American Journal of Epidemiology
Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health 2018. This work is written by (a) US Government employee(s) and is in the public domain in the US.

Vol. 187, No. 8
DOI: 10.1093/aje/kwy017
Advance Access publication:
January 30, 2018

Original Contribution

Serum Metabolomic Profiling of All-Cause Mortality: A Prospective Analysis in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study Cohort

Jiaqi Huang, Stephanie J. Weinstein, Steven C. Moore, Andriy Derkach, Xing Hua, Linda M. Liao, Fangyi Gu, Alison M. Mondul, Joshua N. Sampson, and Demetrius Albanes*

* Correspondence to Dr. Demetrius Albanes, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Drive, Bethesda, MD 20892, USA (e-mail: daa@nih.gov).

Initially submitted September 21, 2017; accepted for publication January 23, 2018.

Tobacco use, hypertension, hyperglycemia, overweight, and inactivity are leading causes of overall and cardiovascular disease (CVD) mortality worldwide, yet the relevant metabolic alterations responsible are largely unknown. We conducted a serum metabolomic analysis of 620 men in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (1985–2013). During 28 years of follow-up, there were 435 deaths (197 CVD and 107 cancer). The analysis included 406 known metabolites measured with ultra-high-performance liquid chromatography/mass spectrometry–gas chromatography/mass spectrometry. We used Cox regression to estimate mortality hazard ratios for a 1-standard-deviation difference in metabolite signals. The strongest associations with overall mortality were *N*-acetylvaline (hazard ratio (HR) = 1.28; $P < 4.1 \times 10^{-5}$, below Bonferroni statistical threshold) and dimethylglycine, 7-methylguanine, *C*-glycosyltryptophan, taurocholate, and *N*-acetyltryptophan ($1.23 \leq \text{HR} \leq 1.32$; $5 \times 10^{-5} \leq P \leq 1 \times 10^{-4}$). *C*-Glycosyltryptophan, 7-methylguanine, and 4-androsten-3β,17β-diol disulfate were statistically significantly associated with CVD mortality ($1.49 \leq \text{HR} \leq 1.62$, $P < 4.1 \times 10^{-5}$). No metabolite was associated with cancer mortality, at a false discovery rate of <0.1. Individuals with a 1-standard-deviation higher metabolite risk score had increased all-cause and CVD mortality in the test set (HR = 1.4, $P = 0.05$; HR = 1.8, $P = 0.003$, respectively). The several serum metabolites and their composite risk score independently associated with all-cause and CVD mortality may provide potential leads regarding the molecular basis of mortality.

7-methylguanine; all-cause mortality; bile acids; cardiovascular disease mortality; *C*-glycosyltryptophan; dimethylglycine; *N*-acetylvaline; serum metabolomics

Abbreviations: ATBC, Alpha-Tocopherol, Beta-Carotene Cancer Prevention; CVD, cardiovascular disease; FDR, false discovery rate; HR, hazard ratio; ICD, *International Classification of Diseases*; SD, standard deviation.

Leading causes of mortality worldwide include tobacco use, overweight, hypertension, hyperglycemia, and physical inactivity (1), with these risk factors contributing to a higher risk of cardiovascular disease (CVD), diabetes, and cancer. However, the underlying biological mechanisms and biochemical actions that could serve as therapeutic or preventive targets are not completely understood. Advances in laboratory technologies of liquid and gas chromatography, mass spectrometry and nuclear magnetic resonance have enabled population-based metabolomic studies to quantify a broader spectrum of low–molecular weight metabolites in biospecimens, including serum. Such metabolomic profiles reflect influence of exogenous and endogenous exposures and, when coupled with health status, may offer insight into biochemical pathways involved in disease pathogenesis and mortality.

Very limited prospective population data exist relating circulating metabolites to overall mortality. One targeted study found that 4 out of 106 plasma biomarkers were associated with all-cause and CVD mortality in the Estonian Biobank and FINRISK study ($n = 17{,}345$) (2), but these were primarily large proteins or lipoproteins with known CVD functions (i.e., α-1-acid glycoprotein, albumin, and very-low-density lipoprotein particle size). Another untargeted serum metabolomic analysis identified that 9 out of 204 metabolites (including cotinine, mannose, and

γ-glutamyl-leucine) were associated with all-cause mortality among 1,887 African Americans in the Atherosclerosis Risk in Communities (ARIC) Study (3).

To evaluate serum metabolites independently associated with mortality risk, we conducted an untargeted, prospective, serum metabolomic analysis of overall mortality in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study cohort, with up to 28 years of outcome follow-up.

METHODS

Study population

Details of ATBC Study have been documented elsewhere (4). Briefly, a total of 29,133 male smokers aged 50–69 years were recruited into the trial from 1985 to 1988 in southwest Finland. Participants were randomly assigned to receive α-tocopheryl-acetate (50 mg), β-carotene (20 mg), both, or placebo daily for 5–8 years. At a baseline, presupplementation visit, fasting blood samples were collected and stored at −70°C; risk-factor questionnaires were completed; and height, weight, and total and high-density lipoprotein cholesterol were measured (4). The ATBC Study was approved by institutional review boards at the US National Cancer Institute and the Finnish National Public Health Institute. All participants provided written informed consent.

The present analysis is based on control participants without cancer (cancer-free at the index date) previously selected for one of 5 metabolomic studies nested within the ATBC Study (Web Figure 1 (available at https://academic.oup.com/aje): metabolomic set 1 ($n = 186$), metabolomic set 2 ($n = 38$), metabolomic set 3 ($n = 67$), metabolomic set 4 ($n = 131$), and metabolomic set 5 ($n = 198$)) (5–7). After excluding duplicate samples, our study included 620 men.

Outcome assessment

All-cause mortality, CVD-related mortality, and cancer-related mortality were ascertained through December 31, 2013, using the Causes of Death registry, Statistics Finland. All-cause mortality was defined as death from any cause, CVD mortality was defined where the underlying cause of death was CVD (*International Classification of Diseases* (ICD), ninth or tenth revision: ICD-9 codes 390–459 or ICD-10 codes I00–I99), and malignant neoplasms as underlying cause of death defined cancer mortality (ICD-9 codes 140–208 or ICD-10 codes C00–C96).

Metabolite assays

Fasting serum metabolites were measured using high-resolution, accurate-mass ultrahigh-performance liquid chromatography/mass spectroscopy and gas chromatography/mass spectroscopy at Metabolon, Inc. (Durham, North Carolina). Methodologic details of sample preparation, quality control, data extraction, and compound identification were described previously (8, 9). Metabolites with fewer than 10 nonmissing values within each metabolomic set were excluded, with 906 metabolites identified in at least one of the 5 metabolomic sets. After further excluding metabolites missing from 2 or more of the metabolomic sets, 406 metabolites remained in the final analysis. Of these, 406 metabolites were categorized into one of 8 chemical classes: amino acids, carbohydrates, cofactors and vitamins, energy metabolites, lipids, nucleotides, peptides, and xenobiotics (Web Table 1). Quality-control samples (9%) were assigned to each batch to evaluate technical reliability, and a coefficient of variation was calculated (median coefficient of variation = 9% (interquartile range, 4–20)) (5–7). In previous studies, we and others have evaluated the within-individual variability of metabolites over time. These studies have found that the median intraclass correlation coefficient of metabolites, based on samples separated by between 4 months and 2 years, was approximately 0.5 (10–12).

Statistical analysis

We batch-normalized each metabolite by dividing by the batch median. Undetected values (missing values) within each metabolite were imputed to the minimum value. The metabolite levels were then processed through log-transformation and normalization. Within each of the 5 data sets, we examined the association between the metabolite level and all-cause, CVD, and cancer mortality using Cox proportional hazard regression, using attained age as the time scale. For Cox regression, among subjects included in the nested case-control studies, start date was the index date at which the individually matched case was diagnosed with cancer. Among subjects included in the evaluation of vitamin supplementation (5), the start date was the baseline enrollment date. We thus removed from our analysis "immortal" person-time (13), which is the person-time during which an event case could not have occurred. In the models, we adjusted for age at blood collection (continuous), body mass index (calculated as weight (kg)/height $(m)^2$; continuous), number of cigarettes per day (continuous), total cholesterol (continuous), high-density lipoprotein cholesterol (continuous), history of hypertension (elevated blood pressure), history of diabetes mellitus, and serum creatinine (continuous). ATBC intervention group (as a categorical variable) was omitted because it was not associated with baseline serum levels. We additionally adjusted for physical activity and dietary factors (total energy intake, fruit intake, vegetable intake, and red meat consumption) as potential confounder factors in the models, and they did not change the effect of any metabolite remarkably, thus they were not included in the final model. We then performed a fixed-effect meta-analysis to obtain single estimated hazard ratios and 95% confidence intervals to describe the association between each metabolite level with all-cause, CVD, and cancer-related mortality. We also fitted crude models for our top metabolite signals that adjusted only for age at blood collection (continuous) in order to evaluate and not overadjust for potential mediators. To account for multiple testing (14), a Bonferroni-corrected threshold of statistical significance was defined as $P < 4.1 \times 10^{-5}$ (across tests for 406 metabolites and 3 outcomes).

