Development and validation of a metabolite score for red meat intake: an observational cohort study and randomized controlled dietary intervention

**Authors**

Chunxiao Li1, Fumiaki Imamura1, Roland Wedekind2, Isobel D Stewart1, Maik Pietzner1,3, Eleanor Wheeler1, Nita G Forouhi1, Claudia Langenberg1,3, Augustin Scalbert2, Nicholas J Wareham1\*

**Affiliations**

1MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge CB2 0QQ, UK

2Nutrition and Metabolism Branch, International Agency for Research on Cancer, World Health Organization, 150 Cours Albert Thomas, 69372 Lyon CEDEX 08, France

3Computational Medicine, Berlin Institute of Health at Charité–Universitätsmedizin Berlin, 10117 Berlin, Germany

\*Corresponding author. Email: nick.wareham@mrc-epid.cam.ac.uk

**The running title**

A metabolite score for red meat intake

**Conflict of Interest**

None of the authors reported a conflict of interest related to the study.

**Disclaimer**

Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

Dr Fumiaki Imamurais a member of the ASN Statistical Review Board.

**Data Share Statement**: The data described in the manuscript and analytic code will be made available upon request, pending application and approval. The meta-data for the EPIC-Norfolk study including the study dictionary are freely available without restriction at <https://www.epic-norfolk.org.uk/>.

**Clinical trial registration number:** NCT03354130

**Sources of support:**

The EPIC-Norfolk study (<https://doi.org/10.22025/2019.10.105.00004>) has received funding from the Medical Research Council (MR/N003284/1 MC-UU\_12015/1 and MC\_UU\_00006/1) and Cancer Research UK (C864/A14136). Metabolite measurements in the EPIC-Norfolk study were supported by the MRC Cambridge Initiative in Metabolic Science (MR/L00002/1) and the Innovative Medicines Initiative Joint Undertaking under EMIF grant agreement no. 115372. N.J.W, N.G.F, F.I, I.D.S, M.P, E.W, and C.L acknowledge funding from the Medical Research Council Epidemiology Unit MC\_UU\_00006/1 and MC\_UU\_00006/3; NJW and NGF, from NIHR Cambridge Biomedical Research Centre: nutrition, diet, and lifestyle research theme (IS-BRC-1215-20014). C.Li was supported by a Jardine-Cambridge Graduate Scholarship.

**Abbreviations used:**

7dDD, 7-day diet diary;

EPIC-Norfolk, European Prospective Investigation into Cancer and Nutrition-Norfolk;

FFQ, food frequency questionnaire;

GPE, glycerophosphoethanolamine;

GPC, glycerophosphocholine;

g/d, grams per day;

HR, hazard ratio;

IARC, International Agency for Research on Cancer;

MS, mass spectrometry;

RCT, randomized controlled trial;

SD, standard deviation;

T2D, type 2 diabetes;

TMAO, trimethylamine N-oxide;

UPLC, ultra-performance liquid chromatography.

# Abstract (298 words)

**Background:** Self-reported meat consumption is associated with disease risk but objective assessment of different dimensions of this heterogeneous dietary exposure in observational and interventional studies remains challenging.

**Objective:** To derive and validate scores based on plasma metabolites for types of meat consumption. For the most predictive score, we aimed to test whether the included metabolites varied with change in meat consumption, and whether the score was associated with incidence of type 2 diabetes (T2D) and other non-communicable diseases.

**Methods:** We derived scores based on 781 plasma metabolites for red meat, processed meat and poultry consumption assessed with 7-day food records among 11,432 participants in the European Prospective Investigation into Cancer and Nutrition-Norfolk (EPIC-Norfolk) cohort. The scores were then tested for internal validity in an independent subset (n=853) of the same cohort. In focused analysis on the red meat metabolite score, we examined whether the metabolites constituting the score were also associated with meat intake in a randomized cross-over dietary intervention trial of meat (n=12, Lyon, France, NCT03354130). In the EPIC-Norfolk Study, we assessed the association of the red meat metabolite score with T2D incidence (n=1,478) and other health endpoints.

**Results:** The best performing score was for red meat, comprising 139 metabolites which accounted for 17% of the explained variance of red meat consumption in the validation set. In the intervention, 11 top-ranked metabolites in the red meat metabolite score increased significantly after red meat consumption. In the EPIC-Norfolk study, the red meat metabolite score was associated with T2D incidence (adjusted hazard ratio per standard deviation=1.17 (95% confidence interval: 1.10, 1.24)).

**Conclusions:** The red meat metabolite score derived and validated in this study contains metabolites directly derived from meat consumption and is associated with T2D risk. These findings suggest the potential for objective assessment of dietary components and their application for understanding diet-disease associations.

# Key words

Metabolomics, meat, prediction, biomarker, diabetes

# Introduction (332 words)

Meat is an important component of the human diet and high consumption is a risk factor for many non-communicable diseases, including type 2 diabetes (T2D) (1–5). However, meat consumption is a heterogeneous exposure and assessing total meat intake and specific subtypes such as red meat, processed meat and poultry in epidemiological studies that evaluate its influence on health outcomes remains challenging.

Metabolite profiling is a promising approach for quantifying habitual meat intake and can be a complementary approach to self-reported dietary assessment methods (e.g., food frequency questionnaires (FFQs) or dietary records) (6,7). Diet is an important determinant of the plasma metabolome and one study estimated that it accounts for 50% of the explainable variance, compared to 2% of the variance explained by lifestyle factors, including smoking status, exercise time, etc. (7). Measurement of metabolites as a complement to self-reported assessment methods has other theoretical advantages, including diminishing social desirability bias and recall bias, and greater comparability across populations (8,9).

Several individual metabolites have previously been reported to be significantly associated with different types of meat consumption (10–13). However, few studies have examined how combinations of metabolites can predict meat consumption. Cuparencu *et al* reported that a combination of six metabolite biomarkers were able to assign people to a binary classification of red meat consumption in a 2-day feeding trial. However, the study was small and the result may be liable to overfitting (12).

In the current study, we aimed to develop and test metabolite scores for different types of meat consumption by combining 781 blood metabolites in the European Prospective Investigation into Cancer and Nutrition-Norfolk (EPIC-Norfolk) cohort and then to take forward the red meat metabolite score to potential replication in a short-term randomized controlled trial (RCT) that measured metabolites after a red meat and a non-meat diet. Finally we tested whether the meat metabolite score was associated with the risk of incident T2D and other non-communicable diseases to explore the potential utility of the score in understanding disease risk.

# Methods (1704 words)

## Data source and study design

The overall design of the project includes a derivation and validation phase in an observational study, a test of change in an RCT and a test of association with incident health outcomes in a prospective study as shown in ***Figure 1***.

