HHS Public Access

Author manuscript

Environ Toxicol Pharmacol. Author manuscript; available in PMC 2023 October 01.

Published in final edited form as:

Environ Toxicol Pharmacol. 2022 October ; 95: 103966. doi:10.1016/j.etap.2022.103966.

Exposure to diesel engine exhaust and alterations to the Cys34/Lys525 adductome of human serum albumin

Jason Y.Y. Wong[*,1], **Partow Imani**[*,2], **Hasmik Grigoryan**[*,2], **Bryan A. Bassig**[¥,1], **Yufei Dai**[3], **Wei Hu**[1], **Batel Blechter**[1], **Mohammad L. Rahman**[1], **Bu-Tian Ji**[1], **Huawei Duan**[3], **Yong Niu**[3],

Corresponding author: Jason Y.Y. Wong, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Drive, Rockville, MD, 20850, USA. Phone:240-276-5149. jason.wong@nih.gov.

[¥]Formerly at the U.S. National Cancer Institute. This author is currently employed by the U.S. Centers for Disease Control and Prevention, National Center for Health Statistics. All work on this study by the author was conducted while employed by the U.S. National Cancer Institute.

[*]Co-first authors

[†]These authors supervised the study

CRediT authorship contribution statement

Jason Y.Y. Wong: Formal analysis; Investigation; Methodology; Project administration; Roles/Writing - original draft.
Partow Imani: Software; Formal analysis; Investigation; Methodology; Project administration; Roles/Writing - original draft.
Hasmik Grigoryan: Software; Formal analysis; Investigation; Methodology; Project administration; Roles/Writing - original draft.
Bryan A. Bassig: Funding acquisition; Data curation; Formal analysis; Investigation; Methodology; Project administration; Roles/Writing - original draft.
Yufei Dai: Investigation; Methodology; Roles/Writing - original draft.
Wei Hu: Project administration; Data curation; Investigation; Methodology; Roles/Writing - original draft.
Batel Blechter: Roles/Writing - original draft.
Mohammad L. Rahman: Roles/Writing - original draft.
Bu-Tian Ji: Roles/Writing - original draft.
Huawei Duan: Investigation; Methodology; Roles/Writing - original draft.
Yong Niu: Investigation; Methodology; Roles/Writing - original draft.
Meng Ye: Investigation; Methodology; Roles/Writing - original draft.
Xiaowei Jia: Investigation; Methodology; Roles/Writing - original draft.
Tao Meng: Investigation; Methodology; Roles/Writing - original draft.
Ping Bin: Investigation; Methodology; Roles/Writing - original draft.
George Downward: Investigation; Methodology; Roles/Writing - original draft.
Kees Meliefste: Investigation; Methodology; Roles/Writing - original draft.
Shuguang Leng: Investigation; Methodology; Roles/Writing - original draft.
Wei Fu: Investigation; Methodology; Roles/Writing - original draft.
Jufang Yang: Investigation; Methodology; Roles/Writing - original draft.
Dianzhi Ren: Investigation; Methodology; Roles/Writing - original draft.
Jun Xu: Investigation; Methodology; Roles/Writing - original draft.
Baosen Zhou: Investigation; Methodology; Roles/Writing - original draft.
H. Dean Hosgood: Investigation; Methodology; Roles/Writing - original draft.
Roel Vermeulen: Investigation; Methodology; Roles/Writing - original draft.
Yuxin Zheng: Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Roles/Writing - original draft.
Debra T. Silverman: Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Roles/Writing - original draft.
Nathaniel Rothman: Investigation; Methodology; Project administration; Supervision; Roles/Writing - original draft.
Stephen M. Rappaport: Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Roles/Writing - original draft.
Qing Lan: Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Roles/Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Meng Ye[3], **Xiaowei Jia**[3], **Tao Meng**[3], **Ping Bin**[3], **George Downward**[4], **Kees Meliefste**[4], **Shuguang Leng**[5], **Wei Fu**[6], **Jufang Yang**[6], **Dianzhi Ren**[6], **Jun Xu**[7], **Baosen Zhou**[8], **H. Dean Hosgood**[9], **Roel Vermeulen**[4,†], **Yuxin Zheng**[10,†], **Debra T. Silverman**[†,1], **Nathaniel Rothman**[†,1], **Stephen M. Rappaport**[†,2], **Qing Lan**[†,1]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, Maryland, USA.

[2]School of Public Health, University of California, Berkeley, California, USA.

[3]National Institute of Occupational Health and Poison Control, Chinese Center for Disease Control and Prevention, Beijing, China

[4]Division of Environmental Epidemiology, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands

[5]Division of Epidemiology, Biostatistics, and Preventive Medicine, Department of Internal Medicine, University of New Mexico School of Medicine; Cancer Control and Population Sciences, University of New Mexico Comprehensive Cancer Center, Albuquerque, New Mexico, USA.

[6]Chaoyang Center for Disease Control and Prevention, Chaoyang, Liaoning, China

[7]School of Public Health, The University of Hong Kong, Hong Kong, Hong Kong

[8]China Medical University, Shenyang, Liaoning, China

[9]Division of Epidemiology, Albert Einstein College of Medicine, New York, NY, USA

[10]School of Public Health, Qingdao University, Qingdao, China

Abstract

We investigated whether exposure to carcinogenic diesel engine exhaust (DEE) was associated with altered adduct levels in human serum albumin (HSA) residues. Nano-liquid chromatography-high resolution mass spectrometry (nLC-HRMS) was used to measure adducts of Cys34 and Lys525 residues in plasma samples from 54 diesel engine factory workers and 55 unexposed controls. An untargeted adductomics and bioinformatics pipeline was used to find signatures of Cys34/Lys525 adductome modifications. To identify adducts that were altered between DEE-exposed and unexposed participants, we used an ensemble feature selection approach that ranks and combines findings from linear regression and penalized logistic regression, then aggregates the important findings with those determined by random forest. We detected 40 Cys34 and 9 Lys525 adducts. Among these findings, we found evidence that 6 Cys34 adducts were altered between DEE-exposed and unexposed participants (i.e., 841.75, 851.76, 856.10, 860.77, 870.43, and 913.45). These adducts were biologically related to antioxidant activity.

1. INTRODUCTION

Diesel engine exhaust (DEE) is a known lung carcinogen based largely on evidence from occupational studies of highly exposed miners and members of the U.S. trucking industry (Garshick et al., 2012; IARC, 2014; Silverman et al., 2012). DEE is a complex mixture of particulates, organic and inorganic gaseous pollutants, as well as nitrated polycyclic

aromatic hydrocarbons (nitro-PAHs), which are suspected to contribute to its carcinogenicity (IARC, 2014). DEE is a significant public health burden, with millions of workers occupationally exposed worldwide across various industries. Furthermore, environmental exposure to DEE from traffic-related air pollution and residential proximity to major roadways threatens the respiratory health of the general population (Gowda et al., 2019; Hamra et al., 2015; Hart et al., 2015; Puett et al., 2014; Turner et al., 2020; Wong et al., 2021c).

Despite strong epidemiologic evidence linking DEE to lung cancer risk, the mechanism by which DEE causes early biologic changes and lung carcinogenesis in humans remains unclear. To investigate further, we conducted a cross-sectional study that included collection of biospecimens and personal air monitoring of workers in a diesel engine manufacturing facility and unexposed controls (Lan et al., 2015). We found that occupational exposure to DEE was associated with alterations to lung cancer-related lymphocyte subsets and inflammatory/immunologic proteins, as well as Alu retroelement copy number (Bassig et al., 2017; Clifford et al., 2017; Lan et al., 2015; Rahman et al., 2021; Shiels et al., 2017; Wong et al., 2021a). However, DEE could potentially operate through other biological mechanisms.

