

Patterns in metabolite profile are associated with risk of more aggressive prostate cancer: A prospective study of 3,057 matched case–control sets from EPIC

Julie A. Schmidt¹, Georgina K. Fensom¹, Sabina Rinaldi², Augustin Scalbert², Paul N. Appleby¹, David Achaintre², Audrey Gicquiau², Marc J. Gunter², Pietro Ferrari², Rudolf Kaaks³, Tilman Kühn³, Heiner Boeing⁴, Antonia Trichopoulou⁵, Anna Karakatsani^{5,6}, Eleni Peppas⁵, Domenico Palli⁷, Sabina Sieri⁸, Rosario Tumino⁹, Bas Bueno-de-Mesquita^{10,11,12,13}, Antonio Agudo¹⁴, Maria-Jose Sánchez^{15,16}, María-Dolores Chirlaque^{15,17,18}, Eva Ardanaz^{15,19,20}, Nerea Larrañaga^{15,21}, Aurora Perez-Cornago¹, Nada Assi², Elio Riboli²², Konstantinos K. Tsilidis^{12,22}, Timothy J. Key¹ and Ruth C. Travis¹

¹Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom

²International Agency for Research on Cancer, Lyon, France

³Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

⁴Department of Epidemiology, German Institute of Human Nutrition (DIfE) Potsdam-Rehbrücke, Nuthetal, Germany

⁵Hellenic Health Foundation, Athens, Greece

⁶2nd Pulmonary Medicine Department, School of Medicine, National and Kapodistrian University of Athens, "ATTIKON" University Hospital, Haidari, Greece

⁷Cancer Risk Factors and Life-Style Epidemiology Unit, Institute for Cancer Research, Prevention and Clinical Network (ISPRO), Florence, Italy

⁸Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy

⁹Cancer Registry and Histopathology Department, "Civic - M.P.Arezzo" Hospital, Azienda Sanitaria Provinciale Di Ragusa (ASP), Ragusa, Italy

¹⁰Department for Determinants of Chronic Diseases (DCD), National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

¹¹Department of Gastroenterology and Hepatology, University Medical Centre, Utrecht, The Netherlands

¹²Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom

¹³Department of Social & Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

¹⁴Unit of Nutrition and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain

¹⁵CIBER in Epidemiology and Public Health (CIBERESP), Madrid, Spain

¹⁶Escuela Andaluza de Salud Pública, Instituto de Investigación Biosanitaria ibs.GRANADA, Hospitales Universitarios de Granada/Universidad de Granada, Granada, Spain

¹⁷Department of Epidemiology, Regional Health Council, IMIB-Arixaca, Murcia, Spain

Key words: epidemiology, metabolomics, prostate cancer risk, treelet transform

Abbreviations: ATBC: Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study; CI: confidence interval; EPIC: European Prospective Investigation into Cancer and Nutrition; IARC: International Agency for Research on Cancer; LysoPC: lysophosphatidylcholines; OR_{1SD}: odds ratio for a one standard deviation increase in treelet component score; PC aa: diacyl phosphatidylcholines; PC ae: acyl-alkyl phosphatidylcholines; PCA: principal component analysis; *r*: correlation coefficient; SD: standard deviation; SM (OH): hydroxysphingomyelins; SM: sphingomyelins; TNM: tumor-node-metastasis

Additional Supporting Information may be found in the online version of this article.

Conflict of interest: The authors declare no conflict of interest.

Grant sponsor: Cancer Research UK; **Grant numbers:** C8221/A19170, C570/A16491, 14136; **Grant sponsor:** World Cancer Research Fund (WCRF); **Grant number:** 2014/1183; **Grant sponsor:** Medical Research Council; **Grant numbers:** MR/M012190/1, 1000143; **Grant sponsor:** ISCIII Health Research Funds; **Grant number:** RD12/0036/0018; **Grant sponsor:** Health Research Fund (FIS); **Grant numbers:** PI13/01162, PI13/00061; **Grant sponsor:** Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra; **Grant sponsor:** Statistics Netherlands; **Grant sponsor:** Dutch ZON (Zorg Onderzoek Nederland); **Grant sponsor:** Dutch Prevention Funds; **Grant sponsor:** LK Research Funds; **Grant sponsor:** Netherlands Cancer Registry (NKR); **Grant sponsor:** Dutch Ministry of Public Health, Welfare and Sports (VWS); **Grant sponsor:** National Research Council; **Grant sponsor:** Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy; **Grant sponsor:** Hellenic Health Foundation; **Grant sponsor:** Federal Ministry of Education and Research (BMBF); **Grant sponsor:** Deutsches Krebsforschungszentrum; **Grant sponsor:** Deutsche Krebshilfe; **Grant sponsor:** Federal Ministry of Education and Research; **Grant sponsor:** German Cancer Research Center (DKFZ); **Grant sponsor:** German Cancer Aid; **Grant sponsor:** European Commission (DG-SANCO); **Grant sponsor:** International Agency for Research on Cancer

DOI: 10.1002/ijc.32314

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

History: Received 2 Dec 2018; Accepted 19 Mar 2019; Online 5 Apr 2019.

Correspondence to: Julie A. Schmidt, Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Richard Doll Building, Oxford, OX3 7LF, United Kingdom, Tel.: +44-1865-289641, Fax: +44-1865-289610, E-mail: julie.schmidt@ndph.ox.ac.uk

¹⁸Department of Health and Social Sciences, Murcia University, Murcia, Spain

¹⁹Navarra Public Health Institute, Pamplona, Spain

²⁰IdiSNA, Navarra Institute for Health Research, Pamplona, Spain

²¹Basque Regional Health Department, Public Health Division of Gipuzkoa-BIODONOSTIA, San Sebastian, Spain

²²Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece

Metabolomics may reveal novel insights into the etiology of prostate cancer, for which few risk factors are established. We investigated the association between patterns in baseline plasma metabolite profile and subsequent prostate cancer risk, using data from 3,057 matched case–control sets from the European Prospective Investigation into Cancer and Nutrition (EPIC). We measured 119 metabolite concentrations in plasma samples, collected on average 9.4 years before diagnosis, by mass spectrometry (Absolute/DQ p180 Kit, Biocrates Life Sciences AG). Metabolite patterns were identified using treelet transform, a statistical method for identification of groups of correlated metabolites. Associations of metabolite patterns with prostate cancer risk (OR_{1SD}) were estimated by conditional logistic regression. Supplementary analyses were conducted for metabolite patterns derived using principal component analysis and for individual metabolites. Men with metabolite profiles characterized by higher concentrations of either phosphatidylcholines or hydroxysphingomyelins ($OR_{1SD} = 0.77$, 95% confidence interval 0.66–0.89), acylcarnitines C18:1 and C18:2, glutamate, ornithine and taurine ($OR_{1SD} = 0.72$, 0.57–0.90), or lysophosphatidylcholines ($OR_{1SD} = 0.81$, 0.69–0.95) had lower risk of advanced stage prostate cancer at diagnosis, with no evidence of heterogeneity by follow-up time. Similar associations were observed for the two former patterns with aggressive disease risk (the more aggressive subset of advanced stage), while the latter pattern was inversely related to risk of prostate cancer death ($OR_{1SD} = 0.77$, 0.61–0.96). No associations were observed for prostate cancer overall or less aggressive tumor subtypes. In conclusion, metabolite patterns may be related to lower risk of more aggressive prostate tumors and prostate cancer death, and might be relevant to etiology of advanced stage prostate cancer.