## Observational data for the derivation and validation of the metabolite scores: the EPIC-Norfolk study

We developed and validated the metabolite scores for three types of meat consumption (red meat, processed meat and poultry), using baseline data from the EPIC-Norfolk study which originally recruited 25,639 men and women aged 40-79 years between 1993 and 1998 in the United Kingdom. Details of the recruitment procedures and data collection have been described previously (14). Briefly, baseline characteristics for all participants were collected, including socio-demographic factors (age, sex, and education level), health behaviors (smoking status, alcohol drinking, and physical activity), and dietary measures. Blood samples were collected at baseline and stored in liquid nitrogen at -175oC. The EPIC-Norfolk study was approved by the Norwich Local Ethics Committee (REC Ref: 98CN01); all participants gave their informed written consent before entering the study.

We developed metabolite scores for different types of meat consumption in an exploratory set which included 11,432 participants who had both untargeted metabolomics and dietary data. We excluded from this exploratory dataset individuals who were part of a nested case-cohort study for incident T2D; those with extreme energy intake measures (< 500 and > 3500 kcal/d for women, < 800 and > 4200 kcal/d for men); or those with prevalent diabetes at baseline.

Participants from the subcohort of an independent nested T2D case-cohort study (15) were used as a validation set, which included 853 participants after exclusions.

### Metabolomics measurement and data processing in the EPIC-Norfolk study

We measured untargeted metabolomics using ultra-performance liquid chromatography-tandem mass spectrometry (UPLC MS/MS) on the Metabolon DiscoveryHD4® platform from plasma samples collected at baseline. The measurement of metabolites was performed in three subsets in March 2015, January 2016 and March 2017 successively. The data quality control and processing methods have been described previously (16) and are summarized in the *Supplementary methods*. After data quality control and data management, three subsets included 1503, 5992 and 5980 individuals, in which 944, 1168 and 1219 metabolites were measured, respectively, and 781 metabolites were identical across all subsets.

Before analysis, we undertook the following steps for each metabolite within each subset: log-transformation, replacement of outliers with 5 standard deviations (SDs) from the mean (Winsorization) and standardization to a mean of 0 and SD of 1. For metabolite concentrations that were below the limit of detection, we imputed them with the lowest values of that metabolite (17). The different subsets in the exploratory dataset were measured at different time points, and we adjusted for measurement time period in the regression analysis.

### Assessment of meat consumption in the EPIC-Norfolk study

Meat consumption and other dietary exposures were assessed with a 7-day diet diary (7dDD) as documented previously (18). Briefly, on the day of a baseline assessment that included blood sampling, participants were asked to record everything they had eaten (food types, amounts, brands, recipes, and cooking methods) prospectively for the following seven consecutive days. The dietary information was then processed into food and nutrient data by programs and databases (DINER and DINERMO) using standard protocols (18,19). The meat related categories were all disaggregated from composite dishes including red meat (unprocessed beef, lamb, pork, veal, rabbit, venison etc.), processed meat (bacon, ham and sausages etc., smoked, cured, salted or chemically-preserved), and poultry (chicken, turkey, goose, duck, guinea fowl, pheasant etc.) in the unit of grams per day (g/d). Participants were also asked whether they followed a special diet (vegetarian, other diet or no special diet).

### Development and validation of metabolite scores of self-reported red, processed meat and poultry consumption

In the EPIC-Norfolk study, 781 metabolites were evaluated simultaneously for the prediction of red meat consumption. In the exploratory set, we applied elastic net regression (20) with a bootstrap approach (21,22) to select a combination of metabolites for prediction of red meat consumption; and ridge regression (23) to estimate penalized weights of these candidate metabolites (***Supplementary Methods***). We applied the weights of all candidate metabolites and constructed a metabolite score for each individual in both of the derivation and validation datasets. The score was standardized to a mean of 0 and SD of 1 for further analysis. The metabolite scores for processed meat and poultry were derived and tested using the same process.

### Randomized controlled trial of meat consumption

Given the availability of trial-based data for meat consumption, we further investigated associations of metabolites in the score from the observational EPIC-Norfolk study with red meat consumption in an RCT previously conducted in Lyon, France in 2018. The details of this RCT have been reported previously (24). In brief, 12 healthy adults consumed in random order 5 different foods (fried pork, hot dogs, bacon, salami and tofu) as part of a controlled diet. For this analysis, we examined the differences in metabolites levels between fried pork (unprocessed red meat) and tofu control arms. In this trial, fasting plasma samples were collected in the morning after the last meal of each test period. Participants provided informed consent and procedures were carried out according to the Declaration of Helsinki. The study was approved by the International Agency for Research on Cancer (IARC) Ethics Committee (IEC Project 17-12) and registered at *clinicaltrials.gov* (NCT03354130).

### Test of candidate metabolites of red meat intake in the RCT

We assessed whether metabolites that were part of the metabolite score for red meat intake were increased after intake of fried pork (red meat) compared to tofu in the RCT. The process of identification of metabolites that make up the red meat metabolite score in the RCT is shown in ***Supplementary Figure 1****.* First, we focused on metabolites that had been annotated successfully in the IARC laboratory and had a positive coefficient in the metabolite score. Corresponding signal intensities were extracted with Agilent Profinder 10.0 (Agilent Technologies, Santa Clara, CA, USA) using the find-by-formula method ([M+H]+ and [M-H]- ions only, exact mass +/- 8ppm, Retention time +/- 0.05 min). Metabolites were carried forward for statistical analysis if they were detected in >75 % of the samples collected after pork intake. Then we used paired Welch’s t-tests to assess whether metabolites were significantly (p < 0.05) elevated in plasma samples collected after pork intake compared to tofu intake. Second, for metabolites not previously identified in the IARC laboratory, we extracted only those with a coefficient of >1.0 in the meat intake score from the raw data by formula only to test for their increase in plasma samples after pork intake. Compounds were confirmed by comparison of MS/MS spectra with those in the literature (annotation confidence level 2 or 3) (25).

## Prospective cohort analysis of the association of the red meat metabolite score with incident disease outcomes in the EPIC-Norfolk study

We examined the association of the red meat metabolite score and the relevant self-reported consumption parameter with the risk of incident T2D in a case-cohort study nested in the EPIC-Norfolk cohort (15). This comprised a total of 659 incident cases of T2D and a comparison subcohort of 846 participants, which had an overlap by design of 27 individuals with the case set, after we excluded participants who had extreme energy intake measures or missing covariate data.

### Ascertainment of T2D cases in the EPIC-Norfolk study

Incident cases of T2D were ascertained by reviewing evidence from multiple sources, including self-report, linkage to primary and secondary care registers, medication use from drug registers, hospital admissions, and mortality data. All self-reported cases were verified with independent evidence. Person time of follow up was determined from the date of baseline assessment to the date of diagnosis, date of death, or 31 December 2006, whichever came first.