DEE can exert biological changes through oxidative stress generated from the production of reactive oxygen and electrophilic species (ROS/RES) (Rappaport et al., 2012). Electrophiles enter the bloodstream primarily via the metabolism of xenobiotics and from inflammation caused by exogenous exposures and endogenous processes (Rappaport et al., 2012). RES can directly react with proteins and nucleic acids to influence processes and genes that are upregulated during periods of environmental and cellular stress, leading to adverse changes including inflammation, mitochondrial dysfunction, and reduced antioxidant capabilities. Given the highly reactive characteristics and short half-lives of RES, they cannot be directly measured *in vivo*. Rather, indirect methods to measure RES have focused on evaluating long-lasting chemical modifications to abundant proteins, such as adducts in human serum albumin (HSA) (Lu et al., 2017), which is the dominant scavenger of RES in the interstitial space. Notably, the cysteine (Cys34) residue, which is located on the third largest tryptic (T3) peptide of HSA, is a nucleophilic hotspot that accounts for nearly 80% of the antioxidant capacity of serum (Aldini et al., 2008).

The Stephen Rappaport laboratory has developed an untargeted HSA-adductomics assay to detect and quantify modifications to Cys34 (Grigoryan et al., 2016) and applied this assay to compare adduct levels between subjects with different particulate and chemical exposures (Grigoryan et al., 2018; Lu et al., 2017) or disease states (Grigoryan et al., 2019; Liu et al., 2018; Yano et al., 2020). This HSA-adductomics assay was recently updated to include simultaneous untargeted detection of both Cys34 and Lys525 adducts, which expands coverage to different classes of electrophilic addition products (Grigoryan et al., 2021). These protein modifications are potential biomarkers of exogenous exposures and endogenous processes (Grigoryan et al., 2016). As such, untargeted measurement of the Cys34/Lys525 adductome potentially offers insight into mechanisms by which environmental exposures contribute to the pathogenesis of chronic diseases.

We have previously found relationships between alterations to the Cys34 adductome and exposure to known lung carcinogens, as well as respiratory diseases. For instance, in a rural population of non-smokers who domestically burned solid fuels, we found differences in plasma levels of several adducts across fuel types, including among those who used the most carcinogenic coal type, namely bituminous ("smoky") coal (Lu et al., 2017). We observed that adduct levels of *S*-glutathione (*S*-GSH) were significantly lower in smoky coal users compared with electricity/gas users, suggesting that the constituents of smoky coal emissions may have depleted this essential antioxidant. With respect to adductome-disease relationships, we previously found significant Cys34 adduct signatures in patients with chronic obstructive pulmonary disease (COPD) (Liu et al., 2018), which is a risk factor for lung cancer that has been linked to occupational exposure to DEE (Ferguson et al., 2020; Hart et al., 2012). Importantly, in a case-control study nested within the prospective European Prospective Investigation into Cancer and Nutrition (EPIC) study, two Cys34 adducts related to N-acetylcysteine and cysteinyl-glycine (CysGly) were found to be associated with lung cancer (Dagnino et al., 2020).

Our primary study aim was to evaluate whether occupational exposure to high levels of carcinogenic DEE, which has some similar chemical constituents as coal combustion emissions and cigarette smoke, is associated with alterations to the Cys34/Lys525 adductome of HSA. Our investigation expands upon the previous adductomic studies of lung cancer risk (Dagnino et al., 2020), COPD (Liu et al., 2018), as well as smoky coal emissions (Lu et al., 2017). Using an untargeted adductomics approach and recently developed bioinformatic methods, we identified differences in adduct levels between highly exposed diesel engine factory workers and unexposed controls (Grigoryan et al., 2016; Grigoryan et al., 2021). Notably, this is the first epidemiologic study to report the use of an ensemble feature selection pipeline to detect alterations in the Cys34/Lys525 adductome by an environmental exposure. Our findings could potentially contribute insight into how oxidative stress pathways operate in the etiologic mechanism underlying DEE-related lung carcinogenesis.

2. MATERIALS AND METHODS

2.1. Study Population and Exposure Assessment:

We used previously collected plasma samples from a cross-sectional molecular epidemiology study that included 54 male workers exposed to a wide range of DEE exposure levels (elemental carbon (EC) median (range) = 49.7 (6.1–107.7 μg/m^3)) and 55 unexposed male control workers in China (Lan et al., 2015). The exposed workers were enrolled from a diesel engine manufacturing facility and control workers were selected from four separate facilities from the same geographic area with no expected occupational exposure to DEE. Extensive site visits and monitoring were conducted at the control facilities to ensure that no occupational sources of DEE were detectable.

Demographic and lifestyle characteristics were obtained for each worker through a questionnaire and peripheral blood samples were collected from all workers as part of a health exam conducted by the local Center for Disease Control. Informed consent was obtained from all participants and the study was approved by Institutional Review Boards

at the US National Cancer Institute and the National Institute of Occupational Health and Poison Control, China CDC. The characteristics of the study population have been previously described in detail (Bassig et al., 2017; Dai et al., 2018; Lan et al., 2015; Wong et al., 2021b) and in Supplementary Table 1. Furthermore, an extensive exposure assessment survey was conducted from October 2012 to March 2013 in the diesel engine manufacturing facility (Lan et al., 2015). This survey included an assessment of specific diesel exhaust constituents including fine particulate matter ($PM_{2.5}$), EC, organic carbon (OC), and soot levels. Repeated full–shift personal air samples of EC, OC, and $PM_{2.5}$ were collected using a portable device attached to the lapel near the breathing zone of each worker.

2.2. Measurement of Cys34/Lys525 adducts using liquid chromatography-high resolution mass spectrometry:

To measure Cys34/Lys525 adduct/feature abundances, the collected plasma biospecimens were processed and analyzed using nano-liquid chromatography-high resolution mass spectrometry (nLC-HRMS) as previously described (Grigoryan et al., 2016; Grigoryan et al., 2021). We show the details of the nLC-HRMS procedure in the Appendix. We analyzed 119 plasma samples from the study participants, including 10 quality control replicate samples for reproducibility assessment. Laboratory investigators were blinded to the quality control samples and DEE exposure status.

2.3. Statistical analysis:

2.3.1. Detection of Cys34/Lys525 adducts in the study population:

Primary statistical analysis for peptide adduct/feature detection and selection was performed as previously described (Grigoryan et al., 2019). We detected 49 adducts in the overall study population (40 adducts in the T3 peptide containing Cys34 and 9 adducts in Lys525; Tables 1 and 2). Among these adducts, 44 had sufficient data for statistical analysis, while 5 were excluded for poor peak integration or detection in only one or two samples. Subsequently, based on adduct levels in duplicate injections, we filtered these 44 adducts for adequate data quality and excluded one adduct due to a low ICC value (Grigoryan et al., 2021). Among the remaining 43 adducts, duplicate injections were averaged, and missing values were ignored. We then excluded 11 adducts with ≥30% missing values because they did not show evidence of being differentially missing between exposed and unexposed groups using Fisher‟s exact tests. After these exclusions, 32 putative adducts were examined for associations with DEE exposure. Missing values were imputed using *k*-nearest neighbor imputation with $k=5$. The data were then normalized using the „scone‟ package in R (Risso et al., 2014) to consider possible adjustment for the following variables: housekeeping peptide (HK), internal standard (IS), and run order as the QC matrix, batch, and DEE exposure as the outcome. Data were normalized according to the top ranked method by scone: DESeq scaling accounting for the QC matrix as well as batch and DEE exposure.

2.3.2. Ensemble machine learning approach to detect alterations to the Cys34/Lys525 adductome among DEE-exposed and unexposed subjects—

To detect alterations to adduct levels among DEE-exposed participants and unexposed controls, we used an ensemble machine learning approach (Cai et al., 2015; Dietterich, 2000; Grigoryan et al., 2019; Mark et al., 2019; Petrick et al., 2019; Sarkar et al.,

2021), which selects important adducts/features by aggregating findings from three separate methods. The three steps of the ensemble machine learning approach are described in the Appendix. Briefly, we fit separate linear regression models with DEE-exposure status as the independent variable and each of the 32 detected adducts/features that passed QC as dependent variables, along with age, BMI, current smoking status, current infection status, and current alcohol consumption status as covariates. Second, we performed penalized logistic regression with DEE-exposure status as the dependent variable and each of the adducts/features as predictors, along with the five previously mentioned covariates. Third, we ranked the findings from linear regression by nominal p-values and penalized logistic regression by the proportion of iterations in which the adduct was kept in the model. We then determined the agreement among these findings using concordance plots, and further aggregated the important combined results with additional important findings selected by random forest using the same input. A variation of this ensemble machine learning approach was previously used to detect adductome alterations by disease states (Grigoryan et al., 2019; Yano et al., 2020). Notably, the current study is the first to use this ensemble approach to detect adductome alterations by an environmental exposure in a human population.

