



Plasma metabolite profiles related to plant-based diets and the risk of type 2 diabetes

Fenglei Wang¹ · Megu Y. Baden¹ · Marta Guasch-Ferré^{1,2} · Clemens Wittenbecher¹ · Jun Li¹ · Yanping Li¹ · Yi Wan¹ · Shilpa N. Bhupathiraju^{1,2} · Deirdre K. Tobias^{1,3} · Clary B. Clish⁴ · Lorelei A. Mucci⁵ · A. Heather Eliassen^{2,5} · Karen H. Costenbader⁶ · Elizabeth W. Karlson⁶ · Alberto Ascherio^{1,2,5} · Eric B. Rimm^{1,2,5} · JoAnn E. Manson^{2,3,5,7} · Liming Liang^{5,8} · Frank B. Hu^{1,2,5}

Received: 20 July 2021 / Accepted: 24 January 2022 / Published online: 8 April 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Aims/hypothesis Plant-based diets, especially when rich in healthy plant foods, have been associated with a lower risk of type 2 diabetes. However, whether plasma metabolite profiles related to plant-based diets reflect this association was unknown. The aim of this study was to identify the plasma metabolite profiles related to plant-based diets, and to evaluate the associations between the identified metabolite profiles and the risk of type 2 diabetes.

Methods Within three prospective cohorts (Nurses' Health Study, Nurses' Health Study II and Health Professionals Follow-up Study), we measured plasma metabolites from 10,684 participants using high-throughput LC MS. Adherence to plant-based diets was assessed by three indices derived from the food frequency questionnaire: an overall Plant-based Diet Index (PDI), a Healthy Plant-based Diet Index (hPDI), and an Unhealthy Plant-based Diet Index (uPDI). Multi-metabolite profiles related to plant-based diet were identified using elastic net regression with a training/testing approach. The prospective associations between metabolite profiles and incident type 2 diabetes were evaluated using multivariable Cox proportional hazards regression. Metabolites potentially mediating the association between plant-based diets and type 2 diabetes risk were further identified.

Results We identified multi-metabolite profiles comprising 55 metabolites for PDI, 93 metabolites for hPDI and 75 metabolites for uPDI. Metabolite profile scores based on the identified metabolite profiles were correlated with the corresponding diet index (Pearson $r = 0.33$ – 0.35 for PDI, 0.41 – 0.45 for hPDI, and 0.37 – 0.38 for uPDI, all $p < 0.001$). Metabolite profile scores of PDI (HR per 1 SD higher = 0.81 [95% CI 0.75 , 0.88]) and hPDI (HR per 1 SD higher = 0.77 [95% CI 0.71 , 0.84]) showed an inverse association with incident type 2 diabetes, whereas the metabolite profile score for uPDI was not associated with the risk. Mutual adjustment for metabolites selected in the metabolite profiles, including trigonelline, hippurate, isoleucine and a subset of triacylglycerols, attenuated the associations of diet indices PDI and hPDI with lower type 2 diabetes risk. The explainable proportion of PDI/hPDI-related diabetes risk by these metabolites ranged between 8.5% and 37.2% (all $p < 0.05$).

Conclusions/interpretation Plasma metabolite profiles related to plant-based diets, especially a healthy plant-based diet, were associated with a lower risk of type 2 diabetes among a generally healthy population. Our findings support the beneficial role of healthy plant-based diets in diabetes prevention and provide new insights for future investigation.

Keywords Hippurate · Intermediates · Isoleucine · Metabolomics · Plant-based diets · Triacylglycerols · Trigonelline · Type 2 diabetes

Fenglei Wang and Megu Y. Baden contributed equally to this study.

✉ Frank B. Hu
fhu@hsph.harvard.edu

Extended author information available on the last page of the article

Abbreviations

FFQ	Food frequency questionnaires
hPDI	Healthy Plant-based Diet Index
HPFS	Health Professionals Follow-up Study
NHS	Nurses' Health Study
PDI	Plant-based Diet Index

Research in context

What is already known about this subject?

- Plant-based diets, especially a healthy plant-based diet, have been associated with a lower risk of type 2 diabetes
- Metabolomics is widely used to identify metabolites related to dietary intake

What is the key question?

- Are the plasma metabolite profiles related to plant-based diets associated with the risk of type 2 diabetes?

What are the new findings?

- Multi-metabolite profiles comprising 55 metabolites for an overall plant-based diet, 93 metabolites for a healthy plant-based diet, and 75 metabolites for an unhealthy plant-based diet were correlated with the adherence to the corresponding diet
- Scores based on the metabolite profile for the overall plant-based diet and the healthy plant-based diet were inversely associated with type 2 diabetes risk
- The inverse associations of overall and healthy plant-based diets with type 2 diabetes risk were attenuated after adjusting for metabolites selected in the metabolite profiles, such as trigonelline, hippurate, isoleucine, and a small set of triacylglycerols

How might this impact on clinical practice in the foreseeable future?

- The identified metabolite profiles could be used to assess the adherence of and metabolic response to a healthy plant-based diet for type 2 diabetes prevention

TAG Triacylglycerol

uPDI Unhealthy Plant-based Diet Index

Introduction

Type 2 diabetes, constituting over 90% of diabetes cases, poses a major threat to global health. The worldwide prevalence of type 2 diabetes in adults has increased from ~150 million in 2000 to >450 million in 2019 and is projected to rise to ~700 million by 2045 [1]. The complications of type 2 diabetes, both macrovascular (for example, CVD) and microvascular (for example, complications affecting the kidney, the retina and the peripheral nervous system), further increase the global burden of type 2 diabetes [2]. The major drivers of the global diabetes epidemic are unhealthy diet, overweight/obesity, genetic predisposition and other lifestyle factors [3]. Plant-based diets, especially healthy plant-based diets rich in high-quality plant foods, such as whole grains, fruits and vegetables, have been associated with a lower risk of type 2 diabetes [4, 5]. However, the underlying mechanisms are not fully understood.

Recent advances in high-throughput metabolomics profiling have brought about a new era for nutritional research. Metabolomics measures downstream components or metabolic products of foods. Identifying metabolites associated with plant-based diets may provide new insights into the biological processes behind the association between plant-based diets

and reduction in type 2 diabetes risk. Many previous studies have identified the metabolites that were correlated with several other dietary patterns, such as the Alternate Mediterranean Diet [6], the Dietary Approaches to Stop Hypertension diet [7], and western and prudent diets [8]. However, most studies analysed metabolites individually. Only a few studies accounted for the high correlations among the metabolites [9, 10] and obtained multi-metabolite profiles of dietary patterns [11].

We therefore analysed the plasma metabolites among 10,684 participants from the Nurses' Health Study (NHS), NHSII, and Health Professionals Follow-up Study (HPFS) and identified multi-metabolite profiles associated with plant-based diets by applying a training and testing approach. The identified metabolite profiles may serve as potential biomarkers of plant-based diets, but also markers of complex metabolic responses to the dietary exposures [12, 13]. We then prospectively examined the association of obtained multi-metabolite profiles with incident type 2 diabetes risk and explored the potential mediating metabolites.

Methods

Study population Our analyses were performed within three prospective cohort studies: NHS, NHSII and HPFS. The NHS was initiated in 1976, enrolling 121,700 female nurses aged 30–

55 years [14]. The NHSII was established in 1989 and recruited 116,429 female nurses aged 25–42 years [14]. The HPFS began in 1986 and enrolled 51,529 male health professionals aged 40–75 years [15]. Using a similar protocol, we collected blood samples from 32,826 participants in the NHS between 1989 and 1990, from 29,611 participants in the NHSII between 1996 and 1999, and from 18,225 participants in the HPFS between 1993 and 1995. Participants were mailed blood collection kits, and after collection samples were shipped via overnight courier with an icepack to our laboratory. Whole blood samples were then processed and separated into plasma, buffy coat and red blood cells and stored in liquid nitrogen freezers [16, 17].