### Assessment of covariates in the EPIC-Norfolk study

Information about health behaviors and clinical risk factors were collected by trained nurses during a health check at baseline. Information obtained included age, sex, education level (primary school or no qualifications, middle school or equivalent, high school or equivalent, college degree and above), smoking status (never, former, and current smokers), alcohol drinking (g/d), physical activity (inactive, moderately inactive, moderately active, active), height (m), weight (kg), and other food group consumption in g/d ( fruits, vegetables, fatty fish, white fish, dairy, legumes, nuts, eggs and sugar sweetened beverages). BMI was calculated as weight divided by the square of height (kg/m2). Total energy intake was calculated from the 7dDD.

### Statistical methods for the assessment of the association with incident T2D

We analyzed the association of a standardized metabolite score for red meat consumption with incident T2D in the case-cohort study using Prentice-weighted Cox regression (26) to estimate the hazard ratio (HR) for T2D and its 95% confidence interval per SD of the exposure.

We considered the effect of potential confounders in a model adjusting for age, sex, and then further adjusted for education, smoking status, alcohol drinking, BMI and dietary factors (consumption of fruits, vegetables, fatty fish and white fish, sugar sweetened beverages, dairy, legumes, nuts, eggs and total energy intake). For alcohol drinking and BMI, we included their linear and squared terms to account for their potential non-linear associations with each outcome.

### Ascertainment of other non-communicable diseases outcomes in the EPIC-Norfolk study

We ascertained the incident outcomes of six health conditions including cardiovascular diseases (including ischemic heart disease, hemorrhagic stroke, cerebral stroke, heart failure, and atrial fibrillation); gastrointestinal cancers (including colon cancer, rectal cancer, stomach cancer); liver disease, renal disease, fractures, and deaths due to any causes (16). Outcome data were obtained by linkage to Hospital Episode Statistics, the cancer registry and the Office of National Statistics. Follow-up ended on March 31st, 2016. Prevalent and incident cases for each disease were identified with the International Classification of Diseases 10th revision as listed in ***Supplementary Table 1***.

### Statistical methods for the assessment of the association with multiple disease outcomes

In an exploratory analysis we tested the association of the red meat metabolite score with incident health outcomes using standard Cox regression after excluding the prevalent cases for each clinical outcome (see *Supplementary Table 1)*. We adjusted for the same sets of potential confounders as considered in the association with T2D.

# Results (808 words)

## Baseline characteristics and meat consumption of study participants in the EPIC-Norfolk study

The baseline characteristics of the participants in the exploratory and validation sets within the EPIC-Norfolk study are shown in ***Table 1***. Among the 11,432 participants in the exploratory set, 46% were male and the mean (SD) age at baseline was 59.6 (9.0) years. The mean (SD) meat consumption in g/d was 34.4 (29.3) for red meat, 22.5 (21.0) for processed meat, and 24.8 (27.5) for poultry. The characteristics in the validation set (n=853) were broadly similar to those in the exploratory set.

## Development and validation of metabolite scores for meat consumption

In the exploratory set in the EPIC-Norfolk study, 139 metabolites were identified to be associated with red meat consumption, and they were assembled into a composite red meat metabolite score. This score was made up of 49 (19.3%) lipids and 30 (22.2%) amino acids, other metabolite classes such as xenobiotics (n=14, 12.5%) and 36 (18.4%) unknown metabolites (***Figure 2****,* ***Supplementary Table 2***). The top 5 metabolites with positive coefficients were 1-(1-enyl-stearoyl)-2-arachidonoyl- glycerophosphoethanolamine (GPE) (P-18:0/20:4), 1-(1-enyl-stearoyl)-2-arachidonoyl- glycerophosphocholine (GPC) (P-18:0/20:4), 1-margaroyl-2-oleoyl-GPC (17:0/18:1), trans-4-hydroxyproline, and Verapamil. The derived metabolite score for red meat consumption achieved an explained variance of 24% and 17% in the exploratory and validation sets. The metabolite score for red meat intake was associated with quintiles of self-reported meat intake (***Figure 3***). It was also significantly higher in the subgroups of self-reported red meat consumers and non-vegetarians, compared to non-consumers of red meat and vegetarians, respectively.

The metabolite scores for processed meat consumption and poultry consumption consisted of 82 and 49 predictive metabolites, respectively, and were made up predominantly of lipids and amino acids (*Figure 2****, Supplementary Tables 3 and 4***). The overlapping and distinct sets of metabolites that were associated with red meat, processed meat and poultry consumption are shown in *Figure 2*. Six metabolites were included in all three metabolite scores: trans-4-hydroxyproline, trimethylamine N-oxide (TMAO), methionine sulfone, sphingomyelin (d18:2/14:0, d18:1/14:1), N-acetylputrescine, and X-11849. Overall the 7dDD meat intake variance explained by the corresponding metabolite scores in the validation set was 15% for processed meat and 13% for poultry (***Supplementary Figure 2***). The details of parameter optimization and metabolites selection in the bootstrapping process are shown in ***Supplementary Tables 5-7.*** In additional analyses, we estimated glomerular filtration rate (eGFR) as an indicator of renal function and assessed the associations between red meat intake and each of 781 metabolites after statistical adjustment for eGFR, showing that the coefficients were unchanged from the analysis that was not adjusted for eGFR.

## Associations of metabolites in the red meat metabolite score with meat intake in an RCT

For the metabolites that were part of the metabolite score for red meat intake, we used untargeted plasma metabolomics data from a meat RCT to investigate the differences of metabolite concentrations after a 3-day red meat intervention compared to a non-meat diet. Of the 50 known metabolites positively associated with self-reported red meat consumption in the observational EPIC-Norfolk study, 11 were identified in the RCT and significantly increased after fried pork (red meat) intake compared to tofu: several glycerophospholipids, 4-hydroxyproline, TMAO, creatine, deoxycarnitine and stearoylcarnitine (***Table 2****,* ***Supplementary Figure 3 and 4***). The correlations between these top-ranked metabolites and types of meat consumption in the EPIC-Norfolk study are shown in ***Supplementary Figure 5***. The correlations between these top-ranked metabolites are reported in ***Supplementary Table 8***. Of the top 8 metabolites that had the highest coefficients in the red meat metabolite score in the EPIC-Norfolk study, 6 were validated in the RCT.

## Association of the red meat metabolite score with T2D

The baseline characteristics of the participants in the T2D case-cohort are presented in ***Table 3***. In the subcohort, participants with higher metabolite scores of red meat consumption were more likely to be male, current smokers, have higher BMI, higher consumption of alcohol, legumes, sugar sweetened beverages and total energy, and have lower levels of fruit, and fish consumption, compared to participants with lower metabolite scores.