Each metabolomic set was divided into a training set and test set (70% and 30%, respectively). Using only the former, we identified metabolites (false discovery rate (FDR) < 0.1) associated with all-cause mortality (number of metabolites = 12), CVD mortality ($n = 12$), and cancer mortality ($n = 0$; no further training-test analysis). In each training set, we performed a Cox regression with all qualifying metabolites and then used fixed-effects meta-analysis to obtain a single set of coefficients. In the test set, we constructed a metabolite risk score, a linear sum of

metabolite levels weighted by their corresponding coefficients. Then the metabolite risk score was normalized and used as both a continuous variable (per standard deviation (SD)) and categorized quartiles to estimate the associations with each outcome (all-cause or CVD mortality), using Cox proportional hazard regression (attained age as time scale). All models adjusted for multiple covariates as described above. If the number of covariates exceeded the size of the training set (or the model fit did not converge) for one of the 5 data sets, results from that training set were not included in the meta-analysis.

Pathway analyses assessed the associations between chemical classes and subclasses of metabolites and mortality. For each pathway, we created a single measure of significance, a *P* value based on Fisher's statistic (e.g., sum of log-*P* values) to combine the *P* values for the score statistics from the Cox regression (after adjustment for multiple covariates). Because of the correlation between metabolites, we calculated the *P* value based on Fisher's statistic using a parametric bootstrap (15, 16). For each bootstrap replication, we generated a vector of score test statistics from a multivariate normal distribution with mean 0 and estimated covariance matrix (15). Fisher's statistic was recalculated for each replication, and the reported *P* value for each pathway is the proportion of the 10^5 permutations where the permuted statistic is more extreme than the observed value.

Correlations between top metabolites for the 3 outcomes were estimated using Pearson's coefficient. Metabolites with *r* values greater than 0.5 or lower than −0.5 were considered highly positively or negatively correlated, respectively. We used SAS, version 9.3 (SAS Institute, Inc., Cary, North Carolina), and R, version 3.2.3 (R Foundation for Statistical Computing, Vienna, Austria), for all analyses. All reported *P* values were 2-sided.

RESULTS

Baseline characteristics of the study population in each metabolomic set are presented in Table 1. A total of 620 participants were included from the 5 metabolomic sets, with a follow-up period of up to 28 years (median, 10.7 years (interquartile range, 5.7–18.5 years)), during which there were 435 deaths, including 197 CVD deaths and 107 cancer deaths. We observed no meaningful differences in these characteristics across sets.

Metabolites associated with all-cause mortality

After adjustment for multiple covariates, metabolites associated with all-cause mortality with a FDR of ≤0.05 are shown in Table 2, sorted by *P* value. The amino acid *N*-acetylvaline,

Table 1. Baseline Characteristics of 620 Men in 5 Metabolomic Sets in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study, Finland, 1985–2013[a]

Characteristic	Metabolomic Set 1 (*n* = 186) Median (IQR)	Metabolomic Set 2 (*n* = 38) Median (IQR)	Metabolomic Set 3 (*n* = 67) Median (IQR)	Metabolomic Set 4 (*n* = 131) Median (IQR)	Metabolomic Set 5 (*n* = 198) Median (IQR)
Age at baseline, years	58 (53–61)	60 (56–63)	56 (53–62)	57 (54–61)	59 (55–64)
Follow-up time, years	20.2 (13.7–26.9)	9.0 (5.4–14.3)	6.2 (3.7–9.7)	8.5 (4.6–13.5)	8.4 (3.6–13.5)
No. of deaths					
All-cause mortality	128	29	39	94	145
CVD mortality	63	13	16	53	52
Cancer mortality	34	9	10	20	34
BMI[b], mean	26.0	26.0	26.3	26.2	26.0
Smoking, cigarettes/day	20 (15–25)	18 (10–20)	20 (15–25)	20 (15–24)	20 (13–25)
Serum HDL cholesterol, mmol/L	1.1 (1.0–1.3)	1.2 (1.0–1.3)	1.1 (1.0–1.4)	1.2 (1.0–1.4)	1.1 (1.0–1.3)
Serum total cholesterol, mmol/L	6.3 (5.6–7.1)	6.3 (5.6–6.8)	6.3 (5.3–7.1)	6.4 (5.6–6.9)	6.2 (5.4–6.8)
History of hypertension, %	16.1	10.5	10.5	21.4	17.7
History of diabetes mellitus, %	4.3	2.6	1.5	3.1	4.0
Physically active, %	24.2	21.1	23.9	24.4	19.2
Dietary intake per day					
Total energy, kcal	2,648 (2,231–3,063)	2,532 (2,060–3,122)	2,614 (2,255–3,084)	2,601 (2,222–3,024)	2,598 (2,070–3,066)
Fruit, g	121 (62–200)	131 (72–182)	124 (73–207)	110 (63–185)	98 (57–170)
Vegetables, g	113 (74–162)	116 (58–158)	106 (65–167)	102 (73–149)	98 (66–146)
Red meat, g	68.5 (52.2–95.5)	63.1 (49.6–83.6)	69.6 (49.5–90.7)	68.3 (49.5–88.9)	67.0 (49.0–88.1)

Abbreviations: BMI, body mass index; CVD, cardiovascular disease; HDL, high-density lipoprotein; IQR, interquartile range.
[a] Data for continuous variables are shown as median (IQR); otherwise as indicated.
[b] Calculated as weight (kg)/height $(m)^2$.

Table 2. Associations Between All-Cause Mortality and Serum Metabolites[a] in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study, Finland, 1985–2013

Metabolite	Chemical Class	Chemical Subclass	HR[b]	95% CI	*P* Value
N-Acetylvaline[c]	Amino acid	Valine, leucine, and isoleucine metabolism	1.28	1.14, 1.43	2.02×10^{-5}
Dimethylglycine	Amino acid	Glycine, serine, and threonine metabolism	1.26	1.13, 1.40	4.56×10^{-5}
7-Methylguanine	Nucleotide	Purine metabolism, guanine-containing	1.31	1.15, 1.51	8.77×10^{-5}
C-Glycosyltryptophan	Amino acid	Tryptophan metabolism	1.32	1.15, 1.51	9.18×10^{-5}
Taurocholate	Lipid	Primary bile acid metabolism	1.23	1.10, 1.36	1.33×10^{-4}
N-Acetyltryptophan	Amino acid	Tryptophan metabolism	1.24	1.11, 1.39	1.38×10^{-4}
Erythronate	Carbohydrate	Aminosugar metabolism	1.23	1.09, 1.39	6.62×10^{-4}
4-Androsten-3β,17β-diol disulfate 1	Lipid	Sterol/steroid	1.21	1.08, 1.36	7.46×10^{-4}
N-Acetylmethionine	Amino acid	Cysteine, methionine, SAM, taurine metabolism	1.23	1.09, 1.39	7.62×10^{-4}
5,6-Dihydrothymine	Nucleotide	Pyrimidine metabolism, thymine-containing	1.28	1.11, 1.48	9.46×10^{-4}
Hexanoylcarnitine	Lipid	Carnitine metabolism	1.20	1.08, 1.34	1.17×10^{-3}
Palmitoleate (16:1n7)	Lipid	Long-chain fatty acid	1.20	1.08, 1.35	1.26×10^{-3}
5-Dodecenoate (12:1n7)	Lipid	Medium-chain fatty acid	1.20	1.08, 1.35	1.29×10^{-3}
N-Acetylphenylalanine	Amino acid	Phenylalanine and tyrosine metabolism	1.20	1.07, 1.34	1.33×10^{-3}
Myristoleate (14:1n5)	Lipid	Long-chain fatty acid	1.20	1.07, 1.35	1.37×10^{-3}
*N*1-Methylguanosine	Nucleotide	Purine metabolism, guanine-containing	1.23	1.08, 1.39	1.37×10^{-3}
Docosadienoate (22:2n6)	Lipid	Long-chain fatty acid; polyunsaturated fatty acid (n3 and n6)	1.18	1.06, 1.32	2.70×10^{-3}
Taurochenodeoxycholate	Lipid	Primary bile acid metabolism	1.22	1.07, 1.39	2.73×10^{-3}
Homocitrulline	Amino acid	Urea cycle; arginine and proline metabolism	1.24	1.08, 1.44	2.81×10^{-3}
3-Hydroxycotinine glucuronide	Xenobiotics	Tobacco metabolite	1.26	1.08, 1.47	2.86×10^{-3}
N-Formylmethionine	Amino acid	Cysteine, methionine, SAM, taurine metabolism	1.20	1.06, 1.36	3.49×10^{-3}
3-Methyl catechol sulfate 1	Xenobiotics	Benzoate metabolism	1.18	1.05, 1.32	3.86×10^{-3}
4-Acetamidobutanoate	Amino acid	Guanidino and acetamido metabolism; polyamine metabolism	1.23	1.07, 1.41	3.88×10^{-3}
γ-Glutamyltryptophan	Peptide	γ-Glutamyl amino acid	1.19	1.06, 1.34	3.96×10^{-3}
Asparagine	Amino acid	Alanine and aspartate metabolism	0.86	0.78, 0.96	4.55×10^{-3}
N-Acetyltyrosine	Amino acid	Phenylalanine and tyrosine metabolism	1.16	1.05, 1.29	5.19×10^{-3}
Xylose	Carbohydrate	Pentose metabolism	1.17	1.05, 1.31	5.31×10^{-3}
3-Methyl catechol sulfate 2	Xenobiotics	Benzoate metabolism	1.22	1.06, 1.39	5.43×10^{-3}
*N*2,*N*2-Dimethylguanosine	Nucleotide	Purine metabolism, guanine-containing	1.20	1.05, 1.36	5.58×10^{-3}
3-Hydroxybutyrate (BHBA)	Lipid	Ketone bodies	1.17	1.05, 1.30	5.78×10^{-3}
4-Vinylphenol sulfate	Xenobiotics	Benzoate metabolism	1.17	1.05, 1.31	6.20×10^{-3}

