







## Article

# HSA Adductomics in the Shanghai Women's Health Study Links Lung Cancer in Never-Smokers with Air Pollution, Redox Biology, and One-Carbon Metabolism

Partow Imani <sup>1,†</sup>, Hasmik Grigoryan <sup>1,†</sup>, Sandrine Dudoit <sup>2</sup>, Xiao-Ou Shu <sup>3</sup>, Jason Wong <sup>4</sup>, Luoping Zhang <sup>1</sup>, Junfeng Zhang <sup>5,6</sup>, Wei Hu <sup>7</sup>, Qiuyin Cai <sup>3</sup>, Yutang Gao <sup>8</sup>, Batel Blechter <sup>7</sup>, Mohammad Rahman <sup>7</sup>, Wei Zheng <sup>3</sup>, Nathaniel Rothman <sup>7</sup>, Qing Lan <sup>7,‡</sup> and Stephen M. Rappaport <sup>1,\*,‡</sup>

<sup>1</sup> School of Public Health, University of California, Berkeley, Berkeley, CA 94720, USA; pimani@berkeley.edu (P.I.); hasgrig@yahoo.com (H.G.); luoping@berkeley.edu (L.Z.)

<sup>2</sup> Department of Statistics, University of California, Berkeley, Berkeley, CA 94720, USA; sandrine@stat.berkeley.edu

<sup>3</sup> Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN 37232, USA; xiao-ou.shu@vanderbilt.edu (X.-O.S.); qiuyin.cai@vumc.org (Q.C.); wei.zheng@vumc.org (W.Z.)

<sup>4</sup> Epidemiology and Community Health Branch, National Heart Lung and Blood Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892, USA; jason.wong@nih.gov

<sup>5</sup> Global Health Research Center, Duke Kunshan University, Kunshan 215316, China; junfeng.zhang@duke.edu

<sup>6</sup> Nicholas School of the Environment, Global Health Institute, Duke University, Durham, NC 27708, USA

<sup>7</sup> Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892, USA; wei.hu@nih.gov (W.H.); batel.blechter@nih.gov (B.B.); mohammad.rahman2@nih.gov (M.R.); rothmann@exchange.nih.gov (N.R.); qingl@mail.nih.gov (Q.L.)

<sup>8</sup> Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai 200025, China; ytgao@vip.sina.com

\* Correspondence: srappaport@berkeley.edu; Tel.: +1-(510)-334-8128

† These authors contributed equally to this work as co-first authors.

‡ These authors contributed equally to this work as co-senior authors.



Academic Editor: Giancarlo Aldini

Received: 8 February 2025

Revised: 5 March 2025

Accepted: 7 March 2025

Published: 13 March 2025

**Citation:** Imani, P.; Grigoryan, H.; Dudoit, S.; Shu, X.-O.; Wong, J.; Zhang, L.; Zhang, J.; Hu, W.; Cai, Q.; Gao, Y.; et al. HSA Adductomics in the Shanghai Women's Health Study Links Lung Cancer in Never-Smokers with Air Pollution, Redox Biology, and One-Carbon Metabolism. *Antioxidants* **2025**, *14*, 335. <https://doi.org/10.3390/antiox14030335>

**Copyright:** © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Nearly one fourth of lung cancers occur among never-smokers and are predominantly lung adenocarcinomas (LUADs) that are distinct from smoking-related cancers. Causal links between LUADs in never-smokers have been attributed to reactive oxygen species (ROS) arising from airborne fine particulate matter (PM<sub>2.5</sub>) and polycyclic aromatic hydrocarbons (PAHs). These effects are pronounced among East Asian women who experience massive exposures to PM<sub>2.5</sub> and PAHs and have the highest incidence of LUADs in the world. We employed untargeted adductomics to establish ROS adduct signatures in human serum albumin (HSA) from lung cancer cases and controls from never-smokers in the Shanghai Women's Health Study. Forty-seven HSA adducts were quantified by mass spectrometry, nine of which were selected for association with lung cancer, including Cys34 sulfoxidation products and disulfides of cysteine and homocysteine and two modifications to Lys525. Associated adducts include constituents of redox biology and one-carbon metabolism (OCM), which are pathways associated with lung cancer. Differences in adduct abundance between cases and controls and correlations of adducts with urinary PAHs and dietary factors provide additional evidence linking air pollutants, OCM, and redox biology with lung cancer in never-smokers.

**Keywords:** lung cancer; never-smokers; human serum albumin (HSA) adducts; cysteine 34 (Cys34); lysine 525 (Lys525); reactive oxygen species (ROS); polycyclic aromatic hydrocarbons (PAHs); fine particulate matter (PM<sub>2.5</sub>)

## 1. Introduction

Lung cancer is the leading cause of cancer mortality worldwide, accounting for 1.82 million deaths in 2022 (<https://gco.iarc.fr/today/en/fact-sheets-cancers>, accessed on 3 March 2024). Although most lung cancers are caused by tobacco smoking, nearly 25% occur among never-smokers and are predominately lung adenocarcinomas (LUADs) that are distinct from smoking-related cancers both histologically and epidemiologically [1]. In never-smokers, LUADs often contain oncogenic *EGFR* mutations, particularly among East Asian women (including those residing in the United States) that differ from those of never-smoking women in other geographic regions [1,2]. This implicates interactions across genetic factors, *EGFR* mutations and environmental effects as potential causes of lung cancer in never-smokers. Among never-smoking Chinese women, environmental risk factors for lung cancer include environmental tobacco smoke (ETS, secondhand smoke) [3], exposure to air pollutants [1,4] including cooking fumes [5], and dietary variables [6,7].

The International Agency for Research on Cancer (IARC) has declared air pollution to be carcinogenic based on evidence that combinations of pollutants and genetic factors cause lung cancers [8]. Across the plethora of air pollutants, fine particulate matter less than 2.5  $\mu\text{m}$  in diameter ( $\text{PM}_{2.5}$ ) is regarded as a primary risk factor for LUADs in never-smokers [9]. Current exposures to  $\text{PM}_{2.5}$  in urban China are among the highest in the world with average  $\text{PM}_{2.5}$  concentrations in Shanghai exceeding 80  $\mu\text{g}/\text{m}^3$ , or roughly 20 times the World Health Organization (WHO) health guideline of 5  $\mu\text{g}/\text{m}^3$  [1]. Following deposition in the deep lung,  $\text{PM}_{2.5}$  mobilizes inflammatory cells (alveolar macrophages and polymorphonuclear neutrophils) that generate reactive oxygen species (ROS) as a defense to kill invading microorganisms [10,11]. The same inflammatory response to nonviable  $\text{PM}_{2.5}$ —including soot, fibers, transition metals, and cooking fumes—damages pulmonary epithelial cells and releases an array of cytokines including interleukin-1 $\beta$  (IL-1 $\beta$ ) [10].

Based on an elegant set of experiments in which genetically engineered mice were induced with oncogenic human *EGFR*<sup>L858R</sup> and then exposed to relevant doses of  $\text{PM}_{2.5}$ , Hill et al. demonstrated that macrophages are a primary source of IL-1 $\beta$  and that the signaling of this cytokine promotes *EGFR*-driven LUADs [9]. Then, using aggregate data from a set of epidemiological databases, including never-smokers, the authors reported an association between “... the frequency of *EGFR* mutant lung cancer incidence and increasing  $\text{PM}_{2.5}$  levels”. Based on these results and the literature, they concluded that the “... data suggest a mechanistic and causative link between air pollutants and lung cancer...”, with  $\text{PM}_{2.5}$  being an important tumor promoter. However, because tumors in mice exposed to  $\text{PM}_{2.5}$  did not contain increased numbers of *EGFR* mutant cells, they inferred that LUADs must be initiated by mutagenic co-exposures.

Prominent mutagenic exposures to air pollutants involve the ubiquitous mixture of polycyclic aromatic hydrocarbons (PAHs) adsorbed on soot particles arising from combustion processes as well as aerosols generated from the pyrolysis of cooking oils (both subsets of  $\text{PM}_{2.5}$ ) [12,13]. PAHs can cause cancer through pathways involving metabolism to mutagenic diol epoxides that form bulky DNA adducts and initiate tumors and to pairs of PAH catechols and quinones, metabolites that undergo redox cycling with the generation of tumor-promoting ROS [14,15]. Diesel engine exhaust (DEE) contains nanoparticles (diameters < 0.1  $\mu\text{m}$ , also a subset of  $\text{PM}_{2.5}$ ) with adsorbed PAHs including mutagenic nitro-PAHs [16].

The hypothesis that exposure to  $\text{PM}_{2.5}$  and PAHs can cause lung cancer in never-smokers is supported by studies on inhabitants in Xuanwei, China, who were exposed to effluents from the indoor combustion of bituminous (“smoky”) coal, which contain massive concentrations of  $\text{PM}_{2.5}$  and PAHs [17]. Although women in Xuanwei are primarily nonsmokers, they nonetheless have a very high incidence of lung cancer [18] with LUAD

tumors that harbor *EGFR* mutations [4,19,20]. Furthermore, Ho et al. conducted in vitro experiments exposing human LUAD cells, with and without *EGFR* mutations, to samples of PM<sub>2.5</sub> from indoor coal combustion in Xuanwei and found a dose-related reduction in the viability of cells containing *EGFR-TKI* mutations [21]. Finally, Lan et al. reported that a single-nucleotide polymorphism in the *AKR1C3-Gln/Gln* genotype, which regulates the production of PAH catechols/quinones that generate ROS, was associated with increased incidence of LUADs in Xuanwei, particularly in never-smoking females [22].

In addition to the production of ROS by inflammatory cells, ROS are generated by nonphagocytic cells as signaling molecules for the oxidation and regulation of target proteins that promote critical cellular functions [23]. Under homeostasis, ROS are removed through reactions with cellular and extracellular antioxidants, of which low-molecular-weight thiols and protein thiols form the basis of redox biology [23]. When ROS production exceeds the body's homeostatic defenses, a state of oxidative stress exists, which can be mitigated by antioxidants derived from cellular processes and diet. Indeed, there is evidence in both smokers and never-smokers for inverse associations between lung cancer and increased fruit and vegetable intake [24,25]. Increased dietary soy, which contains isoflavones with strong anti-inflammatory and antioxidant properties [26], was associated with reduced risk of lung cancer in never-smokers from the Shanghai Women's Health Study (SWHS) [6]. An untargeted urinary metabolomics study of never-smokers in this population identified biological pathways associated with lung cancer, including oxidative stress, inflammation, and one-carbon metabolism (OCM) [27].