We conducted additional analyses independently of and in parallel to the ensemble machine learning approach to qualitatively confirm the DEE-adduct relationships. In separate multiple linear regression models, we assessed the associations between EC tertiles (Lan et al., 2015) (unexposed controls, low: median (range): 23.8 (6.1–39.0), medium: 49.7 (39.1–54.5), and high: 69.4 (54.6–107.7) $\mu g/m^3$) and each of the adducts (continuous normalized/scaled levels), parsimoniously adjusted for age, BMI, and smoking status (never, former, ever) to avoid overfitting. Further, we estimated exposure-response relationships using ordinal EC tertiles and continuous adduct levels. EC is a prominent component of diesel exhaust that is often used as a proxy for DEE levels. Adducts/features that qualitatively showed evidence of non-linear trends in box and whisker plots were further analyzed by including a quadratic term for EC in the models.

Correlations between normalized levels of the selected adducts/features were displayed with agglomerative hierarchical clustering using complete linkage and Spearman correlation („superheat“ function in R). A similar analysis was generated including the selected adducts/features and covariates, as well as additional immunologic/inflammatory markers (Bassig et al., 2017; Dai et al., 2018; Lan et al., 2015). Associations between the selected adducts/features and covariates were further examined with random forests plots to determine the relative rankings of the adducts/features according to mean decrease in Gini index. Associations between all candidate adducts and individual covariates were assessed similarly.

3. RESULTS

3.1. Detected Cys34/Lys525 adducts via mass spectrometry

Forty adducts/features in the Cys34 residue of the T3 peptide and 9 Lys525 adducts/features were detected in our study population (Tables 1 and 2). Accurate masses for 35 T3 peptide adducts/features led to reasonable elemental compositions added to the T3 peptide within 3 ppm of theoretical values from −46 Da to 496 Da, and similarly added to the Lys525

peptide with masses from −128 Da to 163 Da. Negative added masses refer to deletions and truncations. Peak area ratios for the 35 T3 peptide features with at least minimal quantitation covered a 366-fold range (PARx10,000: 1.0 – 351.1), while those of the Lys525 features covered an 820-fold range (PAR: 0.11 – 90.25).

As shown in Supplementary Table 2 for the 38 T3 peptide and 6 Lys525 adducts/features with sufficient data, the median intraclass correlation coefficient (ICC) was 0.678 (range: 0.000 – 0.928), indicating that the technical variation accounted for 32.2% (Range: 7.2% - 100%) of the total variance of adduct abundances. The median coefficient of variation (CV) for adduct features was 0.312 (Range: 0.143 – 0.531).

A subset of 30 T3 peptide adducts/features was annotated (Table 1), including 11 that were confirmed with reference standards in prior investigations (footnoted in Table 1). Since LC-MSMS parameters are the same as from prior investigations, identities of these features in the current study were based upon matches of accurate masses and retention times. Annotations of the other adducts/features are based on elemental compositions derived from accurate masses as well as database searches and should be regarded as putative. Annotations include one truncation [796.43 (Cys34→Gly)], Cys34 sulfoxidation products (816.42, 822.42 and 827.76, representing addition of 1, 2 and 3 oxygens to Cys34, respectively), a reactive carbonyl species (crotonaldehyde, 835.11), a host of mixed Cys34 disulfides, notably those of methanethiol (827.09), Cys (851.43), hCys (856.10), CysGly (870.43), GluCys (894.44) and GSH (913.45), and a product of reaction with allylmethylsulfone (a metabolite of garlic). Three of the adducts/features had MS2 spectra indicating modifications of the T3 peptide at sites other than Cys34, including methylation (816.43). We had previously detected 30 of the T3 adducts/features in at least one of six previous studies with serum or plasma from diverse populations (footnoted in Table 1).

As shown in Table 2, 9 Lys525 adducts/features were detected in this study, five of which were putatively annotated. The much smaller number of Lys525 adducts/features with sufficient data for statistical analysis (n=6) compared to that of Cys34 (n=38) reflects the lower reactivity of the ε-amino group of Lys525 compared to the sulfhydryl of Cys34, which is the primary scavenger of reactive electrophiles in serum, as well as steric effects (Aldini et al., 2008). Annotated Lys525 adducts/features include the tryptic miscleaved Lys525 peptide (564.85), the tryptic native peptide showing loss of the Lys525 residue (500.80) as well as products of acetylation (577.86), carbamylation (586.36), and glycation (645.88), all of which had been reported previously (Grigoryan et al., 2021). Four additional Lys525 features were not annotated (586.33, 587.31, 638.82, and 639.87) and were not detected in our previous investigation (Grigoryan et al., 2021).

The 9 adducts/features that were unique to the current investigation include five from Cys34 (851.77 from allylmethylsulfone and four unknowns: 858.41, 867.10, 869.06, and 910.18) and four unknowns from Lys525 (586.33, 587.31, 638.82, and 639.87). The MS2 spectra and SICs/MS1 spectra of these new adducts are reproduced in Supplementary Figures S1, S2 and S3, S4 for the T3 peptide and Lys525 residue, respectively. The raw fold changes (FC) for adduct abundances in subjects with and without DEE exposure ranged from 0.61 to 1.46 (Tables 1 and 2).

3.2. Alterations to the Cys34/Lys525 adductome among DEE-exposed and unexposed subjects

We found evidence that 6 adducts/features in the Cys34 residue were altered between DEE-exposed subjects and unexposed controls (Figure 1). The agreement of aggregated findings from multiple linear regression and penalized logistic regression was 57%. The highest-ranked adducts/features that differed by DEE exposure status according to the nominal p-values from linear regression were 851.76, S-Cys (NH2→OH), p=0.065; 913.45, S-GSH, p=0.029; 860.77, S-hCys(+CH3), p=0.121; and 841.75, S-mercaptoacetic acid, p=0.293. Adduct 851.76 was also highly ranked by random forest (Figures 1B and 1C). These findings were aggregated with the remaining adducts/features selected to be of high importance by random forest (856.10, S-hCys, p=0.141; and 870.43, S-CysGly, p=0.387) (Figure 1C).

As shown in the volcano plot (Figure 1A), three adducts were less abundant in DEE-exposed subjects compared to unexposed subject (FC estimate < 1): 851.76, FC = 0.88; 860.77, FC = 0.83; and 856.10, FC = 0.91. Conversely, three adducts were more abundant in DEE-exposed subjects compared with unexposed subjects (FC estimate > 1): 913.45, FC = 1.18; 870.43, FC = 1.05; and 841.75, FC = 1.07. None of the Lys525 adducts differed between DEE-exposed and unexposed subjects.

To qualitatively confirm the potential exposure-response relationships between DEE and the adducts/features, we evaluated their associations with tertiles of EC, which is an extensively used surrogate measure that reflects DEE exposure levels (Supplementary Tables 3A and 3B). We observed a negative EC exposure-response relationship for 851.76 that was largely driven by the highest EC tertile (54.6–107.7 μg/m^3) (Supplementary Figure S5). We also observed a positive non-linear exposure-response relationship for 913.45 (Supplementary Figure S5).

Results from hierarchical clustering based on Spearman correlation coefficients between selected adducts/features are shown in Figure 2. Adducts/features 860.77 and 856.10 were modestly correlated (Spearman r = 0.61), while 913.45 and 851.76 were weakly correlated (Spearman r = −0.24). Random forest analysis of the selected adducts/features and all the covariates is shown in Figure 3. Here, adducts/features 870.43, 851.76, and 856.10 were found to be of importance across the decision trees. Additionally, CD4 and CD8 counts, lymphocyte counts, MIP-1D, interleukin (IL)-16, and C-reactive protein (CRP) were found to be of importance in the random forest analysis, which further supports their previously reported associations with DEE(Bassig et al., 2017; Dai et al., 2018; Lan et al., 2015).

