

UC Berkeley

UC Berkeley Previously Published Works

Title

HSA Adductomics Reveals Sex Differences in NHL Incidence and Possible Involvement of Microbial Translocation.

Permalink

<https://escholarship.org/uc/item/4qg540w3>

Journal

Cancer Epidemiology Biomarkers & Prevention, 32(9)

ISSN

1055-9965

Authors

Grigoryan, Hasmik

Imani, Partow

Sacerdote, Carlotta

et al.

Publication Date

2023-09-01

DOI

10.1158/1055-9965.epi-23-0231

Peer reviewed



Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2023 September 01; 32(9): 1217–1226.

doi:10.1158/1055-9965.EPI-23-0231.

HSA Adductomics Reveals Sex Differences in NHL Incidence and Possible Involvement of Microbial Translocation

Hasmik Grigoryan^{1,†}, Partow Imani^{1,†}, Carlotta Sacerdote², Giovanna Masala³, Sara Grioni⁴, Rosario Tumino⁵, Paolo Chiodini⁶, Sandrine Dudoit^{1,7}, Paolo Vineis^{2,8}, Stephen M. Rappaport^{1,*}

¹ School of Public Health, University of California, Berkeley, California, 94720, United States

² Unit of Cancer Epidemiology Città della Salute e della Scienza University-Hospital, 10126, Turin, Italy.

³ Institute of Cancer Research, Prevention and Clinical Network (ISPRO), 50139, Florence, Italy

⁴ Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, 20133 Milan, Italy.

⁵ Hyblean Association for Epidemiological Research, AIRE-ONLUS, 97100, Ragusa, Italy

⁶ Dipartimento di Salute Mentale e Fisica e Medicina Preventiva, Università degli Studi della Campania 'Luigi Vanvitelli', 80138, Naples, Italy

⁷ Department of Statistics, University of California, Berkeley, CA, 94720, United States

⁸ MRC-PHE Centre for Environment and Health, Imperial College, Norfolk Place London W21PG, UK

Abstract

Background: The higher incidence of non-Hodgkin's lymphoma (NHL) in males is not well understood. Although reactive oxygen species (ROS) have been implicated as causes of NHL, they cannot be measured directly in archived blood.

Methods: We performed untargeted adductomics of stable ROS adducts in human serum albumin (HSA) from 67 incident NHL cases and 82 matched controls from the EPIC-Italy cohort.

Regression and classification methods were employed to select features associated with NHL in all subjects and in males and females separately.

* **Corresponding author:** Stephen M. Rappaport, School of Public Health, University of California, Berkeley, California, 94720, United States, srappaport@berkeley.edu, tel: 510 334-8128.

† H.G. and P.I. contributed equally to this work.

Author contributions

H.G. and P.I. contributed equally to this work with H.G. generating the analytical data and annotations and with P.I. performing statistical analyses; both wrote the corresponding methods sections. C.S, G.M., S.G., R.T. and P.C. recruited subjects and collected follow-up and covariate data. P.V. supervised portions of the work and obtained financial support. S.D. supervised statistical analyses, obtained financial support and edited the manuscript. S.R. supervised most of the work, wrote the draft and final manuscripts and obtained financial support.

Conflict of interest disclosure statement

The authors declare no potential conflicts of interest.

Results: Sixty seven HSA-adduct features were quantified by liquid chromatography-high-resolution mass spectrometry at Cys34 ($n=55$) and Lys525 ($n=12$). Three features were selected for association with NHL in all subjects, while seven were selected for males and five for females with minimal overlap. Two selected features were more abundant in cases and seven in controls, suggesting that altered homeostasis of ROS may affect NHL incidence. Heatmaps revealed differential clustering of features between sexes, suggesting differences in operative pathways.

Conclusions: Adduct clusters dominated by Cys34 oxidation products and disulfides further implicate ROS and redox biology in the etiology of NHL. Sex differences in dietary and alcohol consumption also help to explain the limited overlap of feature selection between sexes. Intriguingly, a disulfide of methanethiol from enteric microbial metabolism was more abundant in male cases, thereby implicating microbial translocation as a potential contributor to NHL in males.

Impact: Only two of the ROS adducts associated with NHL overlapped between sexes and one adduct implicates microbial translocation as a risk factor.

Introduction

The non-Hodgkin lymphomas (NHL) comprise a group of malignancies derived primarily from B- and T-cells in lymphoid tissue. According to the Italian Association of Cancer Registries and the European Cancer Information System, NHL accounts for approximately 3% of all cancers in Italy with estimated 2020 incidences of 7,000 cases in males and 6,200 in females (1). From the 1970s to 1990s the age-adjusted incidence of NHL in Italy increased by 3–4% per year and then stabilized in the 2000s, (2) and similar trends were observed in other westernized countries (3). Age-standardized incidence rates varied world-wide from 4 to 19 per 100,000 males and 3 to 12 per 100,000 females, with the higher rates observed in westernized countries (3,4).

