

A Prospective Investigation of Circulating Metabolome Identifies Potential Biomarkers for Gastric Cancer Risk

Xiang Shu^{1,2}, Hui Cai¹, Qing Lan³, Qiuyin Cai¹, Bu-Tian Ji³, Wei Zheng¹, and Xiao-Ou Shu¹



ABSTRACT

Background: Metabolomics is widely used to identify potential novel biomarkers for cancer risk. No investigation, however, has been conducted to prospectively evaluate the role of perturbation of metabolome in gastric cancer development.

Methods: 250 incident cases diagnosed with primary gastric cancer were selected from the Shanghai Women's Health and the Shanghai Men's Health Study, and each was individually matched to one control by incidence density sampling. An untargeted global profiling platform was used to measure approximately 1,000 metabolites in prediagnostic plasma. Conditional logistic regression was utilized to generate ORs and *P* values.

Results: Eighteen metabolites were associated with gastric cancer risk at *P* < 0.01. Among them, 11 metabolites were lysophospholipids or lipids of other classes; for example, 1-(1-enyl-palmitoyl)-GPE (P-16:0) (OR = 1.56; *P* = 1.89 × 10⁻⁴). Levels of methylmalonate,

a suggested biomarker of vitamin B12 deficiency, was correlated with increased gastric cancer risk (OR = 1.42; *P* = 0.004). Inverse associations were found for three biomarkers for coffee/tea consumption (3-hydroxypyridine sulfate, quinate and N-(2-furoyl) glycine), although the associations were only significant when comparing cases that were diagnosed within 5 years after the blood collection to matched controls. Most of the identified associations were more profound in women and never smokers than their male or ever smoking counterparts and some with notable significant interactions.

Conclusions: Our study identified multiple potential risk biomarkers for gastric cancer independent of *Helicobacter pylori* infection and other major risk factors.

Impact: New risk-assessment tools to identify high-risk population could be developed to improve prevention of gastric cancer.

See related commentary by Drew et al., p. 1601

Introduction

Gastric cancer is the third leading cause of cancer-related deaths in the world. The incidence of gastric cancer varies significantly across regions with the highest ones found in Eastern Asia and Eastern Europe (1). *Helicobacter pylori* (*H. pylori*) infection affects nearly 50% of the global population, which is a well-established risk factor for gastric cancer (2). It has been estimated that *H. pylori* infection may contribute to the development of approximately 90% of noncardia gastric cancer (2). Other major risk factors for gastric cancer include male sex, smoking, and alcohol consumption (3). Although a significant improvement in prognosis has been observed during the past decades, the clinical outcome of gastric cancer remains dismal. The 5-year survival rate is approximately 20% to 30% worldwide with a notable exception for Japan and South Korea, which is above 60% due to high rates of screening (4–7). The dismal prognosis of gastric cancer is largely attributable to the fact that a great proportion of patients are diagnosed at advanced stage and lack of effective treatments.

To mitigate the health burden associated with gastric cancer, it is of particular importance to improve the current knowledge on gastric

cancer etiology and developing tools to identify high-risk subgroups. Identifying blood biomarkers, noninvasive in nature, represents a long interest in cancer research. Serum biomarkers such as CEA and CA19–9 have been tested for clinical or screening use but failed to show sufficient effectiveness due to low sensitivity (8). A panel of human antibodies to *H. pylori* recombinantly expressed fusion proteins were developed for characterizing high-risk subgroups among general population (9–11). We previously showed that subjects seropositive to *H. pylori* had a 2- to 4-fold risk of developing incident gastric cancer compared with those who were seronegative (10), confirming the utility of these serologic biomarkers in epidemiologic research. However, since the rate of *H. pylori* infection is decreasing over the years (12), new biomarkers are needed in addition to the conventional ones to identify individuals who are at high risk but not characterized by those traditional risk factors of the disease.

Dysregulated metabolism provides a selective advantageous micro-environment to cancer and is closely associated with tumorigenesis (13). Therefore, we hypothesize that blood metabolome, which can be measured semi-/quantitatively, comprehensively and noninvasively, is an important source for discovering novel risk biomarkers for gastric cancer. Although many previous investigations have focused on human metabolome for gastric cancer, most findings were not consistent across these studies (14, 15). Notably, no study has prospectively investigated the associations between baseline metabolic profiles and gastric cancer development. Herein, we reported a metabolomics study for gastric cancer nested within two large prospective population-based cohorts to search for potential novel biomarkers for risk assessment of gastric cancer and facilitate the understanding of the disease etiology.

Materials and Methods

Study population

The nested case-control study was carried out within two population-based prospective cohorts, the Shanghai Women's Health Study

¹Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, Tennessee.

²Department of Epidemiology & Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York. ³Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, NCI, NIH, Department of Health and Human Services, Bethesda, Maryland.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Corresponding Author: Xiang Shu, Memorial Sloan Kettering Cancer Center, 485 Lexington Avenue, 2nd Floor, New York, NY 10017. E-mail: ShuX@mskcc.org

Cancer Epidemiol Biomarkers Prev 2021;30:1634–42

doi: 10.1158/1055-9965.EPI-20-1633

©2021 American Association for Cancer Research

(SWHS) and the Shanghai Men's Health Study (SMHS). Detailed methodology on study design, participant recruitment, information and biospecimen collection, and longitudinal follow-up for these two cohorts was described previously (16, 17). Briefly, 74,941 women aged 40–70 years were recruited to the SWHS from 1996 to 2000 from urban Shanghai China. From 2002 to 2006, a total of 61,480 men aged 40–74 years were recruited to the SMHS from the same region. Baseline information on sociodemographics, disease history and medication use, family history of cancer, dietary intakes, cigarette smoking, alcohol consumption and other lifestyle factors was collected by in-person interviews. Anthropometric measurements including height, weight, and circumferences of waist and hip were taken following a standard protocol by trained interviewers. A 10 mL blood sample was collected from the participants in EDTA tube during baseline survey. Approximately 75% of the participants provided blood samples at baseline. All blood samples were kept at 4°C during transportation. Within 6 hours of receipt, blood samples were processed, and plasma was aliquoted to 2 mL vials for long-term storage at -75°C. All participants are followed by a combination of annual linkage with the population-based Shanghai Tumor Registry and Vital Statistics Registry and in-person surveys taking place every 2–4 years. Incident cancer cases are verified via reviews of medical charts. Institutional review boards of all involved institutes approved the studies, and written informed consent was obtained from all participants prior to interview.