4. DISCUSSION

Using an untargeted adductomics approach, we investigated whether occupational exposure to DEE was associated with alterations to the Cys34/Lys525 adductome of HSA, which reflects the biological activity of circulating ROS/RES. We found evidence that 6 adducts/features in the Cys34 residue were altered between DEE-exposed subjects and unexposed controls (i.e., 841.75, 851.76, 856.10, 860.77, 870.43, and 913.45). All 6 of the highly ranked adducts/features were previously detected in studies of Cys34 adductomics in diverse

populations. The 6 adducts were biologically related to antioxidant activity, including reactions with cysteine (851.76), homocysteine (856.10 and 860.77), mercaptoacetic acid (841.75), CysGly (870.43), and glutathione (913.45) (Bellamy and McDowell, 1997; Boushey et al., 1995; Dringen et al., 1999; Elias et al., 2005; McBean, 2017). These findings suggest that occupational exposure to DEE could alter downstream biological processes related to detoxification, as reflected by Cys34 adducts.

The three Cys34 adducts/features that were lower among DEE-exposed subjects were derived from reactions with circulating cysteine (851.76) and homocysteine (856.10 and 860.77), which are natural constituents of human blood. We previously observed that Cys34-cysteine adducts were inversely associated with cigarette smoking (Grigoryan et al., 2019), which also has some similar constituents as DEE. In the current study, we did not detect associations between levels of the cysteine adduct and smoking (data not shown), which may be attributed to the narrow variation in smoking in our study population, as most subjects were light-to-moderate smokers (Lan et al., 2015; Wong et al., 2021b). Adducts of homocysteine were previously observed to be positively associated with exposure to benzene (Grigoryan et al., 2018). Homocysteine can be irreversibly degraded to cysteine and metabolized by cystathionine β-synthase (CBS) to cystathionine and ultimately to cysteine by cystathionine γ-lyase (Blom and Smulders, 2011). This process may explain why adducts of these cysteine and homocysteine were highly ranked for association with DEE (Figure 1 B,C).

The three adducts/features that were higher among DEE-exposed subjects were products of reactions with mercaptoacetic acid (841.75), glutathione (913.45) and CysGly (870.43). The mercaptoacetic acid product was previously found to be positively associated with exposure to benzene (Grigoryan et al., 2018), which is a noteworthy component of DEE (Muzyka et al., 1998a; Muzyka et al., 1998b). Since mercaptoacetic acid is commonly used as an organic additive to lubricants (Grigoryan et al., 2018) and to activate hydrotreating catalysts to upgrade oil (Oliviero, 2020), it is possible that diesel fuel may contain mercaptoacetic acid. We have previously shown that the glutathione adduct was negatively associated with exposure to smoky coal (Lu et al., 2017) and with prevalent cases of COPD and ischemic heart disease (Liu et al., 2018). As such, it is interesting that the glutathione adduct was elevated in subjects exposed to DEE in our study, as DEE shares some carcinogenic components with smoky coal and is a risk factor for COPD (Ferguson et al., 2020; Hart et al., 2012).

Our study had various strengths. First, this investigation was conducted in an occupational setting with high DEE concentrations, which improved the likelihood of detecting alterations to the Cys34/Lys525 adductome by DEE exposure status. We had a nearly 18-fold range of exposure (Wong et al., 2021a) and the DEE concentrations in our study were higher than those found in a previous study of trucking industry workers in the U.S. (Davis et al., 2007; Davis et al., 2006; Garshick et al., 2012; Garshick et al., 2008; Smith et al., 2006). Furthermore, we used an ensemble machine learning method that aggregates the findings from three separate models to improve stability, accuracy, and predictive power of the analyses (Cai et al., 2015; Dietterich, 2000; Mark et al., 2019; Sarkar et al., 2021).

Our study had some limitations. First, the sample size was limited; however, the wide difference in DEE concentrations between the DEE-exposed workers and unexposed controls increased the likelihood of detecting effects. Second, we could not discount the possibility of unmeasured or residual confounding. Third, as with most proteomic studies, formation of analytical artifacts is possible and can be problematic when untargeted methods are used to screen for post-translational modifications (Grigoryan et al., 2016). However, to reduce artifact formation, we eliminated steps from our mass spectrometry assay that have been used for adduct enrichment, buffer exchanges, and peptide fractionation, and performed digestion in the absence of tris (2-carboxyethyl) phosphine (Grigoryan et al., 2016). Finally, most of the adduct annotations in Tables 1 and 2 should be regarded as putative, because it was not possible to confirm identities of most adducts/features with reference standards.

5. CONCLUSIONS

In summary, we identified several alterations to the Cys34 adductome among DEE-exposed workers compared with unexposed controls. Our findings suggest that exposure to DEE could promote changes to circulating ROS/RES and support the role of ROS/RES related pathways in the mechanism underlying DEE-related lung carcinogenesis (IARC, 2014). However, caution is recommended when interpreting the findings. We used a hypothesis-generating approach and our results require replication. Further studies are warranted to characterize the interrelationships between exposure to DEE, circulating ROS/RES, inflammatory/immunologic processes (Bassig et al., 2017; Lan et al., 2015), and lung cancer risk (Shiels et al., 2017).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was supported by intramural funding from the National Cancer Institute, as well as grant P42ES04705 from the National Institute for Environmental Health Sciences (to SMR).

REFERENCES

Aldini G, Vistoli G, Regazzoni L, Gamberoni L, Facino RM, Yamaguchi S, Uchida K, Carini M, 2008. Albumin is the main nucleophilic target of human plasma: a protective role against pro-atherogenic electrophilic reactive carbonyl species? Chem Res Toxicol 21, 824–835. [PubMed: 18324789]

Bassig BA, Dai Y, Vermeulen R, Ren D, Hu W, Duan H, Niu Y, Xu J, Shiels MS, Kemp TJ, Pinto LA, Fu W, Meliefste K, Zhou B, Yang J, Ye M, Jia X, Meng T, Wong JYY, Bin P, Hosgood HD 3rd, Hildesheim A, Silverman DT, Rothman N, Zheng Y, Lan Q, 2017. Occupational exposure to diesel engine exhaust and alterations in immune/inflammatory markers: a cross-sectional molecular epidemiology study in China. Carcinogenesis 38, 1104–1111. [PubMed: 28968774]

Bellamy MF, McDowell IF, 1997. Putative mechanisms for vascular damage by homocysteine. J Inherit Metab Dis 20, 307–315. [PubMed: 9211203]

Blom HJ, Smulders Y, 2011. Overview of homocysteine and folate metabolism. With special references to cardiovascular disease and neural tube defects. J Inherit Metab Dis 34, 75–81. [PubMed: 20814827]

Boushey CJ, Beresford SA, Omenn GS, Motulsky AG, 1995. A quantitative assessment of plasma homocysteine as a risk factor for vascular disease. Probable benefits of increasing folic acid intakes. JAMA 274, 1049–1057. [PubMed: 7563456]

Cai Z, Xu D, Zhang Q, Zhang J, Ngai SM, Shao J, 2015. Classification of lung cancer using ensemble-based feature selection and machine learning methods. Mol Biosyst 11, 791–800. [PubMed: 25512221]

Clifford GM, Lise M, Franceschi S, Scherrer AU, 2017. CD4/CD8 ratio and lung cancer risk. Lancet HIV 4, e103.

Dagnino S, Bodinier B, Grigoryan H, Rappaport SM, Karimi M, Guida F, Polidoro S, Edmands WB, Naccarati A, Fiorito G, Sacerdote C, Krogh V, Vermeulen R, Vineis P, Chadeau-Hyam M, 2020. Agnostic Cys34-albumin adductomics and DNA methylation: Implication of N-acetylcysteine in lung carcinogenesis years before diagnosis. Int J Cancer 146, 3294–3303. [PubMed: 31513294]

Dai Y, Ren D, Bassig BA, Vermeulen R, Hu W, Niu Y, Duan H, Ye M, Meng T, Xu J, Bin P, Shen M, Yang J, Fu W, Meliefste K, Silverman D, Rothman N, Lan Q, Zheng Y, 2018. Occupational exposure to diesel engine exhaust and serum cytokine levels. Environ Mol Mutagen 59, 144–150. [PubMed: 29023999]

Davis ME, Smith TJ, Laden F, Hart JE, Blicharz AP, Reaser P, Garshick E, 2007. Driver exposure to combustion particles in the U.S. Trucking industry. J Occup Environ Hyg 4, 848–854. [PubMed: 17885912]

Davis ME, Smith TJ, Laden F, Hart JE, Ryan LM, Garshick E, 2006. Modeling particle exposure in U.S. trucking terminals. Environ Sci Technol 40, 4226–4232. [PubMed: 16856739]

Dietterich TG, 2000. Ensemble Methods in Machine Learning. Multiple Classifier Systems 1857.