In a prospective analysis with a median follow-up of 10 years, the metabolite score for red meat consumption was positively associated with a higher risk of incident T2D (HR per SD =1.17 [1.10, 1.24]) after adjusting for potential confounding factors (***Figure 4***). There was a significant association between self-reported red meat consumption and incident T2D (HR per SD=1.08 [1.03, 1.14]).

## Association of the red meat metabolite score with other health outcomes

In an exploratory analysis, we examined the association of the red meat metabolite score with six health outcomes. In an adjusted analysis, a higher red meat metabolite score was significantly associated with higher risk of incident cardiovascular disease (1.04 [1.00, 1.09] per SD) and gastrointestinal cancers (1.16 [1.03, 1.29]). The estimates of associations for meat intake using 7dDD measurements were similar to those using the derived scores but the *p*-values were generally smaller (*Figure 4*).

# Discussion (1212 words)

In this paper we report the development and validation of metabolite scores for three different types of meat consumption: red meat, processed meat and poultry, based on untargeted plasma metabolomics data and 7dDD data in a large British cohort with comprehensive phenotypes. In focused analysis on the red meat metabolite score, we found that eleven top-ranked metabolites in the score were associated with red meat intake in an RCT suggesting a causal link between red meat intake and change of these metabolites. Finally, we found that the red meat metabolite score was associated with T2D incidence and potentially also associated with other cardiometabolic diseases.

## Metabolite scores of meat consumption

Previous evidence on combining biomarkers into scores to measure meat intake is limited. A trial in Denmark indicated that combinations of several metabolic biomarkers of red meat intake were more efficient than a single biomarker to classify red meat consumers compared to other participants (12). However, previous studies had not evaluated a dose-response association between meat intake and a combination of biomarkers. In this large population-based study, we estimated the absolute amounts of meat intake with 7dDD, which provides more accurate estimates than a FFQ to rank consumption levels (27). Our results indicate the utility of untargeted metabolomics to generate an overall score to predict the level of meat intake rather than only being able to discriminate between consumers and non-consumers.

The metabolite scores of meat consumption were characterized by a wide range of metabolites, including lipids, amino acids, and xenobiotics. Several metabolites that constitute the derived scores have been identified by previous studies, such as TMAO, trans-4-hydroxyproline, creatine, and stearoylcarnitine (10,11,28). Specifically, an RCT in the United States (n=113) reported that TMAO in plasma significantly increased after red meat consumption compared to consumption of poultry or non-meat products. Positive associations of plasma TMAO levels with risk of cardiovascular disease, diabetes and all-cause mortality have been reported in several meta-analyses of clinical studies (29–31). These results suggest that TMAO might be involved as part of underlying mechanisms between red meat intake and the development of chronic disease. In addition to metabolites in the score of red meat intake, several metabolites specific to processed meat (e.g., o-cresol sulphate) (32,33) or poultry consumption (e.g., 3-methylhistidine) (10) in our study were also reported by previous intervention studies.

We also identified several yet unreported metabolites that were associated with red meat consumption in both the observational study and the RCT, in particular several plasmalogens, such as 1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4), 1-margaroyl-2-oleoyl-GPC (17:0/18:1) and 1-palmityl-GPC (O-16:0). Plasmalogens, a subclass of membrane glycerophospholipids, contain a vinyl-ether bond at the sn-1 position and are enriched in polyunsaturated fatty acids at the sn-2 position of the glycerol backbone (34). Mazzilli et al. found several plasmalogens were correlated with self-reported red meat (35). However, most of plasmalogens identified in our study were not reported in that previous study, partly due to different platforms used to measure and annotate metabolites in different studies. These compounds present a very promising group of potential new biomarkers for meat intake. Their role in meat metabolism and disease development is largely unknown and warrants additional investigation. Some drug metabolites were also identified in the red meat metabolite score, such as Verapamil and Ranitidine. These metabolites were detected in only a small number of participants (***Supplementary Table 9***), so are likely to represent a sub-group of patients who have chronic disease and are taking these drugs (Verapamil for cardiac illness and Ranitidine for gastrointestinal illness). Since these conditions could themselves be linked to red meat consumption, it is likely that the association between these drug metabolites and dietary behavior is confounded by indication.

One group of metabolites which make major contributions to the red meat metabolite score are small meat-derived molecules with short half-lives (e.g., TMAO, trans-4 hydroxyproline or creatine) (10). These compounds are unlikely to be good long-term markers of meat intake in people who consume meat occasionally because the metabolites would be cleared from the body relatively quickly. By contrast, these biomarkers may reflect regular red meat intake well. The second group of metabolites that rank highest in the score are lipophilic compounds (e.g., plasmalogens). These compounds have half-lives of days or even weeks and are likely to be good markers of long-term dietary habits (36,37), including the identification of foods that are consumed rarely. These two groups of metabolites in the meat metabolite score ensure that the score reflects not only recent food intake but also dietary intake over a longer time frame. The focus on longer term habitual intake as well as short-term intake is a strength of this study not only in respect of the biomarkers, but also of the 7dDDs which have previously been shown to capture short-term and habitual dietary intakes (38).

## Associations with T2D risk

The red meat metabolite score, as a proxy for red meat intake, showed a positive association with T2D risk consistent with results from several large cohort studies that have reported associations with T2D risk with self-reported intake as dietary exposures (3,4,39,40). The score-derived association appeared to be comparable in magnitude with that using 7dDD-measured meat intake. Similar results were reported in a study on a metabolomics signature of the Mediterranean diet and its association with cardiovascular disease risk (41). Future evaluations of the additional complementary information that can be obtained by measurement of metabolites over and above traditional dietary assessment methods should include investigation of cost-effectiveness and predictive utility for disease outcomes.

## Strengths and limitations

To our knowledge, this study was the first of this kind to develop and validate a metabolite score for red meat intake in a large population study which has comprehensive dietary measurements and metabolomics data. Metabolite profiling provided a complementary approach to assess types of meat consumption objectively. The application of metabolomics to a meat intervention trial provided additional evidence on biological plausibility and reproducibility of the red meat metabolite score. Additionally, in the EPIC-Norfolk study, a long follow-up with detailed information of multiple incident diseases enabled us to examine associations between the meat metabolite score and multiple health outcomes simultaneously.

Several limitations warrant discussion. Firstly, the study was based on a British population so generalizability is limited for other populations and further validation studies should be considered. Secondly, although we have adjusted for a comprehensive set of confounders to examine the association between the red meat metabolite score and risks of non-communicable diseases, the results may be affected by residual confounding. Thirdly, while we have tested the change of metabolites after meat intervention in a trial, the limited number of red meat products and the limited size of the trial hindered a comprehensive validation analysis. The potential causal links between meat intake and most of the candidate metabolites are largely unknown. Many metabolites in the score are probably not directly influenced by meat intake, but affected by factors that are correlated with meat intake, such as BMI or derived from metabolic or physiological processes. Also, we might be unable to validate metabolites that reflect long-term diets because the feeding study tested short-term exposures. However, the most important metabolites were validated in the RCT and the score correlated well with meat intake in the validation set. Further validation studies with a wider range of confirmed metabolites in other populations are needed.

