circumference as described[29]. Serum metabolomics comprised the absolute or relative levels of 116 circulating metabolites determined by GC-MS or UPLC-MS/MS[29,34]. Urine metabolomics included absolute quantification of 47 urine metabolites from ^{1}H-NMR spectra calculated with the IVDr algorithm. Microbiota variables included 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[29,34]. Microbiome (Modules) group included abundance of 116 manually curated bacterial modules[29]. 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)[66], co-opting a strategy from Bar *et. al*[21]. Xgboost consistently outperforms other algorithms in Kaggle competitions for tabular data. For each of our 10 variable groups (**Suppl.Table.5**) we optimized xgboost models using 5-fold cross-validation and two sequential hyperparameter grids searches (972 different parameter combinations for each feature group in total) to predict mean-centred 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 (5 for each round) were saved and used for feature attribution analysis in the left-out group (see below).

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[26] and recently used to objectively evaluate factors driving metabolite plasma levels in humans[21]. 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[26]. Variable SHAP values for each individual in the 5 left-out groups were extracted from corresponding xgboost models trained on the remaining 80% participants for each variable group using SHAPforxgboost (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 Bat *et al.*[21]. For swarm-plots depicting each individual participants SHAP values, variable values were inverse-normalized using the RNOmni (v1.0.0) package.

Clustering

Clustering was performed using the built-in R k-means function with the Hartigan-Wong algorithm using 25 random sets, as described[25].

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 co-variates for 999 iterations were performed with vegan (v2.5.7).

Mediation analysis

We performed mediation analysis[67] 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[68], 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[69] (v.4.1.0) 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

A complete list of antibodies and reagents used in these studies are appended (**Suppl.Table.24)**

***In vitro* experiments**

Cell Culture and Protein Extraction

Primary human adult kidney fibroblasts from a single donor (Cambridge Bioscience) were cultured in 10% FCS low-Glucose DMEM with 1% Pen/Strep antibiotics (Sigma). For acute (up-to 30min) TMAO stimulation, cells (100,000/condition) were serum-starved for 1h in a physiologic serum-free buffer as previously described[43]. Fibroblasts were preincubated with Tramenitib (10nM, Shelleckchem), BAPTA-AM (20µM, Molecular Probes) or vehicle for 30min before TMAO (Sigma) stimulation. For longer-term (24h) stimulation, fibroblasts were in serum-free low-glucose DMEM overnight and

subsequently stimulated with TMAO, TGF-β1 (5ng/ml, R&D), or their combination. At the end of the experiment fibroblasts were lysed and stored at −80°C until further use, as described[42].

Cytosolic [Ca^{2+}] Measurements

Fibroblasts were serum-starved overnight in low-glucose DMEM and subsequently loaded with 5μM fura-2-AM (Molecular Probes) in pH 7.4 Hanks Balanced Salt Solution (HBSS; Sigma) containing Ca^{2+} and Mg^{2+}. Measurements were obtained on an epifluorescence inverted microscope equipped with a ×20 fluorite objective. Single cell intracellular Ca^{2+} ($[Ca^{2+}]_i$) was monitored using excitation at 340 and 380nm, through a monochromator (Cairn Research). Emitted light was reflected through a 515nm 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 6-8weeks of age were fed a control, 0.12% w/w TMAO- or 1% w/w choline-containing diets (Teklad; 6animals/group) for 6weeks[18]. Subsequently, Unilateral Ureter Obstruction (UUO) was established as previously described[42]. Briefly, mice were anesthetized with isoflurane and the abdominal cavity was exposed using midline laparotomy, the right ureter was isolated and tied off 0.5cm from the pelvis. The left ureter was left unclamped and served as the sham-operated control. At day 5 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 16h at 4°C, transferred to a 70% v/v ethanol solution for a further 24h 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[40,41]. 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.

Immunohistochemical Staining

Formalin-fixed kidneys were embedded in paraffin and 4µm sections were cut by the Barts Cancer Institute Pathology Unit. Staining for Sirius red, α-Smooth Muscle Actin (αSMA) and F4/F80 were performed as described[42] with the DISCOVERY XT (Ventana) automated slide processing instrument using the OmniMap reagents (Ventana), according to the manufacturer's recommendations. Images were captured at ×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 percentage of total area, using ImageJ 1.46r (NIH).