One-carbon metabolism represents a network of pathways involving folate, several B vitamins, homocysteine (hCys), and methionine and its derivative S-adenosylmethionine (SAM), a universal methyl donor for DNA and other molecules [28]. The methylation of DNA drives gene expression and the methionine-hCys cycle leads to the production of antioxidants including cysteine (Cys) and glutathione (GSH) [28]. Given the dynamic nature of OCM, various co-dependent molecules can become imbalanced, thereby altering redox homeostasis and the expression of genes that modulate the suppression and progression of cancers [19]. Pathways involving folate and methionine-hCys cycles were reported to reduce risks in lung cancer cases from the European Prospective Investigation into Cancer and Nutrition (EPIC) [29], and meta-analyses found that increased levels of folate and methionine in urine or blood were associated with reduced lung cancer risks [19,30].

The above evidence points to ROS as key intermediates in lung cancer etiology among never-smokers that involve the interplay between air pollutants, antioxidants, and redox biology. This motivates research to compare levels of ROS between never-smoking lung cancer cases and controls in prospective cohorts. Although ROS cannot readily be measured in vivo, blood levels can be inferred from their stable adducts in human serum albumin (HSA), which is the primary scavenger of ROS in the interstitial space. In particular, the Cys residue (Cys34) of the third largest tryptic (T3) peptide of HSA (<sup>21</sup>ALVLIAFAQYLQQC<sup>34</sup>PFEDHVK<sup>41</sup>, *m/z* 811.7593) is a nucleophilic locus that accounts for most of the antioxidant capacity [31]. Two Cys34 oxidation products (the sulfinic and sulfonic acids) are recognized biomarkers of oxidative stress [23,32], and Watanabe et al. investigated the clinical significance of the redox state of Cys34 to clinical findings involving inflammatory diseases of the liver and kidney and diabetes mellitus [33]. Furthermore, measuring adducts of the ε-amino group of the lysine residue Lys525 in another HSA peptide (<sup>525</sup>KQTALVELVK<sup>534</sup>, *m/z* 564.8529) expands coverage to different classes of adducts including Schiff bases and advanced glycation end products [34].

The aim of this study was to determine whether alterations to the Cys34/Lys525 adductome is associated with lung cancer risk in never-smokers from the SWHS, using plasma collected up to 14 years prior to diagnosis. By considering correlations among

Cys34/Lys525 adduct features (hereafter simply called ‘features’) and covariates including biomarkers of PM<sub>2.5</sub> and PAH exposure and dietary factors associated with lung cancer, we sought insight from this unique vantage point into the roles played by air pollution, oxidative stress, and OCM pathways in lung carcinogenesis.

## 2. Materials and Methods

### 2.1. Study Population

The SWHS is a prospective population-based cohort study on 74,942 women living in urban Shanghai, China, who were enrolled between the ages of 40 and 70 years in December 1996–May 2000 [35]. Within this larger prospective cohort study, we conducted a nested case/control study of 573 never-smoking women with available plasma samples including 278 incident lung cancer cases and 293 cancer-free controls. Controls were individually matched to cases on date of birth ( $\pm 2$  years) and date of sample collection ( $\pm 3$  months). Procedures for biospecimen collection and storage were previously described in detail. Participants completed a baseline survey, which collected information on demographics and anthropometric and lifestyle factors; dietary habits were gleaned from a food frequency questionnaire with a 7-day recall preceding blood collection.

### 2.2. Lung Cancer Ascertainment

The study participants were followed-up with using in-person surveys every 2–3 years and periodic linkage to the population-based Shanghai Cancer Registry, Shanghai Vital Statistics Registry, and Shanghai Resident Registry. In-home visits and a review of medical records by a panel of oncologists were used to verify all possible lung cancer diagnoses. Lung cancer was defined as code 162 under the *International Classification of Diseases, Ninth Revision (ICD-9)*. All female lung cancer patients were captured upon diagnosis after recruitment through 31 December 2009. Lung cancer cases diagnosed before recruitment were excluded from the study. Data on lung cancer histological subtypes were extracted from medical records and classified according to code 8140/3 from the *International Classification of Diseases for Oncology, Second Edition (ICD-O-2)*. Of the 278 incident lung cancer cases, we excluded from statistical analyses 72 cases that had not been histologically confirmed (HC), leaving 206 HC cases, of which 148 were either LUADs (133) or adenosquamous carcinomas (18) (combined hereafter as ‘LUADs’). The remaining HC cases included small-cell carcinomas (2), ‘other’ (3), ‘unclassified’ (22), and ‘missing ICD-O code’ (22).

### 2.3. Mass Spectrometry and Data Acquisition

Plasma samples were maintained at  $-80$  °C for up to 14 years prior to analysis. Case/control pairs were analyzed randomly on the same day to reduce technical variation. Samples were processed and analyzed by nano-liquid chromatography–tandem mass spectrometry (nLC-HRMSMS) with duplicate injections, as detailed in Supplementary Method S1. After the tryptic digestion of HSA, an isotopically labeled T3 peptide modified with iodoacetamide (IAA-iT3) was added to normalize data for instrument performance. One microliter of the digest was then analyzed by nLC-HRMSMS to pinpoint Cys34 and Lys525 peptides. Precursor ions were extracted for monoisotopic peptide masses (MIMs) to assign accurate masses and elemental compositions. Peak areas of MIMs were normalized for HSA content by extracting the MIM of the doubly charged HSA peptide <sup>42</sup>LVNEVTEFAK<sup>51</sup> (‘housekeeping peptide’,  $m/z = 575.3111$ ). Peaks were selected and integrated using average MIMs and retention times. Added masses relative to the Cys34 thiolate ion and the Lys525 peptide were estimated, and annotations were based on accurate mass, elemental composition, retention time, database searches, and a few reference standards.

Several urinary metabolites of exogenous PAHs and nitro-PAHs were measured by LC-MSMS, as described in Supplementary Method S2, for use as covariates in statistical analyses. Two composite sets of analytes were generated, namely 'hydroxy-PAHs' (sum of 2-, 3-, and 4-hydroxyphenanthrene, 1,9-dihydroxyphenanthrene, and 10-hydroxypyrene), which are metabolites of unmodified PAHs, and 'amino-PAHs' (sum of 9-aminophenanthrene, 3-aminobenzanthracene, 1- and 2-aminonaphthalene, 2-aminofluorene, and 1-aminopyrene), which are metabolites of nitro-PAHs, arising primarily from diesel emissions [16,36].

#### 2.4. Statistical Analysis

Mass spectrometry data for 34 Cys34-containing features and 13 Lys525-containing features were available from 573 never-smoking subjects with fewer than 20 percent of missing observations. Duplicate injections were available for all but for two individuals who had single injections ( $n = 1144$ ). Statistical analysis was conducted in the R environment (<https://www.R-project.org/>, version 4.2.2) and follows methods described previously [37]. Using a random effects model with the log feature abundance as the outcome, the subject-specific random effect and the random error term (from duplicate injections) were used to estimate the intraclass correlation coefficient (ICC) for each feature representing the proportion of variance that was not attributed to technical variation. As none of the ICC values were below an a priori cutoff of 0.2, no features were excluded. Duplicate injections were averaged while ignoring missing values, and features were sorted across subjects by the percentage of missing values. Three features with more than 30% missing values were removed after using Fisher's exact test to check for differentially missing values by case/control status; this left 44 features for subsequent analysis. Any values that were still missing were imputed using the k-Nearest Neighbor algorithm with  $k$  equal to 4. The *scone* package (<https://rdrr.io/bioc/scone/>, accessed on 6 June 2023, R package version 1.22.0) was then used to evaluate various candidate normalization schemes based on several criteria measuring how well technical noise was removed while maintaining the signal of interest. The inputs to *scone* were the feature abundance matrix, a batch variable, the QC matrix (comprising the house-keeping peptide, internal standard, sample age, and run order), case-control status, and several candidate scaling methods [DESeq, Trimmed Mean of M-values, upper-quartile, and no scaling]. Based on *scone* results, the selected normalization procedure adjusted for all QC measures as well as batch and case/control status.

Following normalization, separate analyses were performed for comparing all HC cases to all controls and only LUAD cases (a subset of the HC cases) to all controls. A previously described ensemble method was applied to select the features that appeared to have the strongest association with the outcome for either HC cases or LUAD cases [37]. First, linear regression was used where each logged feature abundance was regressed on case/control status, batch, QC measures, and age, and the estimated regression coefficient corresponding to the case/control variable as well as the associated nominal  $p$ -value from the  $t$ -test were retained. Next, regularized logistic regression (LASSO) was used with the case/control status as the outcome and all normalized adducts and age as predictors. This was performed over 500 bootstrapped iterations, and the features were ranked based on the number of times their corresponding coefficients were non-zero. Features that ranked highly according to both nominal  $p$ -values from the linear regression and the bootstrapped LASSO were selected. Finally, the same inputs were given to Random Forests, and any additional features that appeared important, based on the mean decrease in Gini Index, were selected, thereby allowing for the detection of nonlinear relationships. Hierarchical clusterings of features using pairwise Spearman correlation were visualized with the R-package 'Superheat' (<https://rlbarter.github.io/superheat/>, accessed on 6 June 2023,



R-package version 1.0.0). Heatmaps were similarly used to visualize feature correlations with various dietary factors and urinary PAH biomarkers as well as age, body mass index (BMI), and education level, and these variables were given to Random Forests to examine relative importance as predictors of case/control status.

Permutation tests were used to evaluate possible associations between relative feature abundance (case/matched control) and time from recruitment to diagnosis (ttd) to discriminate potentially causal effects from effects of disease progression (reverse causality) [37]. For each of 10,000 random permutations of the vector of the ttd (thereby breaking any association with relative case/control feature abundance), the estimated regression coefficient of case/control log-fold change (LFC) on the ttd was recorded. The *p*-value reported is the proportion of the 10,000 random permutations in which the absolute value of the estimated regression coefficient was greater than or equal to the absolute value estimated using the observed data. If a significant linear trend (either positive or negative) in LFC for a given feature was detected with increasing ttd, the feature was considered potentially reactive. The permutation test for each feature used its respective set of HC or LUAD cases and controls.

### 3. Results

Summary statistics are given in Table 1 comparing lung cancer cases and controls for covariates previously associated with lung cancer risk. Evidence of a significant inverse association between lung cancer and BMI was observed (*p*-value < 0.05), consistent with meta-analyses [38]. Soy consumption, age, and education level were also significantly lower in cases than controls. Table 1 also summarizes the ttd (years from recruitment).

**Table 1.** Summary statistics for covariates included with histologically confirmed (HC) cases, LUAD (lung adenocarcinoma) cases and controls.