# Conclusion (38 words)

This study suggests that a metabolite score derived from untargeted metabolomics profile in plasma has the potential to reflect red meat consumption and inform the study of the association of red meat consumption, assessed objectively, with clinical outcomes.

# Acknowledgements

We thank all the participants who have been part of the project and to the many members of the study teams at the University of Cambridge who have enabled this research. We also thank Junqing Xie for insightful discussion.

# Author contributions

C.Li, F.I, and N.J.W designed the research; C.Li and R.W analyzed the data; C.Li, F.I, N.J.W, and R.W drafted the manuscript; R.W conducted the laboratory analyses of the intervention study; C.Li, F.I, R.W, A.S, and N.J.W interpreted the data; I.D.S, M.P, E.W, N.G.F, and C.L provided administrative, technical or material support; N.J.W had primary responsibility for final content; and all authors revised and approved the final manuscript.

# References

1. Hoogendijk EO, Afilalo J, Ensrud KE, Kowal P, Onder G, Fried LP. Frailty: implications for clinical practice and public health. The Lancet 2019;394(10206): 1365–75.

2. Godfray HCJ, Aveyard P, Garnett T, Hall JW, Key TJ, Lorimer J, Pierrehumbert RT, Scarborough P, Springmann M, Jebb SA. Meat consumption, health, and the environment. Science 2018;361(6399).

3. Yang X, Li Y, Wang C, Mao Z, Zhou W, Zhang L, Fan M, Cui S, Li L. Meat and fish intake and type 2 diabetes: Dose–response meta-analysis of prospective cohort studies. Diabetes Metab 2020;46(5):345–52.

4. Neuenschwander M, Ballon A, Weber KS, Norat T, Aune D, Schwingshackl L, Schlesinger S. Role of diet in type 2 diabetes incidence: umbrella review of meta-analyses of prospective observational studies. BMJ 2019;366:l2368.

5. Bouvard V, Loomis D, Guyton KZ, Grosse Y, Ghissassi F El, Benbrahim-Tallaa L, Guha N, Mattock H, Straif K, Stewart BW, et al. Carcinogenicity of consumption of red and processed meat. The Lancet Oncol 2015;1616):1599–600.

6. Nicholson JK, Holmes E, Kinross JM, Darzi AW, Takats Z, Lindon JC. Metabolic phenotyping in clinical and surgical environments. Nature 2012;491(7424):384–92.

7. Bar N, Korem T, Weissbrod O, Zeevi D, Rothschild D, Leviatan S, Kosower N, Lotan-Pompan M, Weinberger A, Le Roy CI, et al. A reference map of potential determinants for the human serum metabolome. Nature 2020;588(7836):135–40.

8. Shim J-S, Oh K, Kim HC. Dietary assessment methods in epidemiologic studies. Epidemiol Health 2014;36:e2014009.

9. Guasch-Ferre M, Bhupathiraju SN, Hu FB. Use of Metabolomics in Improving Assessment of Dietary Intake. Clin Chem 2018;64(1):82–98.

10. Cheung W, Keski-Rahkonen P, Assi N, Ferrari P, Freisling H, Rinaldi S, Slimani N, Zamora-Ros R, Rundle M, Frost G, et al. A metabolomic study of biomarkers of meat and fish intake. Am J Clin Nutr 2017;105(3):600–8.

11. Wedekind R, Kiss A, Keski-Rahkonen P, Viallon V, Rothwell JA, Cross AJ, Rostgaard-Hansen AL, Sandanger TM, Jakszyn P, Schmidt JA, et al. A metabolomic study of red and processed meat intake and acylcarnitine concentrations in human urine and blood. Am J Clin Nutr 2020;112(2):381–8.

12. Cuparencu C, Rinnan A, Dragsted LO. Combined Markers to Assess Meat Intake-Human Metabolomic Studies of Discovery and Validation. Mol Nutr Food Res 2019; 63(17):e1900106.

13. Mitry P, Wawro N, Rohrmann S, Giesbertz P, Daniel H, Linseisen J. Plasma concentrations of anserine, carnosine and pi-methylhistidine as biomarkers of habitual meat consumption. Eur J Clin Nutr 2019;73(5): 692-702;

14. Day N, Oakes S, Luben R, Khaw KT, Bingham S, Welch A, Wareham N. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. Br J Cancer 1999;80:95–103.

15. Forouhi NG, Ye Z, Rickard AP, Khaw KT, Luben R, Langenberg C, Wareham NJ. Circulating 25-hydroxyvitamin D concentration and the risk of type 2 diabetes: results from the European Prospective Investigation into Cancer (EPIC)-Norfolk cohort and updated meta-analysis of prospective studies. Diabetologia 2012; 55(8): 2173-82.

16. Pietzner M, Stewart ID, Raffler J, Khaw K-T, Michelotti GA, Kastenmüller G, Wareham NJ, Langenberg C. Plasma metabolites to profile pathways in noncommunicable disease multimorbidity. Nat Med 2021; 47(3): 471-79.

17. Wang Y, Gapstur SM, Carter BD, Hartman TJ, Stevens VL, Gaudet MM, McCullough ML. Untargeted metabolomics identifies novel potential biomarkers of habitual food intake in a cross-sectional study of postmenopausal women. J Nutr 2018;148(6):932–43.

18. Lentjes MAH, McTaggart A, Mulligan AA, Powell NA, Parry-Smith D, Luben RN, Bhaniani A, Welch AA, Khaw KT. Dietary intake measurement using 7 d diet diaries in British men and women in the European Prospective Investigation into Cancer-Norfolk study: a focus on methodological issues. Br J Nutr 2014;111(3):516–26.

19. Welch AA, McTaggart A, Mulligan AA, Luben R, Walker N, Khaw KT, Day NE, Bingham SA. DINER (Data Into Nutrients for Epidemiological Research) - a new data-entry program for nutritional analysis in the EPIC-Norfolk cohort and the 7-day diary method. Public Health Nutr 2001;4:1253–65.

20. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Statist Soc B 2005;67:301–20.

21. Abram S V., Helwig NE, Moodie CA, DeYoung CG, MacDonald AWI, Waller NG. Bootstrap Enhanced Penalized Regression for Variable Selection with Neuroimaging Data. Front Neurosci 2016;10:344.

22. Bunea F, She Y, Ombao H, Gongvatana A, Devlin K, Cohen R. Penalized least squares regression methods and applications to neuroimaging. NeuroImage 2011;55:1519–27.