What's new?

This prospective study is the largest investigation of metabolite profile and prostate cancer risk, to date. We found that patterns in plasma metabolite profile (characterized by higher concentrations of phosphatidylcholines and hydroxysphingomyelins; specific acylcarnitines, amino acids and a biogenic amine; and lysophosphatidylcholines, respectively) were associated with subsequent lower risk of more aggressive tumor subtypes and prostate cancer death. Moreover, the results suggest that metabolite profile may be relevant to the etiology of advanced stage disease.

Introduction

Prostate cancer is the second most commonly diagnosed cancer in men worldwide, but few risk factors have been established.^{1,2} Using metabolomics, which measures large numbers of small molecules in body fluids reflecting internal (the genome, epigenome, transcriptome and proteome) and external factors (diet, lifestyle, environment and gut microbiota),^{3–5} may help to identify novel risk factors for prostate cancer. Previously, relatively small prospective studies ($n_{cases} = 74–1,077$), including our study in the European Prospective Investigation into Cancer and Nutrition (EPIC),⁶ showed that men with higher circulating concentrations of lipids and/or energy-related metabolites might have lower risk of prostate cancer, especially more aggressive tumor subtypes.^{6–9} Moreover, a prospective study of lethal prostate cancer reported associations with metabolites in redox (inverse), dipeptide (positive), pyrimidine (mostly positive) and gamma-glutamyl amino acid (positively) pathways ($n_{deaths} = 523$).¹⁰ These studies mostly had limited power and have primarily investigated individual metabolites, which do not capture patterns in the metabolite profile reflecting correlations among metabolites¹¹; such patterns might

represent metabolites which are part of the same metabolic pathway¹² or originating from the same dietary sources.¹³ To overcome this and the issue of multiple testing, dimension-reduction methods such as treelet transform can be applied. Treelet transform combines features from principal component analysis (PCA) and cluster analysis to describe latent patterns in the data, which often are easier to interpret than those of PCA because metabolites irrelevant to the pattern are omitted. Treelet transform has for example been used to successfully describe patterns in serum metabolite profile¹² and to identify patterns in plasma fatty acids which were related to subsequent prostate cancer.¹⁴

We report an extension of our previous case–control study nested in EPIC ($n_{case} = 1,077$),⁶ now including 3,057 matched case–control sets, which, to the best of our knowledge, is the largest study to date of metabolites and prostate cancer risk. We aimed to identify patterns in plasma metabolite profiles in men without prostate cancer using treelet transform and to estimate the prospective associations between patterns and subsequent risk of prostate cancer overall, and by tumor characteristics and time to diagnosis, and risk of prostate cancer death.

Materials and Methods

Study population

The EPIC study includes 153,400 men recruited between 1992 and 2000 from 19 centers in eight countries (Denmark, Germany, Greece, Italy, Netherlands, Spain, Sweden and United Kingdom).¹⁵ At recruitment, detailed information was collected on diet and lifestyle, and 139,600 men gave a blood sample. For the current study, men were eligible if they had blood stored at the central biobank at the International Agency of Research on Cancer, Lyon, France (IARC; centers in Germany, Greece, Italy, Netherlands, Spain and United Kingdom), the date of blood collection was known, and if no cancer (except non-melanoma skin cancer) had been diagnosed at the time of blood collection. All participants gave written informed consent and the study was approved by ethical committees of IARC and the participating centers.

Follow-up and case and control selection

In most study centers, information on cancer incidence, tumor subtypes and vital status was obtained via record linkage to regional and national cancer registries. However, in Germany and Greece, a combination of methods was used, including health insurance records, cancer and pathology registries and active follow-up; self-reported incident cancers were verified through medical records.

Cases were men diagnosed with prostate cancer (defined as code C61 in the 10th revision of the International Statistical Classification of Diseases and Related Health Problems [ICD-10]) after blood collection and before the end of follow-up in August 2014. Each case was matched on the study center, length of follow-up and age (± 6 months), time of day (± 1 hr) and fasting status (< 3 , $3-6$, > 6 hr) at blood collection to one control participant, selected randomly among male cohort participants who were alive and free of cancer (except non-melanoma skin cancer) at the time of diagnosis of the case; an incidence density sampling procedure was used.

Prostate cancer subtypes were categorized based on the tumor-node-metastasis (TNM) system and histological grade as follows. Localized ($\leq T_2$ and $N_{0/x}$ and M_0 , or stage coded as localized, $n = 1,306$), advanced (T_{3-4} and/or N_{1-3} and/or M_1 , or coded as advanced, $n = 580$), non-aggressive ($\leq T_3$ and $N_{0/x}$ and M_0 , $n = 1,519$), aggressive (a subset of advanced stage disease defined as T_4 and/or N_{1-3} and/or M_1 , $n = 367$), low-intermediate grade (Gleason score < 8 or coded as well, moderately or poorly differentiated tumors, $n = 2,157$) and high grade (Gleason score ≥ 8 or coded as undifferentiated tumors, $n = 317$). Death from prostate cancer was defined as prostate cancer listed as the underlying cause of death on the death certificate during follow-up ($n = 326$).

Blood collection and laboratory analysis

A standardized protocol for blood collection and processing was followed and fasting was not required (details published elsewhere¹⁵). All plasma samples (citrate anticoagulant) were assayed at

IARC, using the AbsoluteIDQ p180 Kit (Biocrates Life Sciences AG, Innsbruck, Austria) and following the procedure recommended by the vendor. Samples from matched case-control sets were assayed in the same analytical batch along with quality control samples from pooled plasma; laboratory personnel was blinded to sample category, that is, case, study control or quality control. A triple quadrupole mass spectrometer (Triple Quad 4500; AB Sciex, Framingham, MA) was used to quantify 148 metabolites. After exclusions of metabolites and men, 3,057 matched case-control sets had data for 119 metabolites (details in Supporting Information Methods and Table S1). The metabolites comprised 8 acylcarnitines, 21 amino acids, 5 biogenic amines, 72 phosphatidylcholines (lysophosphatidylcholines [lysoPC, $n = 8$], diacyl phosphatidylcholines [PC aa, $n = 31$] and acyl-alkyl phosphatidylcholines [PC ae, $n = 33$]), hexose and 12 sphingomyelins (denoted hydroxysphingomyelins [SM (OH), $n = 5$] and sphingomyelins [SM, $n = 7$]). See Supporting Information Methods for metabolite nomenclature.

Statistical analyses

Patterns in metabolite profile were identified using treelet transform as described by Gorst-Rasmussen *et al.*^{16,17} In brief, treelet transform is a data-driven linear dimension-reduction method, which produces orthogonal components (vectors) from variables, for example, metabolites, based on their correlation or covariance structure. The numeric size of a variable within the component is called the loading and quantifies the contribution of each variable (here metabolite) to the component. A key aspect of treelet transform is its sparsity feature whereby loadings for some variables irrelevant to a given component are set to zero; this simplifies the interpretation of the patterns.

In treelet transform, the two most closely correlated metabolites are joined together by local PCA and the two original variables are replaced with the first principal component. This procedure is repeated until all variables are joined into one and the structure of the data is visualized by a cluster tree (Fig. 1). To obtain the treelet components the cluster tree is cut at a level providing a good trade-off between maximizing both the variance explained (higher cut-level) and the sparsity of the components (lower cut-level); the optimal cut-level for a given number of retained components is determined by data-driven cross-validation. Finally, for each retained component, a score variable is calculated as the linear combination of a participant's concentrations of the metabolites weighted by the loadings. The score reflects how similar a participant's metabolite profile is to the metabolite pattern.