The most prominent risk factor for NHL is immune deficiency – whether congenital or acquired from infections - with incidence rates among immuno-compromised individuals being many times that of the general population (5). Although relative NHL risks of 100 or more have been reported for individuals infected with HIV, more modest risks have been associated with diverse bacterial and viral infections [reviewed by (4)], e.g., a relative risk of 3.59 has been reported following chronic infection with hepatitis-C virus (6).

Infections activate phagocytic cells that release reactive oxygen species (ROS) to kill foreign organisms (7). Yet, ROS are generated naturally, e.g. by mitochondrial membranes in cells and via activation of NADPH oxidases in cytosol, and are detoxified by antioxidants under homeostasis. However, excessive production of ROS can overwhelm natural defenses and promote a cascade of events leading to metastatic cancers (8). The hypothesis that ROS are mechanistically linked to NHL is supported by increased risks with polymorphisms in oxidative-stress genes (9,10) and type 2 diabetes mellitus (11), and by reduced risks with intake of dietary antioxidants (12,13).

Because ROS cannot be measured in archived blood, investigators have studied their stable adducts with nucleophilic loci of blood proteins, particularly Cys34 of HSA, which is the

dominant scavenger of ROS in the interstitial space (14). Not only does Cys34 react directly with ROS, oxidation of its free sulfhydryl group to the unstable sulfenic acid (Cys34-SOH) leads to mixed-Cys34 disulfides of circulating thiols, including cysteine, glutathione and its precursors, glutamyl cysteine (GluCys) and cysteinyl glycine (CysGly), that are sentinels of redox biology (15,16).

We developed an adductomics pipeline that employs nano-liquid chromatography-high-resolution mass spectrometry (nLC-HRMSMS) to measure all detectable Cys34 adducts in tryptic digests of human serum/plasma, which yield the triply-charged T3 peptide ($^{21}\text{ALVLIAFAQYLQQC}^{34}\text{PFEDHVK}^{41}$, m/z 811.7593) and its modifications (17). Previous applications of Cys34 adductomics for cancer etiology employed archived serum from incident cases of colorectal cancer (CRC) (18) and lung cancer (19) and matched controls from the Italian subset of the European Prospective Investigation into Cancer and Nutrition (EPIC-Italy), a cohort of healthy European adults enrolled between 1992 and 2000 (20). Those investigations found associations between adducts and cancer incidence up to 14 years prior to diagnosis and added mechanistic insights. A third application analyzed HSA from neonatal blood spots in childhood leukemia cases and matched controls and reported associations between adducts and incident cases of acute lymphoblastic and myeloid leukemias (21).

Since not all ROS react with Cys34, we recently extended our HSA-adductomics pipeline to include ϵ -amino modifications to Lys525, a hotspot for adduction of glycation end products and Schiff bases (22,23). The combined Cys34/Lys525 pipeline interrogates modifications to both the T3 peptide and the doubly-charged miscleaved Lys525 peptide ($^{525}\text{KQTALVELVK}^{534}$, m/z 564.8529). Here we report the first application of the Cys34/Lys525 pipeline to cancer etiology, using incident NHL cases and matched controls from the EPIC-Italy cohort. We conducted this exploratory study to discover adducts that discriminate for NHL incidence, with particular emphasis on sex differences and potential roles played by ROS and redox biology.

Materials and Methods

Cases and controls

Serum specimens from 67 incident cases of NHL (59 B-cell and 8 T-cell cases tabulated in Supplementary Table S1) and 82 controls were obtained from the EPIC-Italy cohort of 47,749 volunteers recruited in five centers in Italy between 1993 and 1998 (24). Written informed consent was obtained from all participants and the study was conducted in accordance with the World Medical Association Declaration of Helsinki. The study protocol was approved by an institutional review board of the Human Genetics Foundation (Turin, Italy). Participants were followed-up from the date of entry in the cohort until the occurrence of any cancer (except non-melanoma skin cancer), death, emigration, or end of follow-up (between 2009 and 2014, depending on the center). NHL cases were identified through linkage with cancer and mortality registries by the morphology codes morphology 967–972 of the International Classification of Diseases (ICD-O-2). Controls from the cohort were matched to cases (1:1) by age, year and season of enrollment, and sex. After removal of 15 cases due to misclassified cancer type or missing covariate data, the corresponding 15

controls were retained in the analyses for feature selection. Information about diet, body mass index (bmi) and lifestyle factors were collected with questionnaires (including FFQ) (25). Table 1 lists summary statistics for the NHL cases and controls and selected covariates based on previous associations with NHL risk (4,12,13,26). Across our subjects, no evidence of significant differences between NHL cases and controls were observed (nominal p -values 0.23, Table 1). Table 1 also provides statistics for time (from recruitment) to case diagnosis (ttd, days).