We included 250 incident gastric cancers (125 women; 125 men) in this study. Among the 250 patients, 22 were diagnosed with cardia gastric cancer. For each index case, we randomly selected one control from the risk set of cancer-free cohort members using the incidence density sampling method. Controls were matched to cases on age at recruitment (± 2 years), sex, date of sample collection (± 30 days), time of sample collection (morning or afternoon), time to last meal (± 4 hours), and menopausal status (for women). Human IgA, IgM, and IgG antibodies to 15 *H. pylori* recombinantly expressed fusion proteins (UreA, Catalase, GroEL, NapA, CagA, CagM, Cag δ , HP 0231, VacA, HpaA, Cad, HyaA, Omp, HcpC, and HP 0305) were simultaneously measured using Luminex assays as a part of previously reported study (9). Overall seropositivity for *H. pylori* was defined as four or more seropositive results to the 15 *H. pylori* antigens assessed. In this study, *H. pylori* seropositivity status refer to dual positivity based on data of Omp and HP 0305 because the variable was more strongly associated with gastric cancer risk in the study population than overall seropositivity (Table 1).

Global metabolic profiling, metabolite identification, and quantification

Global metabolic profiling was conducted using a LC-MS global metabolomics platform at Metabolon, Inc following a standard protocol. Case and control samples from the same pair were measured next to each other in the same batch. Briefly, samples were prepared using the automated MicroLab STAR system from Hamilton Company. Recovery standards were added prior to the first step in the extraction process for QC purposes. Proteins were precipitated with methanol under vigorous shaking for 2 minutes followed by centrifugation. The sample extracts were stored overnight under nitrogen before preparation for analysis. All metabolite profiling methods utilized a Waters ACQUITY ultra-performance liquid chromatography and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization source and Orbitrap mass analyzer operated at 35,000 mass resolution. Raw data was extracted, peak-identified and QC processed using Metabolon's hardware and software. Biochemical identification was completed

by comparing retention time/index, mass, and the MS/MS forward and reverse scores between experimental data with authentic standards in the in-house library. Peaks were quantified by area-under-the-curve.

Statistical analysis

Metabolite levels below detection limit were imputed with half of the minimum. Natural logarithm transformation was taken to address distribution skewness for metabolites. The transformed values were converted to one SD change to facilitate result interpretation. Scores plot was shown by plotting quality control samples and cohort samples by top two principal components (Supplementary Fig. S1). Only metabolites that passed the quality control described as follows were included in this analysis (before exclusion, $N_{\text{metabolite}} = 1,061$: 1) levels were below the limit of detection in > 50% of samples ($N_{\text{exclusion}} = 159$); 2) coefficient variation < 25% in quality control samples ($N_{\text{exclusion}} = 184$). For this study, we focused on metabolites with known identity. A total of 581 metabolites were included in the downstream analyses. Univariate analyses were performed for each covariate using paired *t* test (continuous) or conditional logistic regression model (non-continuous). For multivariable analyses, time to last meal, glucose levels (measured metabolite), BMI, smoking status, income, *H. pylori* seropositivity status (dual positivity to Omp and HP 0305), gastritis, and family history of gastric cancer were adjusted in the conditional logistic regression models. Metabolite levels were treated as continuous variable in the models. To search for independent associations, backward selection based on Akaike information criterion was conducted for top associated metabolites ($P < 0.01$). Stratified analyses by sex (female/male), time between blood collection and cancer diagnosis (<5 years/ ≥ 5 years), smoking (never/ever), and *H. pylori* seropositivity status were performed to explore group-specific associations. Interaction was examined by likelihood ratio test and heterogeneity was assessed by Cochran Q test. Heatmap was generated based on pairwise Pearson correlation between the identified metabolites. Spearman correlation and linear regression with adjustment of age and sex were both conducted to evaluate the relationship between metabolite levels and recorded time to last meal (hours)/tea consumption (divided by 50 g/month). The time to last meal and tea consumption were treated as continuous variables in these analyses. Additional analysis was performed to show the impact of time to last meal and glucose levels on the identified associations by excluding these two covariates from multivariable regression models. The main analysis was repeated by restricting to case-control pairs with a comparable time to last meal (± 2 hours). Finally, receiver operating characteristic curve (ROC) and area under the curve (AUC) were calculated and compared between two models, that is, demographics and traditional risk factors with/without the top independently associated metabolites (after backward selection). Bootstrapping technique with 2,000 repeats was applied to determine the 95% CIs. *P* value was calculated using DeLong nonparametric approach. All statistical analyses were conducted using R (version 4.0.2) and Stata 14 (StataCorp).

Results

In this study, demographics and most lifestyle factors were evenly balanced between case-control pairs except that male patients tended to have a higher family income compared with their matched controls (Table 1). The majority of ever smokers and drinkers were men, which was expected as only 2.4% and 1.9% of SWHS participants reported a

Table 1. Demographic characteristics for the participants in the present study.