In the current study, treelet transform was based on control participants only using the covariance matrix of logarithmically transformed metabolite concentrations. To inform the choice of number of factors to retain and the corresponding cut-level, we ran treelet transform for different numbers of retained components (1–10), identified the optimal cut-level for each (ranging from 97 to 104) using cross-validation, inspected scree plots (Supporting Information Fig. S1)¹⁷ and assessed pattern interpretability.¹⁸ The final treelet transform was then computed specifying the number of components to retain and the lowest optimal cut-level. Finally,

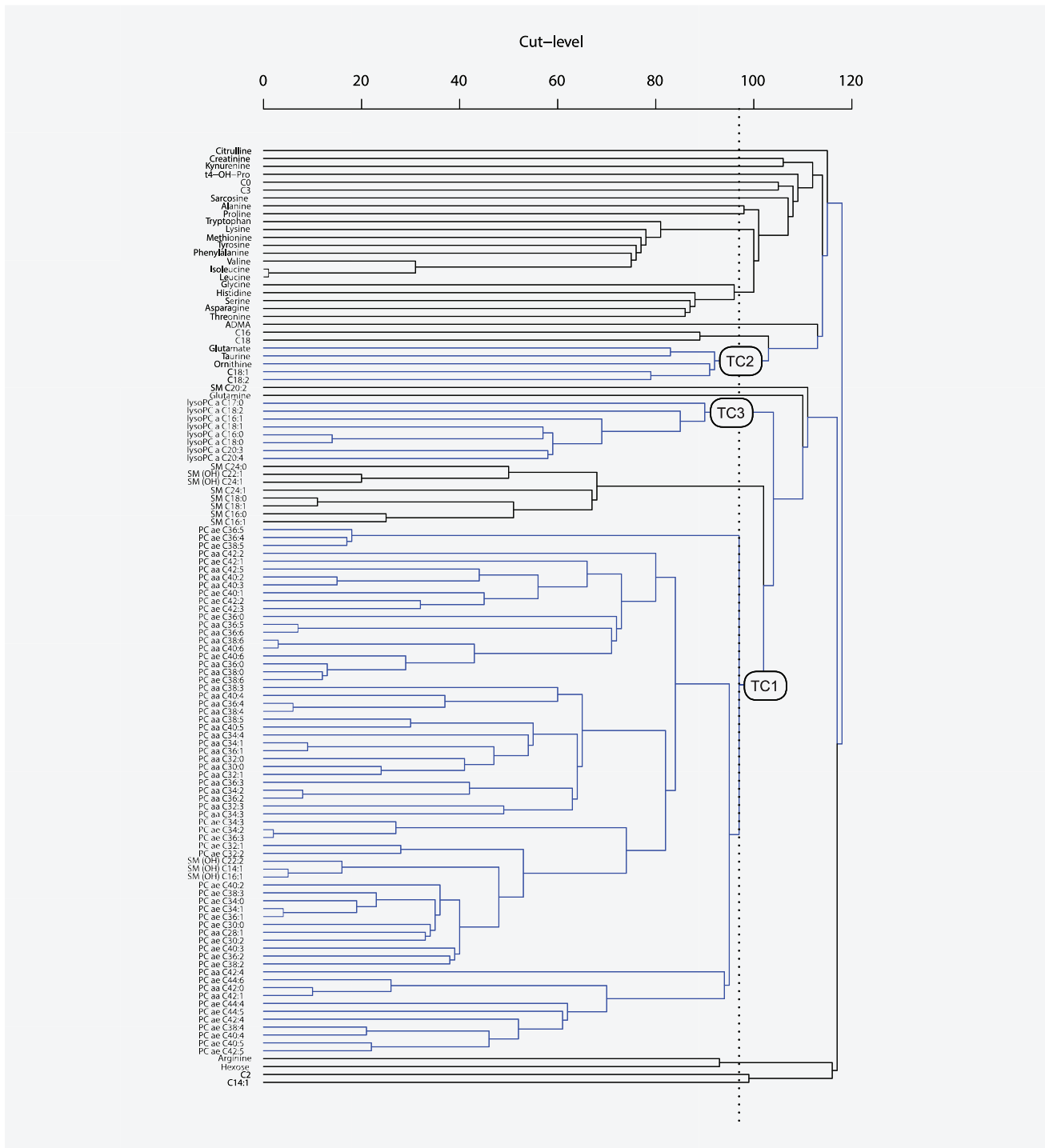


Figure 1. Cluster tree from treelet transform on 119 metabolite concentrations. The dotted line indicates the cut-level at 97. Blue lines indicate joining of correlated metabolites which belong to the three retained treelet components (TC). $n = 3,057$ control participants from EPIC.

component scores were calculated for all participants. A sensitivity analysis of the cut-level and a stability assessment of the treelet components were performed.

We used conditional logistic regression to estimate risk of prostate cancer for quintiles of component scores and per one

SD increase in component score. The analysis was conditioned on the matching variables and further adjusted for exact age at blood collection (continuously), baseline values of body mass index (quartiles; unknown), smoking (never, past, current and unknown), alcohol intake (<10, 10–19, 20–39 and

≥40 g of alcohol/day; unknown), attained education level (primary, secondary, degree level and unknown) and marital status (married or cohabiting, not married or cohabiting, and unknown).

Similar conditional logistic regression models were fitted for subgroup analyses by tumor subtypes (localized/advanced, non-aggressive/aggressive and high/low-intermediate grade), and likelihood ratio tests for heterogeneity in the association by subtypes were conducted. To assess potential presence of reverse causation in the associations of metabolite patterns with risk of more aggressive prostate cancer stages, the analyses for advanced and aggressive prostate cancer were further stratified by time to diagnosis (≤10/>10 years). In the analysis of risk of prostate cancer death, matched sets in which the control died, emigrated or was lost to follow-up before the case died were excluded ($n = 29$).

In a sensitivity analysis, the models examining the association between any treelet component and risk were additionally adjusted for any other treelet components (as continuous variables) with a Pearson correlation higher than 0.2 in absolute value with the treelet component of interest.

As a secondary analysis, we repeated the analysis of treelet components and risk of overall prostate cancer in the subset of participants, who were not included in our previously published analysis ($n_{\text{case}} = 2,018$).⁶

For comparison with the treelet transform method, we additionally identified metabolite patterns using PCA and assessed their association with risk of prostate cancer, in the entire study population.

For the purpose of comparison with previous studies, we also computed the associations of each individual metabolite and risk of prostate cancer, while accounting for multiple testing. This analysis was done for the full data set ($n_{\text{case}} = 3,057$), and separately for the data not previously published⁶ ($n_{\text{case}} = 2,018$).

All tests of statistical significance were two-sided, and 0.05 was considered as the nominal level for statistical significance in analyses of metabolite patterns and prostate cancer risk. All analyses were conducted in Stata Statistical Software Package, version 15 (Stata Corporation, College Station, TX). The “tt” add-on for STATA is available from <http://www.gorst.dk>. See Supporting Information Methods for further details of statistical analyses.

Results

At blood collection, participants were on average 58 years old (SD = 7.3 years) and cases were diagnosed on average 9.4 years later. There were no marked differences between baseline characteristics of the cases and controls (Table 1).