23. Hoerl AE, Kennard RW. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics 1970;12:55–67.

24. Wedekind R, Keski-Rahkonen P, Robinot N, Viallon V, Ferrari P, Engel E, Boutron-Ruault MC, Mahamat-Saleh Y, Mancini FR, Kühn T, et al. Syringol metabolites as new biomarkers for smoked meat intake. Am J Clin Nutr 2019;110:1424–33.

25. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW, Fiehn O, Goodacre R, Griffin JL, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 2007;3:211–21.

26. Onland-Moret NC, van der A DL, van der Schouw YT, Buscher W, Elias SG, van Gils CH, Koerselman J, Roest M, Grobbee DE, Peeters PH. Analysis of case-cohort data: a comparison of different methods. J Clin Epidemiol 2007;60(4):350-5.

27. Bingham SA, Welch AA, McTaggart A, Mulligan AA, Runswick SA, Luben R, Oakes S, Khaw KT, Wareham N, Day NE. Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. Public Health Nutr 2001;4(3):847–58.

28. Cuparencu C, Praticó G, Hemeryck LY, Sri Harsha PSC, Noerman S, Rombouts C, Xi M, Vanhaecke L, Hanhineva K, Brennan L, et al. Biomarkers of meat and seafood intake: An extensive literature review. Genes and Nutrition. 2019;14(35).

29. Heianza Y, Ma W, Manson JAE, Rexrode KM, Qi L. Gut microbiota metabolites and risk of major adverse cardiovascular disease events and death: A systematic review and meta-analysis of prospective studies. J Am Heart Assoc 2017;6(7):e004947

30. Farhangi MA. Gut microbiota-dependent trimethylamine N-oxide and all-cause mortality: Findings from an updated systematic review and meta-analysis. Nutrition 2020; 78:110856.

31. Zhuang R, Ge X, Han L, Yu P, Gong X, Meng Q, Zhang Y, Fan H, Zheng L, Liu Z, et al. Gut microbe–generated metabolite trimethylamine N-oxide and the risk of diabetes: A systematic review and dose-response meta-analysis. Obesity Reviews 2019;20:883–94.

32. Shortt C, Hasselwander O, A Meynie R, Nauta A, Fernández E, P P, Rowland I, Swann J, Türk J, Vermeiren J, et al. Systematic review of the effects of the intestinal microbiota on selected nutrients and non-nutrients. Eur J Nutr 2018;57:25–49.

33. Khodorova N, Rutledge D, Oberli M, Mathiron D, Marcelo P, Benamouzig R, Tomé D, Gaudichon C, Pilard S. Urinary Metabolomics Profiles Associated to Bovine Meat Ingestion in Humans. Mol Nutr Food Res 2019;63(1):e1700834.

34. Braverman NE, Moser AB. Functions of plasmalogen lipids in health and disease. Biochim Biophys Acta 2012;1822(9):1442–52.

35. Mazzilli KM, McClain KM, Lipworth L, Playdon MC, Sampson JN, Clish CB, Gerszten RE, Freedman ND, Moore SC. Identification of 102 Correlations between Serum Metabolites and Habitual Diet in a Metabolomics Study of the Prostate, Lung, Colorectal, and Ovarian Cancer Trial. J Nutr 2020;150:694–703.

36. Thürmann PA, Steffen J, Zwernemann C, Aebischer CP, Cohn W, Wendt G, Schalch W. Plasma concentration response to drinks containing β-carotene as carrot juice or formulated as a water dispersible powder. Eur J Nutr 2002;41:228–35.

37. Skeaff CM, Hodson L, McKenzie JE. Dietary-induced changes in fatty acid composition of human plasma, platelet, and erythrocyte lipids follow a similar time course. J Nutr 2006;136:565–9.

38. Welch AA, Bingham SA, Ive J, Friesen MD, Wareham NJ, Riboli E, Khaw KT. Dietary fish intake and plasma phospholipid n-3 polyunsaturated fatty acid concentrations in men and women in the European Prospective Investigation into Cancer-Norfolk United Kingdom cohort. Am J Clin Nutr 2006;84(6):1330–9.

39. Bendinelli B, Palli D, Masala G, Sharp SJ, Schulze MB, Guevara M, van der A DL, Sera F, Amiano P, Balkau B, et al. Association between dietary meat consumption and incident type 2 diabetes: The EPIC-InterAct study. Diabetologia 2013;56(1):47–59.

40. Pan A, Sun Q, Bernstein AM, Schulze MB, Manson JE, Willett WC, Hu FB. Red meat consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis. Am J Clin Nutr 2011;94(4):1088–96.

41. Li J, Guasch-Ferré M, Chung W, Ruiz-Canela M, Toledo E, Corella D, Bhupathiraju SN, Tobias DK, Tabung FK, Hu J, et al. The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. Eur Heart J 2020;41(28):2645–56.

## Table 1. Baseline characteristics of the study participants for development and validation of meat metabolite scores in the EPIC-Norfolk study1

|  | **Exploratory set (n=11,432)** | **Validation set(n=853)** |
| --- | --- | --- |
| **Age (y)** | 59.6 ± 8.96 | 59.0 ± 9.40 |
| **Female** | 6204 (54 %) | 494 (58 %) |
| **Red meat intake (g/d)** | 34.4 ± 29.3 | 33.6 ± 29.1 |
| **Processed meat intake (g/d)** | 22.5 ± 21.0 | 21.7 ± 19.7 |
| **Poultry intake (g/d)** | 24.8 ± 27.5 | 26.0 ± 25.5 |
| **Education** |  |  |
| No | 4345 (38 %) | 326 (38 %) |
| Olevel | 1155 (10 %) | 79 (9 %) |
| Alevel | 4541 (40 %) | 330 (39 %) |
| Degree | 1385 (12 %) | 117 (14 %) |
| Missing | 6 (0.1%) | 1 (0.1%) |
| **Smoking** |  |  |
| Current | 1290 (11 %) | 112 (13 %) |
| Former | 4826 (42 %) | 329 (39 %) |
| Never | 5224 (46 %) | 407 (48 %) |
| Missing | 92 (0.8%) | 5 (0.6%) |
| **Alcohol intake (g/d)** | 11.9 ± 17.8 | 11.6 ± 16.6 |
| **PA** |  |  |
| Inactive | 3325 (29 %) | 238 (28 %) |
| Moderately inactive | 3243 (28 %) | 246 (29 %) |
| Moderately active | 2658 (23 %) | 206 (24 %) |
| Active | 2206 (19 %) | 163 (19 %) |
| **BMI (kg/m2)** |  |  |
| Mean ± SD | 26.1 ± 3.67 | 26.1 ± 3.71 |
| Missing | 16 (0.1%) | 2 (0.2%) |
| **Total Energy (kcal/d)** | 1950 ± 526 | 1940 ± 517 |
| **Fruit intake (g/d)** | 166 ± 126 | 168 ± 125 |
| **Vegetable intake (g/d)** | 152 ± 76.9 | 150 ± 68.6 |
| **Fatty fish intake (g/d)** | 12.3 ± 20.4 | 13.3 ± 22.3 |
| **White fish intake (g/d)** | 15.5 ± 18.5 | 15.9 ± 17.6 |
| **Legumes intake (g/d)** | 28.6 ± 30.2 | 26.7 ± 26.9 |
| **Nuts intake (g/d)** | 2.31 ± 6.51 | 2.18 ± 5.64 |
| **Dairy intake (g/d)** | 222 ± 146 | 217 ± 142 |
| **Egg intake (g/d)** | 14.3 ± 17.4 | 14.0 ± 17.0 |
| **Sugar sweetened beverages intake (g/d)** | 32.9 ± 78.6 | 30.8 ± 65.5 |