Variable		HC Cases (n = 206)	LUAD Cases (n = 148)	Controls (n = 293)		<i>p</i> -Value (HC) <sup>1</sup>	<i>p</i> -Value (LUAD) <sup>1</sup>
BMI (kg/m <sup>2</sup> )	Min	16.51	16.59	Min	16.45	0.001 *	0.005 *
	Mean	23.82	23.82	Mean	24.90		
	Median	23.82	23.90	Median	24.44		
	Max	35.42	35.42	Max	36.18		
	Not available	0	0	Not available	1		
Amino-PAHs <sup>2</sup> (pg/mL)	Min	555.8	591.6	Min	288.1	0.226	0.449
	Mean	4783	4949	Mean	3916		
	Median	3544	3508	Median	2951		
	Max	57,121	57,121	Max	16,024		
	Not available	116	85	Not available	178		
Hydroxy-PAHs <sup>3</sup> (pg/mL)	Min	95.87	95.87	Min	105.5	0.820	0.840
	Mean	1040	1010	Mean	978.0		
	Median	872.0	854.4	Median	799.9		
	Max	3914	3914	Max	3578		
	Not available	126	95	Not available	189		
Education	College	30	21	College	34	0.024 *	0.003 *
	High School	60	50	High School	60		
	Middle School	57	39	Middle School	80		
	Elementary	59	38	Elementary	119		
Age (years)	Min	40	40	Min	40	0.002 *	0.0003 *
	Mean	56.2	55.5	Mean	58.5		
	Median	57	56.5	Median	62		
	Max	70	70	Max	70		
All Soy Consumption (g/d)	Min	6.29	14.42	Min	2.28	0.014 *	0.270
	Mean	211.3	236.8	Mean	249.6		
	Median	89.10	100.3	Median	151.3		
	Max	1099	1099	Max	1277		

Table 1. Cont.

Variable		HC Cases (n = 206)	LUAD Cases (n = 148)	Controls (n = 293)		p-Value (HC) <sup>1</sup>	p-Value (LUAD) <sup>1</sup>
All Vegetable Consumption (g/d)	Min	52.2	59.5	Min	23.8	0.487	0.075
	Mean	308.7	320.7	Mean	298.1		
	Median	265.0	294.7	Median	258.7		
	Max	1050	983.6	Max	878.6		
Methionine (g/d)	Min	0.414	0.414	Min	0.630	0.781	0.659
	Mean	1.501	1.522	Mean	1.505		
	Median	1.393	1.440	Median	1.454		
	Max	3.986	3.826	Max	3.167		
Folate (µg/d)	Min	45.3	45.3	Min	68.4	0.675	0.182
	Mean	263.6	269.8	Mean	264.3		
	Median	250.0	262.8	Median	244.0		
	Max	692.2	658.3	Max	1155		
ETS (Ever Exposed)	No	44	34	No	75	0.522	0.624
	Yes	141	133	Yes	204		
	Not available	21	15	Not available	14		
Time to Diagnosis (years)	Min	0.183	0.183				
	Mean	7.533	7.568				
	Median	8.355	8.475				
	Max	13.503	13.336				

\*  $p$ -value < 0.05. <sup>1</sup> Nominal  $p$ -values for testing for differences in covariate distributions between cases (HC or LUAD) and controls were computed using a chi-squared test for categorical variables and a Wilcoxon rank-sum test for continuous variables. <sup>2</sup> Sum of 9-aminophenanthrene, 3-aminobenzanthracene, 1- and 2-aminonaphthalene, 2-aminofluorene, and 1-aminopyrene. <sup>3</sup> Sum of urinary 2-, 3-, and 4-hydroxyphenanthrene, 1,9-dihydroxyphenanthrene, and 10-hydroxypyrene.

The 34 Cys34 and 13 Lys525 features are listed in Tables 2 and 3, respectively. For each feature, the following variables are shown: retention time, observed MIM, theoretical MIM, the corresponding mass deviation ( $\Delta$ mass), added mass, elemental composition, annotation, and mean peak area. Supplementary Table S1 shows the values of the ICC (median = 0.55; range: 0.25–0.93) and coefficient of variation (CV; median = 0.37; range: 0.18–1.4) based on the random effects model of duplicate nLC-HRMSMS analyses of feature abundances. The impact of technical variation in feature selection was reduced by averaging duplicate observations for each subject in the analyses. Supplementary Tables S2 and S3 show FCs, where the nominal  $p$ -values for testing whether the case/control coefficient from the multivariable regression are zero using a  $t$ -test for HC and LUAD cases, respectively, and all controls. In total, 34 of these 44 features have been reported previously in our laboratory (footnoted in Tables 2 and 3).

Supplementary Figures S1A–C and S2A–C show the results of feature selection for HC cases/controls and LUAD cases/controls, respectively, and Table 4 combines the results to list the nine selected features. These included seven Cys34 features, i.e., 822.42 (Cys34 sulfinic acid), 827.088 (S-methanethiol), 845.42 (S-Cys-H<sub>2</sub>O), 849.07 (S-S-sulfonic acid trisulfide), 856.10 (S-hCys), 858.75 (Na adduct of S-Cys), and 914.83 (unknown), and two Lys525 features, i.e., 571.84 (Lys oxidation product; -H<sub>2</sub>+O) and 587.31 (unknown). Most of the selected features were common to both HC and LUAD analyses, exceptions being 822.42 (only for HC) and 849.07, 858.75, and 827.088 (only for LUADs). Also, two of the five features from the overlapping list (587.31 and 849.07) had stronger signals in the LUAD analysis despite the smaller sample size (148 cases as opposed to 206).

**Table 2.** T3-peptide (mostly Cys34) features with sufficient data for statistical analysis. (Mean peak areas are based on normalized data after imputation for missing values.)

Feature	RT, min	MIM Observed ( <i>m/z</i> , +3)	MIM Theoret. ( <i>m/z</i> , +3)	$\Delta$ Mass (ppm)	Added Mass (+H, Da)	Elemental Composition	Annotation	Mean Peak Area ( $\times 10^6$ )
796.43 <sup>a,b,c,f,g,h</sup>	27.717	796.4298	796.4301	−0.3767	−45.983	−CH <sub>2</sub> S	Cys34→Gly	6.16
805.76 <sup>a,b,c,f,h</sup>	27.365	805.7614	805.7618	−0.4964	−17.989	−SH <sub>2</sub> , +O	Cys34→oxoalanine or fGly	0.182
808.73 <sup>a,b,c,d,e,f,g,h</sup>	27.868	808.7281			−9.088		Not Cys34 adduct	42.9
810.45 <sup>a,d,f</sup>	28.263	810.4521			−3.916		Not Cys34 adduct	0.168
811.42 <sup>a,b,c,d,e,f,g,h</sup>	30.803	811.4242	811.4234	0.9859	−1.000		T3 dimer *	3.45
811.76 <sup>a,b,c,d,e,f,g,h</sup>	28.682	811.7576	811.7594	−2.2174	0.000		Unmodified T3 *	9.25
816.42 <sup>a,b,c,d,e,f,g,h</sup>	27.390	816.4188	816.4191	−0.3675	13.984	−H <sub>2</sub> , +O	Cys34-Gln crosslink *	1.39
816.43 <sup>a,b,c,d,e,f,g,h</sup>	29.102	816.4306	816.4312	−0.7349	15.027	+CH <sub>3</sub>	Methylation (at Glu37)	3.94
819.09 <sup>b,e,g</sup>	28.514	819.0863			22.994		Na adduct of T3	0.469
822.42 <sup>a,b,c,d,e,f,g,h</sup>	27.207	822.4222	822.4226	−0.4864	33.002	+HO <sub>2</sub>	Cys34 sulfinic acid *	16.4
827.088 <sup>b,c,f,g,h</sup>	29.566	827.0883	827.0886	−0.3627	47.000	+CH <sub>3</sub> S	S-Methanethiol *	40.9
827.094 <sup>c,d,f,g,h</sup>	27.790	827.0942	827.0945	−0.3627	47.018	+CH <sub>3</sub> O <sub>2</sub>	S-(O)-O-CH <sub>3</sub>	1.23
827.75 <sup>a,b,c,d,e,f,g,h</sup>	27.361	827.7537	827.7543	−0.7249	48.996	+HO <sub>3</sub>	Cys34 Sulfonic acid *	2.08
829.396 <sup>a,b,g,h</sup>	28.575	829.3959			53.923		Not Cys34 adduct	0.321
835.11 <sup>a,c,e,g,f,h</sup>	28.809	835.1062	835.1066	−0.4790	71.054	+C <sub>4</sub> H <sub>7</sub> O	Crotonaldehyde *	1.58
841.75 <sup>a,b,c,d,e,f,g,h</sup>	28.411	841.7513	841.7519	−0.7128	90.989	+C <sub>2</sub> H <sub>3</sub> O <sub>2</sub> S	S-Mercaptoacetic acid	0.061
842.07	28.121	842.0734			91.955		Unknown	NA
845.42 <sup>a,b,c,d,e,f,g,h</sup>	27.383	845.4238	845.4239	−0.1183	102.006	+C <sub>3</sub> H <sub>4</sub> NOS	S-Cys (-H <sub>2</sub> O)	0.487
849.07 <sup>a,b,c,h</sup>	28.440	849.0684	849.0689	−0.5889	112.940	+HO <sub>3</sub> S <sub>2</sub>	S-Sulfonic acid trisulfide	0.880
851.43 <sup>a,b,c,d,e,f,g,h</sup>	26.980	851.4266	851.4274	−0.9396	120.015	+C <sub>3</sub> H <sub>6</sub> NO <sub>2</sub> S	S-Cys *	67.2
851.76 <sup>a,b,c,d,e,f,g</sup>	27.632	851.7554	851.7554	0.0000	121.000	+C <sub>3</sub> H <sub>5</sub> O <sub>3</sub> S	S-Cys (NH <sub>2</sub> →OH)	1.23
853.78 <sup>a,b,f</sup>	27.503	853.7823			127.082		Unknown	9.20
856.10 <sup>a,b,c,d,e,f,g,h</sup>	26.746	856.0988	856.0993	−0.5840	134.032	+C <sub>4</sub> H <sub>8</sub> NO <sub>2</sub> S	S-hCys *	200
858.75 <sup>a,b,c,d,e,g,h</sup>	26.391	858.7546			141.999	+C <sub>3</sub> H <sub>5</sub> NO <sub>2</sub> SNa	Na adduct of S-Cys	1.52
860.77 <sup>b,d,e,f,g</sup>	26.866	860.7703	860.7712	−1.0456	148.046	+C <sub>5</sub> H <sub>10</sub> NO <sub>2</sub> S	S-hCys (+CH <sub>3</sub> )	5.10
864.08 <sup>a,b,c,e,f,g</sup>	26.305	864.0763			157.964		Not Cys34 adduct	0.760
870.43 <sup>a,b,c,d,e,f,g,h</sup>	26.031	870.4342	870.4345	−0.3447	177.038	+C <sub>5</sub> H <sub>9</sub> N <sub>2</sub> O <sub>3</sub> S	S-CysGly *	0.751
894.44 <sup>a,b,c,d,e,f,g,h</sup>	26.684	894.4414	894.4416	−0.2236	249.059	+C <sub>8</sub> H <sub>13</sub> N <sub>2</sub> O <sub>5</sub> S	S-γ-GluCys *	0.092
910.18 <sup>g,h</sup>	32.954	910.1767			296.265	+C <sub>18</sub> H <sub>34</sub> NO <sub>2</sub>	Unknown	0.009
913.45 <sup>a,b,c,d,e,f,g,h</sup>	26.554	913.4487	913.4487	0.0000	306.081	+C <sub>10</sub> H <sub>16</sub> N <sub>3</sub> O <sub>6</sub> S	S-Glutathione *	0.250
914.84 <sup>h</sup>	32.931	914.8359			310.243		Unknown	0.552
931.82 <sup>b,c,d,f,g</sup>	24.921	931.8204			361.196		Unknown	0.685
965.49 <sup>b,c,d,e,f,h</sup>	25.207	965.4910			462.208		Unknown	2.24
970.16 <sup>d,f,h</sup>	25.473	970.1631			476.224		Unknown	NA

Legend: MIM, monoisotopic mass; NA, not available; RT, retention time; fGly, formyl glycine. \* Identity confirmed with a synthetic standard. <sup>a</sup> Reference [39]. <sup>b</sup> Reference [40]. <sup>c</sup> Reference [41]. <sup>d</sup> Reference [42]. <sup>e</sup> Reference [43]. <sup>f</sup> Reference [44]. <sup>g</sup> Reference [45]. <sup>h</sup> Reference [37].