In the present study, we included participants who were selected for 14 prior nested case–control sub-studies on

metabolomics (electronic supplementary material [ESM] Table 1). A total of 10,684 participants with available dietary intake data were included in the metabolomics analysis to identify the multi-metabolite profiles that were correlated to plant-based diets. For the analysis of metabolite profiles with incident type 2 diabetes, we excluded participants lost to follow-up after blood collection or participants reporting a history of cancer, CVD, or type 2 diabetes at blood draw. We also excluded participants from the gestational diabetes sub-study, which oversampled type 2 diabetes cases. In all, 8827 participants were included in the prospective analysis (Fig. 1). The study protocols were approved by the institutional review boards of Brigham and Women’s Hospital and Harvard T.H. Chan School of Public

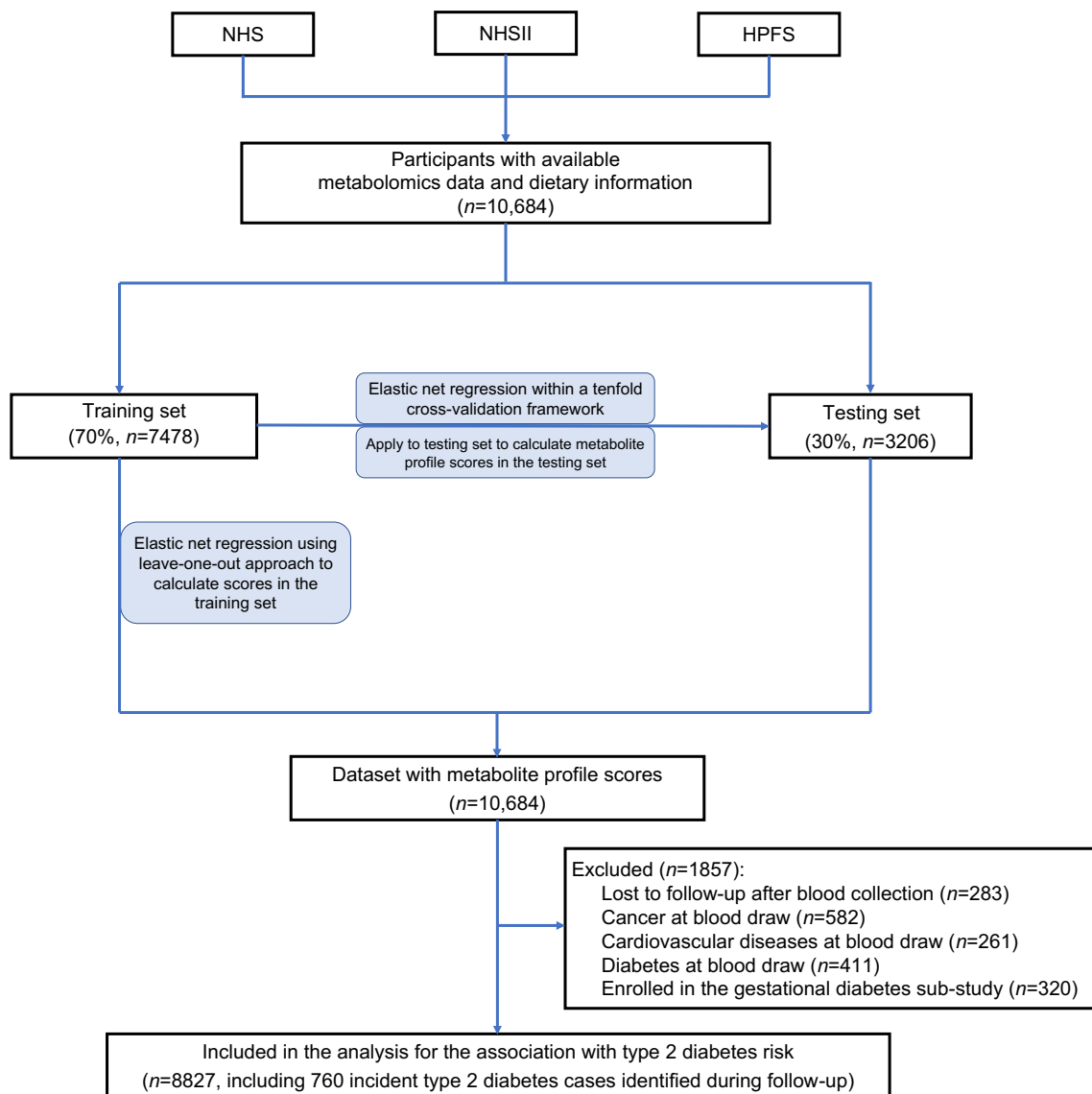


Fig. 1 Schematic of the study design. To make full use of the data, we calculated metabolite profile scores in the training set using a leave-one-out approach. We left one participant out each time and used the remaining participants to train an elastic net model (using tenfold cross

validation). Then we applied this trained model to the participant left out and calculated the metabolite profile score. We repeated this step for every participant in the training set and calculated their metabolite profile scores

Health. Participants' completion of questionnaires was considered as implied consent.

Dietary assessment and plant-based diet indices Self-reported dietary data were ascertained using semiquantitative food frequency questionnaires (FFQs). The validity and reproducibility of the questionnaires have been well documented [18, 19]. To better reflect habitual diets, we calculated the mean of the dietary intakes from the two FFQs closest to the time of blood draw. Three indices were calculated to estimate the adherence to three plant-based diets: an overall Plant-based Diet Index (PDI), a Healthy Plant-based Diet Index (hPDI), and an Unhealthy Plant-based Diet Index (uPDI). The development of these three indices has been described previously [5, 20]. In brief, based on nutrient and culinary similarities, 18 food groups were created within three large categories: healthy plant foods ($n = 7$; whole grains, fruits, vegetables, nuts, legumes, vegetable oils and tea/coffee); unhealthy plant foods ($n = 5$; refined grains, fruit juices, potatoes, sugar-sweetened beverages and sweets/desserts); and animal foods ($n = 6$; animal fats, dairy, eggs, fish/seafood, meat and miscellaneous animal-based foods). We distinguished between healthy and unhealthy plant foods using existing knowledge of associations of the foods with type 2 diabetes, CVD, certain cancers, and intermediate conditions, including obesity, hypertension, hyperlipidaemia and inflammation [5]. Alcoholic beverages were not included in the indices but were adjusted for in the multivariable analyses.

Each of the 18 food groups was divided into quintiles of consumption and given positive or reverse scores. For positive scores, a score of 5 was assigned to the highest quintile and a score of 1 to the lowest; for reverse scores, the scoring pattern was inverted, with 5 being assigned to the lowest quintile and 1 to the highest. All plant foods were given positive scores, and animal food groups were given reverse scores to calculate PDI. When calculating hPDI, only healthy plant foods were given positive scores, whereas only unhealthy plant foods were given positive scores for uPDI calculation. Finally, the 18 food group scores for each participant were summed to obtain the indices, with a possible range of 18 to 90.

Metabolomics measurement The plasma metabolomics profiling data were obtained using high-throughput LC MS techniques at the Broad Institute of MIT and Harvard (Cambridge, MA, USA) [21]. Details of the metabolomics measurement can be found in the ESM [Methods](#). Polar metabolites were separated using hydrophilic interaction LC with positive ionisation mode MS detection (HILIC-pos), whereas C8 chromatography with positive ionisation mode detection (C8-pos) was used to profile lipids. Targeted raw data were processed using TraceFinder software (Thermo Fisher

Scientific, Waltham, MA, USA), and non-targeted data were processed using Progenesis Q1 (Nonlinear Dynamics, Newcastle upon Tyne, UK). Metabolite identification was reported based on the standards proposed by the Metabolomics Standards Initiative [22]. Requirements for 'level 1' and 'level 2' identification were described in the ESM [Methods](#).

We excluded metabolites whose intraclass r across blinded quality control replicates (10% of study samples) were <0.3 ($n = 7$) and metabolites with an overall detection rate $<75\%$ ($n = 139$). Missing data for each metabolite were imputed using the random forest imputation approach as it has been previously recommended for metabolomics analysis [23]. We also excluded metabolites with poor stability (Spearman correlation or intraclass r , 0.75) due to delayed processing that was observed in our pilot study ($n = 61$) [24]. Finally, we included a total of 264 known metabolites in the current analysis (ESM [Table 2](#)). They were primarily lipids ($n = 178$, including 75 glycerolipids, 28 glycerophospholipids, 23 plasmalogens, 21 carnitines, 19 lysophospholipids, and 12 cholesteryl esters), but also included amino acids related metabolites ($n = 37$) and other metabolites ($n = 49$) (Fig. [2a](#)). Metabolites were highly correlated within each category (ESM [Fig. 1](#)).