References

54. Belda E, Voland L, Tremaroli V, *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* 2022; gutjnl-2021-325753. doi: 10.1136/gutjnl-2021-325753 pmid:https://pubmed.ncbi.nlm.nih.gov/35017197

55. Yancy CW, Jessup M, Bozkurt B, *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* 2017;**136**:e137-e161. doi: 10.1161/CIR.0000000000000509 pmid:https://pubmed.ncbi.nlm.nih.gov/28455343

1132 56. The challenge of obesity in the WHO European region and the strategies for response:
1133 summary. (World Health Organization, Regional Office for Europe, 2007)
1134 https://www.euro.who.int/en/publications/abstracts/challenge-of-obesity-in-the-who-european-
1135 region-and-the-strategies-for-response-the
1136 57. Alberti KG, Zimmet P, Shaw J, *et al*. The metabolic syndrome--a new worldwide definition.
1137 *Lancet* 2005;**366**:1059-62.doi: 10.1016/S0140-6736(05)67402-8
1138 pmid:https://pubmed.ncbi.nlm.nih.gov/16182882
1139 58. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes*
1140 *Care* 2004;**27** Suppl 1:S5-S10.doi: 10.2337/diacare.27.2007.s5
1141 pmid:https://pubmed.ncbi.nlm.nih.gov/14693921
1142 59. Verger EO, Armstrong P, Nielsen T, *et al*. Dietary Assessment in the MetaCardis Study:
1143 Development and Relative Validity of an Online Food Frequency Questionnaire. *J Acad Nutr Diet*
1144 2017;**117**:878-888.doi:10.1016/j.jand.2016.10.030 pmid:https://pubmed.ncbi.nlm.nih.gov/2802480
1145 60. Levey AS, Stevens LA, Schmidt CH, *et al*. A new equation to estimate glomerular filtration
1146 rate. *Ann Intern Med* 2009;**150**:604-12.doi: 10.7326/0003-4819-150-9-200905050-00006
1147 pmid:https://pubmed.ncbi.nlm.nih.gov/19414839
1148 61. Vandeputte D, Kathagen G, D'hoe K. *et al*. Quantitative microbiome profiling links gut
1149 community variation to microbial load. *Nature* 2017;**551**:507-511.doi:10.1038/nature24460
1150 pmid:https://pubmed.ncbi.nlm.nih.gov/29143816
1151 62. R Core Team (2020). R: A language and environment for statistical computing. R Foundation
1152 for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.
1153 63. InterAct Consortium. Adherence to predefined dietary patterns and incident type 2 diabetes
1154 in European populations: EPIC-InterAct Study. *Diabetologia* 2014;**57**:321-33.doi:10.1007/s00125-
1155 013-3092-9 pmid:https://pubmed.ncbi.nlm.nih.gov/24196190

1156 64. Tyson CC, Lin PH, Corsino L, *et al*. Short-term effects of the DASH diet in adults with
1157 moderate chronic kidney disease: a pilot feeding study. *Clin Kidney J* 2016;**9**:592-8.
1158 doi:10.1093/ckj/sfw046 pmid:https://pubmed.ncbi.nlm.nih.gov/27478603
1159 65. Jeurnink SM, Büchner FL, Bueno-de-Mesquita HB, *et al*. Variety in vegetable and fruit
1160 consumption and the risk of gastric and esophageal cancer in the European Prospective Investigation
1161 into Cancer and Nutrition. *Int J Cancer* 2012;**131**:E963-73.doi:10.1002/ijc.27517
1162 pmid:https://pubmed.ncbi.nlm.nih.gov/22392502
1163 66. Chen T., & Guestrin C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of
1164 the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–
1165 794). New York, NY, USA: ACM. https://doi.org/10.1145/2939672.2939785
1166 67. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. *Annu Rev Psychol* 2007;**58**:593-
1167 614. doi:10.1146/annurev.psych.58.110405.085542
1168 pmid:https://pubmed.ncbi.nlm.nih.gov/16968208/
1169 68. Tingley D, Yamamoto T, Hirose K *et al*. mediation: R Package for Causal Mediation Analysis. *J*
1170 *Stat Softw* 2014;**59**:1–38. doi:10.18637/jss.v059.i05
1171 69. Ho DE, Imai K, King G, *et al*. MatchIt: Nonparametric Preprocessing for Parametric Causal
1172 Inference. *J Stat Softw* 2011;**42**:1-28. doi:10.18637/jss.v042.i08