	All			SWHS			SMHS		
	Cases N = 250	Control N = 250	P	Cases N = 125	Control N = 125	P	Cases N = 125	Control N = 125	P
Age, years	59.5 ± 9.8	59.5 ± 9.7	0.381	57.8 ± 8.9	57.8 ± 8.9	0.823	61.1 ± 10.4	61.3 ± 10.2	0.168
Ever smoking, n (%)									
No	152 (60.8)	168 (67.2)		117 (93.6)	120 (96.0)		35 (28.0)	48 (38.4)	
Yes	98 (39.2)	82 (32.8)	0.045	8 (6.4)	5 (4.0)	0.410	90 (72.0)	77 (61.6)	0.067
Pack-year smoking	11.0 ± 17.7	10.3 ± 19.8	0.647	0.7 ± 3.8	0.4 ± 3.2	0.525	21.2 ± 20.0	20.2 ± 24.1	0.720
Ever drinking, n (%)									
No	196 (78.4)	199 (79.6)		122 (97.6)	122 (97.6)		74 (59.2)	77 (61.6)	
Yes	54 (21.6)	51 (20.4)	0.714	3 (2.4)	3 (2.4)	1	51 (40.1)	48 (38.4)	0.701
Education level, n (%)									
Elementary school or none	72 (28.8)	70 (28.0)		57 (45.6)	50 (40.0)		15 (12.0)	20 (16.0)	
Before college	166 (66.4)	162 (64.8)		66 (52.8)	69 (55.2)		100 (80.0)	93 (74.4)	
College or above	12 (4.8)	18 (7.2)	0.489	2 (1.6)	6 (4.8)	0.233	10 (8.0)	12 (9.6)	0.517
Family income, n (%)									
Low	33 (13.2)	51 (20.4)		24 (19.2)	31 (24.8)		9 (7.2)	20 (16.0)	
Middle	200 (80.0)	185 (74.0)		95 (76.0)	88 (70.4)		105 (84.0)	97 (77.6)	
High	17 (6.8)	14 (5.6)	0.067	6 (4.8)	6 (4.8)	0.545	11 (8.8)	8 (6.4)	0.044
BMI, n (%)									
<18.5 kg/m ²	11 (4.4)	7 (2.8)		4 (3.2)	1 (0.8)		7 (5.6)	6 (4.8)	
18.5–25 kg/m ²	147 (58.8)	149 (59.6)		69 (55.2)	71 (56.8)		78 (62.4)	78 (62.4)	
25–29.9 kg/m ²	77 (30.8)	77 (30.8)		41 (32.8)	41 (32.8)		36 (28.8)	36 (28.8)	
≥30 kg/m ²	15 (6.0)	17 (6.8)	0.774	11 (8.8)	12 (9.6)	0.585	4 (3.2)	5 (4.0)	0.977
WHR	0.86 ± 0.07	0.86 ± 0.07	0.282	0.82 ± 0.05	0.82 ± 0.05	0.767	0.90 ± 0.06	0.90 ± 0.06	0.243
Physical activity, MET-hour/day	1.1 ± 1.9	1.2 ± 2.2	0.427	0.8 ± 1.4	1.0 ± 1.8	0.152	1.4 ± 2.1	1.4 ± 2.5	0.968
Time to last meal, hours	4.3 ± 3.6	5.8 ± 4.8	6.36 × 10 ⁻⁹	4.1 ± 3.7	4.0 ± 3.7	0.722	4.5 ± 3.6	7.5 ± 5.2	3.62 × 10 ⁻¹⁰
Current tea drinking, n (%)									
No	144 (57.6)	144 (57.6)		99 (79.2)	91 (72.8)		45 (36.0)	53 (42.4)	
Yes	106 (42.4)	106 (42.4)	1.000	26 (20.8)	34 (27.2)	0.230	80 (64.0)	72 (57.6)	0.278
Amount of dry tea leaf consumed, g/month	108.8 ± 184.6	99.2 ± 149.6	0.436	24.2 ± 65.6	34.4 ± 69.8	0.253	193.4 ± 223.0	164 ± 177.9	0.200
History of type II diabetes, n (%)									
No	228 (91.2)	235 (94.0)		114 (91.2)	116 (92.8)		114 (91.2)	119 (95.2)	
Yes	22 (8.8)	15 (6.0)	0.213	11 (8.8)	9 (7.2)	0.618	11 (8.8)	6 (4.8)	0.206
Family history of gastric cancer, n (%)									
No	229 (91.6)	228 (91.2)		115 (92.0)	112 (89.6)		114 (91.2)	116 (92.8)	
Yes	21 (8.4)	22 (8.8)	0.879	10 (8.0)	13 (10.4)	0.533	11 (8.8)	9 (7.2)	0.655
History of gastritis, n (%)									
No	193 (77.2)	196 (78.4)		91 (72.8)	91 (72.8)		102 (81.6)	105 (84.0)	
Yes	57 (22.8)	54 (21.6)	0.753	34 (27.2)	34 (27.2)	1	23 (18.4)	20 (16.0)	0.602
Overall <i>H. pylori</i> seropositivity, n (%)									
No	9 (3.6)	32 (13.1)		5 (4.1)	19 (16.0)		4 (3.2)	13 (10.4)	
Yes	238 (96.4)	212 (86.9)	3.35 × 10 ⁻⁵	117 (95.9)	100 (84.0)	5.60 × 10 ⁻⁴	121 (96.8)	112 (89.6)	0.032
Missing	3 (1.2)	6 (2.4)		3 (2.4)	6 (4.8)		0 (0)	0 (0)	
<i>H. pylori</i> biomarker-dual positivity, n (%)									
Omp- and/or HP 0305-	80 (32.0)	129 (51.6)		46 (36.8)	66 (52.8)		34 (27.2)	63 (50.4)	
Omp+ & HP 0305+	167 (66.8)	115 (46.0)	4.67 × 10 ⁻⁶	76 (60.8)	53 (42.4)	2.98 × 10 ⁻³	91 (72.8)	62 (49.6)	5.35 × 10 ⁻⁴
Missing	3 (1.2)	6 (2.4)		3 (2.4)	6 (4.8)		0 (0)	0 (0)	

Note: Mean ± SD was shown for continuous variable. Family income (yuan per capita per year), for women, low: <4,000; middle: 4,000–8,000; high: ≥8,000; for men, low: <6,000; middle: 6,000–10,000; high: ≥10,000 at baseline.

Table 2. Top metabolites associated with gastric cancer risk, multivariable conditional logistic regression.

Metabolites	SUPER. PATHWAY	SUB.PATHWAY	OR (95% CI) ^a	P ^b
Orotate	Nucleotide	Pyrimidine Metabolism, Orotate containing	1.71 (1.30–2.26)	1.25 × 10 ⁻⁴
1-(1-enyl-palmitoyl)-GPE (P-16:0)	Lipid	Lysoplasmalogen	1.56 (1.24–1.97)	1.89 × 10 ⁻⁴
sphinganine	Lipid	Sphingolipid Synthesis	1.55 (1.23–1.95)	1.92 × 10 ⁻⁴
1-(1-enyl-stearoyl)-GPE (P-18:0)	Lipid	Lysoplasmalogen	1.55 (1.22–1.98)	3.85 × 10 ⁻⁴
glycerophosphoethanolamine	Lipid	Phospholipid Metabolism	1.55 (1.21–2.00)	5.95 × 10 ⁻⁴
5α-pregnan-3β,20β-diol monosulfate	Lipid	Progesterin Steroids	1.45 (1.15–1.82)	0.00157
N6-methyllysine	Amino Acid	Lysine Metabolism	1.40 (1.13–1.74)	0.0023
3-hydroxypyridine sulfate	Xenobiotics	Chemical	0.68 (0.53–0.87)	0.00236
1-(1-enyl-oleoyl)-2-docosahexaenoyl-GPE (P-18:1/22:6)	Lipid	Plasmalogen	1.42 (1.13–1.79)	0.00254
1-palmitoyl-GPC (16:0)	Lipid	Lysophospholipid	1.44 (1.13–1.82)	0.00272
1-(1-enyl-palmitoyl)-GPC (P-16:0)	Lipid	Lysoplasmalogen	1.40 (1.11–1.75)	0.00378
quinate	Xenobiotics	Food Component/Plant	0.70 (0.55–0.89)	0.00391
1-stearoyl-GPC (18:0)	Lipid	Lysophospholipid	1.42 (1.12–1.79)	0.00408
methylmalonate (MMA)	Lipid	Fatty Acid Metabolism (also BCAA Metabolism)	1.42 (1.12–1.80)	0.00419
N-(2-furoyl) glycine	Xenobiotics	Food Component/Plant	0.70 (0.54–0.90)	0.0047
5α-pregnan-diol disulfate	Lipid	Progesterin Steroids	1.39 (1.10–1.75)	0.00638
glutamate	Amino Acid	Glutamate Metabolism	1.38 (1.09–1.74)	0.00793
cysteine	Amino Acid	Methionine, Cysteine, SAM and Taurine Metabolism	0.73 (0.58–0.93)	0.00893