Abbreviations: CI, confidence intervals; HDL, high-density lipoprotein; HR, hazard ratios; SAM, *S*-adenosylmethionine; SD, standard deviation.

[a] Metabolites were natural log-transformed and standardized (mean = 0, variance = 1). All 620 participants were included in each test, except for metabolites 5,6-dihydrothymine, taurochenodeoxycholate, homocitrulline, and 3-hydroxycotinine glucuronide (missing metabolic data from set 1 ($n = 186$), thus a total of 434 participants were included in these tests), for metabolites *N*-acetyltryptophan and 3-methyl catechol sulfate (missing metabolic data from set 2 ($n = 38$): a total of 582 participants were included in these tests). False discovery rate of ≤0.05.

[b] For each study set, we used attained age as time metric in the Cox proportional hazards regression model, and we adjusted for age at baseline, body mass index, number of cigarettes per day, total cholesterol, HDL cholesterol, history of hypertension (elevated blood pressure), history of diabetes mellitus, and serum creatinine. The reported HR (per SD) and *P* value were obtained from meta-analysis, which was conducted using a fixed-effects model to pool the study sets estimates.

[c] *N*-acetylvaline achieved statistical significance after Bonferroni correction for multiple tests.

which achieved statistical significance at the Bonferroni-corrected threshold for multiple tests, along with dimethylglycine, 7-methylguanine, *C*-glycosyltryptophan, taurocholate, and *N*-acetyltryptophan, showed the strongest signals, being positively associated with all-cause mortality after adjusting for conventional risk factors (meta-analysis hazard ratios (HRs) per SD metabolite increase were 1.23–1.32, and $2.02 \times 10^{-5} \leq P < 1.38 \times 10^{-4}$; all subsequent HRs are fully adjusted). The next most significant metabolites related to higher overall mortality included erythronate, 4-androsten-3β,17β-diol disulfate 1,

N-acetylmethionine, and 5,6-dihydrothymine (all $P < 9 \times 10^{-4}$). In addition, increased serum metabolites in long-chain fatty-acid metabolism (e.g., palmitoleate, myristoleate, and docosadienoate), purine metabolism (e.g., *N*1-methylguanosine and *N*2,*N*2-dimethylguanosine), and benzoate metabolism (e.g., 3-hydroxycotinine glucuronide, 3-methyl catechol sulfate 1 and 2) were associated with an increased all-cause mortality (Table 2). Associations for the top metabolite signals and all-cause mortality in the crude models, which adjusted only for age at blood collection, did not differ materially from the multivariable models (data not shown).

We combined the 12 metabolites identified from the training set with an FDR of <0.1 (Table 3) to yield a metabolite risk score and observed an elevated mortality in the test set for men having a higher score (HR = 1.38 per SD, 95% confidence interval: 1.08, 1.75; Table 3). For the categorized metabolite risk score, individuals in the third and fourth quartiles showed 94% and 73% higher mortality, respectively, than those in the lowest quartile (*P* for trend = 0.05; Table 3).

Metabolites associated with CVD mortality

Metabolites related to CVD mortality (after adjustment for multiple covariates) with a FDR of ≤0.05 are shown in Table 4, sorted by *P* value. We found that the following were related to elevated CVD mortality (per SD, HR = 1.38–1.62, and $8.4 \times 10^{-6} \leq P < 3.8 \times 10^{-4}$): higher serum amino acids *C*-glycosyltryptophan, 3-(4-hydroxyphenyl)lactate, *N*-acetylvaline, and dimethylglycine; nucleotide 7-methylguanine; and lipids 4-androsten-3β,17β-diol disulfate, taurocholate, and taurochenodeoxycholate. Of these, *P* values for *C*-glycosyltryptophan, 7-methylguanine, and 4-androsten-3β,17β-diol disulfate achieved statistical significance at the Bonferroni-corrected threshold. The next most significant metabolites related to higher CVD mortality included the peptide ADSGEGDFXAEGGGVR, 4-acetamidobutanoate, erythronate, *N*-acetylphenylalanine, and cortisol, with asparagine being inversely associated (all $P < 10^{-3}$) (Table 4). The crude model estimates that adjusted only for age at blood collection were similar to the multivariable-adjusted associations (data not shown).

For the CVD analysis, the metabolite risk score was based on the following 12 metabolites identified in the training set: *C*-glycosyltryptophan, dimethylglycine, *N*-acetylvaline, 4-androsten-3β,17β-diol disulfate 1, stearoyl-linoleoyl-GPPE, ADSGEGDFXAEGGGVR, 4-hydroxyhippurate, mannose, 3-hydroxyhippurate, *N*-acetylputrescine, asparagine, and oleic acid ethanolamide (Table 5). In the test set, we found an increased mortality in men with a higher metabolite risk score (HR = 1.83, 95% confidence interval: 1.28, 2.62) (Table 5). For the categorized metabolite risk score, individuals in the highest quartile experienced over four times the mortality risk when compared with those in the lowest quartile (HR = 4.35, *P* for trend: 0.003, Table 5).

Metabolites associated with cancer-related mortality

After adjustment for multiple covariates, no metabolite was associated with cancer mortality at either the FDR of ≤0.1 or Bonferroni threshold. Metabolites associated with cancer-related mortality with a nominal *P* value of <0.05 are shown in Table 6. The amino acids dimethylglycine, *N*-acetylvaline, levulinate, and *N*-acetylmethionine were the top metabolites positively associated with cancer-related mortality (per SD, HR = 1.42–1.62, and $6.7 \times 10^{-4} \leq P < 6.1 \times 10^{-3}$), as were the tobacco metabolites hydroxycotinine and cotinine *N*-oxide. By contrast, the amino acids indolepropionate and 3-phenylpropionate and the peptide glutamine-leucine were inversely associated with cancer mortality (per SD, HR = 0.63–0.70, and $8.4 \times 10^{-4} \leq P < 6.2 \times 10^{-3}$) (Table 6). As expected,

Table 3. Hazard Ratios for the Association of Metabolite Risk Score With All-Cause Mortality in the 30% Test Sets in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study, Finland, 1985–2013

Quartile of Metabolite Risk Score[a]	Event	Participant	Person-years	HR[b]	95% CI
1	27	44	690.4	1.00	Referent
2	25	41	570.7	1.30	0.71, 2.38
3	36	45	474.7	1.94	1.08, 3.48
4	37	43	400.6	1.73	0.91, 3.30
P value for trend				0.05	
Metabolite risk score as continuous variable (per SD)				1.38	1.08, 1.75

Abbreviations: CI, confidence intervals; FDR, false discovery rate; HDL, high-density lipoprotein; HR, hazard ratios; SD, standard-deviation.