Laboratory analysis and data acquisition

Serum samples had been stored in liquid nitrogen for approximately 20–25 years prior to shipment to our laboratory where they were maintained at -80°C for a few months prior to analysis in 2018. Laboratory investigators were blinded with regard to disease status except that case/control pairs were analyzed in the same daily analytical batch and in random order to reduce effects of technical variation. Sample processing and analysis by nLC-HRMSMS were performed with duplicate injections as previously described (22,23) and elaborated in Supplementary Methods S1. Briefly, after digesting HSA with trypsin, an isotopically labeled and iodoacetamide-modified T3 peptide (IAA-iT3) was added as an internal standard to normalize data for instrument performance. One microliter of each digest was then analyzed by nLC-HRMSMS to locate spectra from the T3 and Lys525 peptides (17,22). The corresponding precursor ions were extracted to obtain a monoisotopic mass (MIM) for each peptide feature (hereafter simply ‘feature’) and assign accurate masses. To normalize peak areas for the amount of HSA in each sample, the MIM was also extracted for the robust HSA peptide $^{42}\text{LVNEVTEFAK}^{51}$ ($m/z=575.3111$), adjacent to the T3 peptide and referred to as the housekeeping peptide (HKP). The peak area ratio (PAR) of the adduct peak to that of the HKP is a robust measure of the adduct concentration (17). Peak picking and integration were performed with the average MIMs and retention times. Added masses relative to the Cys34 thiolate ion and the Lys525 peptide were estimated as described in Supplementary Methods S1. Putative adducts were annotated based on a combination of retention time, accurate mass, elemental composition, database searches and available reference standards.

Statistical analysis

Data consist of nLC-HRMSMS peak areas for 57 Cys34-containing features and 12 Lys525-containing features from 148 samples analyzed with duplicate injections and one with a single injection ($n=297$). Duplicate injections are used to estimate the amount of technical variation in nLC-HRMSMS analysis and averaging them lessens the impact of this source of variability. As previously described (18), peak areas were log-transformed and intraclass correlation coefficients (ICC) and coefficients of variation (CV) were estimated using a one-way random effects model where the log-abundance of the feature was the subject-specific effect and duplicate injections were sample-specific effects. Using an empirical ICC cutoff of >0.2 to remove features with large proportions of technical variation resulted in removal of five features, leaving 62 for further analysis. For each feature, log-abundances from 148 duplicate injections were averaged, ignoring missing values, and those with more than 30% missing values were removed. This resulted in the additional removal of 27 features, leaving 35 for further analysis. When values were missing from both duplicate injections,

missing values were imputed using k-nearest-neighbor imputation with the number of neighboring features set to five. The *scone* package (<https://rdrr.io/bioc/scone/>) was used to select a normalization scheme with inputs consisting of the imputed abundance matrix, the analytical-batch variable, a binary case/control variable, and the matrix of sample-level quality control (QC) measures (i.e., HKP abundance, internal standard abundance and sample run order). Based on performance measures returned by *scone*, the data were normalized using upper-quartile scaling while adjusting for case/control status, batch, and all three principal components of the QC matrix.

Statistical analyses of the normalized data were performed in the R environment as previously described (18,21). To capture associations between features and case/control status for hypothesis generation, features were selected using an ensemble approach based on multivariate linear regression, regularized logistic regression (LASSO), and Random Forest. First, the following linear regression model was used for a given logged and scaled feature abundance Y in the i th subject:

$$E[Y_i|X_i, Q_i] = \beta_0 + \beta_1 X_{i, \text{case/control}} + \beta_2 X_{i, \text{sex}} + \beta_3 X_{i, \text{age}} + \beta_4 Q_{i,1} + \beta_5 Q_{i,2} + \beta_6 Q_{i,3} + \sum_{j: \{7:20\}} \beta_j X_{i, \text{batch}_j - 6} \quad (1)$$

where: *case/control* denotes binary status, Q represents the matrix of QC principal components, *sex* and *age* (at recruitment) are self-explanatory, and *batch* is the analytical sample batch. Features were ranked by their nominal unadjusted p -values for testing the null hypothesis that the case/control coefficient (β_1) is zero using a t -test, and the case/control fold-change (FC) was estimated as $\exp(\hat{\beta}_1)$. Second, a LASSO model was fit to normalized feature abundances over 500 bootstrapped datasets, with *case/control* status as the outcome variable and the following independent variables: normalized logged abundances for all 35 features, and the covariates *age* and *sex*. The percentage of times that each feature was selected by LASSO was used to rank its association with NHL. The same input was given to Random Forest to rank features by their mean decreases in Gini index. Features highly ranked by concordance plots for both linear regression nominal p -value and LASSO were joined with those of high importance from Random Forest, to provide additional evidence of associations with NHL, including non-linear associations. Given prior evidence that NHL incidence was consistently greater in males, feature selection was repeated separately for males and females.

Correlations between normalized feature abundances were displayed with agglomerative hierarchical clustering using complete linkage of Spearman correlation values ('superheat' function in R from the superheat package). Associations between selected features and covariates (smoking, bmi, caloric intake, consumption of meat, vegetables and alcohol, physical activity, and age) were further ranked with Random Forest.