^aAdjusted for time to last meal, measured glucose levels, BMI, ever smoking, income, HP infection status (based on Omp and HP 0305), gastritis, and family history of gastric cancer, in addition to matching variables [age (±2 years), sex, date of sample collection (±30 days), time of sample collection (morning or afternoon), recent antibiotic use (yes, no), and menopausal status (for SWHS)].

^bThe top five associations had a FDR < 0.1.

habit/history of tobacco and alcohol use (16), respectively. More than half of the individuals had a seropositivity to *H. pylori* infection (dual positivity to Omp and HP 0305). Time to last meal was comparable between female cases and controls at an average of 4 hours, while male controls had a longer fasting time than cases at the blood draw at baseline.

Eighteen metabolites were associated with gastric cancer risk at the $P < 0.01$. Among them, the top five were significant at the Benjamini–Hochberg False Discovery Rate (FDR) < 0.1 (Table 2). Eleven risk-associated metabolites fell broadly into lipid class, especially to sub-classes of lysophospholipid/plasmalogen. For example, 1-(1-enyl-palmitoyl)-GPE (P-16:0) was positively associated with gastric cancer risk in the study population [OR = 1.56; 95% confidence interval (CI) = 1.24–1.97; $P = 1.89 \times 10^{-4}$]. Correlations between these lipids varied substantially (Supplementary Fig. S2). We also found that levels of methylmalonate, a suggested biomarker of vitamin B12 deficiency, was correlated with an increased gastric cancer risk

(OR = 1.42; 95% CI, 1.12–1.80; $P = 0.004$). Two metabolites belonging to sub-pathway of progesterin steroids, 5α-pregnan-3β,20β-diol monosulfate and 5α-pregnan-diol disulfate, were positively associated with gastric cancer risk (OR = 1.45 and 1.39, $P = 0.002$ and 0.006, respectively). Three metabolites categorized as xenobiotics were 3-hydroxypyridine sulfate, quinate, N-(2-furoyl) glycine, all highly correlated (Supplementary Fig. S2) and inversely associated with gastric cancer risk. The ORs for the three xenobiotics ranged between 0.66 and 0.70. Ten metabolites remained in the model after the backward selection procedure, of which eight were associated with gastric cancer at $P < 0.05$ (Table 3).

We evaluated whether time to last meal had any impactful effect on the associations for the identified metabolites. As expected, glucose levels were negatively associated with time to last meal (Spearman $\rho = -0.28$, $P = 2.47 \times 10^{-10}$). Inverse associations were also found for the identified xenobiotic metabolites, especially for N-(2-furoyl) glycine (Supplementary Table S1). In the sensitivity analyses, removal of

Table 3. Associations of metabolites and gastric cancer risk: backward selection and mutual adjustment.

Metabolites	OR (95% CI)	P
Methylmalonate (MMA)	1.31 (0.91–1.87)	0.0017
N-(2-furoyl) glycine	0.75 (0.55–1.03)	0.0022
Orotate	1.34 (1.03–1.75)	0.0043
Cysteine	1.48 (1.05–2.09)	0.0072
1-(1-enyl-oleoyl)-2-docosahexaenoyl-GPE (P-18:1/22:6)	1.40 (1.05–1.86)	0.0171
5α-pregnan-diol disulfate	1.43 (1.07–1.91)	0.0221
1-stearoyl-GPC (18:0)	0.66 (0.49–0.90)	0.0253
N6-methyllysine	1.62 (1.16–2.25)	0.0301
Quinate	0.58 (0.41–0.82)	0.0794
Glycerophosphoethanolamine	1.73 (1.23–2.42)	0.1456

Note: The covariates listed in Table 2 were included in the regression analysis.

Table 4. Stratified analysis by sex for top associated metabolites.

Metabolites	Female, 125 pairs		Male, 125 pairs		<i>P</i> _{interaction}
	OR (95% CI)	<i>P</i>	OR (95% CI)	<i>P</i>	
Orotate	1.97 (1.31-2.97)	0.00124	1.69 (1.13-2.52)	0.01091	0.34856
1-(1-enyl-palmitoyl)-GPE (P-16:0)	1.98 (1.42-2.75)	5.35 × 10 ⁻⁵	1.05 (0.69-1.58)	0.83091	0.01519
Sphinganine	2.73 (1.82-4.08)	1.11 × 10 ⁻⁶	0.79 (0.52-1.19)	0.25627	6.24 × 10 ⁻⁶
1-(1-enyl-stearoyl)-GPE (P-18:0)	2.07 (1.46-2.93)	4.52 × 10 ⁻⁵	0.98 (0.65-1.50)	0.93559	0.00796
Glycerophosphoethanolamine	2.04 (1.43-2.91)	8.93 × 10 ⁻⁵	1.00 (0.65-1.55)	0.98755	0.02216
5α-pregnan-3β,20β-diol monosulfate	1.66 (1.22-2.26)	0.00127	1.30 (0.84-2.02)	0.2349	0.49745
N6-methyllysine	1.50 (1.12-2.02)	0.00645	1.17 (0.82-1.68)	0.38973	0.37326
3-hydroxypyridine sulfate	0.74 (0.53-1.02)	0.06896	0.64 (0.42-0.99)	0.04677	0.7874
1-(1-enyl-oleoyl)-2-docosahexaenoyl-GPE (P-18:1/22:6)	1.63 (1.09-2.43)	0.01655	1.44 (1.04-1.98)	0.02592	0.52726
1-palmitoyl-GPC (16:0)	2.32 (1.57-3.42)	2.25 × 10 ⁻⁵	0.77 (0.51-1.15)	0.20017	1.41 × 10 ⁻⁴
1-(1-enyl-palmitoyl)-GPC (P-16:0)	2.02 (1.41-2.88)	1.18 × 10 ⁻⁴	0.94 (0.66-1.32)	0.70482	0.00647
Quinate	0.63 (0.46-0.88)	0.00578	0.85 (0.57-1.28)	0.44122	0.20962
1-stearoyl-GPC (18:0)	1.99 (1.39-2.85)	1.57 × 10 ⁻⁴	0.86 (0.57-1.30)	0.46564	0.00556
Methylmalonate (MMA)	1.32 (0.95-1.84)	0.10169	1.54 (1.06-2.24)	0.02225	0.67982
N-(2-furoyl) glycine	0.71 (0.52-0.98)	0.0342	0.73 (0.46-1.14)	0.16366	0.76516
5α-pregnan-diol disulfate	1.62 (1.14-2.28)	0.00655	1.26 (0.85-1.87)	0.24925	0.55568
Glutamate	1.87 (1.34-2.62)	2.64 × 10 ⁻⁴	0.71 (0.45-1.13)	0.14598	0.0055
Cysteine	0.74 (0.54-1.02)	0.0663	0.65 (0.45-0.96)	0.03059	0.9852