Ascertainment of type 2 diabetes Participants who reported physician-diagnosed diabetes on the biennial questionnaire were sent a validated supplementary questionnaire to enquire about symptoms, diagnostic tests and medication use and confirm the diagnosis [25, 26]. We only included confirmed type 2 diabetes cases that met at least one of the following National Diabetes Data Group criteria: (1) one or more classic symptoms (excessive thirst, polyuria, weight loss and hunger) plus fasting blood glucose ≥ 7.8 mmol/l (≥ 140 mg/dl) or random blood glucose ≥ 11.1 mmol/l (≥ 200 mg/dl); (2) no symptoms but elevated blood glucose on two separate occasions (fasting blood glucose ≥ 7.8 mmol/l [≥ 140 mg/dl] or random blood glucose ≥ 11.1 mmol/l [≥ 200 mg/dl] or 2 h blood glucose ≥ 11.1 mmol/l [≥ 200 mg/dl] after OGTT); (3) any treatment with insulin or other glucose-lowering medications for diabetes [27]. The threshold for fasting blood glucose was changed to ≥ 7.0 mmol/l (≥ 126 mg/dl) as per the ADA diagnostic criteria after 1998 [28].

Non-dietary covariates Information on body weight, physical activity, smoking status, multivitamin use, ethnicity, family history of type 2 diabetes, hypertension, antihypertensive medication use, hypercholesterolaemia and lipid-lowering medication use was collected from self-reported biennial questionnaires preceding blood collection. BMI was derived using height reported at cohort baseline and body weight reported closest to the blood draw. We collected age and fasting status through questionnaires completed at blood collection.

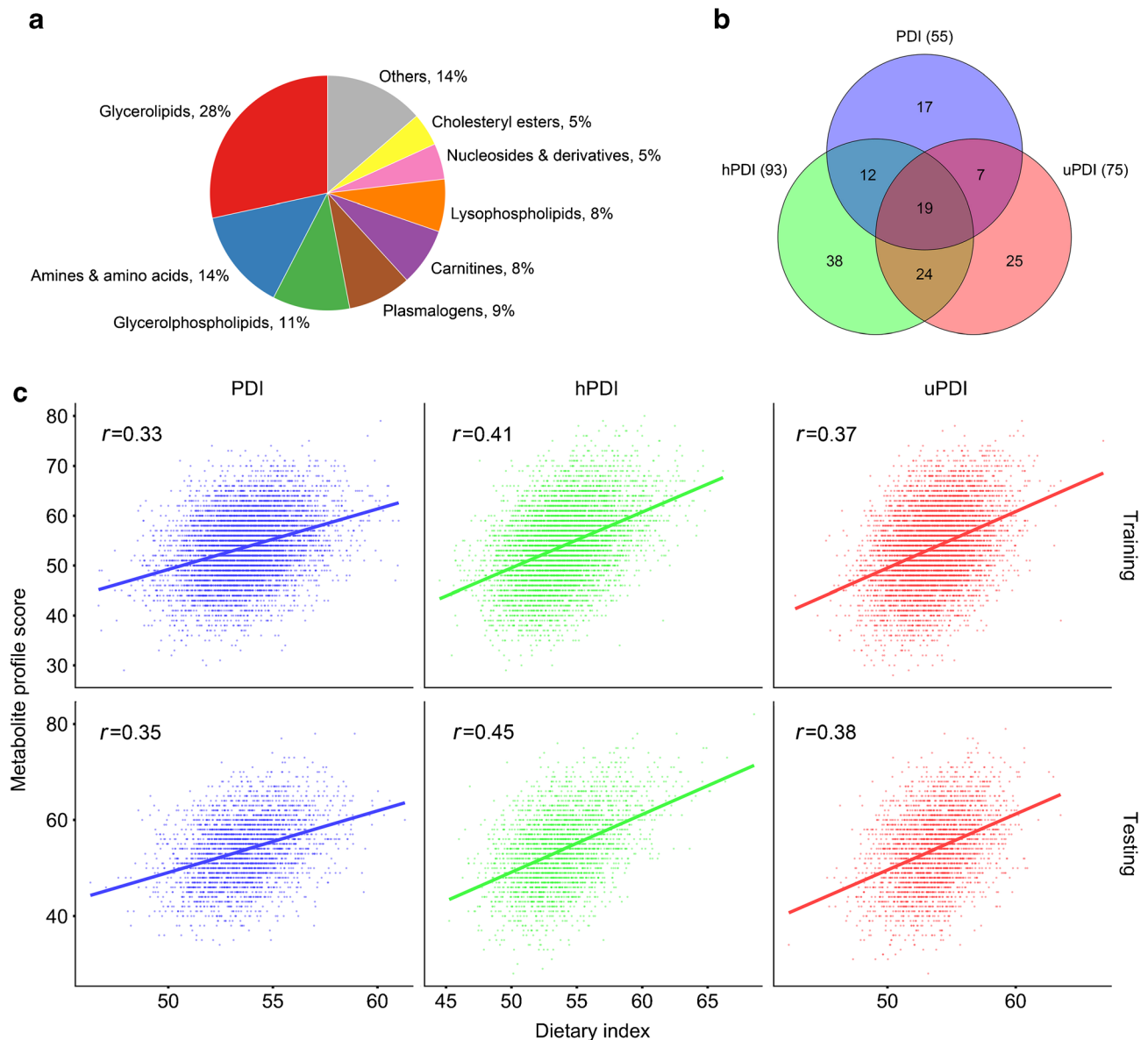


Fig. 2 Known metabolites measured and the metabolite profile score for three plant-based diet indices. **(a)** The distributions of categories for the 264 assayed known metabolites. **(b)** Venn diagram of selected

metabolites for the metabolite profile score of PDI, hPDI and uPDI. **(c)** Correlation between PDI and the corresponding metabolite profile score

Statistical analyses We log-transformed metabolites that were highly skewed (absolute skewness ≥ 2) [29]. All metabolites were then converted to z scores with a mean of 0 and SD of 1 within each sub-study. To identify metabolite profiles correlated to plant-based diets, all the participants were randomised to either the training set or the testing set in a 7 to 3 fashion (Fig. 1). We used an elastic net model within a tenfold cross-validation framework to regress PDI, hPDI or uPDI on the 264 known metabolites in the training set [30]. We then applied the trained model to the testing set to calculate a metabolite profile score. The metabolite profile score was calculated as the weighted sum of the selected metabolites with weights equal to the

elastic net regression coefficients. The score in the training set was obtained using a leave-one-out approach to avoid overfitting. We then calculated r between diet indices and the corresponding metabolite profile score. To evaluate the robustness of the metabolites selected in the metabolite profile score, we examined the associations between these metabolites and the corresponding diet index stratified by training/testing set, cohort, fasting status, and original case–control status. Associations of selected metabolites with diet indices and the dietary components were examined using multivariable linear regression. Plant-based diet indices and intake of each component were standardised before regression analyses.

Associations of plant-based diet indices, metabolite profile scores and selected metabolites with type 2 diabetes risk were assessed by Cox regression using combined data from the training set and testing set (Fig. 1). The Cox regression was stratified by cohort, original sub-study and case–control status and adjusted for age in months, BMI, ethnicity, fasting status, multivitamin use, smoking status, physical activity, hypertension, antihypertensive medication use, hypercholesterolaemia, lipid-lowering medication use, family history of diabetes, total energy intake, alcohol intake and margarine intake. The person-time for each participant was calculated from the blood collection date until the diagnosis of type 2 diabetes, death, loss to follow-up, or end of follow-up (June 2016 in the NHS, June 2017 in the NHSII, and January 2016 in the HPFS), whichever came first.

To test whether an association between the metabolite profile score and risk of type 2 diabetes was due to its correlation with diet index, we further included the diet index and the corresponding score simultaneously in the multivariable model. Among metabolites selected in the profile score, we identified

potential mediating metabolites of the association between plant-based diets and type 2 diabetes risk according to four predefined mediation criteria [31]: (1) existence of an association between PDI/hPDI/uPDI and type 2 diabetes; (2) existence of an association between PDI/hPDI/uPDI and the potential mediating metabolite in all participants and each stratum mentioned above; (3) existence of an association between the potential mediating metabolite and type 2 diabetes in the opposite direction as with PDI/hPDI or in the same direction as with uPDI; and (4) attenuation of the diet–type 2 diabetes association after adjusting for the potential mediating metabolite. Mediation analyses were further performed to estimate the proportion of diet–type 2 diabetes association that could be explained by the intermediate metabolite and the metabolite profile score [32, 33]. All statistical tests were two-sided ($p < 0.05$), and multiple testing was corrected using the Bonferroni method. All analyses were performed in R version 4.0 [34]. The main R packages used were ‘missRanger’ for random forest imputation, ‘glmnet’ for elastic net regression, and ‘survival’ for Cox regression.