Because NHL cases and matched controls were evaluated up to 21 years after recruitment, we used permutation tests to check for associations between relative adduct abundance (case/matched control) and days from recruitment to diagnosis (ttd) to differentiate between potentially causal effects or reactive effects of disease progression (reverse causality) (18,27,28). This was done by randomly permuting the ttd of a given case and the log(FC) 10,000 times (thereby breaking any association between the two), each time recording

the slope of the best fit line when regressing $\log(\text{FC})$ on t_{td} . The p -value reported is the percentage of the 10,000 random permutations that the absolute value of the estimated slope was greater than or equal to the absolute value estimated using the observed data. If a significant linear trend in $\log(\text{FC})$ for a given feature was detected with increasing days to diagnosis, the adduct feature was considered as potentially reactive.

Data Availability

The data generated in this study are available upon request from the corresponding author.

Results

The 57 Cys34 and 12 Lys525 features are listed in Supplementary Tables S2.A and S2.B, respectively. For each feature, the LC retention time, observed and theoretical MIMs and the corresponding mass deviation (Δmass), accurate mass, added mass and elemental composition are shown along with the annotation and PAR value. Supplementary Table S3 shows FCs and the nominal p -values for testing the null hypothesis that the case/control coefficient (β_i) is zero using a t -test for both the unstratified and sex-stratified data.

Case-control-fold changes were both positive and negative, indicating that some features were more abundant in cases and others in controls, respectively. Supplementary Table S4 shows the within- and between-subject variance components from the random effects model of each feature and the corresponding values of the ICC (median=0.753) and CV (median=0.337). Forty seven of the 67 detected features have been reported previously in our laboratory, with citations noted as footnotes in Supplementary Tables S2.A and S2.B.

Supplementary Figures S1.A–C, D–F and G–I show the results of feature selection for all subjects, males and females, respectively, and Table 2 combines the results to list features selected for associations with NHL. Altogether ten features satisfied the selection criteria based on multivariate linear regression plus LASSO and/or Random Forest: eight Cys34 adducts, i.e., 796.43 (Cys→Gly, males), 805.76 (Cys→oxoalanine/formylglycine, males), 816.43 (methylation, males and females), 822.42 (Cys34 sulfinic acid, males), 827.088 (methanethiol, males), 827.094 (Cys34 S -(O)-O-CH₃, all subjects and males), 835.11 (crotonaldehyde, females) and 894.44 (γ -GluCys, all subjects, males and females), and two Lys525 features, i.e., 580.85 (Lys525 oxidation product, females) and 578.32 (unknown, all subjects and females). Two of the three features selected from analyses of all subjects were heavily influenced by either males (827.094) or females (578.32). Eight of the ten selected features were more abundant in controls and two were more abundant in (male) cases.

In order to determine whether the eight T-cell lymphomas substantially altered the results, we repeated data normalization, imputation and feature selection after excluding the T-cell cases (three males and five females). The results, summarized in Supplementary Table S5, are very similar to those in Table 2. Eleven features were selected (rather than ten), including the same seven in males and six (rather than five) in females, and four (rather than three) in all subjects. One new feature was selected in females, 647.34 (unknown) and 835.11 (crotonaldehyde) was now also selected in all subjects.

Heatmaps presenting matrices of Spearman correlations between features and their corresponding clusters are shown in Figures 1A and 1B for males and Figures 2A and 2B for females, respectively. Clusters containing selected features (shown in red) provide clues regarding pathways towards peptide modifications associated with NHL, i.e., clusters C1-C3 for males and D1-D3 for females. Figures 3A and 3B and Figures 4A and 4B show Random Forest plots and the corresponding heatmaps of the selected features and 15 covariates for males and females, respectively. For males, Figure 3A shows that five selected features are ranked as more important by Random Forest than any covariate. Total calories and fruit have somewhat lesser importance by Random Forest while physical activity and smoking have the least importance. The corresponding heatmap (Figure 3B) shows that the five features in cluster E1 (796.43, 805.78, 822.42, 827.088 and 827.094) tend to be strongly negatively correlated with age, total calories and consumption of alcohol and meat (cluster E3) and to a lesser extent with leafy-vegetables. In E2, feature 894.44 clusters with bmi, age and smoking and has a strong negative correlation with physical activity. Regarding females, Figure 4A indicates that consumption of fruit and meat showed importance by Random Forest that was comparable to four of the selected features (578.32, 816.43, 894.44 and 835.11) while other covariates showed substantially less importance. The heatmap (Figure 4B) shows contrasting correlations of these four features with covariates where strong/moderate negative correlations are seen between 835.11 and 816.43 and smoking and root vegetables, and also between 894.44 and meat, physical activity and nuts and seeds, and between 578.32 and physical activity and fruiting vegetables.

To determine whether abundances of selected features may have resulted from disease progression (reverse causality) rather than their causal role in NHL, we examined the relationships between log-fold-changes of adduct abundances of NHL case/control pairs and days from recruitment to diagnosis (ttt). Results are presented in Supplementary Figures S2.A and S2.B as individual plots for the seven selected features in males and five in females, respectively. Of these nine features, only 816.43 (methylation) in male subjects had a sufficiently small p -value for the regression coefficient ($p=0.0026$) to suggest reverse causality (Supplementary Figure S2.A).