Dringen R, Pfeiffer B, Hamprecht B, 1999. Synthesis of the antioxidant glutathione in neurons: supply by astrocytes of CysGly as precursor for neuronal glutathione. J Neurosci 19, 562–569. [PubMed: 9880576]

Elias RJ, McClements DJ, Decker EA, 2005. Antioxidant activity of cysteine, tryptophan, and methionine residues in continuous phase beta-lactoglobulin in oil-in-water emulsions. J Agric Food Chem 53, 10248–10253. [PubMed: 16366723]

Ferguson JM, Costello S, Elser H, Neophytou AM, Picciotto S, Silverman DT, Eisen EA, 2020. Chronic obstructive pulmonary disease mortality: The Diesel Exhaust in Miners Study (DEMS). Environ Res 180, 108876.

Garshick E, Laden F, Hart JE, Davis ME, Eisen EA, Smith TJ, 2012. Lung cancer and elemental carbon exposure in trucking industry workers. Environ Health Perspect 120, 1301–1306. [PubMed: 22739103]

Garshick E, Laden F, Hart JE, Rosner B, Davis ME, Eisen EA, Smith TJ, 2008. Lung cancer and vehicle exhaust in trucking industry workers. Environ Health Perspect 116, 1327–1332. [PubMed: 18941573]

Gowda SN, DeRoos AJ, Hunt RP, Gassett AJ, Mirabelli MC, Bird CE, Margolis HG, Lane D, Bonner MR, Anderson G, Whitsel EA, Kaufman JD, Bhatti P, 2019. Ambient air pollution and lung cancer risk among never-smokers in the Women's Health Initiative. Environ Epidemiol 3, e076. [PubMed: 33778344]

Grigoryan H, Edmands W, Lu SS, Yano Y, Regazzoni L, Iavarone AT, Williams ER, Rappaport SM, 2016. Adductomics Pipeline for Untargeted Analysis of Modifications to Cys34 of Human Serum Albumin. Anal Chem 88, 10504–10512. [PubMed: 27684351]

Grigoryan H, Edmands WMB, Lan Q, Carlsson H, Vermeulen R, Zhang L, Yin SN, Li GL, Smith MT, Rothman N, Rappaport SM, 2018. Adductomic signatures of benzene exposure provide insights into cancer induction. Carcinogenesis 39, 661–668. [PubMed: 29538615]

Grigoryan H, Imani P, Dudoit S, Rappaport SM, 2021. Extending the HSA-Cys34-Adductomics Pipeline to Modifications at Lys525. Chem Res Toxicol 34, 2549–2557. [PubMed: 34788011]

Grigoryan H, Schiffman C, Gunter MJ, Naccarati A, Polidoro S, Dagnino S, Dudoit S, Vineis P, Rappaport SM, 2019. Cys34 Adductomics Links Colorectal Cancer with the Gut Microbiota and Redox Biology. Cancer Res 79, 6024–6031. [PubMed: 31641032]

Hamra GB, Laden F, Cohen AJ, Raaschou-Nielsen O, Brauer M, Loomis D, 2015. Lung Cancer and Exposure to Nitrogen Dioxide and Traffic: A Systematic Review and Meta-Analysis. Environ Health Perspect 123, 1107–1112. [PubMed: 25870974]

Hart JE, Eisen EA, Laden F, 2012. Occupational diesel exhaust exposure as a risk factor for chronic obstructive pulmonary disease. Curr Opin Pulm Med 18, 151–154. [PubMed: 22234274]

Hart JE, Spiegelman D, Beelen R, Hoek G, Brunekreef B, Schouten LJ, van den Brandt P, 2015. Long-Term Ambient Residential Traffic-Related Exposures and Measurement Error-Adjusted Risk of Incident Lung Cancer in the Netherlands Cohort Study on Diet and Cancer. Environ Health Perspect 123, 860–866. [PubMed: 25816363]

IARC, 2014. IARC Monogr Eval Carcinog Risks Hum 105, 9–699. [PubMed: 26442290]

Lan Q, Vermeulen R, Dai Y, Ren D, Hu W, Duan H, Niu Y, Xu J, Fu W, Meliefste K, Zhou B, Yang J, Ye M, Jia X, Meng T, Bin P, Kim C, Bassig BA, Hosgood HD 3rd, Silverman D, Zheng Y, Rothman N, 2015. Occupational exposure to diesel engine exhaust and alterations in lymphocyte subsets. Occup Environ Med 72, 354–359. [PubMed: 25673341]

Liu S, Grigoryan H, Edmands WMB, Dagnino S, Sinharay R, Cullinan P, Collins P, Chung KF, Barratt B, Kelly FJ, Vineis P, Rappaport SM, 2018. Cys34 Adductomes Differ between Patients with Chronic Lung or Heart Disease and Healthy Controls in Central London. Environ Sci Technol 52, 2307–2313. [PubMed: 29350914]

Lu SS, Grigoryan H, Edmands WM, Hu W, Iavarone AT, Hubbard A, Rothman N, Vermeulen R, Lan Q, Rappaport SM, 2017. Profiling the Serum Albumin Cys34 Adductome of Solid Fuel Users in Xuanwei and Fuyuan, China. Environmental science & technology 51, 46–57. [PubMed: 27936627]

Mark E, Goldsman D, Gurbaxani B, Keskinocak P, Sokol J, 2019. Using machine learning and an ensemble of methods to predict kidney transplant survival. PLoS One 14, e0209068.

McBean GJ, 2017. Cysteine, Glutathione, and Thiol Redox Balance in Astrocytes. Antioxidants (Basel) 6.

Muzyka V, Veimer S, Schmidt N, 1998a. On the carcinogenic risk evaluation of diesel exhaust: benzene in airborne particles and alterations of heme metabolism in lymphocytes as markers of exposure. Sci Total Environ 217, 103–111. [PubMed: 9695175]

Muzyka V, Veimer S, Shmidt N, 1998b. Particle-bound benzene from diesel engine exhaust. Scand J Work Environ Health 24, 481–485. [PubMed: 9988090]

Oliviero L, Maugé F, Afanasiev P, Pedraza Parra C, Geantet C, 2020. Organic additives for hydrotreating catalysts: a review of main families and action mechanisms. Catalysis Today.

Petrick LM, Schiffman C, Edmands WMB, Yano Y, Perttula K, Whitehead T, Metayer C, Wheelock CE, Arora M, Grigoryan H, Carlsson H, Dudoit S, Rappaport SM, 2019. Metabolomics of neonatal blood spots reveal distinct phenotypes of pediatric acute lymphoblastic leukemia and potential effects of early-life nutrition. Cancer Lett 452, 71–78. [PubMed: 30904619]

Puett RC, Hart JE, Yanosky JD, Spiegelman D, Wang M, Fisher JA, Hong B, Laden F, 2014. Particulate matter air pollution exposure, distance to road, and incident lung cancer in the nurses' health study cohort. Environ Health Perspect 122, 926–932. [PubMed: 24911062]

Rahman ML, Bassig BA, Dai Y, Hu W, Wong JYY, Blechter B, Hosgood HD, Ren D, Duan H, Niu Y, Xu J, Fu W, Meliefste K, Zhou B, Yang J, Ye M, Jia X, Meng T, Bin P, Silverman DT, Vermeulen R, Rothman N, Zheng Y, Lan Q, 2021. Proteomic analysis of serum in workers exposed to diesel engine exhaust. Environ Mol Mutagen.

Rappaport SM, Li H, Grigoryan H, Funk WE, Williams ER, 2012. Adductomics: characterizing exposures to reactive electrophiles. Toxicol Lett 213, 83–90. [PubMed: 21501670]

Risso D, Ngai J, Speed TP, Dudoit S, 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol 32, 896–902. [PubMed: 25150836]

Sarkar JP, Saha I, Sarkar A, Maulik U, 2021. Machine learning integrated ensemble of feature selection methods followed by survival analysis for predicting breast cancer subtype specific miRNA biomarkers. Comput Biol Med 131, 104244.

Shiels MS, Shu XO, Chaturvedi AK, Gao YT, Xiang YB, Cai Q, Hu W, Shelton G, Ji BT, Pinto LA, Kemp TJ, Rothman N, Zheng W, Hildesheim A, Lan Q, 2017. A prospective study of immune

and inflammation markers and risk of lung cancer among female never smokers in Shanghai. Carcinogenesis 38, 1004–1010. [PubMed: 28981818]

Silverman DT, Samanic CM, Lubin JH, Blair AE, Stewart PA, Vermeulen R, Coble JB, Rothman N, Schleiff PL, Travis WD, Ziegler RG, Wacholder S, Attfield MD, 2012. The Diesel Exhaust in Miners study: a nested case-control study of lung cancer and diesel exhaust. J Natl Cancer Inst 104, 855–868. [PubMed: 22393209]

Smith TJ, Davis ME, Reaser P, Natkin J, Hart JE, Laden F, Heff A, Garshick E, 2006. Overview of particulate exposures in the US trucking industry. J Environ Monit 8, 711–720. [PubMed: 16826284]

Turner MC, Andersen ZJ, Baccarelli A, Diver WR, Gapstur SM, Pope CA 3rd, Prada D, Samet J, Thurston G, Cohen A, 2020. Outdoor air pollution and cancer: An overview of the current evidence and public health recommendations. CA Cancer J Clin.

Wong JYY, Cawthon R, Dai Y, Vermeulen R, Bassig BA, Hu W, Duan H, Niu Y, Downward GS, Leng S, Ji BT, Fu W, Xu J, Meliefste K, Zhou B, Yang J, Ren D, Ye M, Jia X, Meng T, Bin P, Hosgood Iii HD, Silverman DT, Rothman N, Zheng Y, Lan Q, 2021a. Elevated Alu retroelement copy number among workers exposed to diesel engine exhaust. Occup Environ Med.

Wong JYY, Cawthon R, Dai Y, Vermeulen R, Bassig BA, Hu W, Duan H, Niu Y, Downward GS, Leng S, Ji BT, Fu W, Xu J, Meliefste K, Zhou B, Yang J, Ren D, Ye M, Jia X, Meng T, Bin P, Hosgood Iii HD, Silverman DT, Rothman N, Zheng Y, Lan Q, 2021b. Elevated Alu retroelement copy number among workers exposed to diesel engine exhaust. Occup Environ Med 78, 823–828. [PubMed: 34039759]

Wong JYY, Jones RR, Breeze C, Blechter B, Rothman N, Hu W, Ji BT, Bassig BA, Silverman DT, Lan Q, 2021c. Commute patterns, residential traffic-related air pollution, and lung cancer risk in the prospective UK Biobank cohort study. Environ Int 155, 106698.

Yano Y, Schiffman C, Grigoryan H, Hayes J, Edmands W, Petrick L, Whitehead T, Metayer C, Dudoit S, Rappaport S, 2020. Untargeted adductomics of newborn dried blood spots identifies modifications to human serum albumin associated with childhood leukemia. Leuk Res 88, 106268.

HIGHLIGHTS

- Cys34/Lys525 adducts reflect long-lasting effects of reactive oxygen/electrophilic species.
- Investigated Cys34/Lys525 adductome alterations among diesel engine exhaust exposed workers.
- Six Cys34 adducts were altered between exposed and unexposed participants.
- Reactive oxygen/electrophilic species are involved in carcinogenicity of diesel exhaust.

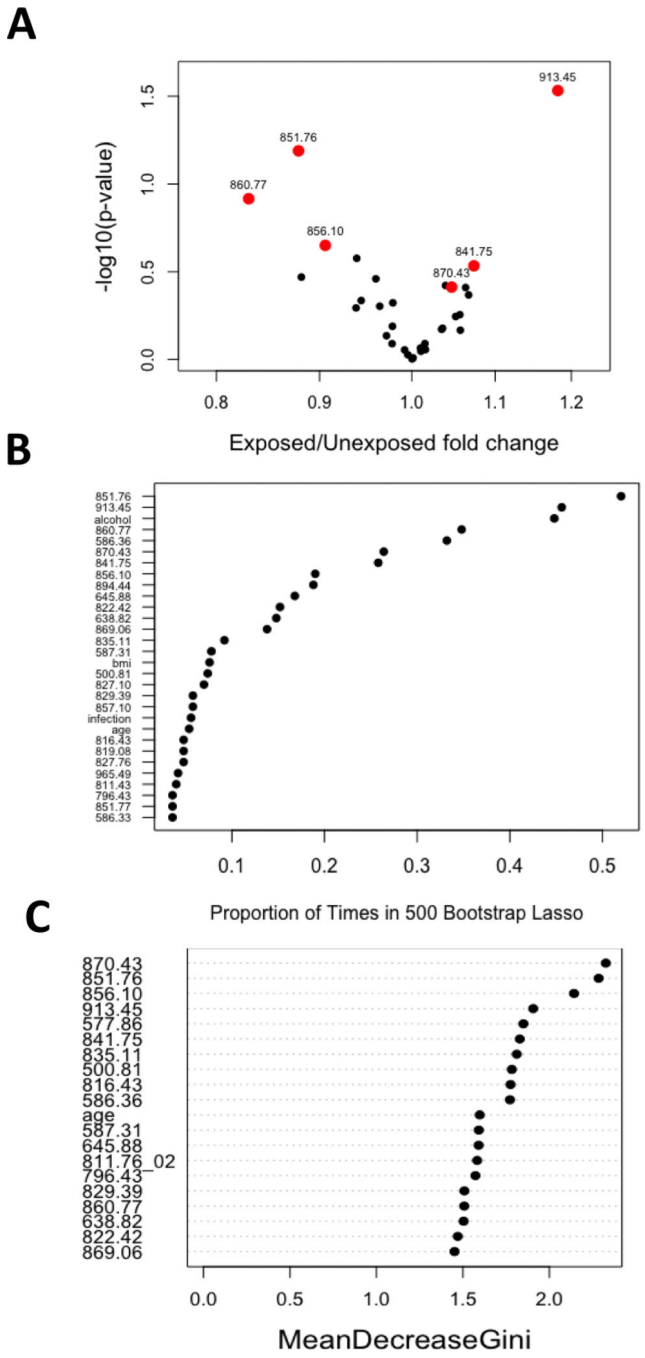

Figure 1. An ensemble machine learning feature selection approach to detect alterations to the Cys34/Lys525 adductome by DEE exposure status.

This ensemble approach aggregates the findings from three individual modeling approaches that estimate different kinds of DEE-adduct associations. The agreement among findings from linear and penalized logistic regression (LASSO) is estimated, then combined with findings from random forest, to select the adducts/features that differ between DEE-exposed workers and unexposed controls. **A.** Volcano plot of nominal p-values for each adduct in the linear regression model comparing DEE-exposed workers to unexposed controls. **B.** Proportion of times that each feature was selected by penalized logistic regression (LASSO)

of DEE-exposure vs. unexposed status. **C.** Ranked variable importance measures for random forest classification of DEE-exposure vs. unexposed status.

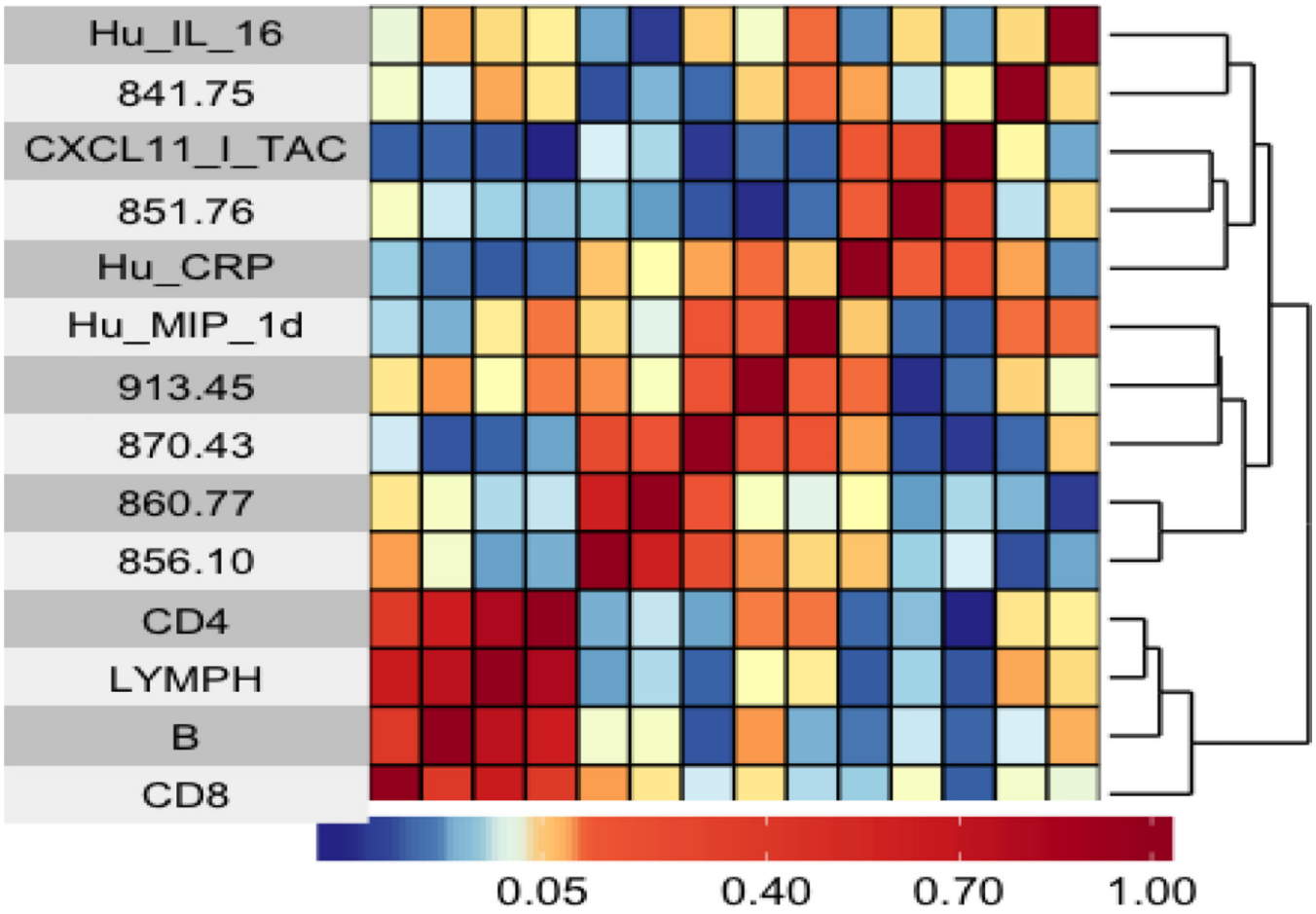

Figure 2. Heatmap showing hierarchical clustering of selected features and covariates using Spearman correlation coefficients.

Darker red cells indicate a stronger positive correlation and darker blue cells indicate a stronger negative correlation

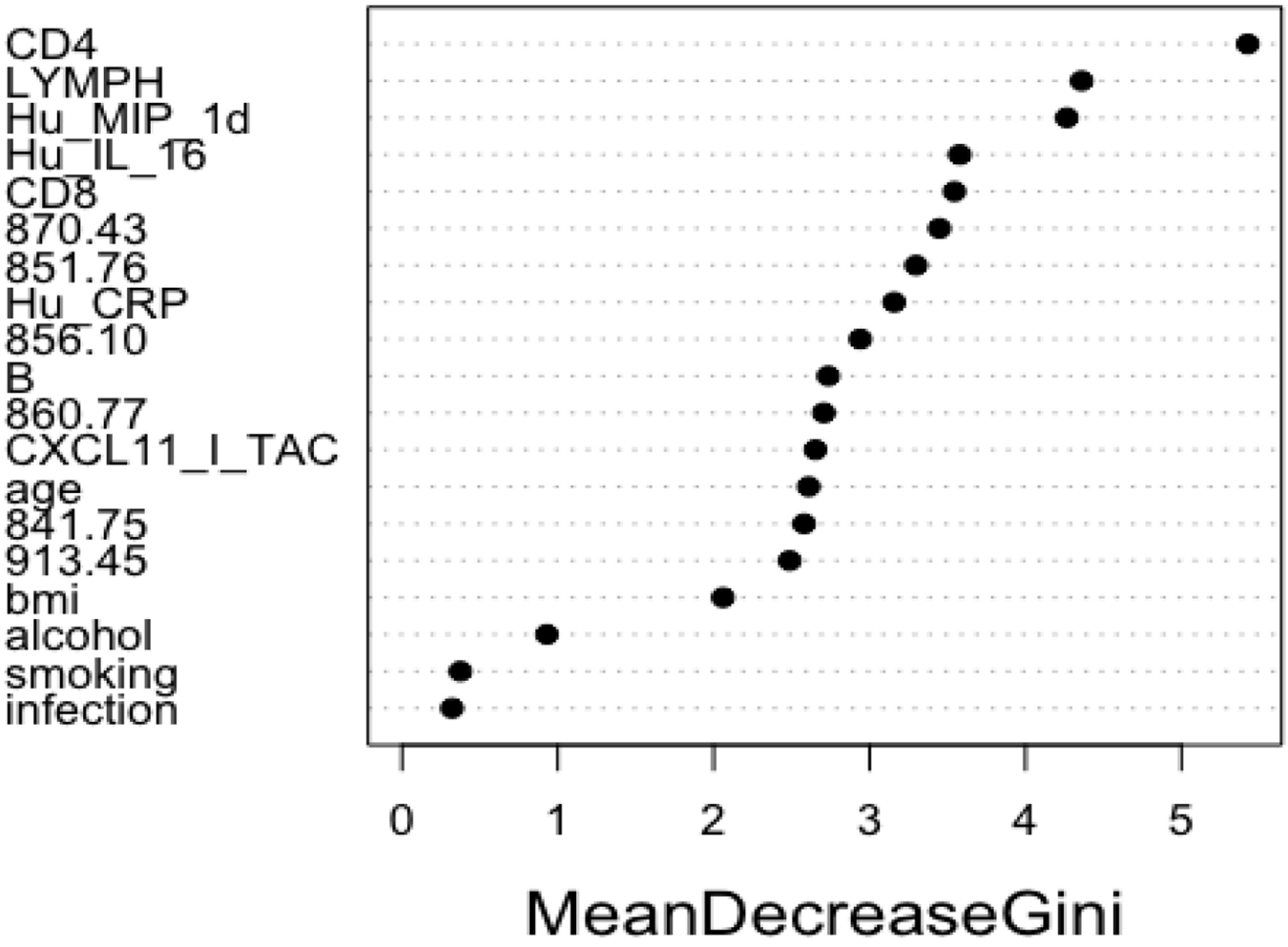

Figure 3.
Random forest variable importance plot for predicting DEE exposure with selected adducts and covariates.

Table 1.

Features of the T3 peptide containing Cys34 and its modifications.