Table 1 Characteristics of the study participants included in the metabolomics analysis

Characteristic	NHS (<i>n</i> =5505)	NHSII (<i>n</i> =3315)	HPFS (<i>n</i> =1864)	Overall (<i>n</i> =10,684)
Age, years	57 (7)	45 (4)	64 (8)	54 (10)
Female (%)	100	100	0	83
White (%)	96	97	95	96
BMI, kg/m ²	25.4 (4.6)	25.8 (5.9)	25.6 (3.1)	25.6 (4.8)
Physical activity, MET h/week	15.9 (20.3)	18.1 (22.0)	31.3 (28.3)	19.3 (23.1)
Fasting at blood collection (%)	70	72	59	69
Smoking status (%)				
Never	45	67	46	52
Past	41	25	50	37
Current	14	8	4	11
Medical history (%)				
Diabetes	4	2	5	3
Hypertension	27	11	30	23
Hypercholesterolaemia	29	25	39	29
Dietary intake				
Energy intake, kJ/day ^a	7418 (1958)	7694 (2088)	8410 (2322)	7678 (2096)
Alcohol intake, g/day	5.8 (9.2)	4.0 (6.7)	11.4 (14.3)	6.2 (10.0)
Healthy plant foods, serving/day	10.4 (3.6)	9.3 (3.8)	10.5 (3.8)	10.0 (3.7)
Unhealthy plant foods, serving/day	4.2 (2.1)	4.7 (2.1)	4.9 (2.4)	4.5 (2.2)
Animal foods, serving/day	5.0 (1.8)	4.7 (1.8)	4.8 (2.0)	4.8 (1.8)
PDI	55.0 (5.8)	55.1 (5.9)	55.6 (6.2)	55.2 (5.9)
hPDI	55.9 (6.6)	55.3 (6.9)	55.5 (6.9)	55.6 (6.8)
uPDI	54.1 (6.7)	54.8 (7.3)	53.9 (6.4)	54.3 (6.8)

Values are means (SDs) for continuous variables and percentages for categorical variables

For PDI/hPDI/uPDI calculation, 18 food groups were used, and the range was 18–90

^a Conversion factor: 1 kcal = 4.184 kJ

MET, metabolic equivalent of task

Table 2 Characteristics of the study participants included in the association analysis of type 2 diabetes

Characteristic	Participants developing type 2 diabetes (<i>n</i> =760)	Participants not developing type 2 diabetes (<i>n</i> =8067)	Overall (<i>n</i> =8827)
Age, year	55 (8)	54 (9)	54 (9)
Female (%)	88	86	87
White (%)	92	96	96
BMI, kg/m ²	29.2 (5.8)	24.9 (4.4)	25.3 (4.7)
Physical activity, MET h/week	14.4 (18.1)	19.2 (23.4)	18.8 (23.0)
Fasting at blood collection (%)	69	69	69
Smoking status (%)			
Never	52	52	52
Past	35	37	37
Current	13	11	11
Medical history (%)			
Hypertension	38	19	20
Hypercholesterolaemia	36	26	27
Medication use (%)			
Antihypertensive	30	14	16
Lipid-lowering	5	3	3
Family history of diabetes (%)	46	27	29
Dietary intake			
Energy intake, kJ/day ^a	7607 (2113)	7632 (2071)	7627 (2075)
Alcohol intake, g/day	4.2 (8.1)	6.3 (9.8)	6.2 (9.7)
Healthy plant foods, serving/day	9.6 (3.6)	10.1 (3.7)	10.0 (3.7)
Unhealthy plant foods, serving/day	4.4 (2.2)	4.5 (2.1)	4.5 (2.1)
Animal foods, serving/day	5.0 (1.8)	4.8 (1.8)	4.8 (1.8)
PDI	54.2 (5.7)	55.3 (5.9)	55.2 (5.9)
hPDI	54.4 (6.8)	55.7 (6.8)	55.6 (6.8)
uPDI	54.8 (7.0)	54.4 (6.8)	54.4 (6.8)

Values are means (SDs) for continuous variables and percentages for categorical variables

For PDI/hPDI/uPDI calculation, 18 food groups were used, and the range was 18–90

^a Conversion factor: 1 kcal = 4.184 kJ

MET, metabolic equivalent of task

Results

Characteristics of the study participants The participants included in the metabolomics analysis were predominantly white, middle-aged (mean age 54 ± 10 years), and the mean BMI was 25.6 ± 4.8 kg/m² (Table 1). Among them, 8827 participants free of diabetes, CVD and cancer at blood collection were further included in the association analysis of type 2 diabetes (Fig. 1). Compared with participants who did not develop type 2 diabetes after the blood draw, participants who were diagnosed with type 2 diabetes had a higher BMI, were more likely to have hypertension and hypercholesterolaemia, use antihypertensive and lipid-lowering medications, and report a family history of diabetes, but less likely to be physically active (Table 2). They also had a lower intake of healthy plant foods and have lower PDI and hPDI.

Metabolite profiles related to plant-based diet indices A panel of 55 metabolites for PDI, a panel of 93 metabolites for hPDI and a panel of 75 metabolites for uPDI were selected by the elastic net regression (Fig. 2b). The metabolite profile score based on the selected metabolites was significantly correlated with each corresponding diet index (Pearson $r = 0.33$ for PDI, 0.41 for hPDI and 0.37 for uPDI in the training set; Pearson $r = 0.35$ for PDI, 0.45 for hPDI and 0.38 for uPDI in the testing set; all p values <0.001) (Fig. 2c). The correlations between the diet index and the corresponding score were consistent across three cohorts (ESM Fig. 2). Among the selected metabolites, 24 out of 55 were consistently associated with PDI in all participants and across training/testing set, three cohorts, fasting/non-fasting, and case/non-case participants (ESM Fig. 3), 37 out of 93 with hPDI (ESM Fig. 4), and 26 out of 75 with uPDI (ESM Fig. 5).