Discussion

As with our previous adductomic investigations of CRC and childhood leukemia (18,21), this study of incident NHL cases and controls relied on an ensemble of regression and classification methods to generate hypotheses about features for likely association with NHL, up to 11 years prior to diagnosis (Table 2). All seven of the T3-peptide features had previously been reported in humans (Supplementary Table S2.A) suggesting that their reactive precursors, including ROS, represent common exposures and/or pathways that increase NHL risks. The two Lys525 features selected for association with NHL (578.32 and 580.85) had not previously been reported in the two prior studies that included modifications at that locus (Supplementary Table S2.B).

Two interesting findings emerged (Table 2). First, only two of the nine selected features overlapped between males and females, supporting empirical evidence that NHL incidence differs substantially between sexes (3,4). Indeed, if we had relied solely on analysis of

all subjects, six of the nine selected features would have been missed. Second, seven of the nine selected features were more abundant in controls (FC range: 0.710–0.826), while only two were more abundant in (male) cases (FC: 1.204 and 1.392) (Figures 1B and 2B). This indicates that adduct features reflect both protective and detrimental effects on NHL incidence.

Focusing first on the features selected in male subjects, four were present in cluster C1, one in C2 and one in C3 (Figure 1A). Cluster C1 contains products of sulfoxidation or oxidative cleavage of Cys34, including selected features: 796.43 (Cys→Gly, FC=0.792), 822.42 (Cys34 sulfinic acid, FC=0.826) and 827.094 (Cys34S-(O)-O-CH₃, FC=0.735). Inclusion of selected feature 827.088 (S-addition of methanethiol, FC=1.392) in C1 is particularly intriguing because methanethiol is a product of microbial-human co-metabolism that is mediated by the gut microbiota via catabolism of methionine and/or methylation of hydrogen sulfide (29). Our previous investigation of CRC found elevated levels of 827.088 in incident cases (18), thereby implicating the gut microbiota as a CRC risk factor, consistent with formal hypotheses (30,31). Here, a similar positive association of NHL with 827.088 years prior to diagnosis suggests that translocation of gut microbes to the circulation may increase NHL risk.

Microbial translocation from the gut to the circulation and the resulting immune activation in healthy humans can result from a host of factors, including gut microbial dysbiosis, viral infections, IgA deficiency, reduced bacterial clearance by the liver, and excessive alcohol consumption (32), and has been associated with B-cell NHL following HIV infection (33). The fact that 827.088 clusters in C1 with features representing Cys34 oxidation products suggests a possible pathway involving systemic inflammation induced by gut microbial translocation. Indeed, the Cys34 sulfinic and sulfonic acids (822.42 and 827.75, respectively) in C1 are recognized biomarkers of systemic inflammation that have been linked to several diseases and syndromes including type 2 diabetes mellitus (34,35) [a risk factor for NHL (11)], kidney disease (36), liver disease and sepsis (37,38), and hemodialysis (39).

Referring next to cluster C2, feature 894.44 (Cys34 disulfide of γ -GluCys) is less abundant in NHL cases (FC=0.710). γ -GluCys is a precursor of GSH that is the predominant intracellular scavenger of ROS (40). Although our previous study of lung cancer found that lower abundance of 894.44 was associated with increased smoking intensity and pack-years of cigarette consumption (19), this adduct shows moderate positive correlation with smoking in this study (Figure 3B). Membrane-bound γ -glutamyltranspeptidase catabolizes conversion of extracellular GSH to CysGly (another constituent of C2) that stimulates production of pro-oxidant species and is upregulated by depletion of intracellular GSH (41,42). Thus, the high correlation in C2 between Cys34 disulfides of γ -GluCys (894.44) and CysGly (870.43) points to dysregulation of GSH metabolism in NHL cases. Interestingly, similar reductions in levels of 894.44 and 870.43, as well as that of the GSH adduct (913.45), were observed in smokers from the lung-cancer study (19) and also among nonsmoking Chinese females exposed to indoor effluents from coal combustion, a major source of ROS (43). It is also noteworthy that 827.088 (methanethiol) is negatively

correlated with both 894.44 and 870.43, suggesting roles for microbial translocation in GSH metabolism and redox biology.

The large C3 cluster contains a mixture of Cys34 and Lys525 features, one of which was selected for association with NHL, i.e., 816.43 [816.43 (methylation of T3 at Glu 37), FC=1.204]. The T3-methylation product 816.43 has been observed in all prior applications of our adductomics pipelines, including inconsistent associations with smoking (17,19). Review of HRMSMS spectra for this investigation confirmed that Glu37 is the methylation site in the T3 peptide. As noted previously, the significant trend of 816.43 abundance with ttd in male subjects (Supplementary Figure 2.A) points to reverse causality and encourages investigation of 816.43 as an early biomarker of NHL, with lower abundance at diagnosis.