[a] The risk score was generated by summing the top 12 metabolites (FDR of ≤0.1) identified in the test set, which were weighed by their regression coefficients in the training set. The risk score is calculated as 0.073 × (log-*N*-acetylvaline) + 0.085 × (log-*C*-glycosyltryptophan) + 0.07 × (log-dimethylglycine) + 0.084 × (log-*N*-acetylputrescine) + 0.077 × (log-erythronate) + 0.07 × (log-ADSGEGDFXAEGGGVR) + 0.071 × (log-mannose) + 0.082 × (log-taurochenodeoxycholate) + 0.065 × (log-taurocholate) + 0.095 × (log-5,6-dihydrothymine) + 0.078 × (log-*N*-acetylmethionine) + 0.085 × (log-acisoga).

[b] Hazard ratio for all-cause mortality generated from Cox proportional hazards regression, and adjusted for age at randomization, body mass index, number of cigarettes per day, history of diabetes, serum cholesterol, HDL, history of hypertension (elevated blood pressure), serum creatinine, and metabolomic sets.

the tobacco metabolites hydroxycotinine and cotinine *N*-oxide showed stronger associations with cancer-related mortality in the age-adjusted model (per SD, HR = 1.83 and 1.72, $P = 6.8 \times 10^{-5}$ and 5.0×10^{-5}, respectively). For the other top signals, associations with cancer mortality were not materially different in the crude models (data not shown).

Table 4. Associations of Cardiovascular Disease-Related Mortality and Serum Metabolites[a] in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study, Finland, 1985–2013

Metabolite	Chemical Class	Chemical Subclass	HR[b]	95% CI	*P* Value
C-Glycosyltryptophan[c]	Amino acid	Tryptophan metabolism	1.60	1.30, 1.96	8.4×10^{-6}
7-Methylguanine[c]	Nucleotide	Purine metabolism, guanine-containing	1.62	1.30, 2.03	1.9×10^{-5}
4-Androsten-3β,17β-diol disulfate 1[c]	Lipid	Sterol/steroid	1.49	1.24, 1.80	2.1×10^{-5}
3-(4-Hydroxyphenyl)lactate	Amino acid	Phenylalanine and tyrosine metabolism	1.43	1.20, 1.71	6.4×10^{-5}
Taurocholate	Lipid	Primary bile acid metabolism	1.38	1.18, 1.62	7.5×10^{-5}
N-Acetylvaline	Amino acid	Valine, leucine, and isoleucine metabolism	1.41	1.19, 1.68	8.6×10^{-5}
Taurochenodeoxycholate	Lipid	Primary bile acid metabolism	1.52	1.22, 1.91	2.7×10^{-4}
Dimethylglycine	Amino acid	Glycine, serine, and threonine metabolism	1.38	1.15, 1.65	3.8×10^{-4}
ADSGEGDFXAEGGGVR	Peptide	Fibrinogen cleavage peptide	1.38	1.15, 1.66	4.4×10^{-4}
4-Acetamidobutanoate	Amino acid	Guanidino and acetamido metabolism; polyamine metabolism	1.55	1.21, 1.98	4.6×10^{-4}
Erythronate	Carbohydrate	Aminosugar metabolism	1.40	1.15, 1.70	7.6×10^{-4}
N-Acetylphenylalanine	Amino acid	Phenylalanine and tyrosine metabolism	1.33	1.13, 1.57	7.9×10^{-4}
Cortisol	Lipid	Sterol/steroid	1.38	1.14, 1.68	9.3×10^{-4}
Asparagine	Amino acid	Alanine and aspartate metabolism	0.76	0.65, 0.90	9.8×10^{-4}
4-Hydroxyhippurate	Xenobiotics	Benzoate metabolism	1.35	1.12, 1.62	1.5×10^{-3}
*N*1-Methylguanosine	Nucleotide	Purine metabolism, guanine-containing	1.39	1.13, 1.71	1.6×10^{-3}
5-Dodecenoate (12:1n7)	Lipid	Medium-chain fatty acid	1.31	1.11, 1.56	1.9×10^{-3}
2-Hydroxyglutarate	Lipid	Fatty acid, dicarboxylate	1.38	1.13, 1.70	1.9×10^{-3}
γ-Glutamyltryptophan	Peptide	γ-Glutamyl amino acid	1.33	1.11, 1.60	2.4×10^{-3}
N-Acetyltyrosine	Amino acid	Phenylalanine and tyrosine metabolism	1.29	1.09, 1.52	2.4×10^{-3}
N-Acetyltryptophan	Amino acid	Tryptophan metabolism	1.30	1.10, 1.55	2.5×10^{-3}
N-Acetylmethionine	Amino acid	Cysteine, methionine, SAM, taurine metabolism	1.38	1.12, 1.69	2.7×10^{-3}
Palmitoleate (16:1n7)	Lipid	Long-chain fatty acid	1.29	1.09, 1.53	2.7×10^{-3}
5-Methylthioadenosine (MTA)	Amino acid	Polyamine metabolism	1.33	1.10, 1.60	2.8×10^{-3}
1-Linoleoylglycerophosphoethanolamine	Lipid	Lysolipid	1.30	1.09, 1.55	3.4×10^{-3}
*N*6-Carbamoylthreonyladenosine	Nucleotide	Purine metabolism, guanine-containing; purine metabolism, adenine-containing	1.39	1.11, 1.73	3.9×10^{-3}
Andro steroid monosulfate 1	Lipid	Sterol/steroid	1.29	1.08, 1.53	4.0×10^{-3}
4-Androsten-3β,17β-diol disulfate 2	Lipid	Sterol/steroid	1.31	1.09, 1.58	4.5×10^{-3}
Mannose	Carbohydrate	Fructose, mannose, galactose, starch, and sucrose metabolism	1.29	1.08, 1.54	4.6×10^{-3}
Glycine	Amino acid	Glycine, serine, and threonine metabolism	0.81	0.70, 0.94	4.7×10^{-3}
1-Oleoylglycerophosphoethanolamine	Lipid	Lysolipid	1.28	1.07, 1.52	5.4×10^{-3}
7-α-hydroxy-3-oxo-4-cholestenoate (7-HOCA)	Lipid	Sterol/steroid	1.27	1.07, 1.50	5.6×10^{-3}
Malate	Energy	Krebs cycle/tricarboxylic acid cycle	1.29	1.07, 1.54	6.2×10^{-3}
3-Hydroxybutyrate (BHBA)	Lipid	Ketone bodies	1.28	1.07, 1.53	6.4×10^{-3}
γ-CEHC glucuronide	Cofactors and vitamins	Tocopherol metabolism	1.26	1.07, 1.49	6.6×10^{-3}
Pseudouridine	Nucleotide	Pyrimidine metabolism, uracil-containing	1.30	1.08, 1.58	6.9×10^{-3}
Acetoacetate	Lipid	Ketone bodies	1.27	1.06, 1.51	7.6×10^{-3}
*N*6-Acetyllysine	Amino acid	Lysine metabolism	1.25	1.06, 1.48	7.6×10^{-3}

Table continues

Table 4. Continued

Metabolite	Chemical Class	Chemical Subclass	HR [b]	95% CI	*P* Value
Pregnen-diol disulfate	Lipid	Sterol/steroid	1.27	1.06, 1.51	8.1×10^{-3}
Myristoleate (14:1n5)	Lipid	Long-chain fatty acid	1.25	1.06, 1.47	8.2×10^{-3}
Glycochenodeoxycholate	Lipid	Primary bile acid metabolism	1.24	1.06, 1.46	8.4×10^{-3}
Histidine	Amino acid	Histidine metabolism	0.80	0.68, 0.95	8.7×10^{-3}
1-Linoleoylglycerophosphoinositol	Lipid	Lysolipid	1.25	1.06, 1.48	8.8×10^{-3}

Abbreviations: CI, confidence intervals; CVD, cardiovascular disease; HDL, high-density lipoprotein; HR, hazard ratios; SAM, *S*-adenosylmethionine; SD, standard-deviation.

[a] Metabolites were natural log-transformed and standardized (mean = 0, variance = 1). All 382 participants were included in each test (analysis included referent individuals ($n = 185$) and individuals with CVD-related death ($n = 197$)), except for metabolites taurochenodeoxycholate (missing metabolic data from set 1: a total of 261 participants were included in the test) and for metabolites ADSGEGDFXAEGGGVR, 2-hydroxyglutarate, *N*-acetyltryptophan, and γ-CEHC glucuronide (missing metabolic data from set 2: a total of 360 participants were included in these tests). False discovery rate of ≤0.05.

[b] For each study set, we used attained age as time metric in Cox proportional hazards regression models that adjusted for age at baseline, body mass index, number of cigarettes per day, total cholesterol, HDL cholesterol, history of hypertension (elevated blood pressure), history of diabetes mellitus, and serum creatinine. The reported HR (per SD) and *P* value were obtained from meta-analysis, which was conducted using a fixed-effects model to pool the study sets estimates.

[c] Achieved statistical significance after Bonferroni correction for multiple tests.

Pathway analysis

In the metabolite pathway analysis, we found that the amino-acid chemical class and primary-bile-acids subclass were significantly associated with all-cause and CVD mortality below the Bonferroni-corrected *P* value threshold (Web Tables 2–5). After Bonferroni correction, however, no chemical class or subclass was associated with cancer mortality (Web Tables 6 and 7).

The pathway analysis also showed that fructose/mannose/galactose/starch/sucrose metabolism, phenylalanine/tyrosine metabolism, glycine/serine/threonine metabolism, purine/guanine metabolism, tryptophan metabolism, and branched-chain amino acids metabolism were top chemical subclasses associated with all-cause and CVD mortality (Web Tables 3 and 5). The subclasses benzoate metabolism and tobacco metabolism ranked as the top pathways related to all-cause and cancer mortality but not for CVD mortality (Web Tables 3, 5 and 7). Of note, all the models adjusted for multiple mortality risk factors, including smoking. The correlation of the top metabolites for the 3 outcomes are presented in Web Figures 2–4. Higher positive correlations were observed among chemical subclasses of lipid fatty acids (myristoleate (14:1n5), 5-dodecenoate (12:1n7), palmitoleate (16:1n7)), bile acids (taurochenodeoxycholate, taurocholate,

Table 5. Hazard Ratios for the Association of Metabolite Risk Score With Cardiovascular Disease-Related Mortality in the 30% Test Sets in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study, Finland, 1985–2013

Quartile of Metabolite Risk Score [a]	Event	Participant	Person-Years	HR [b]	95% CI
1	11	32	462.9	1.00	Referent
2	15	32	450.5	2.13	0.87, 5.23
3	15	32	429.1	1.95	0.79, 4.80
4	21	32	265.2	4.35	1.78, 10.6
P value for trend				0.003	
Metabolite risk score as continuous variable (per SD)				1.83	1.28, 2.62

Abbreviations: CI, confidence intervals; CVD, cardiovascular disease; HDL, high-density lipoprotein; HR, hazard ratios; SD, standard deviation.

[a] The risk score was generated by summing the top 12 metabolites (false discovery rate of ≤0.1) identified in the test set, which were weighed by their regression coefficients in the training set. The risk score is calculated as 0.124 × (log-*C*-glycosyltryptophan) + 0.112 × (log-dimethylglycine) + 0.109 × (log-*N*-acetylvaline) + 0.112 × (log-4-Androsten-3β,17β-Diol Disulfate 1) + 0.145 × (log-stearoyl-linoleoyl-GPPE) + 0.117 × (log-ADSGEGDFXAEGGGVR) + 0.113 × (log-4-hydroxyhippurate) + 0.107 × (log-mannose) + 0.111 × (log-3-hydroxyhippurate) + 0.141 × (log-*N*-acetylputrescine) − 0.1 × (log-asparagine) + 0.177 × (log-oleic acid ethanolamide).

[b] Hazard ratio for CVD mortality generated from Cox proportional hazards regression and adjusted for age at randomization, body mass index, number of cigarettes per day, history of diabetes, serum cholesterol, HDL, history of hypertension (elevated blood pressure), serum creatinine, and metabolomic sets.

Table 6. Associations of Cancer-Related Mortality and Serum Metabolites ($P < 0.05$)[a] in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study, Finland, 1985–2013

Metabolite	Chemical Class	Chemical Subclass	HR[b]	95% CI	*P* Value
Dimethylglycine	Amino acid	Glycine, serine, and threonine metabolism	1.61	1.22, 2.11	6.7×10^{-4}
Indolepropionate	Amino acid	Tryptophan metabolism	0.63	0.48, 0.83	8.4×10^{-4}
N-Acetylvaline	Amino acid	Valine, leucine, and isoleucine metabolism	1.62	1.22, 2.16	9.4×10^{-4}
Levulinate (4-oxovalerate)	Amino acid; xenobiotics	Valine, leucine, and isoleucine metabolism; food component/plant	1.42	1.12, 1.79	3.4×10^{-3}
3-Phenylpropionate (hydrocinnamate)	Amino acid	Phenylalanine and tyrosine metabolism	0.70	0.55, 0.89	3.5×10^{-3}
N-Acetylmethionine	Amino acid	Cysteine, methionine, SAM, and taurine metabolism	1.60	1.14, 2.24	6.1×10^{-3}
Glutamine-leucine	Peptide	Dipeptide	0.70	0.54, 0.90	6.2×10^{-3}
Glycocholate sulfate	Lipid	Secondary bile acid metabolism	1.42	1.10, 1.83	7.0×10^{-3}
N-Acetylphenylalanine	Amino acid	Phenylalanine and tyrosine metabolism	1.50	1.11, 2.02	7.8×10^{-3}
N-Acetyltryptophan	Amino acid	Tryptophan metabolism	1.46	1.10, 1.93	8.9×10^{-3}
Hexanoylcarnitine	Lipid	Carnitine metabolism	1.42	1.08, 1.86	1.1×10^{-2}
O-Cresol sulfate	Amino acid	Phenylalanine and tyrosine metabolism	1.48	1.09, 2.00	1.1×10^{-2}
Ergothioneine	Xenobiotics	Food component/plant	0.74	0.58, 0.94	1.3×10^{-2}
Hydroxycotinine	Xenobiotics	Tobacco metabolite	1.54	1.10, 2.18	1.3×10^{-2}
Betaine	Amino acid	Glycine, serine, and threonine metabolism	1.44	1.08, 1.93	1.4×10^{-2}
Erythronate	Carbohydrate	Aminosugar metabolism	1.45	1.08, 1.95	1.4×10^{-2}
Cotinine *N*-oxide	Xenobiotics	Tobacco metabolite	1.45	1.08, 1.96	1.4×10^{-2}
Andro steroid monosulfate 1	Lipid	Sterol/steroid	1.42	1.07, 1.87	1.5×10^{-2}
1-Oleoylglycerol (1-monoolein)	Lipid	Monoacylglycerol	1.40	1.07, 1.84	1.5×10^{-2}
2-Stearoylglycerophosphoethanolamine	Lipid	Lysolipid	0.77	0.62, 0.95	1.6×10^{-2}
4-Vinylphenol sulfate	Xenobiotics	Benzoate metabolism	1.36	1.06, 1.75	1.6×10^{-2}
C-Glycosyltryptophan	Amino Acid	Tryptophan metabolism	1.43	1.06, 1.92	1.8×10^{-2}
5,6-Dihydrothymine	Nucleotide	Pyrimidine metabolism, thymine-containing	1.53	1.07, 2.19	2.0×10^{-2}
2-Ethylphenylsulfate	Xenobiotics	Benzoate metabolism	1.38	1.05, 1.81	2.0×10^{-2}
N-(2-Furoyl)glycine	Xenobiotics	Food component/plant	1.31	1.03, 1.68	3.0×10^{-2}
Succinate	Energy	Krebs cycle/tricarboxylic acid cycle	1.31	1.03, 1.67	3.1×10^{-2}
Homostachydrine	Xenobiotics	Food component/plant	1.30	1.02, 1.65	3.4×10^{-2}
Xylose	Carbohydrate	Pentose metabolism	1.29	1.02, 1.64	3.7×10^{-2}
Phenylalanylglycine	Peptide	Dipeptide	1.31	1.01, 1.70	4.0×10^{-2}
Phenol sulfate	Amino Acid	Phenylalanine and tyrosine metabolism	1.27	1.01, 1.59	4.2×10^{-2}
4-Hydroxyhippurate	Xenobiotics	Benzoate metabolism	1.31	1.00, 1.72	4.7×10^{-2}
Hydroquinone sulfate	Xenobiotics	Drug	1.32	1.00, 1.73	5.0×10^{-2}

Abbreviations: CI, confidence intervals; HDL, high-density lipoprotein; HR, hazard ratios; SAM, *S*-adenosylmethionine; SD, standard deviation.