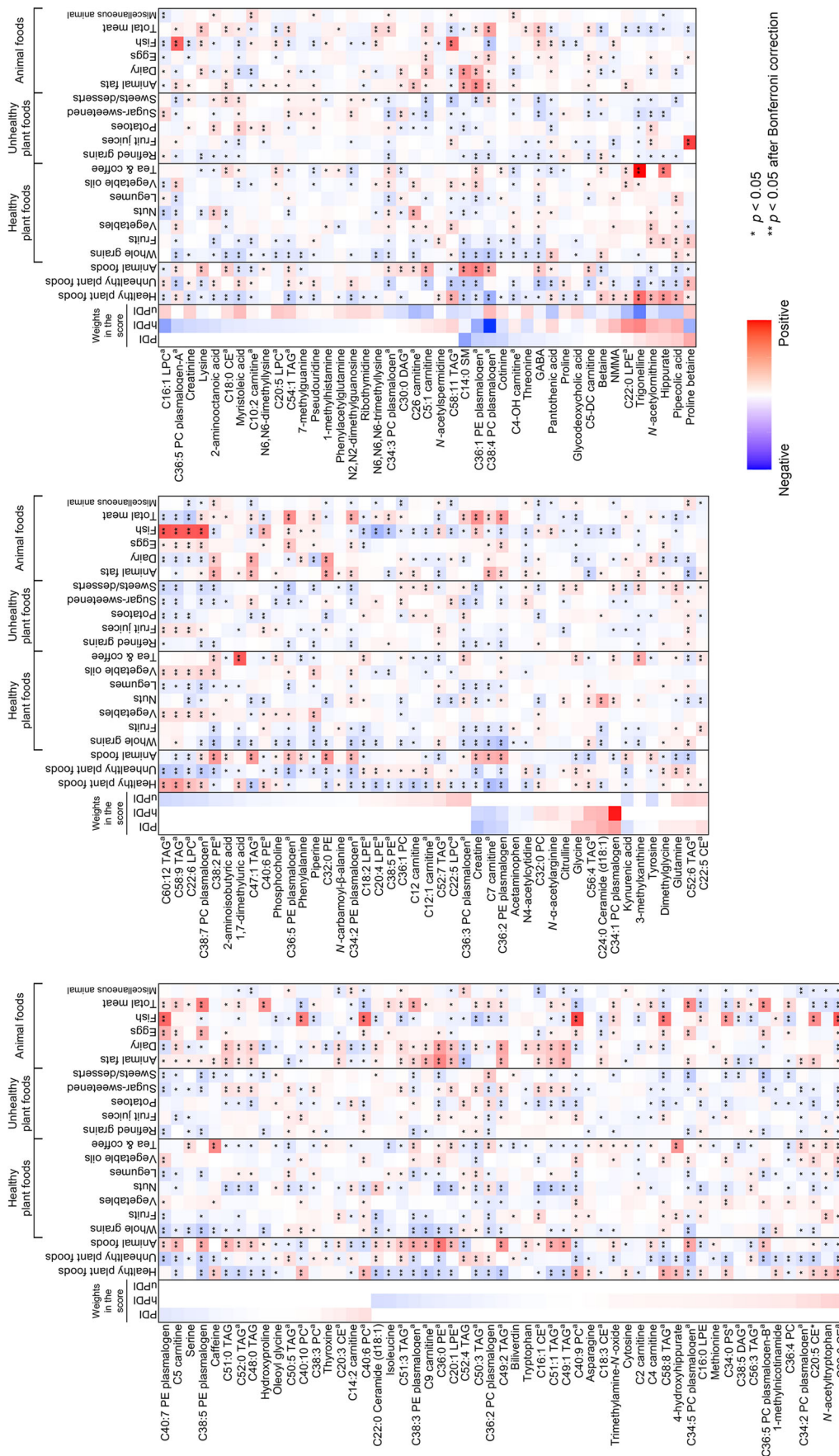


Fig. 3 Associations of the 142 metabolites selected in the metabolite profiles for three plant-based diet indices with food groups. Associations with food groups were conducted among 10,684 participants with available metabolomics data and dietary intake data. Presented from left to right are the metabolites' coefficients obtained from elastic net regression (weights in the calculation of metabolite profile score) and associations with food groups. Coefficients for associations with food

groups indicate the SD changes in metabolites per SD increment in dietary intake. Colours denote the association directions (red, positive; blue, inverse) and magnitudes (the darker the colour, the stronger the magnitude); asterisks in the coloured cell represent association significance (* $p < 0.05$ and ** $p < 0.05$ after Bonferroni-corrected 0.05). ^aLevel 2 identification defined by the Metabolomics Standards Initiative

Table 3 Associations of plant-based diet indices and metabolite profile scores with type 2 diabetes risk

Variable	Diet indices		Metabolite profile scores	
	HR (95% CI) ^a	<i>p</i>	HR (95% CI) ^a	<i>p</i>
PDI				
Age-adjusted ^b	0.83 (0.77, 0.89)	<0.001	0.69 (0.64, 0.75)	<0.001
MV ^c	0.88 (0.81, 0.95)	0.002	0.81 (0.75, 0.88)	<0.001
MV+mutual adjustment ^d	0.93 (0.85, 1.02)	0.13	0.83 (0.76, 0.90)	0.001
hPDI				
Age-adjusted ^b	0.79 (0.73, 0.85)	<0.001	0.65 (0.60, 0.70)	<0.001
MV ^c	0.85 (0.79, 0.93)	<0.001	0.77 (0.71, 0.84)	<0.001
MV+mutual adjustment ^d	0.93 (0.85, 1.01)	0.09	0.80 (0.73, 0.87)	<0.001
uPDI				
Age-adjusted ^b	1.08 (1.01, 1.16)	0.03	1.11 (1.03, 1.20)	0.005
MV ^c	1.07 (0.99, 1.16)	0.08	1.03 (0.95, 1.11)	0.44
MV+mutual adjustment ^d	1.07 (0.98, 1.17)	0.12	1.01 (0.93, 1.09)	0.89

^a HR and 95% CI of type 2 diabetes risk per SD increment in diet indices or the metabolite profile score

^b The basic age-adjusted model was stratified by study cohorts, original sub-studies, and the case–control status in the original sub-study, and was adjusted for age in months

^c Multivariable (MV) model was further adjusted for fasting status, BMI, race, family history of diabetes, multivitamin use, smoking status, physical activity, hypertension, antihypertensive medication use, hypercholesterolaemia, lipid-lowering medication use, total energy intake, alcohol intake and margarine intake

^d We included both plant-based diet indices and metabolite profile scores to examine association independence

Seventeen metabolites were only selected in the metabolite profile for PDI (ESM Fig. 6a). Metabolites inversely correlated with PDI, such as hydroxyproline, C5 carnitine, C38:5 PE plasmalogen, C40:7 PE plasmalogen and three triacylglycerols (TAGs), were negatively associated with intake of healthy plant foods but positively associated with animal foods, especially animal fats, dairy and total meat (Fig. 3). hPDI and uPDI exhibited different metabolite profiles (ESM Fig. 6b and 6c). Several metabolites were selected in the profile for both hPDI and uPDI but with coefficients in the opposite directions (ESM Fig. 6c). Metabolites including hippurate, trigonelline, betaine, pipercolic acid, pantothenic acid, *N*-acetylmethionine, C22:0 LPE and C58:11 TAG were positively associated with hPDI but inversely associated with uPDI. On the contrary, metabolites such as C16:1 LPC, C38:4 PC plasmalogen, N2,N2-dimethylguanosine, C54:1 TAG and myristoleic acid were inversely associated with hPDI but positively associated with uPDI. The positive associations of the above metabolites with hPDI were also observed for whole grains, fruits, vegetables, legumes and tea/coffee; the positive associations of metabolites associated with uPDI were observed for potatoes, sugar-sweetened beverages and sweets/desserts (Fig. 3).

Associations of metabolite profile score with the risk of type 2 diabetes In the prospective analyses of the diet indices with type 2 diabetes risk, inverse associations were observed for PDI and hPDI after adjusting for BMI and other potential confounders (HR per 1 SD higher [the same below]: 0.88

[95% CI 0.81, 0.95] for PDI; HR: 0.85 [95% CI 0.79, 0.93] for hPDI), while no significant association was observed for uPDI (HR: 1.07 [95% CI 0.99, 1.16]). The metabolite profile scores for PDI and hPDI were also both inversely associated with type 2 diabetes incidence after multivariable adjustment (HR: 0.83 [95% CI 0.76, 0.90] for PDI score; HR: 0.80 [95% CI 0.73, 0.87] for hPDI score). No association was observed for the metabolite profile score of uPDI (HR: 1.03 [95% CI 0.95, 1.11]) (Table 3). The inverse associations of metabolite profile scores for PDI and hPDI with type 2 diabetes risk were only slightly attenuated and remained statistically significant after further adjustment for diet indices PDI and hPDI (Table 3). However, the inverse associations for diet indices PDI and hPDI were markedly weakened and became statistically non-significant after adjusting for the metabolite profile score (Table 3). The metabolite profile score for PDI explained 47.8% of the PDI-type 2 diabetes association (Fig. 4a); the proportion of hPDI-type 2 diabetes association explainable by the metabolite profile score for hPDI was 51.0% (Fig. 4b).