Turning now to features selected for association with NHL among females (Table 2), in addition to 816.43 and 894.44 (also selected in males) 835.11 (crotonaldehyde, FC=0.785) has been reported previously (footnoted in Supplementary Table S2), while 578.32 (unknown, FC=0.886) and 580.85 (Lys525 oxidation product) are novel to the current investigation. Crotonaldehyde is a reactive α , β -unsaturated aldehyde produced by ROS oxidation of membrane lipids (44), and we had previously found 835.11 to be elevated in CRC cases (18) as well as in incident cases of childhood acute lymphoblastic leukemia (21) and in workers exposed to benzene (45), a known human leukemogen and promoter of ROS that has been causally linked with NHL (46) and with autoimmune disease as a risk factor for NHL (47). Figure 2A shows that unknown feature 578.32 and the Lys525 oxidation product 580.85 cluster with each other and with unknown feature 587.31 (cluster D3). The fact that neither 578.32 and 580.85 nor 816.43 and 835.11 are appreciably correlated with the oxidation products in D2 that were favored in C1 for association with NHL in males (Figure 1A) emphasizes differences in operative pathways between sexes. However, 894.44 (γ -GluCys adduct) is marginally negatively correlated the oxidation products in D2 (Figure 2A), albeit much weaker than the corresponding correlations observed in males (Figure 1A). Also 827.088 (methane thiol adduct) is moderately positively correlated with the D2 oxidation products suggesting possible influences of microbial translocation/systemic inflammation in females.

Analyses of covariates summarized in Figures 3A and 4A indicate that calories (total energy) and consumption of fruit and meat highly ranked for importance to NHL incidence by Random Forest in both sexes. In males, alcohol was also highly ranked whereas, in females, stalk vegetables and sprouts were highly ranked. In particular, three of the five features representing Cys34 oxidation products in cluster E1 of the male heatmap (Figure 3B: 796.43, 805.76 and 827.088) are all negatively correlated with intake of calories, alcohol and meat. A similar pattern is seen for features 816.43 and 835.31 in the female heatmap (Figure 4B), which are negatively correlated with intake of alcohol, meat and fruiting and root vegetables.

Figure 5 shows a Mean-Difference plot of the case/control ratios of median values of feature abundances for covariates by sex. Since deviations from the dashed line, representing equal ratios, indicate substantial differences between sexes, ratios for consumption of cabbage and root vegetables tend to be greater in females while those for calories and consumption of

alcohol, fruit, meat and mixed and leafy vegetables tend to be greater in males. These sex differences in dietary and alcohol consumption and the heatmaps in Figures 3B and 4B may help to explain the limited overlap of feature selection between males and females (Table 2). Also, the correlations between selected features and consumption of alcohol, meat and certain vegetables are interesting when considering that these same covariates are recognized risk factors for NHL (4,13).

Perhaps our most provocative finding is evidence that the methanethiol adduct (827.088) is elevated in male NHL cases. This biomarker of enteric microbial metabolism was previously found to be elevated in incident CRC cases, where translocation of enteric bacteria into colonic mucosa and the resulting inflammatory response hypothetically contributed to localized tumors (18). Extending this hypothesis to NHL implies that microbial translocation from the gut results in systemic release of microbes and the resulting inflammation/ROS that can transform circulating lymphocytes and/or promote cancers throughout the body. Furthermore, associations between 827.088 and intake of total calories, alcohol, meat and leafy vegetables suggest that prior reports of NHL associations with intake of alcohol, meat and dietary constituents (4) may involve microbial pathways. Follow-up is needed to confirm this finding with additional data from prospective cohorts, preferably in conjunction with fecal metagenomics to determine whether gut microbial dysbiosis may be involved and with screening for immune activation.

Our study has some strengths. First, the adductomics assay provides direct evidence of the disposition of ROS in the interstitial space during the month preceding phlebotomy and thus is highly relevant to ROS-induction of NHL. Second, our hypothesis-free design and ensemble of regression/classification methods permitted unbiased selection of adducts for associations with NHL. Third, expanding the Cys34 adductomics pipeline to Lys525 extended the range of adducts and their underlying chemistries, and included two selected features (578.32 and 580.85). And third, the high quality of serum samples from the EPIC-Italy cohort collected up to 11 years prior to diagnosis reduced the potential for reverse causality.

There are also weaknesses. The sample size was small, particularly after stratification by sex (Table 1) and results will require validation. Although storage of biological specimens for decades can produce artifacts, specimens were stored (for 20–25 years) in liquid nitrogen, and cases and controls were matched by year of enrollment to minimize potential storage effects. While four selected adducts were confirmed with synthetic standards (827.088, 827.094, 835.11 and 894.44), annotations of the others were based on elemental composition and should be regarded as putative. Since HSA adducts have a residence time of one month in humans, adduct levels measured in blood collected at recruitment may not accurately reflect those in following months or years. Likewise, given the seasonality of the Italian diet, correlations between adduct levels and dietary covariates collected by FFQ should be interpreted skeptically. Another limitation was our inability to examine connections between adducts and advanced neoplasms (precursors of NHL) for advanced stage vs. early stage cancers.

In summary, we used untargeted adductomics to detect 68 adduct features in HSA from 67 incident NHL cases and 82 controls, of which nine were selected for association (seven in males and five in females), with three selected in both sexes. The minimal overlap among selected features between males and females reinforces well-known sex differences in NHL risk. In male subjects, decreased abundances of several features representing oxidative pathways as well as γ -GluCys (894.44), a precursor of GSH, add weight to findings that long-term oxidative stress can lead to dysregulation of redox biology. Also, the increased abundance of the methanethiol adduct (827.088) and its correlation with Cys34 oxidation products in males implicates enteric microbial translocation and the resulting systemic inflammation as a potential pathway for NHL.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Financial support for this work was provided from the U.S. National Institutes of Health through grants R33CA191159 from the National Cancer Institute (S. Rappaport, S. Dudoit) and P42ES04705 from the National Institute for Environmental Health Sciences (S. Rappaport) and grant agreement 308610-FP7 from the European Commission (Project Exposomics) (P. Vineis, S. Rappaport). The authors appreciate the helpful comments of Christine Skibola who reviewed a draft of the manuscript.

References

1. ECIS - European Cancer Information System. 2023 Estimates of cancer incidence and mortality in 2020, for all countries. In ECIS - European Cancer Information System. European Union <<https://ecis.jrc.ec.europa.eu>>. Accessed February 27, 2023.
2. Miranda-Filho A, Pineros M, Znaor A, Marcos-Gragera R, Steliarova-Foucher E, Bray F. Global patterns and trends in the incidence of non-Hodgkin lymphoma. *Cancer causes & control : CCC* 2019;30(5):489–99 doi 10.1007/s10552-019-01155-5. [PubMed: 30895415]
3. Muller AM, Ihorst G, Mertelsmann R, Engelhardt M. Epidemiology of non-Hodgkin's lymphoma (NHL): trends, geographic distribution, and etiology. *Ann Hematol* 2005;84(1):1–12 doi 10.1007/s00277-004-0939-7. [PubMed: 15480663]
4. Alexander DD, Mink PJ, Adami HO, Chang ET, Cole P, Mandel JS, et al. The non-Hodgkin lymphomas: a review of the epidemiologic literature. *Int J Cancer* 2007;120 Suppl 12:1–39 doi 10.1002/ijc.22719.
5. Grulich AE, Vajdic CM, Cozen W. Altered immunity as a risk factor for non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev* 2007;16(3):405–8 doi 10.1158/1055-9965.EPI-06-1070. [PubMed: 17337643]
6. Pol S, Vallet-Pichard A, Hermine O. Extrahepatic cancers and chronic HCV infection. *Nat Rev Gastroenterol Hepatol* 2018;15(5):283–90 doi 10.1038/nrgastro.2017.172. [PubMed: 29339810]
7. Halliwell B Free radicals, antioxidants, and human disease: curiosity, cause, or consequence? *Lancet* 1994;344(8924):721–4 doi S0140-6736(94)92211-X [pii]. [PubMed: 7915779]
8. Liou GY, Storz P. Reactive oxygen species in cancer. *Free radical research* 2010;44(5):479–96 doi 10.3109/10715761003667554. [PubMed: 20370557]
9. Lan Q, Zheng T, Shen M, Zhang Y, Wang SS, Zahm SH, et al. Genetic polymorphisms in the oxidative stress pathway and susceptibility to non-Hodgkin lymphoma. *Human genetics* 2007;121(2):161–8 doi 10.1007/s00439-006-0288-9. [PubMed: 17149600]
10. Wang SS, Davis S, Cerhan JR, Hartge P, Severson RK, Cozen W, et al. Polymorphisms in oxidative stress genes and risk for non-Hodgkin lymphoma. *Carcinogenesis* 2006;27(9):1828–34 doi 10.1093/carcin/bgl013. [PubMed: 16543247]