[a] Metabolites were natural log-transformed and standardized (mean = 0, variance = 1). All 292 participants were included in each test (analysis included referent individuals ($n = 185$) and individuals with cancer-related death ($n = 107$)), except for metabolites 2-stearoylglycerophosphoethanolamine and 5,6-dihydrothymine (missing metabolic data from set 1: a total of 200 participants were included in these tests) and for metabolites indolepropionate, glutamine-leucine, N-acetyltryptophan, erythronate, 1-oleoylglycerol (1-monoolein), and homostachydrine (missing metabolic data from set 2: a total of 274 participants were included in these tests).

[b] For each study set, we used attained age as time metric in Cox proportional hazards regression models that adjusted for age at baseline, body mass index, number of cigarettes per day, total cholesterol, HDL cholesterol, history of hypertension (elevated blood pressure), history of diabetes mellitus, and serum creatinine. The reported HR (per SD) and *P* value were obtained from meta-analysis, which was conducted using a fixed-effects model to pool the study sets estimates.

glycochenodeoxycholate), sex steroids (pregnen-diol disulfate, 4-androsten-3β,17β-diol disulfate 1, 4-androsten-3β,17β-diol disulfate 2), purine nucleotides (*N*1-methylguanosine, *N*2, *N*2-dimethylguanosine), benzoate xenobiotics (3-methyl catechol sulfate 1, 3-methyl catechol sulfate 2, 4-vinylphenol sulfate), and tobacco xenobiotics (cotinine *N*-oxide, hydroxycotinine)

(Web Figures 2–4). We also show the metabolite-by-metabolite correlation matrix that constituted metabolite risk scores of all-cause and CVD deaths in Web Figures 5 and 6. Among the 12 metabolites included in the risk score for all-cause mortality, the pairwise correlation ranged from −0.05 to 0.84 (Web Figure 5), and among the 12 metabolites included in the risk score for CVD, the correlated ranged from −0.01 to 0.48 (Web Figure 6). The top positive correlation was seen within chemical subclass bile acids (taurochenodeoxycholate, taurocholate) and benzoate xenobiotics (3-hydroxyhippurate and 4-hydroxyhippurate), respectively.

DISCUSSION

In this study, we prospectively investigated associations between >400 serum metabolites and all-cause, CVD, and cancer mortality among 620 men during a median follow-up of 11 years. The all-cause-mortality–related metabolite *N*-acetylvaline, and CVD-mortality–related metabolites *C*-glycosyltryptophan, 7-methylguanine, and 4-androsten-3β,17β-diol disulfate yielded the strongest signals that exceeded the multiple-comparisons statistical threshold. By contrast, no metabolite was associated with cancer mortality at an FDR of <0.1. Validating in the test set the risk score that was based on the 12 top metabolites revealed that individuals with higher metabolite scores had elevated risks for all-cause and CVD mortality. Of note, men in the highest risk-score quartile experienced quadrupled CVD mortality compared with those in the lowest score quartile.

Serum *N*-acetylvaline was a top metabolite signal positively associated with all-cause and CVD mortality (it was also a top signal for cancer mortality). This metabolite is in the branched-chain amino acids, valine/leucine/isoleucine metabolism pathway. Our data also showed that serum *N*-acetylvaline was strongly correlated with serum valine ($r = 0.44$; $P = 1 \times 10^{-30}$). Branched-chain amino acid metabolites play a role in human health outcomes including cardiovascular disease, stroke, insulin resistance, diabetes, and pancreatic cancer (17–23), and they are associated with obesity and physical activity (24, 25). On the other hand, elevated *N*-acetyl amino acids, including *N*-acetylvaline, may indicate disruptions in acetylation activity that could influence cell homeostasis through histone-chromatin function and gene regulation (26–28). Whether some of these factors mediate the increased mortality–higher circulating *N*-acetylvaline (and other *N*-acetyl amino acids) associations, or a direct biological action influencing risk of death, will require further study.

Also strongly related to mortality was the tertiary amine dimethylglycine, which can be produced from betaine during the transfer of a methyl group from homocysteine to methionine, a reaction catalyzed by betaine-homocysteine methyltransferase (29). Serum dimethylglycine was significantly correlated with serum betaine in our data ($r = 0.26$; $P = 3 \times 10^{-11}$). Elevated plasma dimethylglycine has been associated with mortality risk, and it may enhance risk prediction of all-cause and CVD-related mortality, particularly among coronary heart disease patients (30). Dimethylglycine was also independently associated with incident acute myocardial infarction and improved outcome prediction among patients with stable angina (31). Regarding cancer, higher fecal dimethylglycine has been related to colorectal cancer in China (32), and urinary dimethylglycine has been correlated with clinical stage of hepatocellular carcinoma in West Africa (33). By contrast, circulating dimethylglycine was unrelated to colorectal cancer or prostate cancer risk in nested case-control studies (34–36).

Higher serum *C*-glycosyltryptophan and 7-methylguanine were associated with increased overall and CVD mortality (e.g., odds of overall and CVD mortality increased, respectively, by 30% and 60% with each 1-SD log-metabolite increase). *C*-glycosyltryptophan (also known as *C*-mannosyltryptophan) is a tryptophan glycoconjugate that has been used as a biomarker of kidney function (37–39) and is related to infectious burden and increased inflammation (40). It is strongly correlated with age and has been related to methylation of the promoter region of *WDR85*, a gene that may regulate diphthamide synthesis, important for RNA translation, cell cycle, and embryonic development, thereby supporting a role for *C*-glycosyltryptophan in aging and human development (41). It is possible that its role in aging and inflammation and its relationship to kidney function may partially account for the positive association we observed for overall and CVD mortality. 7-Methylguanine is a by-product of DNA methylation damage repair that is used as a marker of exposure to methylating agents and is potentially related to cancer and aging (42). Tumor tissues exhibit elevated 7-methylguanine levels (43, 44), which might reflect decreased defense against intracellular reactive oxygen species (43–45).

Taurocholate, taurochenodeoxycholate, and the primary bile acid metabolism pathway were also significantly associated with all-cause and CVD mortality. Taurocholate, the bile acid conjugate of cholic acid and taurine, like taurine itself, has been related to lower risks of hypertension, stroke, and other atherosclerotic diseases (46), while hepatic uptake of taurocholate may decline with age (47). Experimental intestinal infusion of taurocholate is related to Akt signaling pathway activation, a possible determinant of cellular senescence (48–50). Other biologically plausible metabolites associated with increased risk of CVD mortality include lower serum histidine and glycine. Histidine may be independently inversely associated with age (51), and it can be metabolized to carnosine, a known antioxidant characterized as an "anti-aging" biochemical based on suppression of oxidative damage, glycation of proteins, and scavenging toxic age-related molecules (52). Glycine is the precursor of several molecular species, including purines and glutathione, and a substantial body of evidence supports its beneficial role in cytoprotection, antioxidation, antiinflammation responses, and metabolic regulation (53–57). Increased CVD mortality was also related to elevated serum mannose, consistent with findings from the Atherosclerosis Risk in Communities study (3), as well as to lysolipids (e.g., 1-linoleoyl-glycerophosphoethanolamine), which are considered important cell-signaling molecules that contribute to regulation of cell differentiation, growth, proliferation, and invasion (58–62), and to steroid hormones in the androgen pathway (e.g., 4-androsten-3β,17β-diol disulfate 1).

Although we identified no metabolites significantly associated with cancer mortality in the model that adjusted for multiple covariates, tobacco metabolism was the top associated pathway, a finding consistent with population studies showing excess cancer mortality is associated with both tobacco smoking and higher circulating cotinine concentrations (63–68).

It is noteworthy but not unexpected that the crude models, not adjusting for tobacco smoking, showed strong cancer mortality associations for tobacco metabolites, including cotinine *N*-oxide and hydroxycotinine.

Several mortality-related metabolites we identified have been associated with chronic aging (41, 51). Considering multiple cellular actions, transformations, and cumulative cellular damage that occur across the life course, with cumulative health deterioration and eventually death, it is biologically plausible that the 2 related but distinct biological traits may be contributed to and regulated by several common molecular functions and biochemical pathways. Also, several of the metabolites we identified are associated with known epidemiologic risk factors related to mortality. These include, for example, physical activity and mannose (24); hypertension risk and tRNA-specific modified nucleoside *N*2,*N*2-dimethylguanosine and tricarboxylic-acid-cycle intermediate malate (69); several tobacco smoke-related metabolites (70); type 2 diabetes/hyperglycemia and glycine (71–73), mannose (71, 72), and ketone bodies 3-hydroxybutyrate and acetoacetate (71); and body mass index–related biochemicals asparagine, 3-(4-hydroxyphenyl)lactate, histidine, and glycine (amino acids), mannose (carbohydrate), and hexanoylcarnitine and 7-HOCA (lipids) (74, 75). Discovery of these specific risk factor-associated metabolites in relation to mortality both validates the clinical risk association and affords hypothesis-generating exploration of underlying biological mechanisms for the factor-outcome associations. Further, with regard to potential clinical value, the elucidation of possible underlying biological mechanisms of action for the risk factor–mortality association affords a more precise understanding with potential therapeutic/preventive implications. Second, the metabolomic approach may identify novel biochemicals associated with heretofore unknown risk exposures.

Limitations and strengths of the present study deserve consideration. Even though the study was not large, substantial and highly statistically significant associations were discovered. All participants were Finnish, aged 50–69 years, male, and smokers, which limits generalizability of our findings when considering other populations (e.g., women, younger individuals, and those of other ethnicities). The analysis was restricted to known compounds that were found in at least 4 of 5 study subsets, making it possible that other associations with mortality exist for excluded or unnamed metabolites that were not evaluated. Although all models adjusted for potential confounding factors such as serum creatinine, body mass index, and history of diabetes, it is possible that metabolites-mortality associations were partly mediated by subclinical diseases, such as renal insufficiency, hepatic dysfunction, or insulin resistance. The reported hazard ratios reflect the association between mortality and a single measure of each metabolite; the association with average lifetime levels are likely to be stronger given their documented within-person variability over time (10–12). Important strengths of the study were its prospective nature, with up to 28 years of follow-up, permitting examination of metabolite profiles years prior to the mortality outcomes, and validated mortality ascertainment from national registries that had little or no loss-to-follow-up.

In summary, we identified a panel of circulating metabolites and their composite risk score that were prospectively independently associated with all-cause and CVD-related mortality and substantiated by pathway analyses. The metabolomic traits were related to branched-chain amino acid metabolism, DNA repair, primary bile acid and androgen metabolism, aging, inflammation, and tobacco smoking. Additional prospective investigations in more diverse populations are warranted to reexamine these associations, which, if replicated, will require elucidation of deeper underlying biological mechanisms. Translation to potential therapeutic and preventive targets should also be pursued.

ACKNOWLEDGMENTS

Author affiliations: Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland (Jiaqi Huang, Stephanie J. Weinstein, Steven C. Moore, Andriy Derkach, Xing Hua, Linda M. Liao, Fangyi Gu, Joshua N. Sampson, Demetrius Albanes); Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, New York (Fangyi Gu); and Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, Michigan (Alison M. Mondul).

This work was supported by the Intramural Research Program of the National Cancer Institute, National Institutes of Health, and by a US Public Health Service contract from the National Cancer Institute, Department of Health and Human Services (contract HHSN261201500005C).

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

Conflict of interest: none declared.

REFERENCES

1. World Health Organization. *Global health risks*. Geneva, Switzerland: Guilford Press; 2009. http://www.who.int/healthinfo/global_burden_disease/global_health_risks/en/.
2. Fischer K, Kettunen J, Würtz P, et al. Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: an observational study of 17,345 persons. *PLoS Med*. 2014;11(2):e1001606.
3. Yu B, Heiss G, Alexander D, et al. Associations between the serum metabolome and all-cause mortality among African Americans in the Atherosclerosis Risk in Communities (ARIC) Study. *Am J Epidemiol*. 2016;183(7):650–656.
4. The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. The ATBC Cancer Prevention Study Group. *Ann Epidemiol*. 1994;4(1):1–10.
5. Mondul AM, Sampson JN, Moore SC, et al. Metabolomic profile of response to supplementation with β-carotene in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study. *Am J Clin Nutr*. 2013;98(2):488–493.
6. 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(5):1036–1041.
7. Mondul AM, Moore SC, Weinstein SJ, et al. Metabolomic analysis of prostate cancer risk in a prospective cohort: the Alpha-Tocolpherol, Beta-Carotene Cancer Prevention (ATBC) Study. *Int J Cancer*. 2015;137(9):2124–2132.

8. Evans AM, DeHaven CD, Barrett T, et al. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal Chem*. 2009; 81(16):6656–6667.
9. Evans AM, Bridgewater BR, Liu Q, et al. High resolution mass spectrometry improves data quantity and quality as compared to unit mass resolution mass spectrometry in high-throughput profiling metabolomics. *Metabolomics*. 2014;4:132.
10. 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(4):631–640.
11. 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(8):e0135437.
12. 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(6):e21103.
13. Rothman K, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
14. Johnson RC, Nelson GW, Troyer JL, et al. Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics*. 2010;11:724.
15. Chen H, Lumley T, Brody J, et al. Sequence kernel association test for survival traits. *Genet Epidemiol*. 2014;38(3):191–197.
16. Davison AC, Hinkley DV. *Bootstrap Methods and Their Application*. Cambridge, United Kingdom: Cambridge University Press; 1997.
17. Lynch CJ, Adams SH. Branched-chain amino acids in metabolic signalling and insulin resistance. *Nat Rev Endocrinol*. 2014;10(12):723–736.
18. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med*. 2011;17(4):448–453.
19. Lotta LA, Scott RA, Sharp SJ, et al. Genetic predisposition to an impaired metabolism of the branched-chain amino acids and risk of type 2 diabetes: a Mendelian randomisation analysis. *PLoS Med*. 2016;13(11):e1002179.
20. Li T, Zhang Z, Kolwicz SC Jr, et al. Defective branched-chain amino acid catabolism disrupts glucose metabolism and sensitizes the heart to ischemia-reperfusion injury. *Cell Metab*. 2017;25(2):374–385.
21. Ruiz-Canela M, Toledo E, Clish CB, et al. Plasma branched-chain amino acids and incident cardiovascular disease in the PREDIMED trial. *Clin Chem*. 2016;62(4):582–592.
22. Mayers JR, Wu C, Clish CB, et al. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nat Med*. 2014; 20(10):1193–1198.
23. Al-Majdoub M, Geidenstam N, Ali A, et al. Branched-chain amino acids are associated with odd-chain fatty acids in normoglycaemic individuals. *Diabetes Metab*. 2017;43(5):475–479.
24. Xiao Q, Moore SC, Keadle SK, et al. Objectively measured physical activity and plasma metabolomics in the Shanghai Physical Activity Study. *Int J Epidemiol*. 2016;45(5): 1433–1444.
25. Mangge H, Zelzer S, Prüller F, et al. Branched-chain amino acids are associated with cardiometabolic risk profiles found already in lean, overweight and obese young. *J Nutr Biochem*. 2016;32:123–127.
26. Aksnes H, Hole K, Arnesen T. Molecular, cellular, and physiological significance of N-terminal acetylation. *Int Rev Cell Mol Biol*. 2015;316:267–305.
27. Eberharter A, Becker PB. Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. *EMBO Rep*. 2002;3(3): 224–229.
28. Verdone L, Caserta M, Di Mauro E. Role of histone acetylation in the control of gene expression. *Biochem Cell Biol*. 2005; 83(3):344–353.
29. Windelberg A, Arseth O, Kvalheim G, et al. Automated assay for the determination of methylmalonic acid, total homocysteine, and related amino acids in human serum or plasma by means of methylchloroformate derivatization and gas chromatography–mass spectrometry. *Clin Chem*. 2005; 51(11):2103–2109.
30. Svingen GF, Schartum-Hansen H, Ueland PM, et al. Elevated plasma dimethylglycine is a risk marker of mortality in patients with coronary heart disease. *Eur J Prev Cardiol*. 2015;22(6): 743–752.
31. Svingen GF, Ueland PM, Pedersen EK, et al. Plasma dimethylglycine and risk of incident acute myocardial infarction in patients with stable angina pectoris. *Arterioscler Thromb Vasc Biol*. 2013;33(8):2041–2048.
32. Lin Y, Ma C, Liu C, et al. NMR-based fecal metabolomics fingerprinting as predictors of earlier diagnosis in patients with colorectal cancer. *Oncotarget*. 2016;7(20):29454–29464.
33. Ladep NG, Dona AC, Lewis MR, et al. Discovery and validation of urinary metabotypes for the diagnosis of hepatocellular carcinoma in West Africans. *Hepatology*. 2014; 60(4):1291–1301.
34. Nitter M, Norgård B, de Vogel S, et al. Plasma methionine, choline, betaine, and dimethylglycine in relation to colorectal cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann Oncol*. 2014;25(8): 1609–1615.
35. Bae S, Ulrich CM, Neuhouser ML, et al. Plasma choline metabolites and colorectal cancer risk in the Women's Health Initiative Observational Study. *Cancer Res*. 2014;74(24): 7442–7452.
36. de Vogel S, Ulvik A, Meyer K, et al. Sarcosine and other metabolites along the choline oxidation pathway in relation to prostate cancer—a large nested case-control study within the JANUS cohort in Norway. *Int J Cancer*. 2014;134(1): 197–206.
37. Hocher B, Adamski J. Metabolomics for clinical use and research in chronic kidney disease. *Nat Rev Nephrol*. 2017; 13(5):269–284.
38. Yonemura K, Takahira R, Yonekawa O, et al. The diagnostic value of serum concentrations of 2-(alpha-mannopyranosyl)-L-tryptophan for normal renal function. *Kidney Int*. 2004;65(4): 1395–1399.
39. Takahira R, Yonemura K, Yonekawa O, et al. Tryptophan glycoconjugate as a novel marker of renal function. *Am J Med*. 2001;110(3):192–197.
40. Lustgarten MS, Fielding RA. Metabolites associated with circulating interleukin-6 in older adults. *J Gerontol A Biol Sci Med Sci*. 2017;72(9):1277–1283.
41. Menni C, Kastenmüller G, Petersen AK, et al. Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int J Epidemiol*. 2013; 42(4):1111–1119.
42. Ames BN. Endogenous DNA damage as related to cancer and aging. *Mutat Res*. 1989;214(1):41–46.
43. Saad AA, O'Connor PJ, Mostafa MH, et al. Bladder tumor contains higher *N*7-methylguanine levels in DNA than adjacent normal bladder epithelium. *Cancer Epidemiol Biomarkers Prev*. 2006;15(4):740–743.