Treelet patterns in metabolite profile

After inspection of scree plots from treelet transform (Supporting Information Fig. S1) and cross-validation results, we retained three patterns at cut-level 97; the hierarchical

grouping of metabolites is shown in Figure 1. Together, the three patterns explained 31.4% of the total variation in metabolite concentrations, with the first to the third pattern (treelet components 1–3) accounting for 21.5, 5.2 and 4.7%, respectively (Supporting Information Table S2). Treelet component 1 was characterized by positive loadings on all diacyl and acyl-alkyl phosphatidylcholines and three hydroxysphingomyelins (Fig. 2; Supporting Information Table S2). Treelet component 2 was characterized by strong, positive loadings on acylcarnitines C18:1 and C18:2, amino acids glutamate and ornithine, and biogenic amine taurine. On treelet component 3, all eight lysophosphatidylcholines loaded strongly, positively. There was a positive correlation between treelet components 1 and 3 ($r = 0.46$), while treelet component 2 was not strongly correlated with the other patterns (Supporting Information Table S3).

Treelet patterns and prostate cancer risk

Associations of metabolite patterns and risk of prostate cancer are shown in Figure 3 and Supporting Information Table S4. No associations were observed for prostate cancer overall, nor with localized, non-aggressive, low-intermediate and high grade prostate cancer.

Treelet component 1 was inversely associated with risk of advanced stage prostate cancer (odds ratio for a SD increase in treelet component score [OR_{1SD}] = 0.77, 95% confidence interval [CI] 0.66–0.89, $p = 0.0007$), suggesting that men with higher concentrations of diacyl and acyl-alkyl phosphatidylcholines and three hydroxysphingomyelins had a lower risk of advanced stage prostate cancer (Fig. 3 and Supporting Information Table S4). The results were unchanged when stratifying at 10 years of follow-up. Furthermore, treelet component 1 was associated with risk of aggressive prostate cancer ($OR_{1SD} = 0.79$, 95%CI 0.66–0.95, $p = 0.013$), with no heterogeneity by follow-up time. In the sensitivity analysis further adjusting for treelet component 3, results for treelet component 1 remained largely unchanged, although the associations with advanced stage with more than 10 years of follow-up and with aggressive prostate cancer were slightly attenuated and no longer statistically significant (Supporting Information Table S5).

Men who scored higher on treelet component 2, characterized by higher acylcarnitines C18:1 and C18:2, glutamate, ornithine and taurine, had a lower risk of advanced stage disease ($OR_{1SD} = 0.72$, 95%CI 0.57–0.90, $p = 0.005$; Figure 3 and Supporting Information Table S4). When stratifying by follow-up time, risk estimates remained unchanged, but the significance was borderline for follow-up beyond 10 years ($OR_{1SD} = 0.70$, 95%CI 0.48–1.02, $p = 0.066$), perhaps due to lower numbers. Moreover, this pattern was inversely associated with risk of aggressive prostate cancer ($OR_{1SD} = 0.70$, 95%CI 0.53–0.93, $p = 0.014$), with no heterogeneity by follow-up time.

Treelet component 3, characterized by positive loadings on lysophosphatidylcholines, was inversely related to risk of

Table 1. Characteristics of prostate cancer cases and controls in EPIC

Characteristic	Cases, <i>n</i> = 3,057	Controls, <i>n</i> = 3,057
Age at blood collection, years (SD)	58.0 (7.3)	58.0 (7.3)
Height, cm (SD) ¹	172.2 (7.0)	172.4 (7.1)
Body Mass Index, kg/m ² (SD) ¹	27.2 (3.4)	27.2 (3.5)
Smoking, <i>n</i> (%) ¹		
Never	1,031 (34.3)	938 (31.1)
Former	1,276 (42.4)	1,316 (43.6)
Current	703 (23.4)	765 (25.3)
Alcohol consumption, <i>n</i> (%) ¹		
<10 g/day	1,271 (41.7)	1,272 (41.6)
10–19 g/day	577 (18.9)	577 (18.9)
20–40 g/day	632 (20.7)	674 (22.1)
≥40 g/day	566 (18.6)	533 (17.4)
Physical activity, <i>n</i> (%) ¹		
Inactive	710 (23.7)	721 (24.1)
Moderately inactive	986 (33.0)	985 (32.9)
Moderately active	716 (23.9)	698 (23.3)
Active	579 (19.4)	592 (19.8)
Marital status, <i>n</i> (%) ¹		
Married or cohabiting	2,051 (88.5)	2,063 (88.9)
Not married or cohabiting	267 (11.5)	257 (11.1)
Educational attainment, <i>n</i> (%) ¹		
Primary or equivalent	1,181 (40.6)	1,195 (40.9)
Secondary	1,001 (34.4)	1,026 (35.1)
Degree	727 (25.0)	704 (24.1)
Cases only		
Age at diagnosis, years (SD) ¹	67.4 (6.9)	–
Time to diagnosis, years (SD) ^{1,2}	9.4 (4.2)	–
Stage, <i>n</i> (%) ^{1,3}		
Localized	1,306 (69.2)	–
Advanced	580 (30.8)	–
Stage (aggressiveness), <i>n</i> (%) ^{1,3}		
Non-aggressive	1,519 (80.5)	–
Aggressive	367 (19.5)	–
Grade, <i>n</i> (%) ^{1,4}		
Low-intermediate grade	2,157 (87.2)	–
High grade	317 (12.8)	–
Death from prostate cancer, <i>n</i> (%) ^{1,5}	297 (9.7)	–

¹Unknown values for some participants; the calculations of percentages exclude missing values.

²Time between blood collection and diagnosis.

³The tumor-node-metastasis (TNM) system was used to categorize stages of prostate cancer; localized: ≤T₂ and N_{0/x} and M₀, or coded as localized; advanced: T_{3–4} and/or N_{1–3} and/or M₁, or coded as advanced; non-aggressive: ≤T₂ and N_{0/x} and M₀, and aggressive: T₄ and/or N_{1–3} and/or M₁. All categories are not mutually exclusive as aggressive is a subset of advanced stage, so numbers do not add up; percentages were calculated separately for localized and advanced, and for non-aggressive and aggressive.

⁴Gleason score <8 or coded as well, moderately or poorly differentiated for low-intermediate grade and Gleason score ≥8 or coded as undifferentiated for high grade.

⁵Death from prostate cancer (prostate cancer listed as the underlying cause of death on the death certificate) during follow-up; 326 died from

advanced stage prostate cancer (OR_{1SD} = 0.81, 95%CI 0.69–0.95, *p* = 0.009; Figure 3 and Supporting Information Table S4). Risk estimates were similar when stratifying by time to diagnosis and remained significant for follow-up beyond 10 years (OR_{1SD} = 0.76, 95%CI 0.60–0.97, *p* = 0.030). An inverse association with risk of death from prostate cancer was also observed (OR_{1SD} = 0.77, 95%CI 0.61–0.96, *p* = 0.023). When further adjusting for treelet component 1, the associations of treelet component 3 with advanced prostate cancer were slightly attenuated and no longer statistically significant, while the association with prostate cancer death remained (Supporting Information Table S5).

When restricting the analysis to 2,018 cases and their matched controls who were not included in our previous publication,⁶ we obtained similar results to those observed for the full study population (*n*_{case} = 3,057; Supporting Information Table S6).