**Table 3.** Lys525 peptide features with sufficient data for statistical analysis. (Mean peak areas are based on normalized data after imputation for missing values.)

Feature	RT, min	MIM Observed ( <i>m/z</i> , +3)	MIM Theoretical ( <i>m/z</i> , +3)	$\Delta$ Mass (ppm)	Added Mass (+H, Da)	Elemental Composition (+H)	Annotation	Mean Peak Area ( $\times 10^8$ )
500.81 <sup>a,b,c</sup>	12.063	500.8045	500.8055	−1.9968	−128.095	−C <sub>6</sub> H <sub>12</sub> N <sub>2</sub> O	Loss of Lysine	156
556.34 <sup>a,b,c</sup>	10.701	556.3396	556.3397	−0.1797	−17.024	−NH <sub>3</sub>	Loss of ammonia	0.037
556.84	10.995	556.8430	556.8436	−1.0775	−16.017	−NH <sub>2</sub>	Deamination	0.309
564.85 <sup>a,b,c</sup>	10.822	564.8518	564.8529	−1.8474	0.000		Lys525 containing peptide	7.96
566.77 <sup>a,c</sup>	10.684	566.7741			4.853		Unknown	0.014
571.84 <sup>c</sup>	11.074	571.8423	571.8426	−0.5246	14.989	−H <sub>2</sub> , +O	Lys525 oxidation product	3.87
577.86 <sup>a</sup>	12.454	577.8607	577.8608	−0.1731	27.026	+C <sub>2</sub> H <sub>3</sub>	Acetaldehyde	1.91
578.32	12.541	578.3164			27.937		Unknown	0.036
580.85 <sup>c</sup>	11.110	580.8476	580.8479	−0.5165	32.999	+O <sub>2</sub> H	Lys525 oxidation product	NA
586.36 <sup>a,b,c</sup>	13.265	586.3557	586.3559	−0.3411	44.016	+CH <sub>2</sub> NO	Carbamylation	9.59
587.31 <sup>c</sup>	13.362	587.3109			45.926		Unknown	1.01
645.88 <sup>a,b,c</sup>	10.904	645.8791	645.8794	−0.4645	163.063	+C <sub>6</sub> H <sub>11</sub> O <sub>5</sub>	Fructosyl lysine (glycation)/Hexose	18.4
647.34 <sup>c</sup>	10.778	647.3378			165.980		Unknown	1.03

Legend: MIM, monoisotopic mass; NA, not available; RT, retention time. <sup>a</sup> Reference [34]. <sup>b</sup> Reference [45]. <sup>c</sup> Reference [37].

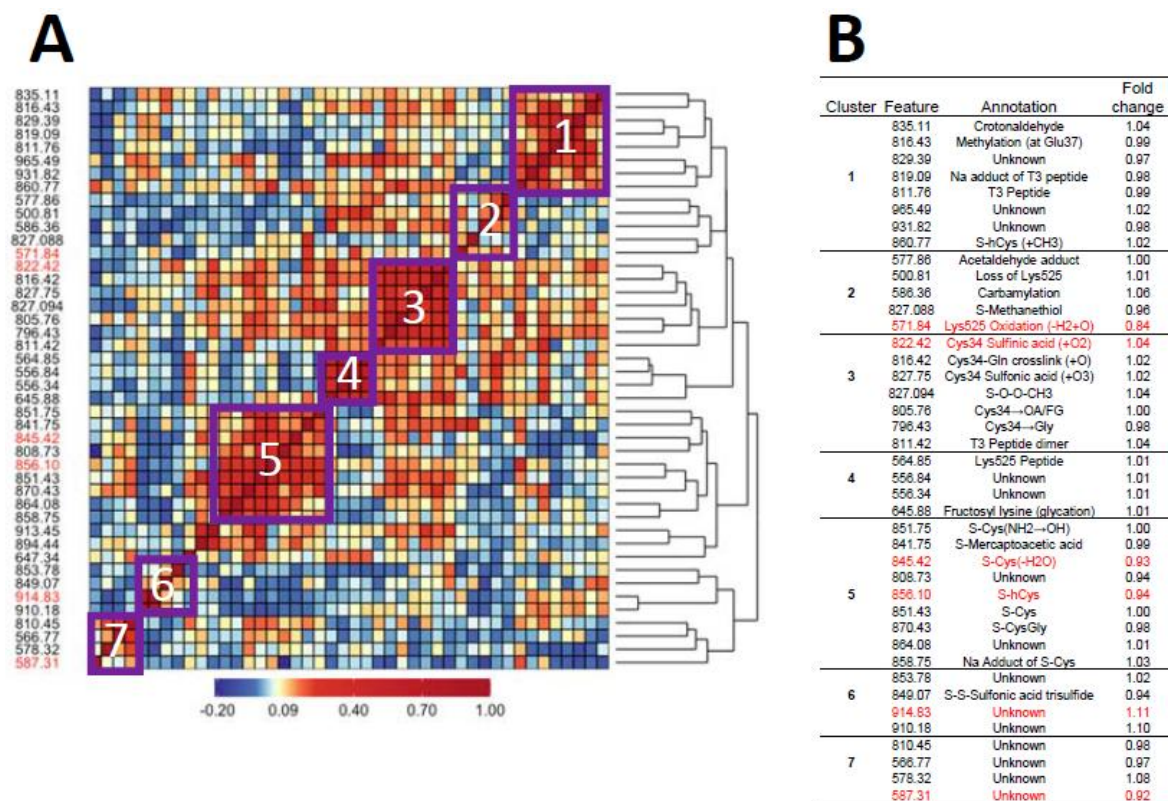
**Table 4.** Features selected by the ensemble pipeline for association with lung cancer. Selection criteria satisfied for each feature are shown in boldface. [Fold-change based on normalized and imputed data (FC); linear regression nominal *p*-value; bootstrap LASSO; Random Forests (RF)].

Feature	Annotation	Histologically Confirmed				LUAD			
		FC	<i>p</i> -Value	LASSO	RF	FC	<i>p</i> -Value	LASSO	RF
571.84	Lys525 oxidation (-H <sub>2</sub> +O)	0.842	<b>0.040</b>	<b>Yes</b>	<b>Yes</b>	0.823	<b>0.095</b>	<b>Yes</b>	<b>Yes</b>
587.31	Unknown	0.921	<b>0.081</b>	<b>Yes</b>	No	0.896	<b>0.040</b>	<b>Yes</b>	No
822.42	Cys34 sulfinic acid	1.042	<b>0.096</b>	<b>Yes</b>	<b>Yes</b>	1.018	0.504	No	No
827.088	S-Methanethiol	0.963	0.453	No	No	0.999	0.922	No	<b>Yes</b>
845.42	S-Cys(-H <sub>2</sub> O)	0.932	<b>0.014</b>	<b>Yes</b>	<b>Yes</b>	0.937	<b>0.047</b>	<b>Yes</b>	<b>Yes</b>
849.07	S-Sulfonic acid trisulfide	0.935	<b>0.079</b>	No	No	0.912	<b>0.028</b>	<b>Yes</b>	No
856.10	S-hCys	0.942	<b>0.044</b>	<b>Yes</b>	<b>Yes</b>	0.946	<b>0.087</b>	<b>Yes</b>	<b>Yes</b>
858.75	Na adduct of S-Cys	1.029	0.400	No	No	1.070	<b>0.087</b>	<b>Yes</b>	No
914.83	Unknown	1.114	<b>0.047</b>	<b>Yes</b>	<b>Yes</b>	1.113	<b>0.073</b>	<b>Yes</b>	<b>Yes</b>

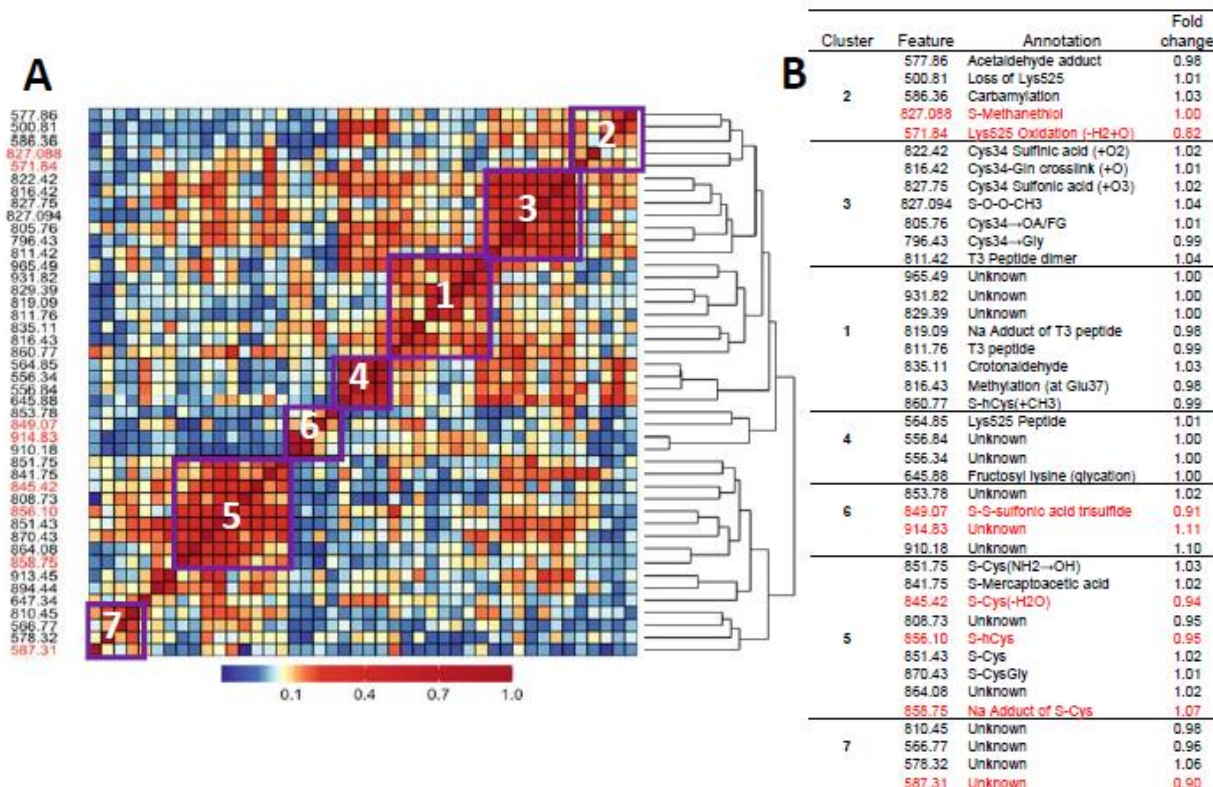
Figures 1A and 2A show heatmaps of pairwise Spearman correlation values for all features in the two analyses with selected features shown in red. Clusters labeled in the heatmaps are annotated in the embedded tables (Figures 1B and 2B). The seven clusters in Figures 1 and 2 between HC features and LUAD features, respectively, contain the same sets, although the relative locations differ in the heatmaps. This concordance of clusters again reflects the fact that LUADs represent 72 percent of HC cases.