1Values are mean ± SD for continuous variables and n (%) for categorical variables.

g/d, grams per day.

## Table 2. Metabolites from the red meat metabolomics score that were positively associated with red meat consumption in both the EPIC-Norfolk and the randomized cross-over trial.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name | Formula | Fold-change1 | p-value | ChromatographicMethod2 | Retention time [min] | Confidence level of identification3 | MS fragments for identification | Rank4 |
| 1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4) | C43H78NO7P | 2.52 | 1.36 x 10-6 | RP | 9.04, 9.43 | Level 2 | 361.2741611.5296392.2934 | 1 |
| 1-(1-enyl-stearoyl)-2-arachidonoyl-GPC (P-18:0/20:4) | C46H84NO7P | 2.00 | 6.69 x 10-6 | RP | 9.1 | Level 3 | 184.0733 | 2 |
| 4-Hydroxyproline | C5H9NO3 | 6.27 | 1.06 x 10-4 | HILIC | 5.74 | Level 1 | 68.049886.0601 | 4 |
| TMAO  | C3H9NO | 1.56 | 6.30 x 10-3 | HILIC | 3.62 | Level 1 | 42.0329 | 7 |
| 1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2) | C42H80NO7P | 1.32 | 1.94 x 10-4 | RP | 8.97 | Level 3 | 184.0733 | 8 |
| 1-palmityl-GPC (O-16:0) | C24H52NO6P | 2.01 | 3.64 x 10-6 | RP | 7.18 | Level 2 | 104.1072184.0770341.3025 | 9 |
| Creatine | C4H9N3O2 | 1.50 | 4.88 x 10-2 | RP | 0.7 | Level 1 | 44.048290.0538 | 13 |
| 1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4) | C44H82NO7P | 2.44 | 4.30 x 10-6 | RP | 9.04 | Level 3 | 184.0733 | 17 |
| 1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2) | C44H84NO7P | 1.96 | 1.00 x 10-3 | RP | 9.19 | Level 3 | 184.0733 | 18 |
| Deoxycarnitine | C7H15NO2 | 1.23 | 6.12 x 10-3 | HILIC | 5.18 | Level 2 | 43.017960.081187.0445 | 21 |
| Stearoylcarnitine | C25H49NO4 | 1.52 | 7.36 x 10-3 | RP | 6.47 | Level 1 | 85.027760.0813 | 57 |

1Fold change in signal intensity in the RCT after fried pork intake compared with the tofu diet. Paired Welch’s t-tests were used to evaluate whether metabolites were significantly increased after pork intake compared to tofu intake. The variation of metabolites intensity after consumption of pork vs tofu is shown in Supplementary Figure 3.

2RP: reverse phase chromatography; HILIC: Hydrophilic Interaction Liquid Chromatography. The chromatographic tracing of selected metabolites in the blood after consumption of pork vs tofu are shown in Supplementary Figure 4.

3Level of confidence in metabolite identification according to Sumner et al.(25): level 1, matching of mass, retention time and mass fragmentation pattern with authentic chemical standard; level 2, matching of accurate mass and mass fragmentation pattern with the corresponding compound in a database; level 3, matching of mass and fragmentation pattern with the corresponding compound a database, due to the non-specific fragment, only the functional group, but not the length of each carbon chains can be determined.

4Rank: The rank of coefficients out of 139 metabolites in the red meat metabolite score in the EPIC-Norfolk study.