44. Newberne PM, Broitman S, Schrager TF. Esophageal carcinogenesis in the rat: zinc deficiency, DNA methylation and alkyltransferase activity. *Pathobiology*. 1997;65(5):253–263.
45. Jung K, Reszka R, Kamlage B, et al. Tissue metabolite profiling identifies differentiating and prognostic biomarkers for prostate carcinoma. *Int J Cancer*. 2013;133(12):2914–2924.
46. Yamori Y, Taguchi T, Hamada A, et al. Taurine in health and diseases: consistent evidence from experimental and epidemiological studies. *J Biomed Sci*. 2010;17(suppl 1):S6.
47. Kitani K. What really declines with age? The Hayflick Lecture for 2006 35th American Aging Association. *Age (Dordr)*. 2007;29(1):1–14.
48. Cao R, Cronk ZX, Zha W, et al. Bile acids regulate hepatic gluconeogenic genes and farnesoid X receptor via G(alpha)i-protein-coupled receptors and the AKT pathway. *J Lipid Res*. 2010;51(8):2234–2244.
49. Minamino T, Miyauchi H, Tateno K, et al. Akt-induced cellular senescence: implication for human disease. *Cell Cycle*. 2004;3(4):449–451.
50. Nogueira V, Park Y, Chen CC, et al. Akt determines replicative senescence and oxidative or oncogenic premature senescence and sensitizes cells to oxidative apoptosis. *Cancer Cell*. 2008; 14(6):458–470.
51. Yu Z, Zhai G, Singmann P, et al. Human serum metabolic profiles are age dependent. *Aging Cell*. 2012;11(6): 960–967.
52. Hipkiss AR. Aging, proteotoxicity, mitochondria, glycation, NAD and carnosine: possible inter-relationships and resolution of the oxygen paradox. *Front Aging Neurosci*. 2010;2:10.
53. Alhasawi A, Castonguay Z, Appanna ND, et al. Glycine metabolism and anti-oxidative defence mechanisms in Pseudomonas fluorescens. *Microbiol Res* 2015;171:26–31.
54. Howard A, Hirst BH. The glycine transporter GLYT1 in human intestine: expression and function. *Biol Pharm Bull*. 2011;34(6):784–788.
55. Aki T, Egashira N, Yamauchi Y, et al. Protective effects of amino acids against gabexate mesilate-induced cell injury in porcine aorta endothelial cells. *J Pharmacol Sci*. 2008;107(3):238–245.
56. Lu Y, Zhu X, Li J, et al. Glycine prevents pressure overload induced cardiac hypertrophy mediated by glycine receptor. *Biochem Pharmacol*. 2017;123:40–51.
57. Van den Eynden J, Ali SS, Horwood N, et al. Glycine and glycine receptor signalling in non-neuronal cells. *Front Mol Neurosci*. 2009;2:9.
58. Goetzl EJ, Dolezalova H, Kong Y, et al. Dual mechanisms for lysophospholipid induction of proliferation of human breast carcinoma cells. *Cancer Res*. 1999;59(18):4732–4737.
59. Xu Y, Xiao YJ, Baudhuin LM, et al. The role and clinical applications of bioactive lysolipids in ovarian cancer. *J Soc Gynecol Investig*. 2001;8(1):1–13.
60. Mills GB, Moolenaar WH. The emerging role of lysophosphatidic acid in cancer. *Nat Rev Cancer*. 2003;3(8): 582–591.
61. Mukai M, Imamura F, Ayaki M, et al. Inhibition of tumor invasion and metastasis by a novel lysophosphatidic acid (cyclic LPA). *Int J Cancer*. 1999;81(6):918–922.
62. Manning TJ Jr, Parker JC, Sontheimer H. Role of lysophosphatidic acid and rho in glioma cell motility. *Cell Motil Cytoskeleton*. 2000;45(3):185–199.
63. Yuan C, Morales-Oyarvide V, Babic A, et al. Cigarette smoking and pancreatic cancer survival. *J Clin Oncol*. 2017; 35(16):1822–1828.
64. Wilson KM, Markt SC, Fang F, et al. Snus use, smoking and survival among prostate cancer patients. *Int J Cancer*. 2016; 139(12):2753–2759.
65. Braithwaite D, Izano M, Moore DH, et al. Smoking and survival after breast cancer diagnosis: a prospective observational study and systematic review. *Breast Cancer Res Treat*. 2012;136(2):521–533.
66. Thun MJ, Carter BD, Feskanich D, et al. 50-year trends in smoking-related mortality in the United States. *N Engl J Med*. 2013;368(4):351–364.
67. Chao A, Thun MJ, Jacobs EJ, et al. Cigarette smoking and colorectal cancer mortality in the cancer prevention study II. *J Natl Cancer Inst*. 2000;92(23):1888–1896.
68. Parajuli R, Bjerkaas E, Tverdal A, et al. Cigarette smoking and colorectal cancer mortality among 602,242 Norwegian males and females. *Clin Epidemiol*. 2014;6:137–145.
69. Rhodes CJ, Ghataorhe P, Wharton J, et al. Plasma metabolomics implicates modified transfer RNAs and altered bioenergetics in the outcomes of pulmonary arterial hypertension. *Circulation*. 2017;135(5):460–475.
70. Gu F, Derkach A, Freedman ND, et al. Cigarette smoking behaviour and blood metabolomics. *Int J Epidemiol*. 2016; 45(5):1421–1432.
71. Guasch-Ferre M, Hruby A, Toledo E, et al. Metabolomics in prediabetes and diabetes: a systematic review and meta-analysis. *Diabetes Care*. 2016;39(5):833–846.
72. Yu D, Moore SC, Matthews CE, et al. Plasma metabolomic profiles in association with type 2 diabetes risk and prevalence in Chinese adults. *Metabolomics*. 2016;12:3.
73. Wang-Sattler R, Yu Z, Herder C, et al. Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol*. 2012;8:615.
74. Zhao H, Shen J, Djukovic D, et al. Metabolomics-identified metabolites associated with body mass index and prospective weight gain among Mexican American women. *Obes Sci Pract*. 2016;2(3):309–317.
75. Moore SC, Matthews CE, Sampson JN, et al. Human metabolic correlates of body mass index. *Metabolomics*. 2014; 10(2):259–269.