Figures 3A and 4A show variable importance plots created using the Random Forests of the selected features and covariates from HC and LUAD analyses side-by-side with heatmaps of pairwise Spearman correlation values (Figures 3B and 4B). The BMI and methionine intake were the most important predictors of HC lung cancers followed closely by features 587.31, 914.83, and 822.42 (Figure 3A). The adjoining heatmap (Figure 3B) shows that all of the selected features except 914.83 were weakly to moderately correlated with urinary hydroxy-PAHs and/or amino-PAHs and that these correlations were negative in all cases. Feature 914.83 was negatively correlated with ‘all soy’ consumption, while 822.42 was negatively correlated with the entire dietary cluster (methionine, all soy, all vegetables and folate) as well as the ‘personal’ cluster representing education, age, and BMI. Turning now to the Random Forests plot for the features and covariates across LUADs in Figure 4A, BMI was the most predictive followed by 587.31, 849.07, and 571.84. As shown in Figure 4B, all selected features except 914.93 and 849.07 were weakly to moderately negatively correlated with one or both of the PAH classes. Also, the dietary cluster was positively correlated with 845.42, while 914.83 was negatively correlated with soy intake.

To determine whether potential associations of features with lung cancer may have resulted from reverse causality, we examined linear regressions between LFC of abundances of lung cancer case/control pairs and ttd. The results are presented in Supplementary Figures S3 and S4 as individual plots for the six selected features in HC cases/controls and eight selected features in LUAD cases/controls, respectively. There was no apparent linear trend for the ttd for any of the selected features in their respective subsets, indicating little evidence of reverse causality.

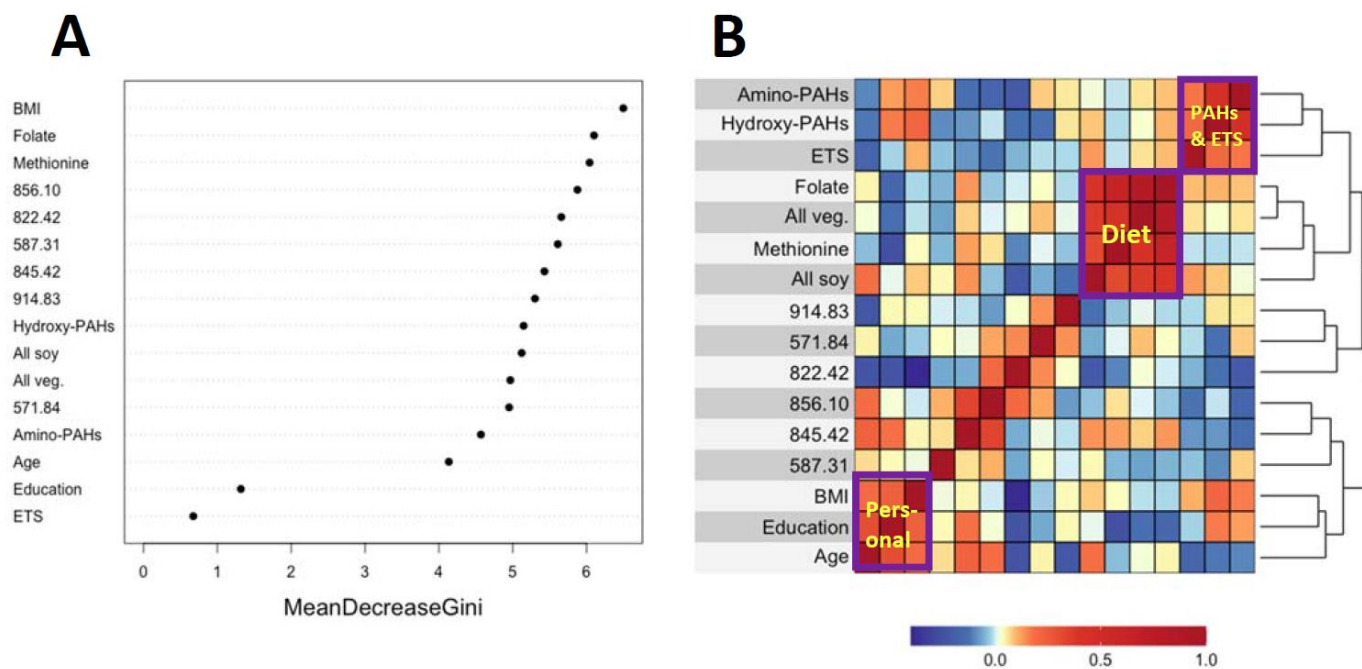


**Figure 1.** Heatmap of pairwise correlation matrix for all features in HC cases and controls (A) with annotations and case/control fold changes shown in (B). Features selected for association with lung cancer are highlighted in red.

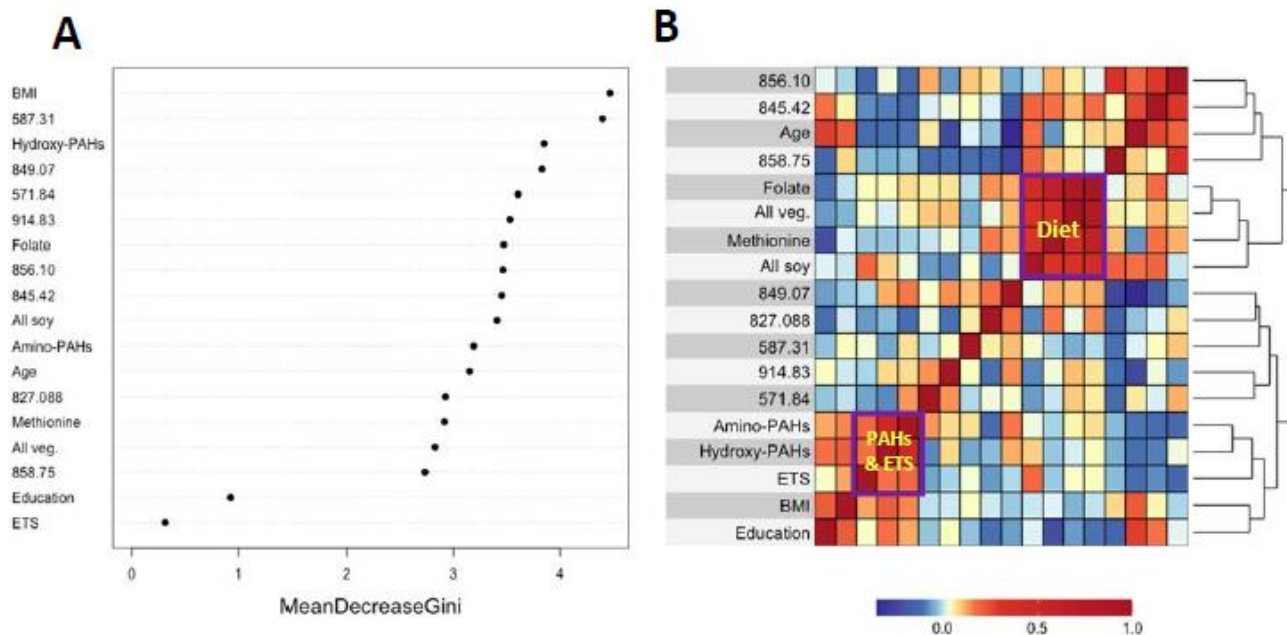


**Figure 2.** Heatmap of pairwise correlation matrix for all features in LUAD cases and controls (A) with annotations and case/control fold changes shown in (B). Features selected for association with lung cancer are highlighted in red.





**Figure 3.** Variable importance of covariates and selected features as ranked by Random Forests in HC cases and controls (A) and heatmap showing the corresponding pairwise correlations (B).



**Figure 4.** Variable importance of covariates and selected features as ranked by Random Forests in LUAD cases and controls (A) and heatmap showing the corresponding correlations (B).

### 4. Discussion

#### 4.1. Air Pollution and Lung Cancer in Never-Smokers

An increasing body of observational, mechanistic, and experimental evidence points to air pollution as a cause of lung cancer in never-smokers. A key element linking these lines of inquiry is the production of ROS by PM<sub>2.5</sub> and PAHs, two ubiquitous classes of air pollutants that collectively initiate and promote lung cancers. A missing component of this amalgamation of findings is direct evidence that the production and removal of ROS in lung cancer cases is imbalanced in never-smokers. The Cys34/Lys525 adductome is an emerging biomarker of exogenous exposures and endogenous processes that provides a

unique vantage point for linking adducts generated by ROS in vivo to lung cancer. Cys34 adducts of *N*-acetylcysteine, cysteinyl glycine (CysGly), and hCys were associated with lung cancer and smoking histories in the EPIC cohort [43] and in patients with chronic obstructive pulmonary disease (COPD, a risk factor for lung cancer) [42] but have not been investigated before in never-smoking lung cancer cases. In this study, we provide evidence that Cys34 adduct features do indeed link lung cancer in never-smokers with air pollution and altered redox biology and further implicate OCM as an important mechanistic pathway.

Our study focused on Chinese women because they have remarkably high lung-cancer incidence despite being predominately never-smokers. This has led to speculation that lung cancer risk across Chinese women reflects the confluence of high exposures to PM<sub>2.5</sub> and PAHs—arising from the indoor/outdoor combustion of fossil fuels and household frying with cooking oils—and their interactions with genetic factors, somatic events, and tumor biology [1]. A key component to this conjecture is chronic inflammation resulting from the imbalance between the production of ROS by phagocytic cells (in response to PM<sub>2.5</sub>) [9,10] plus the redox cycling of PAH catechols/quinones [14,15] and the removal of ROS by cellular and extracellular antioxidants [46].