Note: The covariates listed in **Table 2** were included in the regression analysis.

glucose levels and time to last meal from the adjusted covariates had a minimal or moderate impact on the significant associations mentioned above except for N-(2-furoyl) glycine (Supplementary Table S2). Restricting analysis to case-control pairs whose time to last meal were well-matched (± 2 hours) did not change the association estimates appreciably (Supplementary Table S3). The three xenobiotic metabolites were positively correlated with self-reported amount of dry tea leaves consumed each month in the overall population and the correlations were generally consistent in direction by sex (Supplementary Table S4).

Stratified and interaction analysis identified several significant interactions across subgroups ($P_{\text{interaction}} < 0.01$); for those metabolites, the above reported overall significant associations were

primarily driven by findings in women, including sphinganine, 1-(1-enyl-stearoyl)-GPE (P-18:0), 1-palmitoyl-GPC (16:0), 1-(1-enyl-palmitoyl)-GPC (P-16:0), 1-stearoyl-GPC (18:0), and glutamate (**Table 4**). When stratified by time to cancer diagnosis (< 5 years vs. ≥ 5 years), most of the associations were consistent in the two strata except the xenobiotic metabolites. The associations for these xenobiotic metabolites diminished when restricting to patients who were diagnosed beyond 5 years of follow-up (Supplementary Table S5). Similarly, the identified associations were more profound in never smokers (Supplementary Table S5). Although sample size was small, the associations in male never smokers and overall never smokers followed the same direction for 15 of the 18 top associated metabolites (Supplementary Table S5). The associations were largely consistent

Table 5. Stratified analysis by *H. pylori* biomarker status for top associated metabolites.

Metabolites	Omp- and/or HP 0305-, <i>N</i> = 209		Omp+ and HP 0305+, <i>N</i> = 282		<i>P</i> _{interaction}
	OR (95% CI)	<i>P</i>	OR (95% CI)	<i>P</i>	
Orotate	1.97 (1.27-3.05)	0.00237	1.60 (1.15-2.22)	0.00518	0.42924
1-(1-enyl-palmitoyl)-GPE (P-16:0)	1.53 (1.11-2.11)	0.00921	1.65 (1.20-2.27)	0.00196	0.72353
Sphinganine	1.51 (1.10-2.05)	0.00998	1.61 (1.15-2.25)	0.00544	0.77056
1-(1-enyl-stearoyl)-GPE (P-18:0)	1.60 (1.13-2.25)	0.00727	1.56 (1.13-2.15)	0.00644	0.91327
Glycerophosphoethanolamine	1.58 (1.14-2.20)	0.0067	1.52 (1.09-2.12)	0.01285	0.86561
5α-pregnan-3β,20β-diol monosulfate	1.43 (1.05-1.95)	0.0225	1.44 (1.07-1.93)	0.01579	0.97956
N6-methyllysine	1.37 (0.99-1.91)	0.06112	1.39 (1.04-1.87)	0.02791	0.94965
3-hydroxypyridine sulfate	0.68 (0.48-0.97)	0.03273	0.66 (0.49-0.89)	0.00648	0.91352
1-(1-enyl-oleoyl)-2-docosahexaenoyl-GPE (P-18:1/22:6)	1.65 (1.11-2.45)	0.0132	1.38 (1.06-1.79)	0.01634	0.41561
1-palmitoyl-GPC (16:0)	1.82 (1.30-2.55)	4.61 × 10 ⁻⁴	1.12 (0.81-1.56)	0.48253	0.03265
1-(1-enyl-palmitoyl)-GPC (P-16:0)	1.70 (1.19-2.43)	0.00346	1.21 (0.91-1.61)	0.18623	0.11854
quinate	0.76 (0.55-1.04)	0.08682	0.66 (0.49-0.89)	0.0068	0.47547
1-stearoyl-GPC (18:0)	1.98 (1.36-2.87)	3.25 × 10 ⁻⁴	1.11 (0.81-1.51)	0.52083	0.01361
Methylmalonate (MMA)	1.53 (1.11-2.12)	0.00926	1.33 (0.95-1.86)	0.09316	0.54087
N-(2-furoyl) glycine	0.75 (0.52-1.07)	0.112	0.67 (0.49-0.92)	0.01453	0.64005
5α-pregnan-diol disulfate	1.52 (1.10-2.11)	0.01103	1.30 (0.98-1.72)	0.06981	0.41333
Glutamate	1.31 (0.96-1.80)	0.08963	1.50 (1.08-2.10)	0.01665	0.54464
Cysteine	0.77 (0.55-1.06)	0.10464	0.71 (0.52-0.95)	0.02192	0.69915

Note: The covariates listed in **Table 2** were included in the regression analysis.

across *H. pylori* blood biomarker status (Table 5). Two metabolites, 1-palmitoyl-GPC (16:0) and 1-stearoyl-GPC (18:0), showed a suggestive interaction with *H. pylori* seropositivity ($P_{\text{interaction}} < 0.05$). Most of the associations were consistently associated with risk of the

malignancy regardless of clinical stage at presentation (grouped by stage I–II and III–IV). The association of 5 α -pregnan-diol disulfate might be stronger in patients with an early-stage gastric cancer, although the interaction was not statistically significant