## Table 3. Characteristics of the study participants from baseline of the T2D case-cohort in the EPIC-Norfolk cohort1

|  |  |  |
| --- | --- | --- |
|  | **Subcohort**  | **T2D Cases(n=659)** |
|  | **Total****(n=846)** | **Q12(n=169)** | **Q2(n=169)** | **Q3(n=169)** | **Q4(n=169)** | **Q5(n=170)** |
| **Red meat intake (g/d)** | 33.6 ± 29.1 | 20.7 ± 20.7 | 25.8 ± 22.2 | 29.4 ± 22.7 | 40.9 ± 28.0 | 51.3 ± 37.8 | 39.3 ± 30.6 |
| **Age (y)** | 59.0 ± 9.4 | 59.3 ± 9.5 | 58.5 ± 9.4 | 58.9 ± 9.5 | 59.4 ± 9.3 | 58.7 ± 9.3 | 61.8 ± 8.3 |
| **Female** | 489 (58 %) | 133 (79 %) | 115 (68 %) | 89 (53 %) | 90 (53 %) | 62 (36 %) | 275 (42 %) |
| **Education** |  |  |  |  |  |  |  |
| No | 321 (38 %) | 69 (41 %) | 62 (37 %) | 59 (35 %) | 72 (43 %) | 59 (35 %) | 309 (47 %) |
| Olevel | 79 (9 %) | 20 (12 %) | 14 (8 %) | 17 (10 %) | 11 (7 %) | 17 (10 %) | 54 (8 %) |
| Alevel | 329 (39 %) | 60 (36 %) | 64 (38 %) | 72 (43 %) | 66 (39 %) | 67 (39 %) | 229 (35 %) |
| Degree | 117 (14 %) | 20 (12 %) | 29 (17 %) | 21 (12 %) | 20 (12 %) | 27 (16 %) | 67 (10 %) |
| **Smoking** |  |  |  |  |  |  |  |
| Current | 112 (13 %) | 15 (9 %) | 17 (10 %) | 19 (11 %) | 27 (16 %) | 34 (20 %) | 79 (12 %) |
| Former | 328 (39 %) | 53 (31 %) | 59 (35 %) | 63 (37 %) | 69 (41 %) | 84 (49 %) | 320 (49 %) |
| Never | 406 (48 %) | 101 (60 %) | 93 (55 %) | 87 (51 %) | 73 (43 %) | 52 (31 %) | 260 (39 %) |
| **Alcohol intake (g/d)** | 11.7 ± 16.7 | 6.33 ± 8.71 | 11.0 ± 16.3 | 12.8 ± 17.0 | 10.6 ± 15.7 | 17.8 ± 21.2 | 11.4 ± 19.0 |
| **PA** |  |  |  |  |  |  |  |
| Inactive | 234 (28 %) | 54 (32 %) | 37 (22 %) | 51 (30 %) | 42 (25 %) | 50 (29 %) | 290 (44 %) |
| Moderately inactive | 244 (29 %) | 46 (27 %) | 65 (38 %) | 39 (23 %) | 46 (27 %) | 48 (28 %) | 157 (24 %) |
| Moderately active | 206 (24 %) | 39 (23 %) | 37 (22 %) | 45 (27 %) | 44 (26 %) | 41 (24 %) | 122 (19 %) |
| Active | 162 (19 %) | 30 (18 %) | 30 (18 %) | 34 (20 %) | 37 (22 %) | 31 (18 %) | 90 (14 %) |
| **BMI (kg/m2)** | 26.0 ± 3.71 | 25.3 ± 3.37 | 26.1 ± 3.85 | 26.6 ± 3.79 | 26.0 ± 3.72 | 26.2 ± 3.71 | 29.6 ± 4.51 |
| **Total Energy (kcal/d)** | 1940 ± 516 | 1790 ± 434 | 1850 ± 444 | 1980 ± 543 | 2030 ± 560 | 2060 ± 537 | 1940 ± 538 |
| **Processed meat intake (g/d)** | 21.7 ± 19.7 | 16.3 ± 19.2 | 19.1 ± 17.2 | 19.5 ± 17.2 | 25.7 ± 21.5 | 28.0 ± 20.9 | 25.1 ± 21.1 |
| **Poultry intake (g/d)** | 25.8 ± 25.3 | 19.6 ± 21.7 | 27.0 ± 25.6 | 26.2 ± 25.1 | 28.0 ± 24.5 | 28.2 ± 28.3 | 24.0 ± 26.5 |
| **Fruit intake (g/d)** | 167 ± 124 | 205 ± 138 | 177 ± 117 | 171 ± 119 | 158 ± 128 | 124 ± 99.6 | 151 ± 137 |
| **Vegetable intake (g/d)** | 150 ± 68.6 | 152 ± 63.5 | 149 ± 69.8 | 152 ± 72.1 | 148 ± 67.3 | 147 ± 70.5 | 146 ± 80.9 |
| **Fatty fish intake (g/d)** | 13.3 ± 22.3 | 15.9 ± 22.5 | 15.1 ± 28.9 | 12.5 ± 17.6 | 12.2 ± 22.7 | 10.7 ± 17.7 | 13.9 ± 27.6 |
| **White fish intake (g/d)** | 15.9 ± 17.6 | 15.1 ± 17.0 | 13.5 ± 15.0 | 16.7 ± 18.5 | 18.4 ± 21.0 | 15.7 ± 15.8 | 16.3 ± 22.3 |
| **Legumes intake (g/d)** | 26.7 ± 26.9 | 26.7 ± 27.7 | 22.8 ± 23.4 | 26.4 ± 26.3 | 29.9 ± 31.4 | 27.6 ± 25.0 | 28.7 ± 29.8 |
| **Nuts intake (g/d)** | 2.20 ± 5.66 | 2.36 ± 6.15 | 2.15 ± 4.84 | 2.22 ± 5.61 | 1.63 ± 4.51 | 2.62 ± 6.87 | 2.04 ± 7.25 |
| **Dairy intake (g/d)** | 218 ± 142 | 220 ± 140 | 210 ± 146 | 215 ± 133 | 246 ± 135 | 197 ± 152 | 216 ± 159 |
| **Egg intake (g/d)** | 14.0 ± 17.0 | 11.9 ± 17.6 | 12.2 ± 13.3 | 14.3 ± 17.4 | 13.9 ± 15.3 | 17.8 ± 20.3 | 15.3 ± 17.3 |
| **Sugar sweetened beverages intake (g/d)** | 30.9 ± 65.7 | 19.9 ± 51.6 | 31.2 ± 62.1 | 37.9 ± 74.6 | 29.3 ± 56.1 | 36.4 ± 78.8 | 45.1 ± 127 |

1Values are mean ± SD for continuous variables and n (%) for categorical variables.

2Q: the red meat metabolite score in quintiles.

g/d, grams per day.

# Figure Legends

**Figure 1. Flow chart for the overall analytic approach for development and validation of the meat metabolomics score.**

\*the visualization simplifies the design of RCT as only two out of five arms are shown.

**Figure 2. Coefficients of metabolites with self-reported red and processed meat and poultry intake: the EPIC-Norfolk study (n=11,432)**

The colors represent the coefficients (weights) of each metabolite in each metabolite score; red means positive association and blue means negative association. A single asterisk next to the metabolite name indicates that the metabolite was annotated based on in-silico predictions which indicates the compound has not been confirmed based on a standard but its identity is confident.

**Figure 3. Volcano plot of candidate metabolites for red meat intake (n=139) with self-reported red meat intake and comparison of the red meat metabolite score across different categories of meat consumer groups: the EPIC Norfolk study (n=11,432)**

3A. The top 5 metabolites with the strongest association with self-reported red meat intake after adjustment for age and sex are annotated in the volcano plot. A single asterisk next to the metabolite name indicates that the metabolite was annotated based on in-silico predictions which indicates the compound has not been confirmed based on a standard but its identity is confident. 3B. A red meat non-consumer was defined as a participant with red meat consumption equal to zero (n=1,569) and a red meat consumer was a participant with red meat consumption over zero (n=9,863). Participants who reported consuming a vegetarian diet, other diet or no special diet were identified from self-reported questionnaires

**Figure 4. The associations of the red meat metabolite score and self-reported red meat intake with incident type 2 diabetes in a nested case-cohort study and exploratory analyses of multiple other health outcomes in the EPIC-Norfolk study**

The regression model 1 adjusted for age and sex; the regression model 2 adjusted for the following potential confounders: age, sex, education, smoking status, alcohol drinking, alcohol drinking squared, body-mass index, body-mass index squared, and dietary factors (consumption of fruits, vegetables, fatty fish and white fish, sugar sweetened beverages, dairy, legumes, nuts, eggs and total energy intake). The definition of incident cases and exclusion of prevalent cases are reported in *Supplementary Table 1*. Abbreviations: 7dDD, 7-day diet diary; CI, confidence interval; Mscore, red meat metabolite score; SD, standard deviation. \*, the association with incident type 2 diabetes was conducted in a nested case-cohort study in the EPIC-Norfolk study; associations with other exploratory health outcomes were conducted in the EPIC-Norfolk study after exclusion of participants involved in the case-cohort.