Metabolites mediating the associations of plant-based diets with type 2 diabetes Following four predefined criteria, we identified eight potential intermediate metabolites of the PDI-type 2 diabetes association (ESM Fig. 7a). γ -aminobutyric acid (GABA), C5 carnitine and three TAGs (C48:0 TAG, C51:0 TAG and C52:0 TAG) were inversely associated with PDI but positively associated

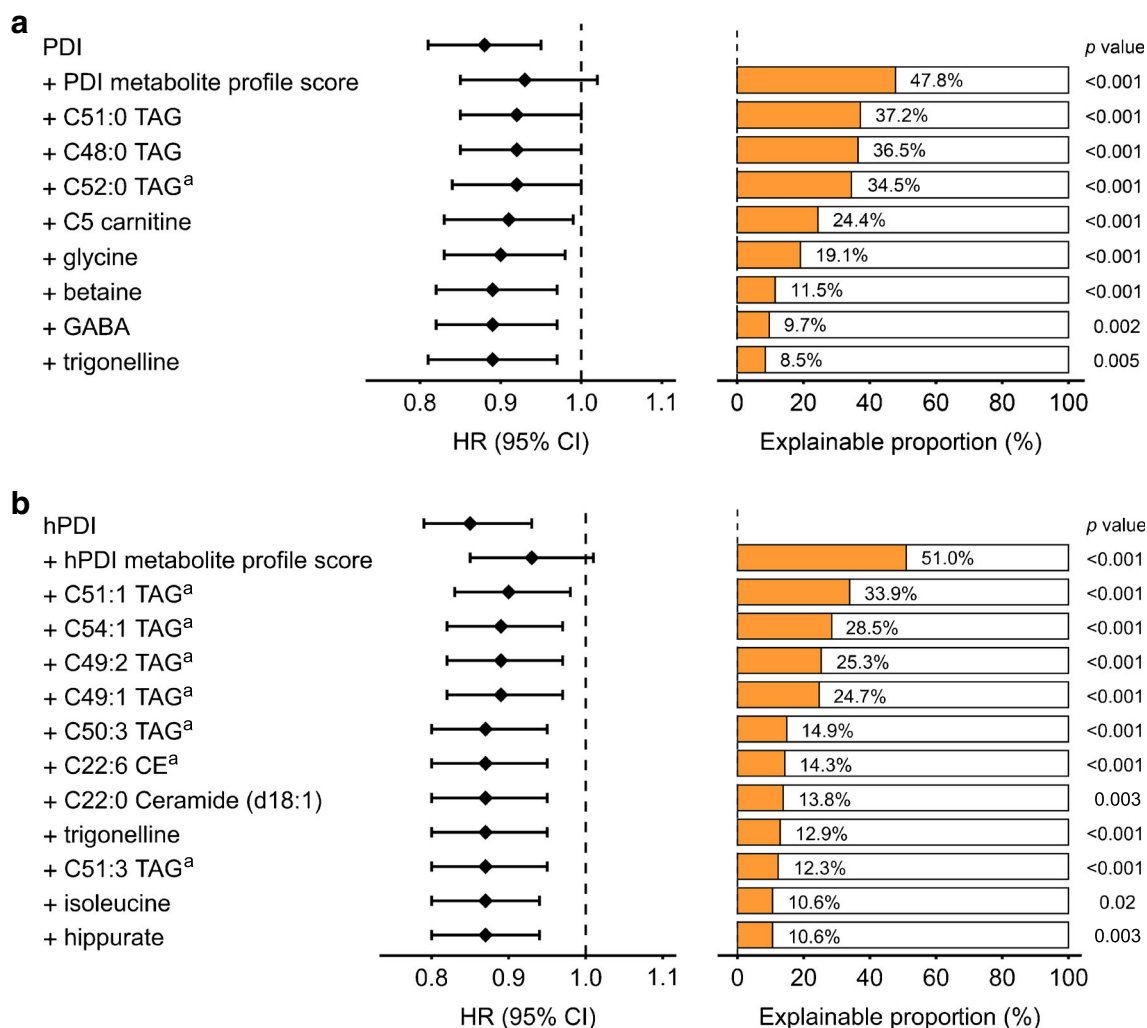


Fig. 4 Associations of PDI (**a**) and hPDI (**b**) with the risk of type 2 diabetes, adjusting for intermediate metabolites. All models were stratified by study cohorts, original sub-studies, and the case–control status in the original sub-study, and adjusted for age in months (continuous), fasting status (yes or no), BMI (continuous), race (white or non-white), family history of diabetes (yes or no), multivitamin use (yes or no),

smoking status (never, past or current), physical activity (continuous), hypertension (yes or no), antihypertensive medication use (yes or no), hypercholesterolaemia (yes or no), lipid-lowering medication use (yes or no), total energy intake (continuous), alcohol intake (continuous) and margarine intake (continuous). ^aLevel 2 identification defined by the Metabolomics Standards Initiative. GABA, γ -aminobutyric acid

with type 2 diabetes, whereas trigonelline, betaine and glycine were positively associated with PDI but inversely associated with type 2 diabetes (ESM Table 3). The proportion of PDI-type 2 diabetes association explainable by these single metabolites ranged between 8.5% and 37.2% (Fig. 4a). Similarly, 11 metabolites were identified as potential intermediates of the hPDI-type 2 diabetes association (ESM Fig. 7b). Isoleucine, C22:0 Ceramide (d18:1) and six TAGs (C49:1 TAG, C49:2 TAG, C50:3 TAG, C51:1 TAG, C51:3 TAG and C54:1 TAG) were inversely associated with hPDI but positively associated with type 2 diabetes. In contrast, trigonelline, hippurate and C22:6 CE were positively associated with hPDI but inversely associated with type 2 diabetes (ESM Table 4). The explainable proportion of hPDI-type 2 diabetes

association by these metabolites ranged between 10.6% and 33.9% (Fig. 4b).

Discussion

Leveraging metabolomics data from 10,684 participants, we identified multi-metabolite profiles associated with plant-based diets, showing differential profiles for healthy vs unhealthy plant-based diets. The metabolite profile scores for both the overall plant-based diet and a healthy plant-based diet were inversely associated with incident type 2 diabetes, independent of BMI, the corresponding diet index, and other diabetes risk factors among a generally healthy population. We further identified several potential

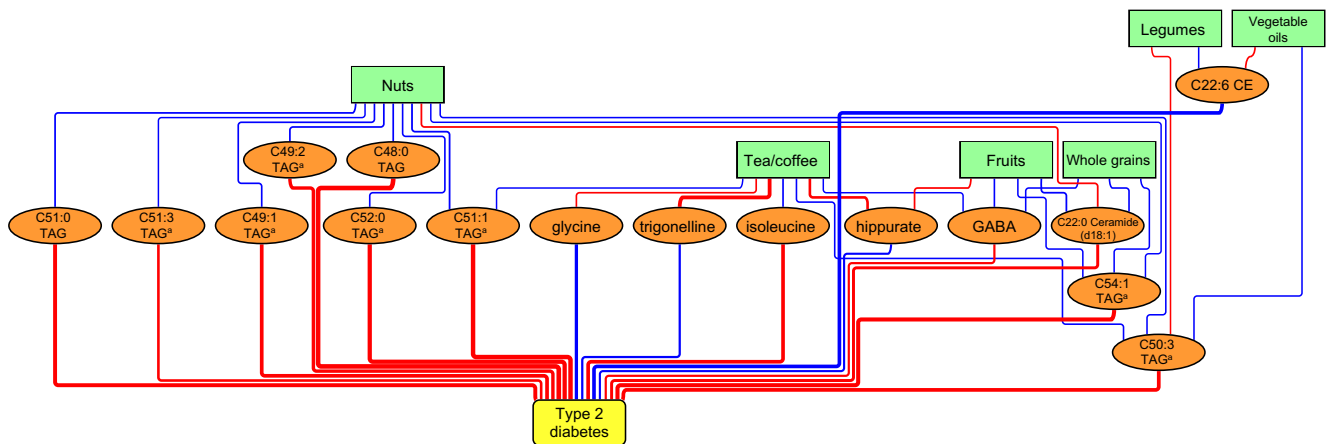


Fig. 5 Association between healthy plant foods, potential intermediate metabolites and type 2 diabetes risk. Associations between healthy plant foods and intermediate metabolites were examined by multivariable linear regression. Associations between the intermediate metabolites and type 2 diabetes risk were examined by Cox regression. Only associations significant at Bonferroni $p < 0.05$ were shown. The lines between healthy

plant foods, metabolites and type 2 diabetes represent regression coefficients. The line thickness indicates the strength of the association and colour indicates the direction (red, positive; blue, negative). ^aLevel 2 identification defined by the Metabolomics Standards Initiative. GABA, γ -aminobutyric acid

intermediate metabolites (trigonelline, hippurate, isoleucine and a subset of TAGs) of the association of overall and healthy plant-based diets with type 2 diabetes risk.