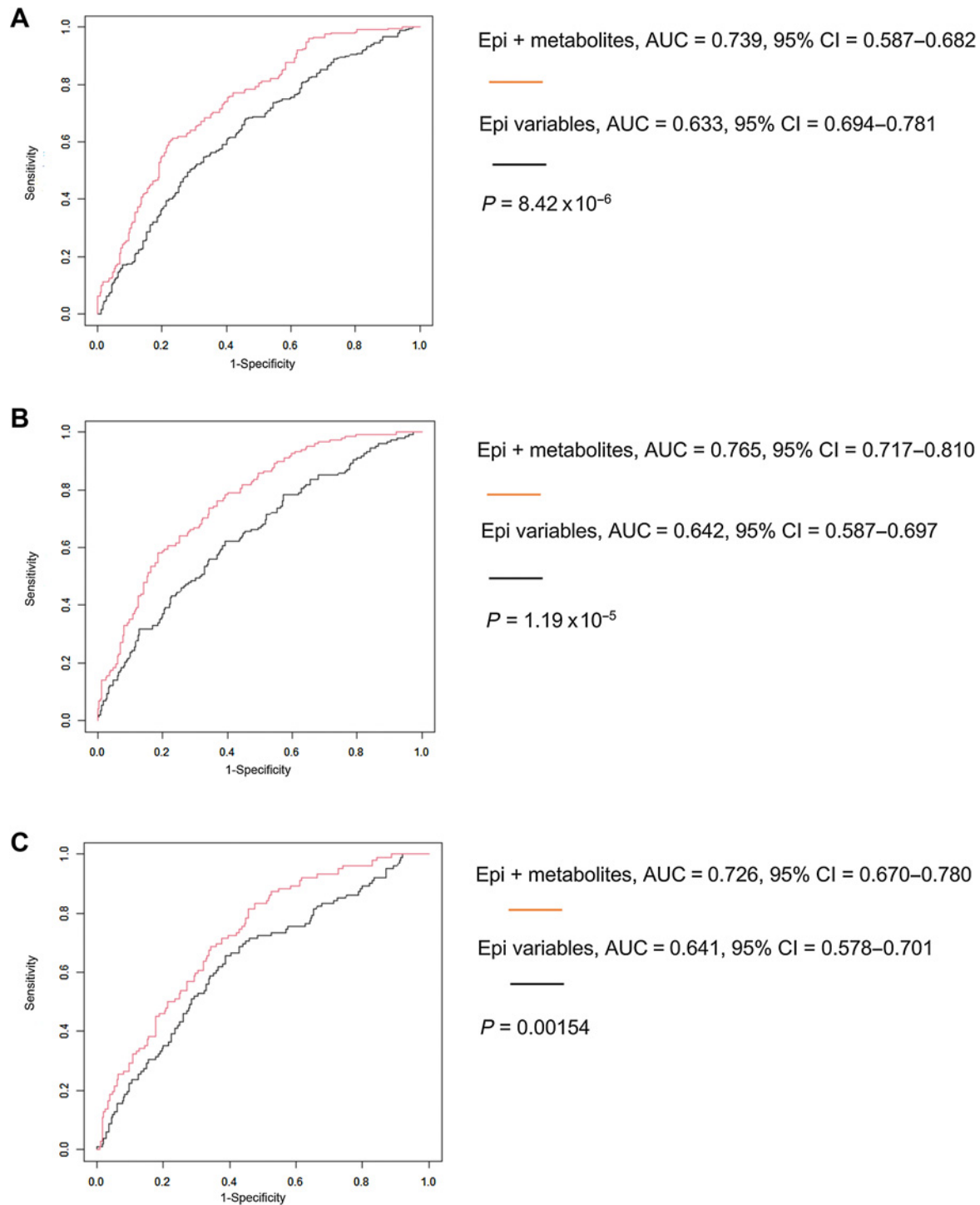


Figure 1.

ROC plot: epidemiologic variables versus epidemiologic variables + metabolites after backward selection. **A**, Full data. **B**, Only cases diagnosed within 5 years of enrollment. **C**, Only cases diagnosed after 5 years of enrollment. Epidemiologic variables included in the base model are BMI, smoking status, income, HP infection status, gastritis, and family history of gastric cancer. P value was calculated using DeLong nonparametric approach for the point estimate, bootstrapping technique with 2,000 repeats was applied to determine the 95% CI.

(Supplementary Table S5). Finally, the prediction accuracy was significantly improved when adding independently associated metabolites to the model that was built upon traditional demographics and lifestyle factors, regardless of length of time to cancer diagnosis (Fig. 1, 13%–19% increase of AUC).

Discussion

In this study, we identified 18 metabolites as potential novel biomarkers for gastric cancer risk, of which the top five were significant at FDR < 0.1. We further showed that 8 of the 18 metabolites were independently associated with the risk after backward selection and mutual adjustment. Among the top associations, lipids or lipid-like molecules were heavily represented, followed by xenobiotics and amino acids (or amino acid derivatives). These associations were primarily seen among women and never smokers. The identified associations vary little by time to cancer diagnosis (except for the three xenobiotic metabolites) and *H. pylori* seropositivity status. Furthermore, the identified associations were independent of traditional risk factors of gastric cancer such as *H. pylori* infection and smoking, etc.

To our knowledge, this is the first prospective metabolomics investigation of gastric cancer risk conducted using pre-diagnostic samples. In this study, we identified independent associations for several lipids with distinct functions in physiologic and pathologic processes. Levels of identified lysophospholipids and glycerophospholipids were found higher in gastric cancer cases. Glycerophospholipids serve as structural constituents of biological membranes and maintain the integrity. One major source of lysophospholipids is from remodeling of glycerophospholipids (18). Lysophospholipids especially lysophosphatidylcholines could modulate inflammatory chemokine expression from endothelial cells, increase oxidative stress, and are considered as a group of proinflammatory lipids (19). Our study therefore supports a promoting role of lysophospholipids played in gastric cancer development.

Methylmalonate (MMA) is a malonic acid derivative. It has been suggested that increased levels of MMA could be a sensitive indicator of vitamin B12 deficiency (20), although multiple other health-related conditions could also lead to elevated levels of MMA (21, 22). Another potential biomarker of vitamin B12 deficiency, homocysteine (20), was not measured in the current study. Vitamin B12 is rich in animal-sourced food therefore vitamin B12 deficiency usually occurs in patients with cancer who have a poor appetite. B-vitamins including B12 are essential to the one-carbon metabolism pathway, having critical functions in DNA synthesis and repair and supplying methyl groups needed for methylation of DNA, RNA, and protein. In this study, we did not find that the association of MMA with gastric cancer risk were statistically differed by time to cancer diagnosis. In addition, *H. pylori* infection and history of gastritis were taken into account in the analysis and no statistical differences were found for other chronic conditions between the cases and controls. Together, these results suggest that vitamin B12 deficiency, indicated by elevated levels of MMA, may play an important role in gastric cancer development. Our findings are in line with previous report that low vitamin B12 increases risk of gastric cancer, which suggest that gastric atrophy is the underlying condition linking vitamin B12 deficiency to the malignancy (23).

Three xenobiotic metabolites, 3-hydroxypyridine sulfate, quinate and N-(2-furoyl) glycine, were inversely associated with gastric cancer risk. In a recent study, quinate and 3-hydroxypyridine sulfate were shown to have a positive association with coffee consumption (24). To

corroborate with this finding, we also found that both metabolites were positively correlated with amount of dry tea leaf consumption in the our study population, although data for coffee intake was lacking because coffee consumption was relatively lower in China, particularly in our study population and during our sample collection period, than other Asia countries (25, 26). N-(2-furoyl) glycine was also shown as a candidate biomarker for coffee consumption in a dietary intervention study (27). In this study, we only found nominal association between tea intake and N-(2-furoyl) glycine levels in blood. Coffee and tea have both been linked to reduced risks of cardiovascular diseases and overall and cause-specific mortality (28–31). Earlier work from our group has also shown that tea drinking was associated reduced risk of digestive system cancers especially the risk of colorectal and gastric cancers (32). The associations observed for the tea-associated metabolites corroborated with our earlier findings of inverse relationship between self-reported tea drinking and gastric cancer risk. However, the associations observed for the three tea-associated metabolites were limited to the patients who were diagnosed with gastric cancer within a relatively short time after blood collection (<5 years). Because stomach upset is one of main reasons for quitting tea drinking, a concern is raised that the observed association might be caused by individual's change of tea drinking habit related to symptoms of cancer development.

Both 5 α -pregnan-3 β ,20beta-diol monosulfate and 5 α -pregnan-diol disulfate are involved in the pathway of progestin steroids, of which the levels were higher in women than men in this study. For example, the geometric mean and SD of transformed 5 α -pregnan-3 β ,20 β -diol monosulfate levels in women and men was 1.03 \pm 1.74 and 0.76 \pm 1.18, respectively. Not much is known about the two metabolites and their functions in biological processes. Nevertheless, a small study nested within Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study reported that levels of 5 α -pregnan-3 β ,20 β -diol monosulfate was correlated with reduced risk of high-grade glioma (33). Finally, limited evidence linking orotate and N6-methyllysine to cancer risk was found in literature, while cysteine as an essential component of the one-carbon metabolism pathway has been extensively investigated for its role in DNA methylation and carcinogenesis. Our results are in line with previous findings showing that cysteine levels were inversely associated with risks of incident gastric cancer and esophageal cancers in a Chinese population (34).

Gastric cancer is nearly two times more prevalent in men than women (3). This sex difference is possibly resulted from higher rates of *H. pylori* infection, smoking, drinking among men, which are known risk factors for this malignancy (3). Our findings indicated that most of the metabolites identified were stronger in women and never smokers. We speculated that sex and smoking could be important effect modifiers, although the exact mechanisms for the observed sex and smoking difference are unclear. Noteworthy, association direction and effect size were comparable between male and female never smokers, as well as between *H. pylori* seronegativity and seropositivity status. If our findings are validated, these biomarkers could be used to identify high-risk subgroups among individuals who do not have traditional risk factors for gastric cancer, that is, females, never smokers, or those free of *H. pylori* infection.

We noted that prediction accuracy could be significantly improved when adding the identified metabolites to the base prediction model of gastric cancer risk with only demographics and traditional risk factors included. Importantly, such improvement was not considerably varied

by length of time to cancer diagnosis after enrollment in the cohorts. However, because these estimates were not validated in additional independent population, model overfitting might be a concern. Therefore, this finding should be interpreted with cautions.

Despite the strengths mentioned earlier, we acknowledge several limitations for this study. The findings may only be generalizable to East Asians. Whether the identified associations were ancestry-specific or shared across populations with different racial background will need to be evaluated using resources of national/international consortia. Another limitation is that fasting time (time to last meal) was not matched well in men between cases and controls. However, we conducted sensitivity analyses to show that our findings were robust and not affected by time of fasting. Furthermore, information regarding specific histologic subtypes (i.e., intestinal vs. diffuse types) were not available for the current study. Finally, we cannot rule out the possibility that some of the findings might be due to chance. Further studies are warranted to validate our findings, especially for those remained significant after correction for multiple comparisons.

In conclusion, we conducted the first prospective metabolomics investigation to date to search for potential risk biomarkers for gastric cancer. The study generated substantial new data for a better understanding of gastric cancer etiology and provided potential novel biomarkers for risk assessment.

References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68: 394–424.
- Lee YC, Chiang TH, Chou CK, Tu YK, Liao WC, Wu MS, et al. Association between helicobacter pylori eradication and gastric cancer incidence: a systematic review and meta-analysis. *Gastroenterology* 2016;150:1113–24 e5.
- Rawla P, Barsouk A. Epidemiology of gastric cancer: global trends, risk factors and prevention. *Prz Gastroenterol* 2019;14:26–38.
- SEER 18 2010–2016, all races, both sexes by SEER summary stage 2000. <https://seer.cancer.gov/statfacts/html/stomach.html>.
- Jim MA, Pinheiro PS, Carreira H, Espey DK, Wiggins CL, Weir HK. Stomach cancer survival in the United States by race and stage (2001–2009): findings from the CONCORD-2 study. *Cancer* 2017;123Suppl 24:4994–5013.
- Collaborators GBDSC. The global, regional, and national burden of stomach cancer in 195 countries, 1990–2017: a systematic analysis for the global burden of disease study 2017. *Lancet Gastroenterol Hepatol* 2020;5:42–54.
- Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. *Cancer Epidemiol Biomarkers Prev* 2014;23:700–13.
- He CZ, Zhang KH, Li Q, Liu XH, Hong Y, Lv NH. Combined use of AFP, CEA, CA125 and CA19-9 improves the sensitivity for the diagnosis of gastric cancer. *BMC Gastroenterol* 2013;13:87.
- Michel A, Waterboer T, Kist M, Pawlita M. Helicobacter pylori multiplex serology. *Helicobacter* 2009;14:525–35.
- Epplein M, Zheng W, Xiang YB, Peek RM Jr, Li H, Correa P, et al. Prospective study of Helicobacter pylori biomarkers for gastric cancer risk among Chinese men. *Cancer Epidemiol Biomarkers Prev* 2012;21:2185–92.
- Song H, Michel A, Nyren O, Ekstrom AM, Pawlita M, Ye W. A CagA-independent cluster of antigens related to the risk of noncardia gastric cancer: associations between Helicobacter pylori antibodies and gastric adenocarcinoma explored by multiplex serology. *Int J Cancer* 2014;134: 2942–50.
- Zamani M, Ebrahimitabar F, Zamani V, Miller WH, Alizadeh-Navaei R, Shokri-Shirvani J, et al. Systematic review with meta-analysis: the worldwide prevalence of Helicobacter pylori infection. *Aliment Pharmacol Ther* 2018; 47:868–76.
- DeBerardinis RJ, Chandel NS. Fundamentals of cancer metabolism. *Sci Adv* 2016;2:e1600200.

Authors' Disclosures

No disclosures were reported.

Authors' Contributions

X. Shu: Formal analysis, investigation, writing—original draft, writing—review and editing. **H. Cai:** Data curation, formal analysis, writing—review and editing. **Q. Lan:** Writing—review and editing. **Q. Cai:** Data curation, writing—review and editing. **B.-T. Ji:** Writing—review and editing. **W. Zheng:** Conceptualization, resources, data curation, funding acquisition, investigation, writing—review and editing. **X.-O. Shu:** Conceptualization, resources, data curation, supervision, funding acquisition, investigation, writing—review and editing.

Acknowledgments

The Shanghai Women's Health Study was supported by NIH/NCI grant UM1 CA182910. The Shanghai Men's Health Study was supported by NIH/NCI grant UM1 CA173640. Sample preparation was conducted at the Survey and Biospecimen Shared Resources, which is supported in part by the Vanderbilt-Ingram Cancer Center (P30CA068485). X. Shu was partially supported by NCI grants K99 CA230205 and R00 CA230205 during the project period.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 15, 2020; revised January 17, 2021; accepted March 19, 2021; published first April 1, 2021.

- Xiao S, Zhou L. Gastric cancer: metabolic and metabolomics perspectives (review). *Int J Oncol* 2017;51:5–17.
- Huang S, Guo Y, Li Z, Zhang Y, Zhou T, You W, et al. A systematic review of metabolomic profiling of gastric cancer and esophageal cancer. *Cancer Biol Med* 2020;17:181–98.
- Zheng W, Chow WH, Yang G, Jin F, Rothman N, Blair A, et al. The Shanghai women's health study: rationale, study design, and baseline characteristics. *Am J Epidemiol* 2005;162:1123–31.
- Shu XO, Li H, Yang G, Gao J, Cai H, Takata Y, et al. Cohort profile: the shanghai men's health study. *Int J Epidemiol* 2015;44:810–8.
- Hishikawa D, Hashidate T, Shimizu T, Shindou H. Diversity and function of membrane glycerophospholipids generated by the remodeling pathway in mammalian cells. *J Lipid Res* 2014;55:799–807.
- Law SH, Chan ML, Marathe GK, Parveen F, Chen CH, Ke LY. An updated review of lysophosphatidylcholine metabolism in human diseases. *Int J Mol Sci* 2019;20: 1149.
- Hannibal L, Lysne V, Bjorke-Monsen AL, Behringer S, Grunert SC, Spiekerkoetter U, et al. Biomarkers and algorithms for the diagnosis of vitamin B12 deficiency. *Front Mol Biosci* 2016;3:27.
- Vashi P, Edwin P, Popiel B, Lammersfeld C, Gupta D. Methylmalonic acid and homocysteine as indicators of vitamin b-12 deficiency in cancer. *PLoS One* 2016; 11:e0147843.
- Lee SM, Oh J, Chun MR, Lee SY. Methylmalonic acid and homocysteine as indicators of vitamin B12 deficiency in patients with gastric cancer after gastrectomy. *Nutrients* 2019;11:450.
- Miranti EH, Stolzenberg-Solomon R, Weinstein SJ, Selhub J, Mannisto S, Taylor PR, et al. Low vitamin B12 increases risk of gastric cancer: A prospective study of one-carbon metabolism nutrients and risk of upper gastrointestinal tract cancer. *Int J Cancer* 2017;141:1120–9.
- Chau YP, Au PCM, Li GHY, Sing CW, Cheng VKF, Tan KCB, et al. Serum metabolome of coffee consumption and its association with bone mineral density: the hong kong osteoporosis study. *J Clin Endocrinol Metab* 2020; 105:dgz210.
- Zhao LG, Li HL, Sun JW, Yang Y, Ma X, Shu XO, et al. Green tea consumption and cause-specific mortality: results from two prospective cohort studies in China. *J Epidemiol* 2017;27:36–41.
- Coffee consumption in East and Southeast Asia: 1990–2012. International coffee council, 112th session; 2014. Available from: <https://www.ico.org/news/icc-112-4e-consumption-asia.pdf>.

27. Heinzmann SS, Holmes E, Kochhar S, Nicholson JK, Schmitt-kopplin P. 2-furoylglycine as a candidate biomarker of coffee consumption. *J Agric Food Chem* 2015;63:8615–21.
28. Kim Y, Je Y, Giovannucci E. Coffee consumption and all-cause and cause-specific mortality: a meta-analysis by potential modifiers. *Eur J Epidemiol* 2019;34:731–52.
29. Tang J, Zheng JS, Fang L, Jin Y, Cai W, Li D. Tea consumption and mortality of all cancers, CVD and all causes: a meta-analysis of eighteen prospective cohort studies. *Br J Nutr* 2015;114:673–83.
30. Ding M, Bhupathiraju SN, Satija A, van Dam RM, Hu FB. Long-term coffee consumption and risk of cardiovascular disease: a systematic review and a dose-response meta-analysis of prospective cohort studies. *Circulation* 2014; 129:643–59.
31. Zhang C, Qin YY, Wei X, Yu FF, Zhou YH, He J. Tea consumption and risk of cardiovascular outcomes and total mortality: a systematic review and meta-analysis of prospective observational studies. *Eur J Epidemiol* 2015;30:103–13.
32. Nechuta S, Shu XO, Li HL, Yang G, Ji BT, Xiang YB, et al. Prospective cohort study of tea consumption and risk of digestive system cancers: results from the Shanghai Women's Health Study. *Am J Clin Nutr* 2012;96:1056–63.
33. Huang J, Weinstein SJ, Kitahara CM, Karoly ED, Sampson JN, Albanes D. A prospective study of serum metabolites and glioma risk. *Oncotarget* 2017;8: 70366–77.
34. Murphy G, Fan JH, Mark SD, Dawsey SM, Selhub J, Wang J, et al. Prospective study of serum cysteine levels and oesophageal and gastric cancers in China. *Gut* 2011;60:618–23.