#### 4.2. ROS Production and Removal

To investigate such imbalance between ROS production and removal in the SWHS, we monitored modifications to the Cys34/Lys525 adductome in 573 never-smoking women, including 278 incident lung cancer cases. Nine features were selected for association with lung cancer (Table 4) using a set of regression and classification methods, as applied previously with incident cases of non-Hodgkin's lymphoma (NHL) and matched controls from the EPIC cohort [37]. Interestingly, the abundances of seven of these nine features were lower in cases than controls ( $0.823 \leq FC \leq 0.946$ ), suggesting that enhanced ROS homeostasis in control subjects was protective for lung cancer.

Since seven of the nine selected features represent adducts to Cys34, we briefly differentiate between Cys34 sulfoxidation products and Cys34 disulfides with circulating thiols. When mercaptalbumin (HSA-Cys34-SH) encounters ROS, the sulfhydryl group is oxidized to its unstable sulfenic acid (Cys34-SOH), which can be further oxidized to sulfinic acid (Cys34-SO<sub>2</sub>H) and sulfonic acid (Cys34-SO<sub>3</sub>H). These sequential oxidations are largely irreversible and represent biomarkers of protein damage arising from oxidative stress. Indeed, the loss of the antioxidant capacity of mercaptalbumin due to the oxidation of Cys34 is a harbinger of chronic diseases of the liver and kidney as well as diabetes mellitus [33,47,48]. Thus, it is notable that the sulfinic acid feature (822.42; FC = 1.04) was selected for association with HC lung cancers and was among the most abundant features measured (Table 2). In cross-sectional investigations into Cys34 adductomics among residents of central London with high exposures to DEE, 822.42 was associated with COPD (FC = 0.85,  $p = 0.038$ ), a risk factor for lung cancer, and ischemic heart disease (FC = 0.78,  $p = 0.027$ ) [42]. The apparent reversal in FC values for 822.42 from negative in the COPD study to positive in our lung cancer investigation probably reflects differences in disease endpoints as well as the subjects' geographic locations, age, ethnicity, smoking history, and diet. The other Cys34 sulfoxidation product selected for association with lung cancers in this study, namely, 849.07 (S-sulfonic acid trisulfide, FC = 0.91 for LUADs and 0.935 for HC cases) was observed in some of our previous investigations (footnoted in Table 2) but without notable associations. Although the origin of 849.07 is obscure, Kim et al. reported similar structures arising from reactions between a protein sulfinic acid and Cys with subsequent oxidations and degradation reactions [23].

In contrast to irreversible oxidations to its sulfinic and sulfonic acids, the sulfenic acid (Cys34-SOH) can react reversibly with circulating thiols (R-SH) to form disulfides



(Cys34-S-S-R). Because Cys and hCys are among the most abundant low-molecular-weight thiols in the serum, it is not surprising that their Cys34 disulfides would be prominent features in our samples (Table 2), and three of them were selected for association with lung cancer (Table 4). Furthermore, the lower abundance of 845.42 [S-Cys(-H<sub>2</sub>O), FC = 0.93] and 856.10 (S-hCys, FC = 0.94) in never-smoking lung cancer cases from our study reinforces similar findings in both smokers and nonsmokers from the EPIC investigation, i.e., for S-Cys(-H<sub>2</sub>O) (FC = 0.71,  $p = 0.15$ ) and S-hCys (FC = 0.74,  $p = 0.09$ ) [43]. Interestingly, we previously used our classification methods to select S-Cys(-H<sub>2</sub>O) and S-hCys for association with exposure to DEE (FC = 0.88 and 0.91, respectively) [45] and also found evidence in never-smoking women in Xuanwei for associations between S-Cys(-H<sub>2</sub>O) and exposure to emissions from smoky coal (FC = 0.37,  $p = 0.07$ ) and the corresponding air concentration of the carcinogenic PAH, benzo[a]pyrene (FC = 1.46,  $p = 0.022$ ) [40]. These findings further suggest that exposure to PM<sub>2.5</sub> and PAHs is a potential risk factor for lung cancer in never-smokers.

#### 4.3. The Role of OCM in Lung Cancer

Since hCys is a key OCM intermediate involved in the production of methionine, and Cys and GSH together form much of the basis for redox homeostasis [28], the dysregulation of hCys metabolism may be a risk factor for lung cancer. Indeed, hCys is an important component of desulfuration and transsulfuration pathways that involve cystathionine  $\beta$ -synthase (CBS), an enzyme that catalyzes reactions producing Cys, methionine, GSH, and hydrogen sulfide (H<sub>2</sub>S) [49]. Acting through the modulation of hCys and H<sub>2</sub>S metabolism, CBS contributes to biological processes controlling redox homeostasis, DNA methylation and protein modification that, when perturbed, can lead to pathological consequences [49]. As noted previously, Seow et al. linked lung cancer among never-smokers from the SWHS to pathways involving methionine and Cys metabolism [27], and Baltar et al. reported protective effects of pathways involving methionine-hCys on lung cancer in EPIC [29]. Furthermore, the lower abundance of two of the three selected Cys and hCys features 845.42 [S-Cys(-H<sub>2</sub>O), FC = 0.93] and 856.10 (S-hCys, FC = 0.94) in never-smoking cases from our study as well as smokers and nonsmokers from the previous EPIC investigation [43] is consistent with the hypothesis that depletion of serum thiols is a risk factor for lung cancer. Interestingly, the Cys34 S-sulfonic acid trisulfide (849.07) does not cluster with the other Cys34 sulfoxidation products (cluster 3, Figures 1 and 2) but is negatively correlated with the Cys and hCys disulfides in cluster 5, suggesting possible influence in depleting serum thiols and the dysregulation of hCys metabolism.

#### 4.4. A Connection to Microbial Translocation?

The other annotated Cys34 feature selected for association with HC lung cancers was 827.088 (FC = 1.00), which is the Cys34 disulfide of methane thiol that is produced by enteric microbiota and was previously selected for associations with CRC (FC = 1.20) [44] and NHL (FC = 1.39, males only) [37] in EPIC. Given the elevation of this biomarker in NHL cases, it was hypothesized that the translocation of enteric bacteria from the gut resulted in the systemic release of ROS that can promote cancers throughout the body [37] (perhaps including the lung). However, this result from the current study should be regarded cautiously because the selection of 827.088 was based solely on Random Forests in the LUAD cases/controls and was not elevated in cases (Supplementary Figure S1C).

#### 4.5. Influence of Covariates

Analyses of covariates by Random Forests, summarized in Figures 3A and 4A, indicate that BMI was ranked highest in lung cancer cases and controls. This reflects the significantly lower values of BMI in cases here (Table 1) and in the literature [38]. Dietary methionine

was also highly ranked in HC cases and controls (Figure 3A), which is consistent with elevated plasma methionine in lung cancer cases across observational studies [19] and with dysregulated methionine and Cys metabolism in never-smoking lung cancer cases in the SWHS [27]. Also, the high rank of the Cys34 sulfinic acid (822.42) in HC cases (Figure 3A) points to oxidative stress as a potentially causal factor and is reinforced in LUAD cases by the high ranks of the Cys34 S-sulfonic acid trisulfide (849.07) and hydroxy-PAHs (Figure 4A).

The heatmaps in Figures 3B and 4B show correlations of selected features with covariates for HC and LUAD cases, respectively. All of the dietary variables (folate, vegetables, methionine, and soy) are positively correlated with each other, as are the two sets of PAH biomarkers and ETS (which contains PM<sub>2.5</sub> and adsorbed PAHs). Several selected features were negatively correlated with the PAH/ETS cluster, namely the Cys34 disulfides of Cys (845.42), hCys (856.10), and Cys34 sulfinic acid (822.42) with HC cases (Figure 3B), suggesting that ROS generated by inhaled PM<sub>2.5</sub> and PAH metabolism perturbed redox homeostasis. Again, of particular interest is 856.10 (hCys, FC = 0.94), which is negatively correlated with ETS and the amino-PAHs (metabolites of nitro-PAHs) that are biomarkers of DEE [16] but not with the hydroxy-PAHs derived from other combustion processes (Figures 3B and 4B). Since we had previously found levels of 856.10 to be reduced in workers and urban residents exposed to DEE [42,45], this suggests that DEE may have a particular connection with lung cancer in urban never-smokers, possibly due to inflammation arising from nanoparticles and nitro-PAHs that are produced primarily by diesel engines [16]. The DEE connection with lung cancer may be explained by the strong mutagenicity of nitro-PAHs, as observed in our previous work [50]. Of the dietary variables, the most noteworthy finding is the strong negative correlation between the Cys34 sulfinic acid (822.42) and consumption of soy and methionine in HC cases (Figure 3B), which links dietary antioxidants with reduced systemic oxidative stress. Dietary methionine may be especially important because it plays a role in the formation of GSH via OCM and is itself a scavenger of ROS [51]. The personal cluster (BMI, age, and education) in HC cases also shows strong negative correlations with the Cys34 sulfinic acid (822.42) (Figure 3B), suggesting inverse associations with systemic oxidative stress as subjects become older, heavier, and better educated.

#### 4.6. Strengths and Weaknesses

This investigation has several strengths. First, by focusing on never-smokers in the SWHS, we were able to investigate lung cancer risks in a particularly vulnerable population without the confounding effects of smoking. Second, the potential for reverse causality was reduced because plasma samples were collected prior to diagnosis by up to 14 years. Third, the untargeted HSA adductomics assay provides an unbiased and unique vantage point with which to evaluate ROS disposition during the month preceding phlebotomy and to determine whether ROS may have contributed to lung cancer. Fourth, using a hypothesis-free design and a set of complementary regression and classification methods, we reduced biases in selecting features for associations with lung cancer, including both linear and nonlinear relationships. Fourth, expanding the adductomics pipeline to Lys525 extended the underlying chemistries of adducts and included two selected features (571.84, Lys oxidation, and 587.32, unknown).

We also recognize some weaknesses. Plasma specimens were stored at  $-80^{\circ}\text{C}$  for up to 14 years prior to analysis, leading to the possibility that adducts may have been modified during storage. Such potential production of artifacts during storage was minimized by matching cases and controls by year of enrollment. Since only four of the nine selected features were confirmed with synthetic standards (822.42, 827.088, 856.10, and 858.75), the

annotations of the others were based on elemental composition and should be regarded as putative. Also, two seemingly important features (587.31 and 914.83) were unannotated, limiting our ability to address their implications. Given a residence time of one month for human HSA adducts, levels measured in blood collected at recruitment may not accurately reflect those produced in subsequent years.

## 5. Conclusions

In conclusion, this untargeted adductomics investigation selected nine HSA adduct features that provided a unique vantage point for considering associations between ROS and lung cancer up to 14 years prior to diagnosis in never-smoking females (206 cases and 293 controls). The annotations of seven of these features point to both direct and indirect links among PM<sub>2.5</sub> and PAH exposures, redox biology, OCM, and dietary antioxidants, consistent with the literature. Several of these features have previously been selected with similar effects in adductomic investigations into lung cancer and COPD and with environmental exposures to PM<sub>2.5</sub>, PAHs, and DEE. Future work should seek the validation of these findings with other lung cancer cohorts involving never-smokers and should apply this untargeted adductomic approach more generally to discover causes of other inflammatory diseases including atherosclerosis.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/antiox14030335/s1>. References [52,53] are cited in Supplementary Materials.

**Author Contributions:** P.I. and S.D. conceived and conducted statistical analyses, interpreted results, drafted statistical methods, reviewed and edited the manuscript, and approved the final version. H.G. performed adductomics analyses, the preprocessing and curation of mass spectrometric data, reviewed the manuscript, and approved the final version. X.-O.S., J.W., L.Z., W.H., B.B., and M.R. acquired data, interpreted results, reviewed the manuscript, and approved the final version. J.Z. performed mass spectrometric analyses of urinary PAH metabolites, drafted the corresponding methods, reviewed and edited the manuscript, and approved the final version. Q.C., Y.G., W.Z., N.R., and Q.L. conceived portions of the work, acquired data, interpreted results, reviewed the manuscript, and approved the final version. S.M.R. conceived and designed the work, interpreted results, drafted and revised the manuscript, and approved the final version. All authors have read and agreed to the published version of the manuscript.

**Funding:** Financial support for this work was provided by the U.S. National Institutes of Health, Department of Health and Human Services through grants P42ES004705 from the National Institute for Environmental Health Sciences (to S.M.R.), from the National Cancer Institute through grants R33CA191159 (to S.M.R. and S.D.), R01CA70867 (to W.Z.) and UM1CA182910 (to W.Z.), and the intramural research program (N.R. and Q.L.).

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Boards of National Cancer Institute (OH98CN006, last approval: 4 April 2019) and the Vanderbilt University (000340; last approval: 9 September 2021).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study. This manuscript contains no individual person's data.

**Data Availability Statement:** The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Conflicts of Interest:** The authors declare that there are no competing financial and/or non-financial interests in relation to the work described.

## Abbreviations

BMI, body mass index; CBS, cystathionine  $\beta$ -synthase; COPD, chronic obstructive pulmonary disease; Cys34, cysteine residue at position 34 of HSA; CysGly, cysteinyl glycine; DEE, diesel engine exhaust; *EGFR*, Epidermal Growth Factor Receptor; EPIC, European Prospective Investigation into Cancer and Nutrition; ETS, environmental tobacco smoke; FC, fold change; GSH, glutathione; HC, histologically confirmed lung cancer cases; hCys; homocysteine; HSA, human serum albumin; IARC, International Agency for Research on Cancer; IL-1 $\beta$ , interleukin-1 $\beta$ ; ICC, intraclass correlation coefficient; LFC, log-fold change; LUAD, lung adenomas; LASSO, regularized logistic regression; Lys525, lysine residue at position 525 of HSA; MIM, monoisotopic mass; nLC-HRMSMS, nano-liquid chromatography–tandem mass spectrometry; OCM, one-carbon metabolism; ROS, reactive oxygen species; PAHs, polycyclic aromatic hydrocarbons; PM<sub>2.5</sub>, airborne fine particulate matter less than 2.5  $\mu$ m in diameter; RF, Random Forests; SAM, S-adenosylmethionine; SWHS, Shanghai Women’s Health Study; T3 peptide; third largest tryptic peptide of HSA; ttd, time from subject recruitment to diagnosis; WHO, World Health Organization.

## References

- LoPiccolo, J.; Gusev, A.; Christiani, D.C.; Janne, P.A. Lung cancer in patients who have never smoked—An emerging disease. *Nat. Rev. Clin. Oncol.* **2024**, *21*, 121–146. [[CrossRef](#)]
- Epplein, M.; Schwartz, S.M.; Potter, J.D.; Weiss, N.S. Smoking-adjusted lung cancer incidence among Asian-Americans (United States). *Cancer Causes Control CCC* **2005**, *16*, 1085–1090. [[CrossRef](#)]
- Zhong, L.; Goldberg, M.S.; Gao, Y.T.; Jin, F. A case-control study of lung cancer and environmental tobacco smoke among nonsmoking women living in Shanghai, China. *Cancer Causes Control CCC* **1999**, *10*, 607–616. [[CrossRef](#)] [[PubMed](#)]
- Hosgood, H.D., 3rd; Pao, W.; Rothman, N.; Hu, W.; Pan, Y.H.; Kuchinsky, K.; Jones, K.D.; Xu, J.; Vermeulen, R.; Simko, J.; et al. Driver mutations among never smoking female lung cancer tissues in China identify unique EGFR and KRAS mutation pattern associated with household coal burning. *Respir. Med.* **2013**, *107*, 1755–1762. [[CrossRef](#)] [[PubMed](#)]
- Yu, I.T.; Chiu, Y.L.; Au, J.S.; Wong, T.W.; Tang, J.L. Dose-response relationship between cooking fumes exposures and lung cancer among Chinese nonsmoking women. *Cancer Res.* **2006**, *66*, 4961–4967. [[CrossRef](#)] [[PubMed](#)]
- Yang, G.; Shu, X.O.; Chow, W.H.; Zhang, X.; Li, H.L.; Ji, B.T.; Cai, H.; Wu, S.; Gao, Y.T.; Zheng, W. 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*, 846–855. [[CrossRef](#)]
- Takata, Y.; Shu, X.O.; Yang, G.; Li, H.; Dai, Q.; Gao, J.; Cai, Q.; Gao, Y.T.; Zheng, W. Calcium intake and lung cancer risk among female nonsmokers: A report from the Shanghai Women’s Health Study. *Cancer Epidemiol. Biomark. Prev.* **2013**, *22*, 50–57. [[CrossRef](#)]
- Loomis, D.; Grosse, Y.; Lauby-Secretan, B.; El Ghissassi, F.; Bouvard, V.; Benbrahim-Tallaa, L.; Guha, N.; Baan, R.; Mattock, H.; Straif, K.; et al. The carcinogenicity of outdoor air pollution. *Lancet Oncol.* **2013**, *14*, 1262–1263. [[CrossRef](#)]
- Hill, W.; Lim, E.L.; Weeden, C.E.; Lee, C.; Augustine, M.; Chen, K.; Kuan, F.C.; Marongiu, F.; Evans, E.J., Jr.; Moore, D.A.; et al. Lung adenocarcinoma promotion by air pollutants. *Nature* **2023**, *616*, 159–167. [[CrossRef](#)]
- Hogg, J.C.; van Eeden, S. Pulmonary and systemic response to atmospheric pollution. *Respirology* **2009**, *14*, 336–346. [[CrossRef](#)]
- Valavanidis, A.; Vlachogianni, T.; Fiotakis, K.; Loridas, S. Pulmonary oxidative stress, inflammation and cancer: Respirable particulate matter, fibrous dusts and ozone as major causes of lung carcinogenesis through reactive oxygen species mechanisms. *Int. J. Environ. Res. Public Health* **2013**, *10*, 3886–3907. [[CrossRef](#)]
- Straif, K.; Baan, R.; Grosse, Y.; Secretan, B.; El Ghissassi, F.; Coglianò, V.; WHO International Agency for Research on Cancer Monograph Working Group. Carcinogenicity of polycyclic aromatic hydrocarbons. *Lancet Oncol.* **2005**, *6*, 931–932. [[CrossRef](#)] [[PubMed](#)]
- Chiang, T.A.; Wu, P.F.; Ko, Y.C. Identification of carcinogens in cooking oil fumes. *Environ. Res.* **1999**, *81*, 18–22. [[CrossRef](#)]
- Moorthy, B.; Chu, C.; Carlin, D.J. Polycyclic aromatic hydrocarbons: From metabolism to lung cancer. *Toxicol. Sci.* **2015**, *145*, 5–15. [[CrossRef](#)] [[PubMed](#)]
- Palackal, N.T.; Lee, S.H.; Harvey, R.G.; Blair, I.A.; Penning, T.M. Activation of polycyclic aromatic hydrocarbon trans-dihydrodiol proximate carcinogens by human aldo-keto reductase (AKR1C) enzymes and their functional overexpression in human lung carcinoma (A549) cells. *J. Biol. Chem.* **2002**, *277*, 24799–24808. [[CrossRef](#)] [[PubMed](#)]
- IARC. Diesel and Gasoline Engine Exhausts and Some Nitroarenes. Iarc Monographs on the Evaluation of Carcinogenic Risks to Humans. *IARC Monogr. Eval. Carcinog. Risks Hum.* **2014**, *105*, 9–699.

17. Lui, K.H.; Bandowe, B.A.; Tian, L.; Chan, C.S.; Cao, J.J.; Ning, Z.; Lee, S.C.; Ho, K.F. Cancer risk from polycyclic aromatic compounds in fine particulate matter generated from household coal combustion in Xuanwei, China. *Chemosphere* **2017**, *169*, 660–668. [[CrossRef](#)]
18. Barone-Adesi, F.; Chapman, R.S.; Silverman, D.T.; He, X.; Hu, W.; Vermeulen, R.; Ning, B.; Fraumeni, J.F., Jr.; Rothman, N.; Lan, Q. Risk of lung cancer associated with domestic use of coal in Xuanwei, China: Retrospective cohort study. *BMJ* **2012**, *345*, e5414. [[CrossRef](#)]
19. Lee, K.B.; Ang, L.; Yau, W.P.; Seow, W.J. Association between Metabolites and the Risk of Lung Cancer: A Systematic Literature Review and Meta-Analysis of Observational Studies. *Metabolites* **2020**, *10*, 362. [[CrossRef](#)]
20. Zhang, T.; Hoang, P.H.; Wong, J.Y.; Yang, K.; Chen, K.; Wong, M.P.; Vermeulen, R.C.; Xuanwei Study Team; Huang, Y.; Chanock, S.J.; et al. Distinct Genomic Landscape of Lung Adenocarcinoma from Household Use of Smoky Coal. *Am. J. Respir. Crit. Care Med.* **2023**, *208*, 733–736. [[CrossRef](#)]
21. Ho, K.F.; Chang, C.C.; Tian, L.; Chan, C.S.; Musa Bandowe, B.A.; Lui, K.H.; Lee, K.Y.; Chuang, K.J.; Liu, C.Y.; Ning, Z.; et al. Effects of polycyclic aromatic compounds in fine particulate matter generated from household coal combustion on response to EGFR mutations in vitro. *Environ. Pollut.* **2016**, *218*, 1262–1269. [[CrossRef](#)] [[PubMed](#)]
22. Lan, Q.; Mumford, J.L.; Shen, M.; Demarini, D.M.; Bonner, M.R.; He, X.; Yeager, M.; Welch, R.; Chanock, S.; Tian, L.; et al. Oxidative damage-related genes AKR1C3 and OGG1 modulate risks for lung cancer due to exposure to PAH-rich coal combustion emissions. *Carcinogenesis* **2004**, *25*, 2177–2181. [[CrossRef](#)] [[PubMed](#)]
23. Kim, H.J.; Ha, S.; Lee, H.Y.; Lee, K.J. ROSics: Chemistry and proteomics of cysteine modifications in redox biology. *Mass Spectrom. Rev.* **2015**, *34*, 184–208. [[CrossRef](#)]
24. Key, T.J.; Allen, N.E.; Spencer, E.A.; Travis, R.C. The effect of diet on risk of cancer. *Lancet* **2002**, *360*, 861–868. [[CrossRef](#)] [[PubMed](#)]
25. Wei, X.; Zhu, C.; Ji, M.; Fan, J.; Xie, J.; Huang, Y.; Jiang, X.; Xu, J.; Yin, R.; Du, L.; et al. Diet and Risk of Incident Lung Cancer: A Large Prospective Cohort Study in UK Biobank. *Am. J. Clin. Nutr.* **2021**, *114*, 2043–2051. [[CrossRef](#)]
26. Yu, J.; Bi, X.; Yu, B.; Chen, D. Isoflavones: Anti-Inflammatory Benefit and Possible Caveats. *Nutrients* **2016**, *8*, 361. [[CrossRef](#)]
27. Seow, W.J.; Shu, X.O.; Nicholson, J.K.; Holmes, E.; Walker, D.I.; Hu, W.; Cai, Q.; Gao, Y.T.; Xiang, Y.B.; Moore, S.C.; et al. Association of Untargeted Urinary Metabolomics and Lung Cancer Risk Among Never-Smoking Women in China. *JAMA Netw. Open* **2019**, *2*, e1911970. [[CrossRef](#)]
28. Ducker, G.S.; Rabinowitz, J.D. One-Carbon Metabolism in Health and Disease. *Cell Metab.* **2017**, *25*, 27–42. [[CrossRef](#)]
29. Baltar, V.T.; Xun, W.W.; Johansson, M.; Ferrari, P.; Chuang, S.C.; Relton, C.; Ueland, P.M.; Middtun, O.; Slimani, N.; Jenab, M.; et al. A structural equation modelling approach to explore the role of B vitamins and immune markers in lung cancer risk. *Eur. J. Epidemiol.* **2013**, *28*, 677–688. [[CrossRef](#)]
30. Yang, J.; Li, H.; Deng, H.; Wang, Z. Association of One-Carbon Metabolism-Related Vitamins (Folate, B6, B12), Homocysteine and Methionine With the Risk of Lung Cancer: Systematic Review and Meta-Analysis. *Front. Oncol.* **2018**, *8*, 493. [[CrossRef](#)]
31. Aldini, G.; Vistoli, G.; Regazzoni, L.; Gamberoni, L.; Facino, R.M.; Yamaguchi, S.; Uchida, K.; Carini, M. Albumin is the main nucleophilic target of human plasma: A protective role against pro-atherogenic electrophilic reactive carbonyl species? *Chem. Res. Toxicol.* **2008**, *21*, 824–835. [[CrossRef](#)] [[PubMed](#)]
32. Dalle-Donne, I.; Rossi, R.; Colombo, R.; Giustarini, D.; Milzani, A. Biomarkers of oxidative damage in human disease. *Clin. Chem.* **2006**, *52*, 601–623. [[CrossRef](#)]
33. Watanabe, H.; Imafuku, T.; Otagiri, M.; Maruyama, T. Clinical Implications Associated with the Posttranslational Modification-Induced Functional Impairment of Albumin in Oxidative Stress-Related Diseases. *J. Pharm. Sci.* **2017**, *106*, 2195–2203. [[CrossRef](#)]
34. Grigoryan, H.; Imani, P.; Dudoit, S.; Rappaport, S.M. Extending the HSA-Cys34-Adductomics Pipeline to Modifications at Lys525. *Chem. Res. Toxicol.* **2021**, *34*, 2549–2557. [[CrossRef](#)] [[PubMed](#)]
35. Zheng, W.; Chow, W.H.; Yang, G.; Jin, F.; Rothman, N.; Blair, A.; Li, H.L.; Wen, W.; Ji, B.T.; Li, Q.; et al. The Shanghai Women’s Health Study: Rationale, study design, and baseline characteristics. *Am. J. Epidemiol.* **2005**, *162*, 1123–1131. [[CrossRef](#)] [[PubMed](#)]
36. Yang, Z.; Lin, Y.; Wang, S.; Liu, X.; Cullinan, P.; Chung, K.F.; Zhang, J. Urinary Amino-Polycyclic Aromatic Hydrocarbons in Urban Residents: Finding a Biomarker for Residential Exposure to Diesel Traffic. *Environ. Sci. Technol.* **2021**, *55*, 10569–10577. [[CrossRef](#)]
37. Grigoryan, H.; Imani, P.; Sacerdote, C.; Masala, G.; Grioni, S.; Tumino, R.; Chiodini, P.; Dudoit, S.; Vineis, P.; Rappaport, S.M. HSA Adductomics Reveals Sex Differences in NHL Incidence and Possible Involvement of Microbial Translocation. *Cancer Epidemiol. Biomark. Prev.* **2023**, *32*, 1217–1226. [[CrossRef](#)]
38. Zhu, H.; Zhang, S. Body mass index and lung cancer risk in never smokers: A meta-analysis. *BMC Cancer* **2018**, *18*, 635. [[CrossRef](#)]
39. Grigoryan, H.; Edmands, W.; Lu, S.S.; Yano, Y.; Regazzoni, L.; Iavarone, A.T.; Williams, E.R.; Rappaport, S.M. Adductomics Pipeline for Untargeted Analysis of Modifications to Cys34 of Human Serum Albumin. *Anal. Chem.* **2016**, *88*, 10504–10512. [[CrossRef](#)]



40. Lu, S.S.; Grigoryan, H.; Edmands, W.M.; Hu, W.; Iavarone, A.T.; Hubbard, A.; Rothman, N.; Vermeulen, R.; Lan, Q.; Rappaport, S.M. Profiling the Serum Albumin Cys34 Adductome of Solid Fuel Users in Xuanwei and Fuyuan, China. *Environ. Sci. Technol.* **2017**, *51*, 46–57. [[CrossRef](#)]
41. Grigoryan, H.; Edmands, W.M.B.; Lan, Q.; Carlsson, H.; Vermeulen, R.; Zhang, L.; Yin, S.N.; Li, G.L.; Smith, M.T.; Rothman, N.; et al. Adductomic signatures of benzene exposure provide insights into cancer induction. *Carcinogenesis* **2018**, *39*, 661–668. [[CrossRef](#)]
42. Liu, S.; Grigoryan, H.; Edmands, W.M.B.; Dagnino, S.; Sinharay, R.; Cullinan, P.; Collins, P.; Chung, K.F.; Barratt, B.; Kelly, F.J.; et al. Cys34 Adductomes Differ between Patients with Chronic Lung or Heart Disease and Healthy Controls in Central London. *Environ. Sci. Technol.* **2018**, *52*, 2307–2313. [[CrossRef](#)] [[PubMed](#)]
43. Dagnino, S.; Bodinier, B.; Grigoryan, H.; Rappaport, S.M.; Karimi, M.; Guida, F.; Polidoro, S.; Edmands, W.B.; Naccarati, A.; Fiorito, G.; et al. Agnostic Cys34-albumin adductomics and DNA methylation: Implication of N-acetylcysteine in lung carcinogenesis years before diagnosis. *Int. J. Cancer* **2019**, *146*, 3294–3303. [[CrossRef](#)] [[PubMed](#)]
44. Grigoryan, H.; Schiffman, C.; Gunter, M.J.; Naccarati, A.; Polidoro, S.; Dagnino, S.; Dudoit, S.; Vineis, P.; Rappaport, S.M. Cys34 Adductomics Links Colorectal Cancer with the Gut Microbiota and Redox Biology. *Cancer Res.* **2019**, *79*, 6024–6031. [[CrossRef](#)]
45. Wong, J.Y.Y.; Imani, P.; Grigoryan, H.; Bassig, B.A.; Dai, Y.; Hu, W.; Blechter, B.; Rahman, M.L.; Ji, B.T.; Duan, H.; et al. Exposure to diesel engine exhaust and alterations to the Cys34/Lys525 adductome of human serum albumin. *Environ. Toxicol. Pharmacol.* **2022**, *95*, 103966. [[CrossRef](#)]
46. Yuan, K.; Liu, Y.; Chen, H.N.; Zhang, L.; Lan, J.; Gao, W.; Dou, Q.; Nice, E.C.; Huang, C. Thiol-based redox proteomics in cancer research. *Proteomics* **2015**, *15*, 287–299. [[CrossRef](#)] [[PubMed](#)]
47. Kobayashi, Y.; Suzuki, R.; Yasukawa, K.; Oba, K.; Yamauchi, T.; Yatomi, Y.; Kadowaki, T. Oxidized albumin in blood reflects the severity of multiple vascular complications in diabetes mellitus. *Metabol. Open* **2020**, *6*, 100032. [[CrossRef](#)]
48. Paramasivan, S.; Adav, S.S.; Ngan, S.C.; Dalan, R.; Leow, M.K.; Ho, H.H.; Sze, S.K. Serum albumin cysteine trioxidation is a potential oxidative stress biomarker of type 2 diabetes mellitus. *Sci. Rep.* **2020**, *10*, 6475. [[CrossRef](#)]
49. Zhu, H.; Blake, S.; Chan, K.T.; Pearson, R.B.; Kang, J. Cystathionine beta-Synthase in Physiology and Cancer. *Biomed. Res. Int.* **2018**, *2018*, 3205125. [[CrossRef](#)]
50. Zhang, J.J.; Zheng, Y.; Vermeulen, R.; Liu, X.L.; Dai, Y.; Hu, W.; He, L.; Lin, Y.; Ren, D.; Duan, H.; et al. Urinary Amino-PAHs in relation to diesel engine emissions and urinary mutagenicity. *Int. J. Hyg. Environ. Health* **2023**, *253*, 114223. [[CrossRef](#)]
51. Campbell, K.; Vowinckel, J.; Keller, M.A.; Ralser, M. Methionine Metabolism Alters Oxidative Stress Resistance via the Pentose Phosphate Pathway. *Antioxid. Redox Signal.* **2016**, *24*, 543–547. [[CrossRef](#)] [[PubMed](#)]
52. Keller, B.O.; Sui, J.; Young, A.B.; Whittall, R.M. Interferences and contaminants encountered in modern mass spectrometry. *Anal. Chim. Acta* **2008**, *627*, 71–81. [[CrossRef](#)] [[PubMed](#)]
53. Xu, X.; Zhang, J.; Zhang, L.; Liu, W.; Weisel, C.P. Selective detection of monohydroxy metabolites of polycyclic aromatic hydrocarbons in urine using liquid chromatography/triple quadrupole tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2299–2308. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.