Plant-based diets, by definition, are higher in plant foods and lower in animal foods, which is reflected in our metabolomics readouts. We observed that hydroxyproline, an amino acid highly abundant in meat and negligible in many plant foods [35], was inversely correlated with PDI and positively associated with total meat intake. A general plant-based diet could encompass unhealthy foods such as refined grains, potatoes and sugar-sweetened beverages, which are associated with higher cardiometabolic risk [36, 37]. Therefore, an overall plant-based dietary pattern does not appear sufficient to differentiate the quality of the diet. As such, diet indices hPDI and uPDI were correlated with unique patterns of metabolites. Among 93 metabolites selected in the metabolite profile for hPDI, 50 were unique to hPDI; similarly, 32 out of 75 were unique to uPDI. A plant-based diet captured by hPDI is scored for being higher in whole grains, fruits, vegetables, nuts, legumes, vegetable oils and tea/coffee, whereas the uPDI is lower in these foods. Thus, metabolites related to coffee intake such as trigonelline and hippurate [38], metabolites related to legume intake such as pipercolic acid [39], and a potential novel marker of vegetable intake – *N*-acetylmethionine [6], were positively correlated with hPDI but inversely with uPDI.

Consistent with our previous findings on diet indices PDI and hPDI, the metabolite profile scores for PDI and hPDI were also inversely associated with the risk of type 2 diabetes [5]. Although our previous study observed that participants with higher uPDI had a higher type 2 diabetes risk than those with a lower uPDI, we did not observe an association between the metabolite profile score for uPDI and type 2 diabetes risk. One

explanation could be that the strength of the association between diet index uPDI and type 2 diabetes risk was weaker than that for PDI and hPDI [5]. With a smaller sample size compared with our previous study, which used participants from the whole cohorts, the association between uPDI and risk of type 2 diabetes was not significant in the present study. It could also be that our metabolomics platforms did not adequately capture the deleterious metabolic effects of unhealthy plant foods, such as high glycaemic starchy foods and sugary beverages.

Our metabolomics results depict the links between healthy plant foods, potential intermediate metabolites, and subsequent risk of type 2 diabetes (Fig. 5). The attenuation of the association after adjustment for these metabolites suggests their potential mediating role in linking plant-based diets to type 2 diabetes risk. Consumption of healthy plant foods, especially nuts, was associated with lower levels of TAGs with ≤ 56 carbon atoms and ≤ 3 double bonds. These TAGs have been associated with an increased risk of type 2 diabetes [40, 41]. Physiologic and pharmacologic experiments also showed that levels of these TAGs decreased in response to insulin action and were elevated in the setting of insulin resistance [41]. Trigonelline and hippurate were two other intermediate metabolites. Trigonelline is a phytochemical found in coffee [42], and its favourable effects on insulin resistance have been reported in animal studies [43]. Hippurate is typically increased after consuming phenolic compounds [44], and an increased level of hippurate was associated with better glycaemic control and enhanced insulin secretion [45, 46]. An association between reduced level of hippurate and higher risk of type 2 diabetes has been reported as well [47]. Another notable intermediate metabolite was the branched-chain

amino acid isoleucine, which was inversely correlated with tea/coffee but positively associated with the risk of type 2 diabetes. Branched-chain amino acids are consistently associated with increased insulin resistance [48], and multiple mechanisms have been proposed to explain how the elevations of these amino acids promote insulin resistance [49].

The main strengths of our study include the prospective examination of metabolomics with incident type 2 diabetes, the large sample size, and the use of means of two validated FFQs to minimise the within-person variability of diet. In addition, we used elastic net regression modelling, which performs well in high-dimensional data, where there are high correlations among the predictors [50]. Nevertheless, several limitations should be discussed as well. First, the metabolomics platform used in the present study was not able to measure all established biomarkers of food intake. More comprehensive approaches [51] will be helpful in understanding the metabolomic link between diet and health. Second, within the metabolites selected in the metabolite profiles, it is difficult to differentiate the metabolites that directly come from the diet and the metabolites that come from the metabolic response to the dietary intake or other metabolic influences. Future studies, especially human feeding trials, can benefit from measuring both the food metabolome and the human metabolome [12, 13]. Third, our findings regarding the intermediate metabolites are hypothesis-generating. We are not able to confirm their causal role in the associations of plant-based diets and type 2 diabetes risk. In addition, we only collected blood samples at one time point for metabolomics measurement. Long-term repeated metabolomics data are needed to understand how dietary changes relate to changes in metabolome, thereby influencing type 2 diabetes risk. Furthermore, we excluded metabolites that were not detected in at least 75% of all samples for quality control. It is possible that we might have omitted some metabolites related to the foods that were not commonly consumed. Finally, further studies are needed to evaluate external validity and replicate our findings in other racial and ethnic groups.

In conclusion, we identified multi-metabolite profiles of plant-based diets. Scores based on identified metabolites for an overall plant-based diet and a healthy plant-based diet showed an inverse association with the risk of type 2 diabetes among a generally healthy population. The large attenuation of inverse associations between plant-based diets and type 2 diabetes after adjustment for trigonelline, hippurate, isoleucine, a small set of TAGs, and several other intermediate metabolites suggests that these metabolites could be further investigated to provide mechanistic explanations for the benefits of plant-based diets on diabetes risk.

Supplementary Information The online version contains peer-reviewed but unedited supplementary material available at <https://doi.org/10.1007/s00125-022-05692-8>.

Acknowledgements The authors thank the participants and staff of the NHS, NHSII and HPFS for their valuable contributions.

Data availability The datasets generated and analysed during the current study are available from the corresponding author on request.

Funding This work was supported by grants from the National Institutes of Health (UM1 CA186107, U01 CA176726, U01 CA167552, U01 HL145386, R01 CA49449, R01 HL034594, R01 HL088521, R01 CA67262, R01 HL35464, R01 HL60712, R01 CA50385, P01 CA87969, R01 AR049880, R01 DK112940, R01 DK119268, R01 DK120870 and P30 DK046200). MGF was supported by ADA grant #1–18-PMF-029. The funders had no role in the design, conduct, analysis or reporting of this study. The funding sources did not participate in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript; and decision to submit the manuscript for publication.

Authors' relationships and activities YL has received research supports from the California Walnut Commission and Swiss Re Management Ltd. SNB is a scientific consultant to LayerIV. All authors declare that there are no relationships or activities that might bias, or be perceived to bias, their work.

Contribution statement FW, MYB and FBH were involved in the study concept and design. FW, MYB and CBC participated in acquisition of data. FW, MYB, MGF, CW, JL, YW and LL were involved in statistical analysis. FW, MYB, YL, YW, SNB, DKT, LAM, AHE, KHC, EWK, AA, EBR, JEM, LL and FBH participated in interpretation of data. FW performed drafting of the manuscript. All authors participated in critical revision of the manuscript and approved the final version of the manuscript. The corresponding author (FBH) takes full responsibility for the work and/or the conduct of the study, had access to the data and controlled the decision to publish.

References

1. Magkos F, Hjorth MF, Astrup A (2020) Diet and exercise in the prevention and treatment of type 2 diabetes mellitus. *Nat Rev Endocrinol* 16:545–555
2. Zheng Y, Ley SH, Hu FB (2018) Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol* 14:88–98
3. Neuenschwander M, Ballon A, Weber KS et al (2019) Role of diet in type 2 diabetes incidence: umbrella review of meta-analyses of prospective observational studies. *BMJ* 366:l2368
4. Tonstad S, Stewart K, Oda K, Batech M, Herring RP, Fraser GE (2013) Vegetarian diets and incidence of diabetes in the Adventist health Study-2. *Nutr Metab Cardiovasc Dis* 23:292–299
5. Satija A, Bhupathiraju SN, Rimm EB et al (2016) Plant-based dietary patterns and incidence of type 2 diabetes in US men and women: results from three prospective cohort studies. *PLoS Med* 13:e1002039
6. Playdon MC, Moore SC, Derkach A et al (2017) Identifying biomarkers of dietary patterns by using metabolomics. *Am J Clin Nutr* 105:450–465
7. Rebholz CM, Lichtenstein AH, Zheng Z, Appel LJ, Coresh J (2018) Serum untargeted metabolomic profile of the dietary approaches to stop hypertension (DASH) dietary pattern. *Am J Clin Nutr* 108:243–255