T3 Peptide feature	RT, min	MIM observed (*m/z*, +3)	MIM theoretical (*m/z*, +3)	ΔMass (ppm)	Added mass (Da)	Elemental composition	Annotation	PAR x10k	Fold change (DEE-exposed/ unexposed)
I) Detected features with altered levels between DEE-exposed workers and unexposed controls [*]									
841.75[a,b,c,d,e,f]	24.68	841.7512	841.7519	−0.832	90.986	$+C_2H_3O_2S$	S-Mercaptoacetic acid	1.5	1.13
851.76[a,b,c,d,e,f]	24.23	851.7549	851.7554	−0.587	120.997	$+C_3H_5O_3S$	S-Cys(NH2→OH)	4.4	0.91
856.10[a,b,c,d,e,f]	22.98	856.099	856.0993	−0.350	134.029	$+C_4H_8NO_2S$	S-hCys[g]	58.5	0.92
860.77[b,d,e,f]	23.27	860.7688	860.7712	−2.788	148.039	$+C_5H_{10}NO_2S$	S-hCys (+CH3)	2.1	0.77
870.43[a,b,c,d,e,f]	21.96	870.4342	870.4345	−0.345	177.035	$+C_5H_9N_2O_3S$	S-CysGly[g]	32.9	1.11
913.45[a,b,c,d,e,f]	23.12	913.4485	913.4487	−0.219	306.078	$+C_{10}H_{16}N_3O_6S$	S-Glutathione[g] (GSH)	2.8	1.22
II) Detected features that were not altered between DEE-exposed workers and unexposed controls									
796.43[a,b,c, f]	23.93	796.4290	796.4301	−1.356	−45.992	$-CH_2S$	Cys34→Gly	1.8	1.04
808.73[a,b,c,d,e,f]	25.00	808.7288			−9.089		Not Cys34 adduct	1.6	1.08
811.43[a,b,c,d,e,f]	27.19	811.4242	811.4234	0.986			T3 dimer[g,h]	22.1	1.01
811.76_01[a,b,c,d,f]	22.73	811.7597			0.004		T3 labile adduct	1.9	0.61
811.76_02[a,b,c,d,e,f]	24.75	811.7585	811.7594	−1.109	0.000		Unmodified T3[g]	124.3	1.03
816.42[a,b,c,d,e,f]	24.29	816.4185	816.4191	−0.712	13.977	$-H_2$, +O	Cys34-Gln[g]	ND	ND
816.43[a,b,c,d,e,f]	25.09	816.4308	816.4312	−0.489	15.025	$+CH_3$	Methylation (not Cys34)	6.9	1.03
819.08[b,e]	24.75	819.0858	819.0867	−1.038	22.990	−H, +Na	Na adduct of T3	3.3	1.01
822.42[a,b,c,d,e,f]	24.34	822.4221	822.4226	−0.608	32.999	$+HO_2$	Cys34 sulfinic acid[g]	36.5	1.06
826.43[a,f]	23.86	826.4343	826.4348	−0.605	45.035	$+C_2H_5O$	Ethylene oxide	2.7	1.46
827.09[b,c,f]	25.87	827.0876	827.0886	−1.227	46.992	$+CH_3S$	S-Methanethiol[g]	1.0	0.94
827.10[c,d,f]	24.69	827.0944	827.0945	−0.121	47.015	$+CH_3O_2$	S-(O)-O-CH3	2.5	1.01
827.76[a,b,c,d,e,f]	24.68	827.7539	827.7543	−0.483	48.994	$+HO_3$	Cys34 sulfonic acid[g]	2.1	1.03
829.39[a,b]	24.78	829.3962			53.921		Unknown	2.5	1.00
829.43[a,f]	24.94	829.4348	829.4349	−0.121	54.037	$+C_3H_4N$	Acrylonitrile	1.2	0.82
835.11[a,c,e,f]	24.97	835.1057	835.1066	−1.062	71.048	$+C_4H_7O$	Crotonaldehyde[g]	1.0	0.97
845.42[a,b,c,d,e,f]	23.71	845.4232	845.4239	−0.828	102.002	$+C_3H_4NOS$	S-Cys (−H2O)	3.6	1.06
850.10[a,b,c,f]	24.51	850.0952	850.0958	−0.706	116.018	$+C_4H_6NOS$	S-hCys (-H2O)	2.1	0.99

T3 Peptide feature	RT, min	MIM observed (*m/z*, +3)	MIM theoretical (*m/z*, +3)	ΔMass (ppm)	Added mass (Da)	Elemental composition	Annotation	PAR x10k	Fold change (DEE-exposed/ une xposed)
851.43[a,b,c,d,e,f]	22.65	851.4264	851.4274	−1.174	120.012	$+C_3H_6NO_2S$	S-Cys[g]	351.1	1.07
851.77	24.77	851.7673	851.7675	−0.235	121.034	$C_4H_9O_2S$	Allylmethylsulfone	5.1	0.99
856.43[c,f]	24.44	856.4269	856.4273	−0.467	135.013	$+C_4H_7O_3S$	S-hCys (NH2→OH)	ND	ND
857.10[a,b,c,f]	23.83	857.0986	857.0992	−0.700	137.028	$+C_4H_9O_3S$	Unknown	58.5	1.06
858.75[a,b,c,d,e]	22.70	858.7542			141.995	$C_3H_5NO_2SNa$	Na adduct of S-Cys	7.5	1.07
862.09[a,c,e]	22.78	862.0914			152.007		not Cys34 adduct	2.9	1.13
863.43[e]	23.00	863.4266	863.4266	0.035	156.012	$+C_4H_7NO_2SNa$	Na adduct of S-hCys	2.4	0.86
864.08[a,b,c,e,f]	22.76	864.0762			157.961		Not Cys34 adduct	4.1	1.21
867.10	25.95	867.1012			167.036		Unknown	1.1	1.00
869.06	22.71	869.0644			172.926		Not Cys34 adduct	5.3	1.15
875.11[b,d,f]	22.55	875.1078			191.056		Not Cys34 adduct	ND	ND
894.44[a,b,c,d,e,f]	23.28	894.4412	894.4416	−0.447	249.056	$+C_8H_{13}N_2O_5S$	S-γ-GluCys[g]	2.7	1.14
910.18	30.59	910.1769			296.263	$C_{18}H_{34}NO_2$	Unknown	1.2	0.81
931.82[b,c,d,f]	20.88	931.8187			361.188		Unknown	ND	ND
965.49[b,c,d,e,f]	21.25	965.4906			462.204		Unknown	9.0	1.02
976.82[b,d]	23.24	976.8197			496.191		Unknown	ND	ND

[*] See Figure 1 for the adducts/features determined by our ensemble machine learning feature selection analysis to be highly ranked for differences between DEE-exposed and unexposed subjects.

Legend: MIM, monoisotopic mass; ND, not determined; PAR, ratio of adduct peak abundance to housekeeping peptide peak abundance; RT, retention time

[a] Grigoryan et al, Anal. Chem. 2016, 88, 10504−10512

[b] Lu et al, Environ Sci Technol. 2017 Jan 3;51(1):46–57

[c] Grigoryan et al, Carcinogenesis, 2018, Vol. 39, No. 5, 661–668

[d] Liu et al, Environ. Sci. Technol. 2018, 52, 2307−2313

[e] Dagnino et al, Int. J. Cancer: 2020, 146, 3294–3303

[f] Grigoryan et al, Cancer Res 2019;79:6024–31

[g] Annotation confirmed with a synthetic standard.

[h] +6 charge state.

Table 2.

Features of the Lys525 containing peptide and its modifications.

Lys525 feature	RT, min	MIM observed (*m/z*, +2)	MIM theoretical (*m/z*, +2)	ΔMass (ppm)	Added mass (Da)	Elemental composition	Annotation	PAR	Fold change (DEE-exposed/ unexposed)
500.81[a]	8.68	500.8050	500.8055	0.998	−128.093	$-C_6H_{12}N_2O$	Loss of N-terminus lysine	0.54	1.01
564.85[a]	6.31	564.8521	564.853	1.593	0		Lys525 containing peptide	90.25	1.02
577.86[a]	9.96	577.8606	577.8608	0.346	27.025	$+C_2H_3$	Acetylation	1.81	0.97
586.33	13.96	586.3319					Unknown	5.56	0.99
586.36[a]	10.68	586.3556	586.3559	0.511	44.014	$+CH_2NO$	Carbamylation	0.38	1.01
587.31	10.71	587.3110					Unknown	25.69	0.98
638.82	11.02	638.8195					Unknown	0.11	0.99
639.87	15.47	639.8681					Unknown	0.33	0.98
645.88[a]	6.49	645.8793	645.8794	0.154	163.062	$+C_6H_{11}O_5$	Fructosyl lysine (glycation)	40.42	0.98

Legend: MIM, monoisotopic mass; PAR, ratio of adduct peak abundance to housekeeping peptide peak abundance; RT, retention time

[a]Grigoryan et al, Chem. Res. Toxicol., 2021.