Principal component analysis

Treelet components 1 and 2, were similar to principal components 1 (*r* = 0.99) and 2 (*r* = 0.88), respectively (Supporting Information Tables S7 and S8). Treelet component 3 (lysophosphatidylcholines) was correlated with principal components 1 (*r* = 0.54), 5 (*r* = 0.43) and 9 (*r* = 0.38) but qualitatively most similar to principal component 9, which was mostly characterized by positive loadings on the lysophosphatidylcholines, some acylcarnitines (especially C14:1), amino acids arginine and t4-hydroxyproline, and sphingomyelin SM C20:2, while having negative loadings on amino acid glutamate and several phosphatidylcholines.

The associations of principal components 1 and 2 with prostate cancer risk (Supporting Information Table S9) were similar to those observed for treelet components 1 and 2 (Fig. 3), respectively. Unlike treelet component 3, principal component 9 was not associated with advanced stage prostate cancer, but in line with the main results, a trend toward an inverse association with risk of prostate cancer death was observed (OR_{1SD} = 0.85, 95%CI 0.70–1.03, *p* = 0.095). Principal component 1 was inversely associated with risk of advanced (overall and by follow-up time) and aggressive prostate cancer (with no heterogeneity by length of follow-up), but not prostate cancer death, while principal component 5 was not related to risk of any prostate cancer outcome.

In addition, principal component 4, characterized by positive loadings on several acylcarnitines, amino acid t4-hydroxyproline, biogenic amine sarcosine and several phosphatidylcholines, and inverse loadings on other phosphatidylcholines and sphingolipids, was inversely associated with advanced (with no heterogeneity by follow-up time), aggressive (with a somewhat stronger association in men who had been followed up for more than 10 years) and high grade prostate cancer. Finally, principal component 8, mainly characterized by positive loadings on some acylcarnitines, arginine,

prostate cancer, but 29 were excluded from further analysis as their matched control either died, emigrated or was lost to follow-up before they died.

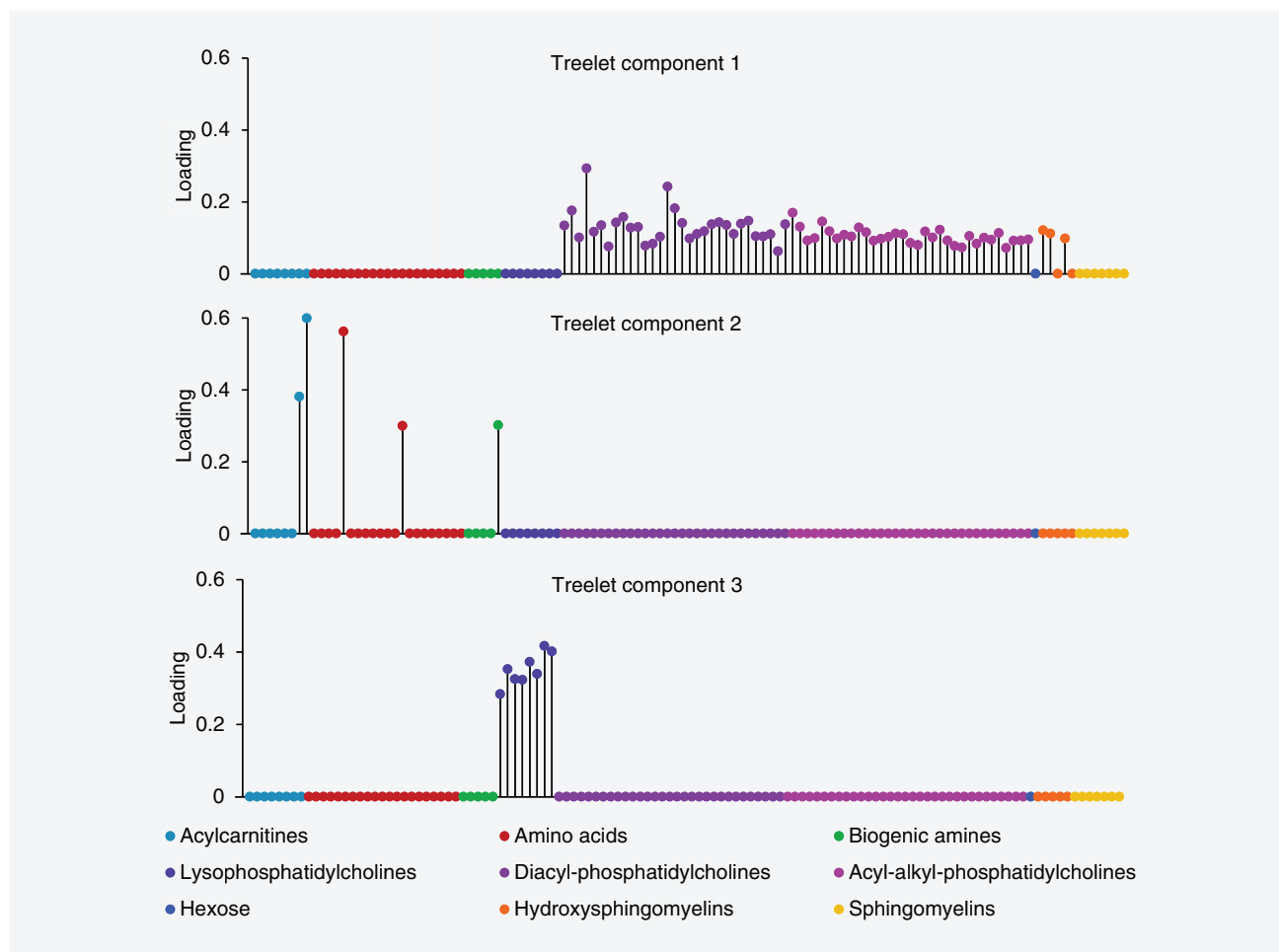


Figure 2. Loading plots for treelet components 1, 2 and 3 from treelet transform on 119 metabolite concentrations, using cut-level 97. The loading is the numeric size of a metabolite within the component, and it quantifies the contribution of each metabolite to the component. $n = 3,057$ control participants from EPIC.

t4-hydroxyproline and SM C20:2, mixed loadings on the glycerophospholipids, and inverse loadings on all the hydroxysphingomyelins, was inversely associated with all prostate cancer outcomes, except advanced prostate cancer at longer follow-up and prostate cancer death.

Analysis of individual metabolites

Results of the associations between individual metabolite concentrations and prostate cancer risk in the full dataset ($n_{\text{case}} = 3,057$) and in the data not previously published⁶ ($n_{\text{case}} = 2,018$) are shown in Supporting Information Tables S10–S23. Results from the new dataset were in line with those of the full dataset, although slightly weaker, with two obvious exceptions; acyl-alkyl phosphatidylcholines were more often associated with overall prostate cancer in the new dataset than in the full dataset (Supporting Information Tables S10 and S17), and the same was the case for diacyl phosphatidylcholines and risk of low-intermediate grade prostate cancer (Supporting Information Tables S15 and S22).

Discussion

In this large prospective study, treelet transform identified three easily interpretable patterns in men's metabolite profile. We observed a lower risk of advanced stage prostate cancer among men with metabolite profiles akin to those described by treelet components 1 and 2, characterised by higher concentrations of phosphatidylcholines and hydroxysphingomyelins (treelet component 1), or the combination of acylcarnitines C18:1 and C18:2, glutamate, ornithine and taurine (treelet component 2). Importantly for our understanding of potential etiological factors for prostate cancer, the risk estimates remained similar in case-sets in which the case was diagnosed more than 10 years after blood collection. These two metabolite patterns showed similar associations with aggressive prostate cancer (the more aggressive subset of advanced stage prostate cancer), also with no heterogeneity by follow-up time. Moreover, similar results were observed for men with higher concentrations of lysophosphatidylcholines (treelet component 3) but these associations were partly explained by confounding from the phosphatidylcholine and hydroxysphingomyelin

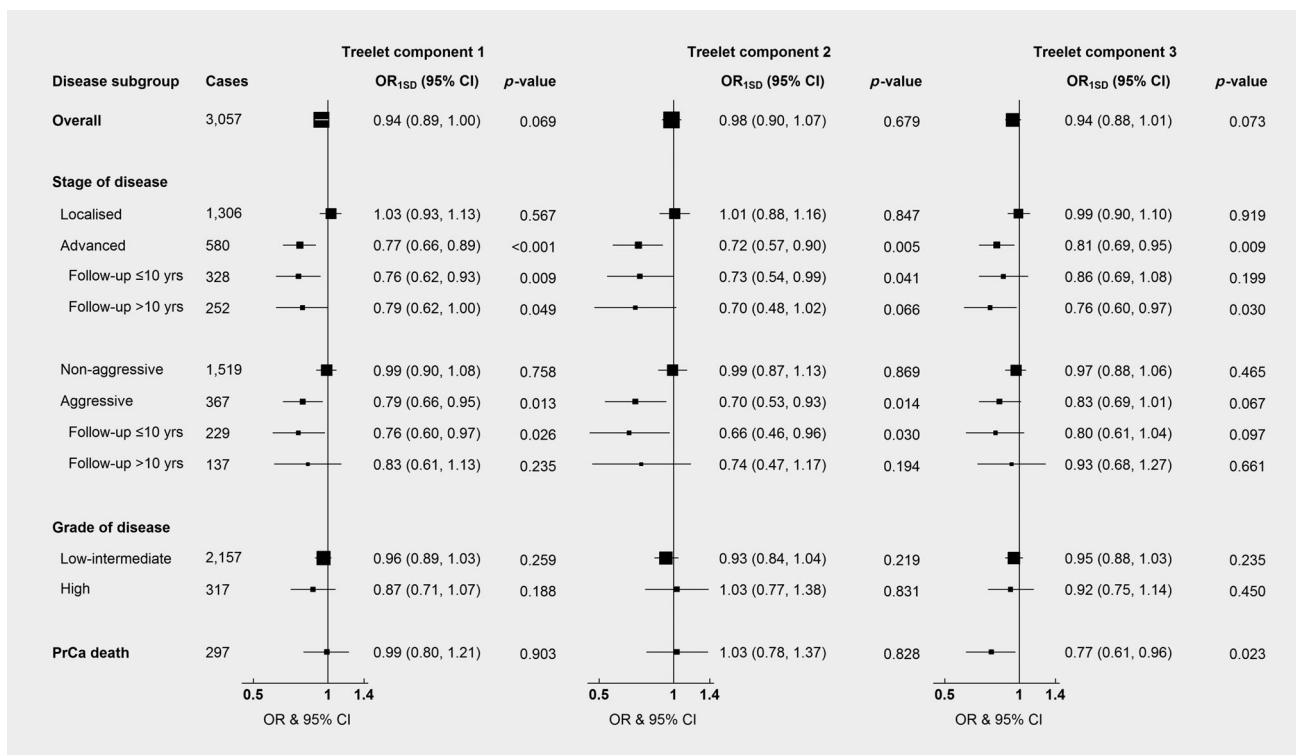


Figure 3. Odds ratio of prostate cancer overall and subgroups associated with a one standard deviation increase in treelet component scores. Stage and grade of prostate cancer were categorized using the tumor-node-metastasis (TNM) system and Gleason score, respectively; localized ($\leq T_2$ and $N_{0/x}$ and M_0 , or stage coded as localized), advanced (T_{3-4} and/or N_{1-3} and/or M_1 , or coded as advanced), non-aggressive ($\leq T_3$ and $N_{0/x}$ and M_0), aggressive (a subset of advanced stage disease defined as T_4 and/or N_{1-3} and/or M_1), low-intermediate grade (Gleason <8 or coded as well, moderately or poorly differentiated tumors) and high grade (Gleason ≥ 8 or coded as undifferentiated tumors). Death from prostate cancer during follow-up was defined as prostate cancer listed as the underlying cause of death on the death certificate. One matched set was excluded from the analysis of aggressive prostate cancer with follow-up >10 years; the case was the only individual with unknown smoking status in this subgroup and the model was thus not stable. Abbreviations: OR_{1SD}, odds ratio for a one standard deviation increase in treelet component score; CI, confidence interval; yrs, years; PrCa, prostate cancer.

pattern (treelet component 1), while the association of treelet component 3 with prostate cancer death may be independent of treelet component 1.

To the best of our knowledge, only one other prospective study of metabolites and prostate cancer risk has applied dimension-reduction methods to the full set of metabolites. None of 10 metabolite patterns derived using PCA were related to risk of overall or aggressive prostate cancer (defined by combining stage and grade information) in observational analyses within the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; 695 metabolites measured using mass spectrometry were included, but the metabolites which characterized each pattern were not described.¹⁹ Use of different metabolomics platforms, dimension-reduction methods, study populations and tumor subtype categorization may partly explain the differences in results.

The clustering of phosphatidylcholine and sphingomyelins (treelet component 1) as well as the positive correlation between treelet components 1 and 3 (lysophosphatidylcholines) observed here may be explained by shared metabolic pathways; sphingomyelins are synthesized from phosphatidylcholines,²⁰

and lysophosphatidylcholines are a subtype of phosphatidylcholines with one fatty acid side chain. In line with our results, treelet transform applied to untargeted metabolomics and lipidomics data from EPIC-Potsdam also identified a pattern characterized by positive loadings on a sphingomyelin, three phosphatidylcholines and other glycerophospholipids¹² (a class including phosphatidylcholines²¹). Diet, fitness level, alcohol intake and body mass index might be modifiable determinants of these metabolites.^{22–26}

The current results on phosphatidylcholines and sphingomyelins (treelet component 1) and risk are broadly consistent with those of other prospective studies investigating individual metabolites, including our previously published results,⁶ which were conducted in a subset (~35%) of the current study population. For example, we previously reported inverse associations of many individual phosphatidylcholines, all of which were included with positive loadings in treelet component 1 or 3, with risk of advanced stage prostate cancer, and some associations remained when restricting the analysis to aggressive prostate cancer. We observed generally similar but less strong associations when restricting the analysis to just the new data included in the current analysis. Comparison with results from other

studies is complex because of the use of different analytical platforms, for which only a limited number of metabolites (acylcarnitines, amino acids and biogenic amines) perfectly overlap with the metabolites studied here, and of a different classification of prostate cancer subtypes (aggressive prostate cancer was defined by combining information on stage and grade, in previous publications from other groups). Nonetheless, observational analyses within the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC) found that men with higher concentrations of some glycerophospholipids had lower risk of aggressive prostate cancer⁸ and possibly tumor stage T3²⁷; these results appear similar to our findings. In contrast, little is known about the association of lysophosphatidylcholines (treelet component 3) with risk of dying from prostate cancer. Another study in ATBC reported an inverse association between lysolipid 1-linoleoyl-glycerophosphatidylcholine (18:2), which is comparable to lysoPC a C18:2 included in TC3 of the current study, and risk of death from prostate cancer ($n_{\text{deaths}} = 523$),¹⁰ but death was not an outcome in the other previous papers.

The metabolite classes included in treelet components 1 and 3 are structural components of cell membranes, and along with the related enzymes and intermediates they are involved in proliferation, cell signaling and cell survival, and thus may play key roles in carcinogenesis, progression and migration of tumors.^{20,21,28} Their potential association may apply to cancer in general rather than being prostate cancer-specific; some prospective studies have reported lower risk of breast cancer,^{9,29} colorectal cancer,³⁰ hepatocellular carcinoma^{31,32} and pancreatic cancer³³ in relation to higher levels of phosphatidylcholines and related metabolites, while another study did not observe an association between the few included phospholipids and breast cancer risk.³⁴

Involvement in shared metabolic pathways may also explain the grouping of metabolites in treelet component 2. Acylcarnitines are involved in β -oxidation of fatty acids, which produces acetyl coenzyme A for energy production via the Krebs cycle.³⁵ Glutamate also feeds into this cycle by conversion to α -ketoglutarate but can also be converted to ornithine. Moreover, taurine has been proposed to stimulate β -oxidation and support mitochondrial function.³⁶ Thus, alteration in energy metabolism could be one possible underlying mechanism of the observed associations between treelet component 2 and reduced risk of advanced stage and aggressive prostate cancer. In line with our results, analyses in ATBC showed an inverse association between metabolites related to energy metabolism (α -ketoglutarate and citrate) and risk of aggressive prostate cancer.^{7,8} Our previous study in EPIC also suggested inverse associations of acylcarnitines and taurine with advanced prostate cancer.⁶ There are also some published data on dietary and lifestyle correlates of the metabolites in treelet component 2.^{23,26,37,38}

Altered levels of phosphatidylcholines, lysophospholipids (mostly lower), sphingomyelins, acylcarnitines, glutamate, ornithine (mostly lower) and taurine in postdiagnostic samples from prostate cancer patients have been reported.^{4,39,40} The most recent of these studies, used Mendelian randomization to assess the

causality of cross-sectionally observed associations between 14 serum metabolites (including tyrosine), measured using nuclear magnetic resonance spectroscopy, and overall risk of prostate cancer, detected via prostate-specific antigen test.⁴⁰ No associations between genetic markers of these metabolites (explaining a median of 6.8% [range 0.4–11.2%] of the variance in metabolite concentrations⁴¹) and risk of prostate cancer were observed.⁴⁰ This suggests no evidence of causality in the associations between these specific metabolites and risk of prostate cancer overall. While these case-control studies could imply that our results may be driven by tumors already present at blood collection, our prospective analysis excluding the first 10 years of follow-up suggests that differences in metabolite concentrations may precede carcinogenesis and thus potentially be relevant for the etiology of prostate cancer.

This study benefits from the large, mature and well-described cohort, which provided more statistical power than previous papers, and limited potential reverse causation and confounding. Additionally, treelet transform was applied to reduce issues of multiple testing and correlated variables pertaining to univariate analyses, while producing interpretable and stable metabolite patterns. For comparison, we also derive metabolite patterns using the more common PCA, and associations of the first and second principal components with prostate cancer risk were similar to those of the first and second treelet components, respectively. The advantage of treelet transform for analysis of metabolomics data in epidemiological studies has further been highlighted by a study also observing that treelet components were similar to and easier to interpret than components from PCA, albeit with somewhat less variance explained.¹² A limitation of treelet transform is the need to select a cut-level to produce components. However, the choice is aided by data-driven cross-validation, and the sensitivity analysis of the cut-level produced similar components.

Further limitations of this study are related to the blood sampling, handling and assay procedure. Only one blood sample was available per participant, which may lead to attenuation of risk estimates if a single measurement does not represent long-term exposure. Although studies of short-term reproducibility suggest that one measurement may be adequate for most metabolites (median intercorrelation coefficients of 0.45–0.70 over 0.3–2.3 years),^{42–45} long-term reproducibility is lower (mean $r = 0.28, 0.13$ and 0.16 over 15 years for all measured metabolites, diacyl and acyl-alkyl phosphatidylcholines, respectively).⁴⁴ Moreover, differences in sample handling procedures between study centers and the use of nonfasting blood samples might potentially have attenuated our risk estimates. To minimize this, cases and controls were matched on study center and fasting status, and the latter has been shown to explain only a small amount of variability in metabolite concentrations.^{23,46–48}

In conclusion, in this the largest study to date, we found evidence that easily interpretable patterns in baseline plasma metabolite profile, identified using treelet transform, are associated with subsequent risk of more aggressive prostate cancer subtypes. Men with metabolite profiles characterized by higher

concentrations of either phosphatidylcholines and hydroxysphingomyelins, or acylcarnitines C18:1 and C18:2, glutamate, ornithine and taurine, had lower risk of advanced and aggressive stage prostate cancer. Differences in scores on these metabolite patterns between cases and controls may precede diagnosis of advanced stage prostate cancer by more than 10 years and thus might be relevant to further understanding of prostate cancer etiology and prevention. Moreover, men with metabolite profiles high in lysophosphatidylcholines may be at lower risk of dying from prostate cancer.

Acknowledgements

We thank all participants in the EPIC cohort for their contribution to the study.

References

1. Ferlay J, Ervik M, Lam F, et al. *Global Cancer Observatory: Cancer Today*. Lyon, France: International Agency for Research on Cancer, 2018. Available from: <http://gco.iarc.fr/> [date last accessed 14 September 2018].
2. Price AJ, Key TJ. Epidemiology of prostate cancer. In: Hamdy FC, Eardley I, eds *Oxford textbook of urological surgery*. Oxford: Oxford University Press, 2017. 445–73.
3. Scalbert A, Brennan L, Fiehn O, et al. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics* 2009;5:435–58.
4. Kelly RS, Vander Heiden MG, Giovannucci E, et al. Metabolomic biomarkers of prostate cancer: prediction, diagnosis, progression, prognosis, and recurrence. *Cancer Epidemiol Biomarkers Prev* 2016;25:887–906.
5. Dunn WB, Goodacre R, Broadhurst DI, et al. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev* 2011; 40:387–426.
6. Schmidt JA, Fensom GK, Rinaldi S, et al. Pre-diagnostic metabolite concentrations and prostate cancer risk in 1077 cases and 1077 matched controls in the European Prospective Investigation into Cancer and Nutrition. *BMC Med* 2017; 15:122.
7. Mondul AM, Moore SC, Weinstein SJ, et al. 1-stearoylglycerol is associated with risk of prostate cancer: results from serum metabolomic profiling. *Metabolomics* 2014;10:1036–41.
8. Mondul AM, Moore SC, Weinstein SJ, et al. Metabolomic analysis of prostate cancer risk in a prospective cohort: the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study. *Int J Cancer* 2015;137:2124–32.
9. Kühn T, Floegel A, Sookthai D, et al. Higher plasma levels of lysophosphatidylcholine 18:0 are related to lower risk of common cancers in a prospective metabolomics study. *BMC Med* 2016;14:13.
10. Huang J, Mondul AM, Weinstein SJ, et al. Prospective serum metabolomic profiling of lethal prostate cancer. *Int J Cancer* 2019. <https://doi.org/10.1002/ijc.32218>.
11. Chadeau-Hyam M, Campanella G, Jombart T, et al. Deciphering the complex: methodological overview of statistical models to derive OMICS-based biomarkers. *Environ Mol Mutagen* 2013;54: 542–57.
12. Foerster J, Hyötyläinen T, Oresic M, et al. Serum lipid and serum metabolite components in relation to anthropometric parameters in EPIC-Potsdam participants. *Metabolism* 2015;64: 1348–58.
13. Edmands WMB, Ferrari P, Rothwell JA, et al. Polyphenol metabolome in human urine and its association with intake of polyphenol-rich foods across European countries. *Am J Clin Nutr* 2015; 102:905–13.
14. Dahm CC, Gorst-Rasmussen A, Crowe FL, et al. Fatty acid patterns and risk of prostate cancer in a case-control study nested within the European Prospective Investigation into Cancer and Nutrition. *Am J Clin Nutr* 2012;96:1354–61.
15. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002;5:1113–24.
16. Gorst-Rasmussen A, Dahm CC, Dethlefsen C, et al. Exploring dietary patterns by using the treelet transform. *Am J Epidemiol* 2011;173: 1097–104.
17. Gorst-Rasmussen A. tt: Treelet transform with Stata. *Stata J* 2012;12:130–46.
18. Assi N, Moskal A, Slimani N, et al. A treelet transform analysis to relate nutrient patterns to the risk of hormonal receptor-defined breast cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Public Health Nutr* 2016;19:242–54.
19. Huang J, Mondul AM, Weinstein SJ, et al. Serum metabolomic profiling of prostate cancer risk in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. *Br J Cancer* 2016;115: 1087–95.
20. Taniguchi M, Okazaki T. The role of sphingomyelin and sphingomyelin synthases in cell death, proliferation and migration—from cell and animal models to human disorders. *Biochim Biophys Acta* 2014;1841:692–703.
21. Ridgway ND. The role of phosphatidylcholine and choline metabolites to cell proliferation and survival. *Crit Rev Biochem Mol Biol* 2013;48: 20–38.
22. Floegel A, Wientzek A, Bachlechner U, et al. Linking diet, physical activity, cardiorespiratory fitness and obesity to serum metabolite networks: findings from a population-based study. *Int J Obes (Lond)* 2014;38:1388–96.
23. Schmidt JA, Rinaldi S, Ferrari P, et al. Metabolic profiles of male meat eaters, fish eaters, vegetarians, and vegans from the EPIC-Oxford cohort. *Am J Clin Nutr* 2015;102:1518–26.
24. Wittenbecher C, Mühlenbruch K, Kroger J, et al. Amino acids, lipid metabolites, and ferritin as potential mediators linking red meat consumption to type 2 diabetes. *Am J Clin Nutr* 2015;101: 1241–50.
25. Carayol M, Leitzmann MF, Ferrari P, et al. Blood metabolic signatures of body mass index: a targeted metabolomics study in the EPIC cohort. *J Proteome Res* 2017;16:3137–46.
26. van Roekel EH, Trijbsburg L, Assi N, et al. Circulating metabolites associated with alcohol intake in the European Prospective Investigation into Cancer and Nutrition Cohort. *Nutrients* 2018; 10:654.
27. Huang J, Mondul AM, Weinstein SJ, et al. Prospective serum metabolomic profile of prostate cancer by size and extent of primary tumor. *Oncotarget* 2017;8:45190–9.
28. Cheng M, Bhujwala ZM, Glunde K. Targeting phospholipid metabolism in cancer. *Front Oncol* 2016;6:266.
29. Lécuyer L, Victor Bala A, Deschasaux M, et al. NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer. *Int J Epidemiol* 2018; 47:484–94.
30. Xiang S, Yong-Bing X, Nathaniel R, et al. Prospective study of blood metabolites associated with colorectal cancer risk. *Int J Cancer* 2018;143:527–34.
31. Fages A, Duarte-Salles T, Stepien M, et al. Metabolomic profiles of hepatocellular carcinoma in a European prospective cohort. *BMC Med* 2015;13:242.
32. Jee SH, Kim M, Kim M, et al. Metabolomics profiles of hepatocellular carcinoma in a Korean prospective cohort: the Korean Cancer Prevention Study-II. *Cancer Prev Res (Phila)* 2018;11:303–12.
33. Shu X, Zheng W, Yu D, et al. Prospective metabolomics study identifies potential novel blood metabolites associated with pancreatic cancer risk. *Int J Cancer* 2018;143:2161–7.
34. Playdon MC, Ziegler RG, Sampson JN, et al. Nutritional metabolomics and breast cancer risk in a prospective study. *Am J Clin Nutr* 2017;106:637–49.

35. McKee T, McKee J. *Biochemistry—the molecular basis of life*, 3rd edn. New York: McGraw-Hill, 2003. 378–9.
36. Shigeru M. Role of taurine in the pathogenesis of obesity. *Mol Nutr Food Res* 2015;59:1353–63.
37. Bouchard-Mercier A, Rudkowska I, Lemieux S, et al. The metabolic signature associated with the Western dietary pattern: a cross-sectional study. *Nutr J* 2013;12:158.
38. Schmidt JA, Rinaldi S, Scalbert A, et al. Plasma concentrations and intakes of amino acids in male meat-eaters, fish-eaters, vegetarians and vegans: a cross-sectional analysis in the EPIC-Oxford cohort. *Eur J Clin Nutr* 2016;70:306–12.
39. Lima AR, Bastos Mde L, Carvalho M, et al. Biomarker discovery in human prostate cancer: an update in metabolomics studies. *Transl Oncol* 2016;9:357–70.
40. Adams C, Richmond RC, Santos Ferreira DL, et al. Circulating metabolic biomarkers of screen-detected prostate cancer in the ProtecT study. *Cancer Epidemiol Biomarkers Prev* 2019;28:208–16.
41. Kettunen J, Demirkan A, Wurtz P, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* 2016;7:11122.
42. Floegel A, Drogan D, Wang-Sattler R, et al. Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. *PLoS One* 2011;6:e21103.
43. Townsend MK, Clish CB, Kraft P, et al. Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. *Clin Chem* 2013;59:1657–67.
44. Kühn T, Sookthai D, Rolle-Kampczyk U, et al. Mid- and long-term correlations of plasma metabolite concentrations measured by a targeted metabolomics approach. *Metabolomics* 2016;12:184.
45. Carayol M, Licaj I, Achaintre D, et al. Reliability of serum metabolites over a two-year period: a targeted metabolomic approach in fasting and non-fasting samples from EPIC. *PLoS One* 2015;10:e0135437.
46. Sampson JN, Boca SM, Shu XO, et al. Metabolomics in epidemiology: sources of variability in metabolite measurements and implications. *Cancer Epidemiol Biomarkers Prev* 2013;22:631–40.
47. Fages A, Ferrari P, Monni S, et al. Investigating sources of variability in metabolomic data in the EPIC study: the principal component partial R-square (PC-PR2) method. *Metabolomics* 2014;10:1074–83.
48. Townsend MK, Bao Y, Poole EM, et al. Impact of pre-analytic blood sample collection factors on metabolomics. *Cancer Epidemiol Biomarkers Prev* 2016;25:823–9.