8. Chandler PD, Balasubramanian R, Paynter N et al (2020) Metabolic signatures associated with Western and prudent dietary patterns in women. *Am J Clin Nutr* 112:268–283
9. Shi L, Brunius C, Bergdahl IA et al (2019) Joint analysis of metabolite markers of fish intake and persistent organic pollutants in relation to type 2 diabetes risk in Swedish adults. *J Nutr* 149:1413–1423
10. Shi L, Brunius C, Johansson I et al (2020) Plasma metabolite biomarkers of boiled and filtered coffee intake and their association with type 2 diabetes risk. *J Intern Med* 287:405–421
11. Shi L, Brunius C, Johansson I et al (2018) Plasma metabolites associated with healthy Nordic dietary indexes and risk of type 2 diabetes—a nested case-control study in a Swedish population. *Am J Clin Nutr* 108:564–575
12. Brouwer-Brolsma EM, Brennan L, Drevon CA et al (2017) Combining traditional dietary assessment methods with novel metabolomics techniques: present efforts by the food biomarker Alliance. *Proc Nutr Soc* 76:619–627
13. Dragsted LO, Gao Q, Pratico G et al (2017) Dietary and health biomarkers—time for an update. *Genes Nutr* 12:24
14. Bao Y, Bertoia ML, Lenart EB et al (2016) Origin, methods, and evolution of the three Nurses' health studies. *Am J Public Health* 106:1573–1581
15. Rimm EB, Giovannucci EL, Willett WC et al (1991) Prospective study of alcohol consumption and risk of coronary disease in men. *Lancet* 338:464–468
16. Hankinson SE, Willett WC, Manson JE et al (1998) Plasma sex steroid hormone levels and risk of breast cancer in postmenopausal women. *J Natl Cancer Inst* 90:1292–1299
17. Chu NF, Spiegelman D, Yu J, Rifai N, Hotamisligil GS, Rimm EB (2001) Plasma leptin concentrations and four-year weight gain among US men. *Int J Obes Relat Metab Disord* 25:346–353
18. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC (1992) Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *Am J Epidemiol* 135:1114–1126 discussion 1127–1136
19. Yuan C, Spiegelman D, Rimm EB et al (2017) Validity of a dietary questionnaire assessed by comparison with multiple weighed dietary records or 24-hour recalls. *Am J Epidemiol* 185:570–584
20. Satija A, Bhupathiraju SN, Spiegelman D et al (2017) Healthful and unhealthful plant-based diets and the risk of coronary heart disease in U.S. adults. *J Am Coll Cardiol* 70:411–422
21. Paynter NP, Balasubramanian R, Giulianini F et al (2018) Metabolic predictors of incident coronary heart disease in women. *Circulation* 137:841–853
22. Sumner LW, Amberg A, Barrett D et al (2007) Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics* 3:211–221
23. Wei R, Wang J, Su M et al (2018) Missing value imputation approach for mass spectrometry-based metabolomics data. *Sci Rep* 8:663
24. Townsend MK, Clish CB, Kraft P et al (2013) Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. *Clin Chem* 59:1657–1667
25. Manson JE, Rimm EB, Stampfer MJ et al (1991) Physical activity and incidence of non-insulin-dependent diabetes mellitus in women. *Lancet* 338:774–778
26. Hu FB, Leitzmann MF, Stampfer MJ, Colditz GA, Willett WC, Rimm EB (2001) Physical activity and television watching in relation to risk for type 2 diabetes mellitus in men. *Arch Intern Med* 161:1542–1548
27. National Diabetes Data Group (1979) Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes* 28:1039–1057
28. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (1997) Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 20:1183–1197
29. Kim HY (2013) Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod* 38:52–54
30. Li J, Guasch-Ferre M, Chung W et al (2020) The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. *Eur Heart J* 41:2645–2656
31. Wittenbecher C, Muhlenbruch K, Kroger J et al (2015) Amino acids, lipid metabolites, and ferritin as potential mediators linking red meat consumption to type 2 diabetes. *Am J Clin Nutr* 101:1241–1250
32. Lin DY, Fleming TR, De Gruttola V (1997) Estimating the proportion of treatment effect explained by a surrogate marker. *Stat Med* 16:1515–1527
33. Jun HJ, Austin SB, Wylie SA et al (2010) The mediating effect of childhood abuse in sexual orientation disparities in tobacco and alcohol use during adolescence: results from the Nurses' health study II. *Cancer Causes Control* 21:1817–1828
34. R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
35. Wu G (2020) Important roles of dietary taurine, creatine, carnosine, anserine and 4-hydroxyproline in human nutrition and health. *Amino Acids* 52:329–360
36. Halton TL, Willett WC, Liu S, Manson JE, Stampfer MJ, Hu FB (2006) Potato and french fry consumption and risk of type 2 diabetes in women. *Am J Clin Nutr* 83:284–290
37. Huang C, Huang J, Tian Y, Yang X, Gu D (2014) Sugar sweetened beverages consumption and risk of coronary heart disease: a meta-analysis of prospective studies. *Atherosclerosis* 234:11–16
38. Rothwell JA, Keski-Rahkonen P, Robinot N et al (2019) A Metabolomic study of biomarkers of habitual coffee intake in four European countries. *Mol Nutr Food Res* 63:e1900659
39. Madrid-Gambin F, Brunius C, Garcia-Aloy M, Estruel-Amades S, Landberg R, Andres-Lacueva C (2018) Untargeted (1)H NMR-based metabolomics analysis of urine and serum profiles after consumption of lentils, chickpeas, and beans: an extended meal study to discover dietary biomarkers of pulses. *J Agric Food Chem* 66:6997–7005
40. Razquin C, Toledo E, Clish CB et al (2018) Plasma Lipidomic profiling and risk of type 2 diabetes in the PREDIMED trial. *Diabetes Care* 41:2617–2624
41. Rhee EP, Cheng S, Larson MG et al (2011) Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J Clin Invest* 121:1402–1411
42. Acidri R, Sawai Y, Sugimoto Y et al (2020) Phytochemical profile and antioxidant capacity of coffee plant organs compared to green and roasted coffee beans. *Antioxidants (Basel)* 9:93
43. Li Y, Li Q, Wang C, Lou Z, Li Q (2019) Trigonelline reduced diabetic nephropathy and insulin resistance in type 2 diabetic rats through peroxisome proliferator-activated receptor-gamma. *Exp Ther Med* 18:1331–1337
44. Wishart DS, Feunang YD, Marcu A et al (2018) HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res* 46:D608–D617
45. de Mello VD, Lankinen MA, Lindstrom J et al (2017) Fasting serum hippuric acid is elevated after bilberry (*Vaccinium myrtillus*) consumption and associates with improvement of fasting glucose levels and insulin secretion in persons at high risk of developing type 2 diabetes. *Mol Nutr Food Res* 61
46. Brial F, Chilloux J, Nielsen T et al (2021) Human and preclinical studies of the host-gut microbiome co-metabolite hippurate as a marker and mediator of metabolic health. *Gut* 70:2105–2114

47. Urpi-Sarda M, Almanza-Aguilera E, Llorach R et al (2019) Non-targeted metabolomic biomarkers and metabolotypes of type 2 diabetes: a cross-sectional study of PREDIMED trial participants. *Diabetes Metab* 45:167–174
48. Wurtz P, Soininen P, Kangas AJ et al (2013) Branched-chain and aromatic amino acids are predictors of insulin resistance in young adults. *Diabetes Care* 36:648–655
49. Arany Z, Neinast M (2018) Branched chain amino acids in metabolic disease. *Curr Diab Rep* 18:76
50. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *J R Stat Soc Ser B (Stat Methodol)* 67:301–320
51. Gonzalez-Dominguez R, Jauregui O, Mena P et al (2020) Quantifying the human diet in the crosstalk between nutrition and health by multi-targeted metabolomics of food and microbiota-derived metabolites. *Int J Obes* 44:2372–2381

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

Affiliations

Fenglei Wang¹ · Megu Y. Baden¹ · Marta Guasch-Ferré^{1,2} · Clemens Wittenbecher¹ · Jun Li¹ · Yanping Li¹ · Yi Wan¹ · Shilpa N. Bhupathiraju^{1,2} · Deirdre K. Tobias^{1,3} · Clary B. Clish⁴ · Lorelei A. Mucci⁵ · A. Heather Eliassen^{2,5} · Karen H. Costenbader⁶ · Elizabeth W. Karlson⁶ · Alberto Ascherio^{1,2,5} · Eric B. Rimm^{1,2,5} · JoAnn E. Manson^{2,3,5,7} · Liming Liang^{5,8} · Frank B. Hu^{1,2,5}

¹ Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA

² Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

³ Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

⁴ Metabolomics Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA

⁵ Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

⁶ Division of Rheumatology, Inflammation and Immunity, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

⁷ Mary Horrigan Connors Center for Women's Health and Gender Biology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

⁸ Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA