



Original Investigation | Oncology

# Association of Untargeted Urinary Metabolomics and Lung Cancer Risk Among Never-Smoking Women in China

Wei Jie Seow, ScD; Xiao-Ou Shu, PhD; Jeremy K. Nicholson, PhD; Elaine Holmes, PhD; Douglas I. Walker, PhD; Wei Hu, PhD; Qiuyin Cai, PhD; Yu-Tang Gao, MD; Yong-Bing Xiang, MD; Steven C. Moore, PhD; Bryan A. Bassig, PhD; Jason Y. Y. Wong, ScD; Jinming Zhang, ScD; Bu-Tian Ji, MD; Claire L. Boulangé, PhD; Manuja Kaluarachchi, PhD; Anisha Wijeyesekera, PhD; Wei Zheng, PhD; Paul Elliott, PhD; Nathaniel Rothman, MD; Qing Lan, PhD

## Abstract

**IMPORTANCE** Chinese women have the highest rate of lung cancer among female never-smokers in the world, and the etiology is poorly understood.

**OBJECTIVE** To assess the association between metabolomics and lung cancer risk among never-smoking women.

**DESIGN, SETTING, AND PARTICIPANTS** This nested case-control study included 275 never-smoking female patients with lung cancer and 289 never-smoking cancer-free control participants from the prospective Shanghai Women's Health Study recruited from December 28, 1996, to May 23, 2000. Validated food frequency questionnaires were used for the collection of dietary information. Metabolomic analysis was conducted from November 13, 2015, to January 6, 2016. Data analysis was conducted from January 6, 2016, to November 29, 2018.

**EXPOSURES** Untargeted ultra-high-performance liquid chromatography–tandem mass spectrometry and nuclear magnetic resonance metabolomic profiles were characterized using prediagnosis urine samples. A total of 39 416 metabolites were measured.

**MAIN OUTCOMES AND MEASURES** Incident lung cancer.

**RESULTS** Among the 564 women, those who developed lung cancer (275 participants; median [interquartile range] age, 61.0 [52-65] years) and those who did not develop lung cancer (289 participants; median [interquartile range] age, 62.0 [53-66] years) at follow-up (median [interquartile range] follow-up, 10.9 [9.0-11.7] years) were similar in terms of their secondhand smoke exposure, history of respiratory diseases, and body mass index. A peak metabolite, identified as 5-methyl-2-furoic acid, was significantly associated with lower lung cancer risk (odds ratio, 0.57 [95% CI, 0.46-0.72];  $P < .001$ ; false discovery rate = 0.039). Furthermore, this peak was weakly correlated with self-reported dietary soy intake ( $\rho = 0.21$ ;  $P < .001$ ). Increasing tertiles of this metabolite were associated with lower lung cancer risk (in comparison with first tertile, odds ratio for second tertile, 0.52 [95% CI, 0.34-0.80]; and odds ratio for third tertile, 0.46 [95% CI, 0.30-0.70]), and the association was consistent across different histological subtypes and follow-up times. Additionally, metabolic pathway analysis found several systemic biological alterations that were associated with lung cancer risk, including 1-carbon metabolism, nucleotide metabolism, oxidative stress, and inflammation.

**CONCLUSIONS AND RELEVANCE** This prospective study of the untargeted urinary metabolome and lung cancer among never-smoking women in China provides support for the hypothesis that

(continued)

## Key Points

**Question** What is the association of metabolomic biomarkers with lung cancer in women who have never smoked?

**Findings** In this case-control study of 275 never-smoking women with lung cancer and 289 never-smoking cancer-free women in China, the metabolite 5-methyl-2-furoic acid was correlated with dietary soy consumption and pathways including 1-carbon metabolism, oxidative stress, nucleotide metabolism, and inflammation and was significantly associated with lower lung cancer risk.

**Meaning** These findings suggest that certain metabolites and pathways are associated with lower lung cancer risk in never-smoking women and biological processes linked to air pollution may be associated with higher lung cancer risk in this population.

## + Supplemental content

Author affiliations and article information are listed at the end of this article.

**Open Access.** This is an open access article distributed under the terms of the CC-BY License.

Abstract (continued)

soy-based metabolites are associated with lower lung cancer risk in never-smoking women and suggests that biological processes linked to air pollution may be associated with higher lung cancer risk in this population.

JAMA Network Open. 2019;2(9):e1911970. doi:10.1001/jamanetworkopen.2019.11970

## Introduction

Lung cancer is the leading cause of cancer mortality among men and women, accounting for 27% of all cancer deaths worldwide and nearly 158 040 deaths in the United States in 2015.<sup>1</sup> Although most cases of lung cancer are caused by active tobacco smoking, approximately 25% occur in never-smokers.<sup>2</sup> Asian individuals (particularly Chinese women) have a higher rate of lung cancer among never-smokers compared with individuals of European descent<sup>3</sup>; however, the etiology and risk factors of this malignant neoplasm independent of smoking are poorly understood.<sup>4</sup> Therefore, it is imperative to understand the etiology of lung cancer and identify new biomarker targets or metabolomic profiles as potential diagnostic and screening tools for lung cancer.

Altered tumor metabolism, a distinctive hallmark of many cancers, is required to acquire sufficient nutrients to maintain deregulated and unlimited proliferation of the cancer cells.<sup>5</sup> An agnostic approach using untargeted metabolic profiling allows more comprehensive coverage of downstream biological responses to environmental exposures, and can provide insight into underlying molecular disease mechanisms.<sup>6</sup> Metabolomic techniques have been used to identify metabolites in various cancers.<sup>7-11</sup> However, most of these metabolomic studies of cancer were conducted using case-control study designs that compared the metabolite levels between individuals diagnosed with cancer and healthy controls, which could be prone to reverse causation bias. Targeted prospective studies have found some urinary metabolites to be associated with higher lung cancer risk, including creatine riboside,<sup>12</sup> N-acetylneuraminic acid,<sup>12</sup> and 7-methylguanine.<sup>13</sup> Both studies were conducted in populations of European descent and had small numbers of never-smokers.

The Shanghai Women's Health Study (SWHS) is a population-based prospective cohort of 73 363 female participants who were predominantly never-smokers. This study provides a unique opportunity to prospectively evaluate the role of the metabolome in lung cancer development. To our knowledge, this is the first prospective study to investigate prediagnostic metabolic biomarkers and pathway alterations in urine samples using an untargeted approach among never-smoking women, which may provide novel insights into the etiology and mechanisms associated with lung cancer development in never-smokers.

## Methods

### Study Population

From December 28, 1996, to May 23, 2000, the SWHS recruited 73 363 Chinese women aged 40 to 70 years who were primarily never-smokers. In this nested case-control study, we performed metabolomic profiling of all 275 female patients with incident lung cancer and 289 female healthy controls who were individually matched on date of birth ( $\pm 2$  years) and date of spot urine sample collection ( $\pm 3$  months). We followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline. Information on demographic characteristics, lifestyle, diet, and medical history was obtained using a detailed survey administered in person at baseline.<sup>14</sup> A comprehensive and validated food frequency questionnaire was used to assess dietary intake during the year prior to the interview. Urine samples were collected into a sterilized cup containing 125 mg of ascorbic acid to prevent oxidation of labile metabolites. After collection, the samples were kept at

approximately 0°C to -4°C and were processed within 6 hours for long-term storage at -70°C.<sup>14</sup> All study participants provided written informed consent prior to participation, and the study protocols were approved by institutional review boards of all participating institutes and the National Cancer Institute.

### Follow-up and Lung Cancer Identification

The participants were followed up with a combination of in-person surveys and periodic linkage to cancer and vital statistics registries.<sup>14</sup> Briefly, cohort members were followed up for diagnosis of cancer by in-person surveys every 2 to 3 years and annual record linkage with databases from the population-based Shanghai Cancer Registry, Shanghai Vital Statistics Registry, and Shanghai Resident Registry. All possible cancer diagnoses were verified through home visits and a review of medical records by a panel of oncologists. Lung cancer was defined as code 162 under the *International Classification of Diseases, Ninth Revision*. All female patients with lung cancer were included as soon as they were diagnosed after the baseline recruitment through December 31, 2009. Lung cancer diagnosed before the recruitment was excluded. Information on histological subtypes of lung cancer was extracted from medical records. Histological groups of lung cancer were classified according to the morphology code of *International Classification of Diseases for Oncology, Second Edition* (8140/3).<sup>15</sup>

### Urinary Metabolomics

All urine samples (patient and pooled quality control samples) were analyzed in an untargeted fashion by ultra-high-performance liquid chromatography–tandem mass spectrometry (UPLC-MS) and 600-MHz hydrogen 1 nuclear magnetic resonance (NMR) spectroscopy at Imperial College London, United Kingdom, acquired between November 13, 2015, and January 6, 2016. The spectral data for the pooled quality control sample was collected after every 5 patient samples. The measurements were carried out according to established protocols used for metabolic phenotyping of biofluids.<sup>16,17</sup> For the UPLC-MS measurements, metabolic profiling was performed on an ACQUITY UPLC System (Waters Corp) coupled to a Xevo Triple Quadrupole Mass Spectrometer (Waters Corp) following a standard set of protocols.<sup>18</sup> Mass spectrometry was performed using electrospray in both positive and negative ionization (electrospray ionization [ESI]<sup>+</sup> and ESI<sup>-</sup>) modes under conditions described previously.<sup>19</sup> This UPLC-MS method allows for a high-resolution, accurate, and simultaneous detection of thousands of peaks that correspond to individual ions with a unique mass to charge (m/z) ratio and retention time.<sup>20</sup>

Each UPLC-MS data set was preprocessed using Progenesis (Non Linear Dynamics) to convert the 3-dimensional UPLC-MS raw data into time-aligned detected features of retention time, m/z ratio, and intensity (peak area). The ion intensities for each peak detected were then normalized, within each sample, to the median peak intensity in that sample. The metabolic profiling data were preprocessed using the freely available XCMS software (Scripps Research Institute) to detect and identify peaks, align retention time, and perform normalization of the raw mass spectrometry data before subsequent statistical analysis.<sup>19,21</sup> The resulting features were filtered to remove features with a coefficient of variation greater than 30% across the quality control samples. The NMR data were imported into MATLAB R2012b (Natick) using the Metaspectra program script, using a resolution of 0.00055. The data were then aligned using recursive segmentwise peak alignment<sup>22</sup> and normalized by probabilistic quotient to account for and exclude systematic sources of bias within samples not due to biological processes or the environment. Statistical recoupling of variables<sup>23</sup> was then used for peak detection. The resulting tables were imported to SIMCA-P+ software version 13.0.2 (Umetrics) to perform multivariate statistical analysis. In total, there were 20 555 UPLC-MS peaks in positive mode, 18 475 in negative mode, and 386 NMR peaks.

## Statistical Analysis

A small constant ( $1 \times 10^{-9}$ ) was added to all metabolite levels, which were then natural log-transformed to approximate normal distributions and subsequently analyzed as both continuous and categorical variables (based on tertiles among controls). Principal component analysis, partial least-squares-discriminant analysis, and orthogonal projection on latent structures-discriminant analysis (OPLS-DA) were performed on all data to check for batch effects and clustering of samples.

The characteristics of patients with lung cancer and control individuals were compared using the Wilcoxon rank sum test for continuous variables and Fisher exact test for categorical variables. Analyses were restricted to lifetime never-smokers (275 female patients with lung cancer and 289 controls). Separate unconditional logistic regression models were used to estimate the odds ratios (ORs) and 95% confidence intervals for the associations between each metabolite and lung cancer risk. Tests of trend were calculated by treating the level of each metabolite as a continuous variable. We also stratified the analysis by follow-up time ( $<9$  years [the median duration of follow-up] or  $\geq 9$  years) and major lung cancer subtypes (adenocarcinoma, nonadenocarcinoma). Because dietary intake has previously been reported to be associated with altered metabolites in urine samples, using a multivariable linear regression model, we evaluated the associations between the most significant metabolite and prediagnosis dietary factors among female patients with lung cancer and controls. In addition, because soy food intake has been previously found to be associated with risk of lung cancer in the SWHS,<sup>24</sup> we further looked at grouped soy intake, which included 11 soy food items that are commonly consumed in Shanghai, including soy milk, tofu, fried tofu, dried or pressed tofu, fresh green soy beans, dry soy beans, soy sprouts, and other soy products. In addition, mediation analysis was used to estimate the proportion of the mediated effect via the significant metabolites on the association between soy consumption and lung cancer risk.<sup>25</sup> Missing data were excluded from all analyses. All models were adjusted for age (continuous), body mass index, history of respiratory diseases (categorical; ever or never) and environmental tobacco smoke (categorical; ever or never). Correlations among significant metabolites, as well as between significant metabolites and 17 dietary factors, were evaluated using Spearman correlation coefficients.

All statistical analyses were conducted using R statistical software version 3.2.2 (R Project for Statistical Computing) and SAS statistical software version 4.3 (SAS Institute). All tests with a false discovery rate (FDR) of less than 0.10 were considered statistically significant. All *P* values were reported based on 2-sided tests and *P* < .05 was considered statistically significant. We calculated FDRs using the Benjamini-Hochberg method to account for multiple comparisons.<sup>26</sup>

## Metabolite Identification

The metabolic phenotypes of female patients with lung cancer and controls were further evaluated using pathway enrichment to identify biological processes associated with lung cancer diagnosis. In this analysis, we tested all metabolic features for their association with lung cancer diagnosis. Pathway enrichment analysis was completed using all metabolic features with *P* < .05. This threshold minimizes type II error (false-negatives) at the increased risk of type I error (false-positives). The pathway enrichment approach uses a random sampling framework to identify a null distribution of metabolites across a metabolic network; comparison of the null distribution with annotated metabolites from signals with *P* < .05 removes false-positives, which would be randomly distributed across the metabolic map, and allows isolation of biological effects. The pathway enrichment was completed separately for positive- and negative-mode data using the mummichog analysis approach<sup>27</sup>; pathways with *P* < .05 were selected for further characterization. Metabolites were identified and annotated based on accurate mass (for UPLC-MS) and by using in-house and available online databases such as the Human Metabolome Database and METLIN.

Results

Among the 564 women, those who developed lung cancer (275 participants; median [interquartile range {IQR}] age, 61.0 [52-65] years) and those who did not develop lung cancer (289 participants; median [IQR] age, 62.0 [53-66] years) at follow-up were similar in terms of secondhand smoke exposure, history of respiratory diseases, and body mass index (Table 1). The median (IQR) time from the baseline questionnaire to lung cancer diagnosis was 9.26 (6.69-11.05) years.

Using an FDR cutoff of 10% to adjust for multiple comparisons, 3 urinary metabolites by UPLC-MS were significantly associated with lower lung cancer risk (Table 2). The statistically significant metabolites were pos\_2.61\_127.0382m/z (OR, 0.57 [95% CI, 0.46-0.72];  $P < .001$ ; FDR = 0.039), neg\_2.60\_369.0408m/z (OR, 0.97 [95% CI, 0.96-0.98];  $P < .001$ ; FDR = 0.039), and pos\_2.61\_184.0325n (OR, 0.55 [95% CI, 0.43-0.71];  $P < .001$ ; FDR = 0.065). The 3 significant metabolites were highly correlated with each other (Spearman correlations ranged from 0.67-0.89) (eTable 1 in the Supplement). In addition, results were comparable when stratified by follow-up time (<9 years of follow-up and  $\geq 9$  years of follow-up) after urine sample collection (Table 2). Similar results were also obtained after excluding 4 women who were diagnosed within 2 years after the baseline interview (data not shown). Furthermore, similar associations were observed among patients with lung adenocarcinoma (135 patients [49.1%]) and those with unclassified lung cancer (104 patients [37.8%]) (eTable 2 and eTable 3 in the Supplement). None of the NMR features were significantly associated with lung cancer risk (data not shown).

Table 1. Characteristics of the Nested Case-Control Study of Never-Smoking Women in the Shanghai Women’s Health Study

Characteristic	Lung Cancer (n = 275)	No Lung Cancer (n = 289)	P Value <sup>a</sup>
Environmental tobacco smoke, No. (%)			
Never	70 (25.5)	71 (24.6)	.86
Ever	187 (68)	200 (69.2)	
Missing	18 (6.5)	18 (6.2)	
History of respiratory diseases, No. (%)			
Yes	28 (10.2)	22 (7.6)	.36
No	247 (89.8)	267 (92.4)	
Age at baseline, median (IQR), y	61.0 (52-65)	62.0 (53-66)	.42
BMI, median (IQR)	24.3 (22.4-26.6)	24.3 (22.0-26.9)	.84
Time to diagnosis, median (IQR), y	9.26 (6.69-11.1)	NA	NA
Histological subtypes, No. (%)			
Adenocarcinoma	135 (49.1)	NA	NA
Squamous cell carcinoma	9 (3.3)	NA	
Other histology subtypes	19 (6.9)	NA	
Unclassified	104 (37.8)	NA	
No ICD-O-2 code	8 (2.9)	NA	

Abbreviations: BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); ICD-O-2, International Classification of Diseases for Oncology, Second Revision; IQR, interquartile range; NA, not applicable.

<sup>a</sup> P values were obtained from Wilcoxon signed-rank test for continuous variables and Fisher exact test for categorical variables.

Table 2. Associations Between Urinary Metabolites and Lung Cancer Risk, Stratified by Follow-up Time

Metabolite <sup>a</sup>	OR (95% CI) <sup>b</sup>	P Value	FDR	Follow-up			
				<9 y <sup>c</sup>		$\geq 9$ y <sup>d</sup>	
				OR (95% CI) <sup>b</sup>	P Value	OR (95% CI) <sup>b</sup>	P Value
pos_2.61_127.0382m/z	0.57 (0.46-0.72)	<.001	0.039	0.51 (0.38-0.70)	<.001	0.62 (0.48-0.81)	<.001
neg_2.60_369.0408m/z	0.97 (0.96-0.98)	<.001	0.039	0.97 (0.95-0.98)	<.001	0.97 (0.96-0.99)	<.001
pos_2.61_184.0325n	0.55 (0.43-0.71)	<.001	0.065	0.49 (0.35-0.69)	<.001	0.60 (0.44-0.82)	.001

Abbreviations: FDR, false discovery rate; OR, odds ratio.

<sup>a</sup> Metabolites were natural log-transformed.

<sup>b</sup> All models were adjusted for age, body mass index, history of respiratory diseases, and secondhand smoke exposure.

<sup>c</sup> Sample sizes were 131 with lung cancer and 289 without lung cancer.

<sup>d</sup> Sample sizes were 144 with lung cancer and 289 without lung cancer.

One of the significant metabolites (pos\_2.61\_127.0382m/z) was identified as 5-methyl-2-furoic acid based on its fragmentation pattern by tandem mass spectrometry (eTable 4 in the Supplement). There was a significant monotonic dose-response association between 5-methyl-2-furoic acid and lung cancer risk when analyzed as tertiles (Table 3). Increasing tertiles of the metabolite level were significantly associated with lower lung cancer risk (in comparison with first tertile, OR for second tertile, 0.52 [95% CI, 0.34-0.80]; and OR for third tertile, 0.46 [95% CI, 0.30-0.70]; *P* for trend < .001). Similar trends were observed for the 2 other significant metabolites (eTable 5 in the Supplement).

Because dietary intake has been previously reported to be associated with altered metabolites in urine samples, we also assessed the correlation of each available dietary factor with the level of 5-methyl-2-furoic acid. Consumption of soy food, including soy milk, tofu, fried tofu, dried or pressed tofu, fresh green soy beans, dry soy beans, soy sprouts, and other soy products, in the past year had a weak positive correlation with levels of 5-methyl-2-furoic acid among controls ( $\rho = 0.21$ ; *P* < .001) (eTable 6 in the Supplement). Using mediation analysis, we found that 25% of the association between soy consumption and higher lung cancer risk was significantly mediated via 5-methyl-2-furoic acid.

Regression analysis included 152 positive-mode and 186 negative-mode features associated with lung cancer risk at *P* < .05; only 4 features were present with Bonferroni correction less than .05. Pathway enrichment identified 40 metabolic pathways associated with lung cancer risk at *P* < .05, including 8 pathways meeting a Bonferroni-corrected *P* < .05 (Figure). Eight were detected in both modes and included amino acid metabolism, nucleotide metabolism, nitrogen catabolism, and B vitamin pathways. Metabolic alterations associated with lung cancer risk suggest systemic changes to key biological processes, including 1-carbon metabolism (vitamin B<sub>9</sub> [folate] metabolism; glycine, serine, alanine, and threonine metabolism), oxidative stress pathways (ascorbate [vitamin C] and aldarate metabolism, methionine and cysteine metabolism, glutamate metabolism), nucleotide metabolism (purine metabolism), and inflammation (leukotriene metabolism).

## Discussion

To our knowledge, this is the first large-scale prospective study of untargeted de novo metabolomics in association with lung cancer among never-smoking Chinese women. We found 3 metabolites associated with a reduction of lung cancer risk after accounting for multiple comparisons. These metabolites were highly correlated with each other and may represent the same chemical family and/or source. One of these metabolites (pos\_2.61\_127.0382m/z) was identified as 5-methyl-2-furoic acid, which belongs to the family of furoic acid derivatives. It was weakly correlated with dietary soy

**Table 3. Associations Between 5-Methyl-2-Furoic Acid and Lung Cancer, by Tertiles and Stratified by Follow-up Time**

Log Metabolite Tertiles	Patients With Lung Cancer, No.	Patients Without Lung Cancer, No.	OR (95% CI) <sup>a</sup>	<i>P</i> Value	<i>P</i> for Trend
First tertile, median = 10.5	139	96	1 [Reference]	NA	
Second tertile, median = 11.3	71	95	0.52 (0.34-0.80)	.003	<.001
Third tertile, median = 12.1	64	96	0.46 (0.30-0.70)	<.001	
<b>&lt;9-y follow-up</b>					
First tertile, median = 10.5	69	96	1 [Reference]	NA	
Second tertile, median = 11.3	34	96	0.51 (0.30-0.86)	.01	<.001
Third tertile, median = 12.1	27	96	0.36 (0.20-0.63)	<.001	
<b>≥9-y follow-up</b>					
First tertile, median = 10.4	70	96	1 [Reference]	NA	
Second tertile, median = 11.2	37	96	0.53 (0.32-0.89)	.02	.02
Third tertile, median = 12.1	37	96	0.56 (0.34-0.93)	.03	

Abbreviations: OR, odds ratio; NA, not applicable.

<sup>a</sup> All models were adjusted for age, body mass index, history of respiratory diseases, and secondhand smoke exposure.

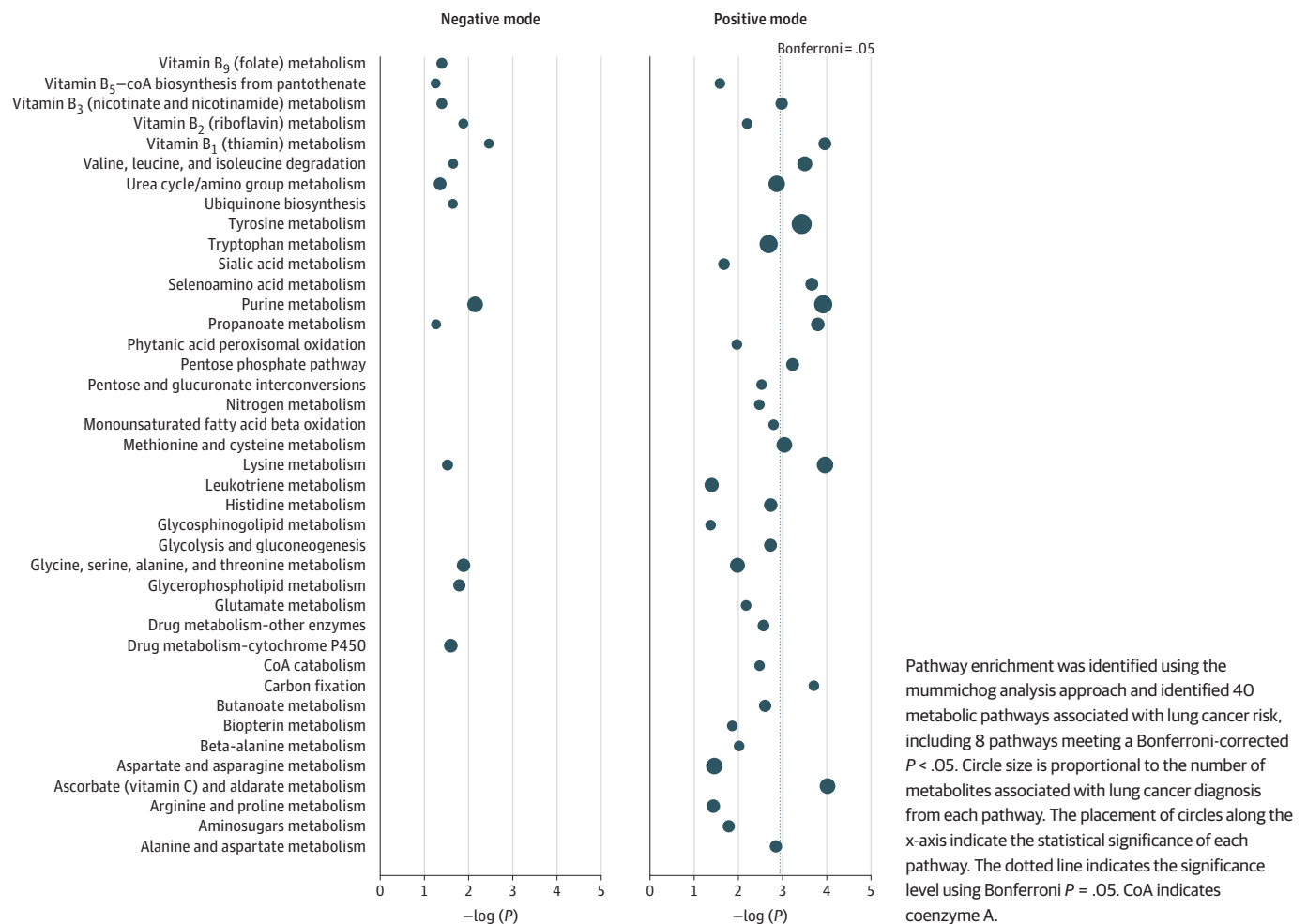


intake in our study population, suggesting that this metabolite may be partially derived from soy-related dietary products. Metabolic alterations identified in the pathway enrichment analysis suggest lung cancer risk is associated with an increased oxidative state and depletion of systemic antioxidant pathways.

Furan fatty acids are related compounds that occur naturally in foods such as soybean oil<sup>28</sup> and have also been shown to have anti-inflammatory and antioxidant properties in vivo.<sup>29</sup> Methylated and furan-containing fatty acids have also been demonstrated to protect against reactive oxygen species-mediated damage in vivo by acting as scavengers of reactive oxygen species.<sup>30</sup> Interestingly, we found higher levels of the furoic acid metabolite in urine to be associated with a lower risk of lung cancer. Another furoic acid derivative, 5-tetradecyl-oxy-2-furoic acid (TOFA), has been reported to reduce fatty acid synthesis by inhibiting acetyl coenzyme A carboxylase- $\alpha$  (ACCA), thereby increasing apoptosis rates and eradicating prostate cancer cells via the mitochondrial pathway.<sup>31</sup> One study<sup>32</sup> identified TOFA as a potent inhibitor of stearyl coenzyme A desaturase 1 (SCD1), which is an enzyme in the fatty-acid synthesis pathway that depletes monounsaturated fatty acids and therefore restricts cancer cell proliferation. Furoic acid is derived from furans, which are commonly used as food additives, and preservatives, which are usually present in high levels in heat-treated foods such as soy sauce, baked beans, and brewed coffee.<sup>33</sup>

Dietary soy intake was previously shown to be associated with a 37% lower risk of lung cancer among a population of Chinese women in Shanghai, comparing the highest quintile with the lowest

Figure. Metabolic Pathways Associated With Lung Cancer Risk Using Features With  $P < .05$



quintile.<sup>24</sup> This is consistent with our findings in the same population of a significant protective association between 5-methyl furan-2-carboxylic acid and lung cancer risk, and a significant weak correlation between this metabolite and dietary soy intake. Soy foods contain high concentrations of isoflavones, a type of phytoestrogen, which possess antiestrogenic and antioxidative effects.<sup>34</sup> It has been hypothesized that the bioactive compounds of dietary soy may contain antioxidant and anti-inflammatory properties that inhibit tumor growth and invasion and induce apoptosis.<sup>35-37</sup>

Evaluating pathway (rather than individual metabolite) changes associated with disease risk provides a systems biology approach that can be used to understand the metabolic effects of lung cancer pathobiology. In the present study, we identified alterations in pathways that suggest dysregulation of 1-carbon metabolism and related pathways, including folate, methionine and cysteine metabolism, purine metabolism, and glycine, serine, alanine, and threonine metabolism. One-carbon metabolism is critical in intermediary metabolism and supports multiple biological processes, including nucleotide metabolism, amino acid metabolism, and homeostasis (methionine and cysteine, serine, and glycine).<sup>38</sup>

Our study suggests that metabolic alterations in oxidative stress and inflammatory pathways, which could arise from multiple exogenous and endogenous exposures and biological processes, are associated with lung cancer risk in never-smoking women in China. It is, however, intriguing that these same pathways have also been linked to exposure to environmental pollutants that are suspected to be associated with higher lung cancer disease risk,<sup>39,40</sup> including traffic-related air pollution.<sup>41-44</sup> Taken together, the overlap in oxidative stress and inflammation metabolites suggest environmental exposures and lung cancer risk in this population may share similarities in metabolic alterations and provide an indirect link between the two. Additional studies are needed to evaluate these possible associations.

One major strength of this study is that the metabolites were measured in never-smoking women, thereby eliminating the possibility of confounding by active smoking, which has occurred in other studies. Other strengths of this study include the prospective design, in which the collection of urinary metabolomic biomarkers preceded lung cancer diagnosis, which provides exposure-outcome temporality, and the untargeted measurement of metabolites, which provides increased opportunities for making novel discoveries.

### Limitations

There were limitations to this study. One is that a notable proportion of the cohort with lung cancer (40.7%) had an unclassified histological subtype. However, analysis of unclassified lung cancer yielded results that were similar to analysis of all lung cancer and of lung adenocarcinoma only. The source of the dietary data was a food frequency questionnaire designed to assess dietary patterns in the past year when the baseline interview was conducted. The exact question asked in the questionnaire was, "Would you please tell me if you ate those foods and how much you ate in general over the past year?" Given that most metabolites have relatively short half-lives, the weak correlation between 5-methyl-2-furoic acid and dietary soy intake may have been stronger if we had questionnaire or diary data for the last few days before urine was collected. In addition, the urine samples were only collected once and we were unable to assess the temporal trends of metabolites over time.

### Conclusions

This nested case-control study found that a urinary metabolite, 5-methyl-2-furoic acid, was associated with decreased risk of lung cancer in a prospective cohort of never-smoking women. The identified significant metabolite (5-methyl-2-furoic acid) was weakly correlated with self-reported usual dietary intake of soy. Furthermore, our finding is consistent with recent epidemiological evidence showing that higher dietary soy consumption may be associated with decreased lung cancer risk among never-smokers, strengthening the biological relevance of the identified



metabolite. In addition, metabolic pathways associated with lung cancer risk suggest systemic biological alterations were present, and a number of pathways identified in the present study have been linked to results of previous metabolomic studies of populations exposed to air pollution, a suspected risk factor for lung cancer. Additional studies are warranted to replicate and further characterize the associations reported in our study, which provide additional clues about the exposures and biological processes that may contribute to the risk of lung cancer in never-smoking populations in Asia.

---

## ARTICLE INFORMATION

**Accepted for Publication:** August 4, 2019.

**Published:** September 20, 2019. doi:10.1001/jamanetworkopen.2019.11970

**Open Access:** This is an open access article distributed under the terms of the [CC-BY License](#). © 2019 Seow WJ et al. *JAMA Network Open*.

**Corresponding Author:** Wei Jie Seow, ScD, Saw Swee Hock School of Public Health, National University of Singapore, 12 Science Drive 2, 09-01H, Singapore 117549 ([ephswj@nus.edu.sg](mailto:ephswj@nus.edu.sg)).

**Author Affiliations:** Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore (Seow); Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore (Seow); Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, Maryland (Seow, Hu, Moore, Bassig, Wong, Zhang, Ji, Rothman, Lan); Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center and Vanderbilt-Ingram Cancer Center, Nashville, Tennessee (Shu, Cai, Zheng); Biomolecular Medicine, Division of Computational and Systems Medicine, Medical Research Council–National Institute for Health Research National Phenome Centre, Imperial College London, United Kingdom (Nicholson, Holmes, Boulangé, Kaluarachchi, Wijeyesekera, Elliott); Medical Research Council–PHE Centre for Environment and Health, Department of Surgery and Cancer, Imperial College London, United Kingdom (Nicholson, Holmes, Boulangé, Kaluarachchi, Wijeyesekera, Elliott); Health Futures Institute, Murdoch University, Murdoch, Western Australia, Australia (Nicholson, Holmes); Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York (Walker); Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China (Gao, Xiang); State Key Laboratory of Oncogene and Related Genes, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China (Xiang); MRC–PHE Centre for Environment and Health, School of Public Health, Department of Epidemiology and Biostatistics, Imperial College London, United Kingdom (Elliott); National Institute for Health Research, Imperial College Biomedical Research Centre, London, United Kingdom (Elliott); Health Data Research UK London at Imperial College London, United Kingdom (Elliott).

**Author Contributions:** Drs Seow and Lan had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Drs Seow, Shu, Nicholson, Holmes, and Walker contributed equally to the study.

*Concept and design:* Seow, Shu, Elliott, Rothman, Lan.

*Acquisition, analysis, or interpretation of data:* Seow, Nicholson, Holmes, Walker, Hu, Cai, Gao, Xiang, Moore, Bassig, Wong, Zhang, Ji, Boulangé, Kaluarachchi, Wijeyesekera, Zheng, Rothman, Lan.

*Drafting of the manuscript:* Seow, Walker, Lan.

*Critical revision of the manuscript for important intellectual content:* All authors.

*Statistical analysis:* Seow, Holmes, Walker, Moore, Wong, Boulangé, Kaluarachchi, Lan.

*Obtained funding:* Shu, Rothman, Lan.

*Administrative, technical, or material support:* Shu, Hu, Gao, Xiang, Wong, Ji, Boulangé, Kaluarachchi, Wijeyesekera.

*Supervision:* Nicholson, Gao, Elliott, Rothman, Lan.

**Conflict of Interest Disclosures:** None reported.

**Funding/Support:** This work was supported by intramural funds from the National Cancer Institute and the National University of Singapore Start-up Grant. The Shanghai Women's Health Study is supported by the US National Institutes of Health/National Cancer Institute (grant UMI CA182910).

**Role of the Funder/Sponsor:** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

## REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA Cancer J Clin*. 2015;65(1):5-29. doi:10.3322/caac.21254
2. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers—a different disease. *Nat Rev Cancer*. 2007;7(10):778-790. doi:10.1038/nrc2190
3. Epplein M, Schwartz SM, Potter JD, Weiss NS. Smoking-adjusted lung cancer incidence among Asian-Americans (United States). *Cancer Causes Control*. 2005;16(9):1085-1090. doi:10.1007/s10552-005-0330-6
4. Thun MJ, Hannan LM, Adams-Campbell LL, et al. Lung cancer occurrence in never-smokers: an analysis of 13 cohorts and 22 cancer registry studies. *PLoS Med*. 2008;5(9):e185. doi:10.1371/journal.pmed.0050185
5. Pavlova NN, Thompson CB. The emerging hallmarks of cancer metabolism. *Cell Metab*. 2016;23(1):27-47. doi:10.1016/j.cmet.2015.12.006
6. Vrijheid M. The exposome: a new paradigm to study the impact of environment on health. *Thorax*. 2014;69(9):876-878. doi:10.1136/thoraxjnl-2013-204949
7. Hori S, Nishiumi S, Kobayashi K, et al. A metabolomic approach to lung cancer. *Lung Cancer*. 2011;74(2):284-292. doi:10.1016/j.lungcan.2011.02.008
8. Wittmann BM, Stirdivant SM, Mitchell MW, et al. Bladder cancer biomarker discovery using global metabolomic profiling of urine. *PLoS One*. 2014;9(12):e115870. doi:10.1371/journal.pone.0115870
9. Yu L, Jiang C, Huang S, Gong X, Wang S, Shen P. Analysis of urinary metabolites for breast cancer patients receiving chemotherapy by CE-MS coupled with on-line concentration. *Clin Biochem*. 2013;46(12):1065-1073. doi:10.1016/j.clinbiochem.2013.05.049
10. Kim K, Taylor SL, Ganti S, Guo L, Osier MV, Weiss RH. Urine metabolomic analysis identifies potential biomarkers and pathogenic pathways in kidney cancer. *OMICS*. 2011;15(5):293-303. doi:10.1089/omi.2010.0094
11. Playdon MC, Ziegler RG, Sampson JN, et al. Nutritional metabolomics and breast cancer risk in a prospective study. *Am J Clin Nutr*. 2017;106(2):637-649. doi:10.3945/ajcn.116.150912
12. Haznadar M, Cai Q, Krausz KW, et al. Urinary metabolite risk biomarkers of lung cancer: a prospective cohort study. *Cancer Epidemiol Biomarkers Prev*. 2016;25(6):978-986. doi:10.1158/1055-9965.EPI-15-1191
13. Loft S, Svoboda P, Kasai H, et al. Prospective study of urinary excretion of 7-methylguanine and the risk of lung cancer: effect modification by mu class glutathione-S-transferases. *Int J Cancer*. 2007;121(7):1579-1584. doi:10.1002/ijc.22863
14. Zheng W, Chow WH, Yang G, et al. The Shanghai Women's Health Study: rationale, study design, and baseline characteristics. *Am J Epidemiol*. 2005;162(11):1123-1131. doi:10.1093/aje/kwi322
15. Wu QJ, Xiang YB, Yang G, et al. Vitamin E intake and the lung cancer risk among female nonsmokers: a report from the Shanghai Women's Health Study. *Int J Cancer*. 2015;136(3):610-617.
16. Dona AC, Jiménez B, Schäfer H, et al. Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. *Anal Chem*. 2014;86(19):9887-9894. doi:10.1021/ac5025039
17. Lewis MR, Pearce JT, Spagou K, et al. Development and application of ultra-performance liquid chromatography-TOF MS for precision large scale urinary metabolic phenotyping. *Anal Chem*. 2016;88(18):9004-9013. doi:10.1021/acs.analchem.6b01481
18. Wijeyesekera A, Clarke PA, Bictash M, et al. Quantitative UPLC-MS/MS analysis of the gut microbial co-metabolites phenylacetylglutamine, 4-cresyl sulphate and hippurate in human urine: INTERMAP Study. *Anal Methods*. 2012;4(1):65-72. doi:10.1039/C1AY05427A
19. Swann JR, Spagou K, Lewis M, et al. Microbial-mammalian cometabolites dominate the age-associated urinary metabolic phenotype in Taiwanese and American populations. *J Proteome Res*. 2013;12(7):3166-3180. doi:10.1021/pr4000152
20. Broeckling CD, Heuberger AL, Prenni JE. Large scale non-targeted metabolomic profiling of serum by ultra performance liquid chromatography-mass spectrometry (UPLC-MS). *J Vis Exp*. 2013;(73):e50242.
21. Want EJ, Masson P, Michopoulos F, et al. Global metabolic profiling of animal and human tissues via UPLC-MS. *Nat Protoc*. 2013;8(1):17-32. doi:10.1038/nprot.2012.135
22. Veselkov KA, Lindon JC, Ebbels TM, et al. Recursive segment-wise peak alignment of biological (1)h NMR spectra for improved metabolic biomarker recovery. *Anal Chem*. 2009;81(1):56-66. doi:10.1021/ac8011544

23. Blaise BJ, Shintu L, Elena B, Emsley L, Dumas ME, Toulhoat P. Statistical recoupling prior to significance testing in nuclear magnetic resonance based metabolomics. *Anal Chem*. 2009;81(15):6242-6251. doi:10.1021/ac9007754
24. Yang G, Shu XO, Chow WH, et al. Soy food intake and risk of lung cancer: evidence from the Shanghai Women's Health Study and a meta-analysis. *Am J Epidemiol*. 2012;176(10):846-855. doi:10.1093/aje/kws168
25. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R package for causal mediation analysis. *J Stat Softw*. 2014;59(5). doi:10.18637/jss.v059.i05
26. Benjamini Y, Hochberg Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;57(1):289-300.
27. Li S, Park Y, Duraisingham S, et al. Predicting network activity from high throughput metabolomics. *PLoS Comput Biol*. 2013;9(7):e1003123. doi:10.1371/journal.pcbi.1003123
28. Wendlinger C, Vetter W. High concentrations of furan fatty acids in organic butter samples from the German market. *J Agric Food Chem*. 2014;62(34):8740-8744. doi:10.1021/jf502975b
29. Wakimoto T, Kondo H, Nii H, et al. Furan fatty acid as an anti-inflammatory component from the green-lipped mussel *Perna canaliculus*. *Proc Natl Acad Sci U S A*. 2011;108(42):17533-17537. doi:10.1073/pnas.1110577108
30. Lemke RA, Peterson AC, Ziegelhoffer EC, et al. Synthesis and scavenging role of furan fatty acids. *Proc Natl Acad Sci U S A*. 2014;111(33):E3450-E3457. doi:10.1073/pnas.1405520111
31. Guseva NV, Rokhlin OW, Glover RA, Cohen MB. TOFA (5-tetradecyl-oxy-2-furoic acid) reduces fatty acid synthesis, inhibits expression of AR, neuropilin-1 and Mcl-1 and kills prostate cancer cells independent of p53 status. *Cancer Biol Ther*. 2011;12(1):80-85. doi:10.4161/cbt.12.1.15721
32. Mason P, Liang B, Li L, et al. SCD1 inhibition causes cancer cell death by depleting mono-unsaturated fatty acids. *PLoS One*. 2012;7(3):e33823. doi:10.1371/journal.pone.0033823
33. Morehouse KM, Nyman PJ, McNeal TP, Dinovi MJ, Perfetti GA. Survey of furan in heat processed foods by headspace gas chromatography/mass spectrometry and estimated adult exposure. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess*. 2008;25(3):259-264. doi:10.1080/02652030701552949
34. Franke AA, Custer LJ, Wang W, Shi CY. HPLC analysis of isoflavonoids and other phenolic agents from foods and from human fluids. *Proc Soc Exp Biol Med*. 1998;217(3):263-273. doi:10.3181/00379727-217-44231
35. Djuric Z, Chen G, Doerge DR, Heilbrun LK, Kucuk O. Effect of soy isoflavone supplementation on markers of oxidative stress in men and women. *Cancer Lett*. 2001;172(1):1-6. doi:10.1016/S0304-3835(01)00627-9
36. Yu J, Bi X, Yu B, Chen D. Isoflavones: anti-inflammatory benefit and possible caveats. *Nutrients*. 2016;8(6):E361. doi:10.3390/nu8060361
37. Yoon GA, Park S. Antioxidant action of soy isoflavones on oxidative stress and antioxidant enzyme activities in exercised rats. *Nutr Res Pract*. 2014;8(6):618-624. doi:10.4162/nrp.2014.8.6.618
38. Ducker GS, Rabinowitz JD. One-carbon metabolism in health and disease. *Cell Metab*. 2017;25(1):27-42. doi:10.1016/j.cmet.2016.08.009
39. Ladva CN, Golan R, Liang D, et al. Particulate metal exposures induce plasma metabolome changes in a commuter panel study. *PLoS One*. 2018;13(9):e0203468. doi:10.1371/journal.pone.0203468
40. Wang Z, Zheng Y, Zhao B, et al. Human metabolic responses to chronic environmental polycyclic aromatic hydrocarbon exposure by a metabolomic approach. *J Proteome Res*. 2015;14(6):2583-2593. doi:10.1021/acs.jproteome.5b00134
41. Donaldson K, Stone V, Seaton A, MacNee W. Ambient particle inhalation and the cardiovascular system: potential mechanisms. *Environ Health Perspect*. 2001;109(suppl 4):523-527.
42. Bates JT, Weber RJ, Abrams J, et al. Reactive oxygen species generation linked to sources of atmospheric particulate matter and cardiorespiratory effects. *Environ Sci Technol*. 2015;49(22):13605-13612. doi:10.1021/acs.est.5b02967
43. Chuang KJ, Chan CC, Su TC, Lee CT, Tang CS. The effect of urban air pollution on inflammation, oxidative stress, coagulation, and autonomic dysfunction in young adults. *Am J Respir Crit Care Med*. 2007;176(4):370-376. doi:10.1164/rccm.200611-1627OC
44. Sørensen M, Daneshvar B, Hansen M, et al. Personal PM2.5 exposure and markers of oxidative stress in blood. *Environ Health Perspect*. 2003;111(2):161-166. doi:10.1289/ehp.111-1241344

#### SUPPLEMENT.

**eTable 1.** Spearman Correlations Between Significant Metabolites (<10% FDR) in Controls Only, Cases Only and Overall Study Population

**eTable 2.** Associations Between Urinary Metabolites and Lung Cancer Risk, Stratified by Lung Cancer Subtypes

**eTable 3.** Associations Between the Most Significant Metabolite (5-Methyl-2-Furoic Acid) and Lung Cancer in Tertiles, Stratified by Lung Cancer Subtypes

**eTable 4.** Identification of the Top Three Metabolites

**eTable 5.** Associations Between neg\_2.60\_369.0408mz, pos\_2.61\_184.0325n and Lung Cancer, by Tertiles

**eTable 6.** Univariate Linear Regression and Correlation Between the Most Significant Metabolite (5-Methyl-2-Furoic Acid) and Diet Variables Among Controls