11. Castillo JJ, Mull N, Reagan JL, Nemr S, Mitri J. Increased incidence of non-Hodgkin lymphoma, leukemia, and myeloma in patients with diabetes mellitus type 2: a meta-analysis of observational studies. *Blood* 2012;119(21):4845–50 doi 10.1182/blood-2011-06-362830. [PubMed: 22496152]
12. Holtan SG, O'Connor HM, Fredericksen ZS, Liebow M, Thompson CA, Macon WR, et al. Food-frequency questionnaire-based estimates of total antioxidant capacity and risk of non-Hodgkin lymphoma. *Int J Cancer* 2012;131(5):1158–68 doi 10.1002/ijc.26491. [PubMed: 22038870]
13. Kelemen LE, Cerhan JR, Lim U, Davis S, Cozen W, Schenk M, et al. Vegetables, fruit, and antioxidant-related nutrients and risk of non-Hodgkin lymphoma: a National Cancer Institute-Surveillance, Epidemiology, and End Results population-based case-control study. *Am J Clin Nutr* 2006;83(6):1401–10 doi 10.1093/ajcn/83.6.1401. [PubMed: 16762953]
14. Sabbioni G, Turesky RJ. Biomonitoring Human Albumin Adducts: The Past, the Present, and the Future. *Chem Res Toxicol* 2017;30(1):332–66 doi 10.1021/acs.chemrestox.6b00366. [PubMed: 27989119]
15. Go YM, Jones DP. The redox proteome. *J Biol Chem* 2013;288(37):26512–20 doi 10.1074/jbc.R113.464131. [PubMed: 23861437]
16. Nagumo K, Tanaka M, Chuang VT, Setoyama H, Watanabe H, Yamada N, et al. Cys34-cysteinylated human serum albumin is a sensitive plasma marker in oxidative stress-related chronic diseases. *PLoS One* 2014;9(1):e85216 doi 10.1371/journal.pone.0085216. [PubMed: 24416365]
17. Grigoryan H, Edmands W, Lu SS, Yano Y, Regazzoni L, Iavarone AT, et al. Adductomics Pipeline for Untargeted Analysis of Modifications to Cys34 of Human Serum Albumin. *Anal Chem* 2016;88(21):10504–12 doi 10.1021/acs.analchem.6b02553. [PubMed: 27684351]
18. Grigoryan H, Schiffman C, Gunter MJ, Naccarati A, Polidoro S, Dagnino S, et al. Cys34 Adductomics Links Colorectal Cancer with the Gut Microbiota and Redox Biology. *Cancer research* 2019;79(23):6024–31 doi 10.1158/0008-5472.CAN-19-1529. [PubMed: 31641032]
19. Dagnino S, Bodinier B, Grigoryan H, Rappaport SM, Karimi M, Guida F, et al. Agnostic Cys34-albumin adductomics and DNA methylation: Implication of N-acetylcysteine in lung carcinogenesis years before diagnosis. *Int J Cancer* 2019;3294–303 doi 10.1002/ijc.32680. [PubMed: 31513294]
20. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002;5(6B):1113–24 doi 10.1079/PHN2002394. [PubMed: 12639222]
21. Yano Y, Schiffman C, Grigoryan H, Hayes J, Edmands W, Petrick L, et al. Untargeted adductomics of newborn dried blood spots identifies modifications to human serum albumin associated with childhood leukemia. *Leuk Res* 2019;88:106268 doi 10.1016/j.leukres.2019.106268. [PubMed: 31760269]
22. Grigoryan H, Imani P, Dudoit S, Rappaport SM. Extending the HSA-Cys34-Adductomics Pipeline to Modifications at Lys525. *Chem Res Toxicol* 2021;34(12):2549–57 doi 10.1021/acs.chemrestox.1c00311. [PubMed: 34788011]
23. Wong JYY, Imani P, Grigoryan H, Bassig BA, Dai Y, Hu W, et al. Exposure to diesel engine exhaust and alterations to the Cys34/Lys525 adductome of human serum albumin. *Environ Toxicol Pharmacol* 2022;95:103966 doi 10.1016/j.etap.2022.103966. [PubMed: 36067935]
24. Palli D, Berrino F, Vineis P, Tumino R, Panico S, Masala G, et al. A molecular epidemiology project on diet and cancer: the EPIC-Italy Prospective Study. Design and baseline characteristics of participants. *Tumori* 2003;89(6):586–93 doi 10.1177/030089160308900602. [PubMed: 14870823]
25. Chajes V, Jenab M, Romieu I, Ferrari P, Dahm CC, Overvad K, et al. Plasma phospholipid fatty acid concentrations and risk of gastric adenocarcinomas in the European Prospective Investigation into Cancer and Nutrition (EPIC-EURGAST). *Am J Clin Nutr* 2011;94(5):1304–13 doi ajcn.110.005892 [pii] 10.3945/ajcn.110.005892. [PubMed: 21993438]
26. Huang Y, Cao D, Chen Z, Chen B, Li J, Guo J, et al. Red and processed meat consumption and cancer outcomes: Umbrella review. *Food chemistry* 2021;356:129697 doi 10.1016/j.foodchem.2021.129697. [PubMed: 33838606]
27. Perttula K, Edmands WM, Grigoryan H, Cai X, Iavarone AT, Gunter MJ, et al. Evaluating Ultra-long-Chain Fatty Acids as Biomarkers of Colorectal Cancer Risk. *Cancer Epidemiol Biomarkers Prev* 2016;25(8):1216–23 doi 10.1158/1055-9965.EPI-16-0204. [PubMed: 27257090]

28. Perttula K, Schiffman C, Edmands WMB, Petrick L, Grigoryan H, Cai X, et al. Untargeted lipidomic features associated with colorectal cancer in a prospective cohort. *BMC cancer* 2018;18(1):996 doi 10.1186/s12885-018-4894-4. [PubMed: 30340609]
29. He X, Slupsky CM. Metabolic fingerprint of dimethyl sulfone (DMSO) in microbial-mammalian co-metabolism. *J Proteome Res* 2014;13(12):5281–92 doi 10.1021/pr500629t. [PubMed: 25245235]
30. O’Keefe SJ. Diet, microorganisms and their metabolites, and colon cancer. *Nat Rev Gastroenterol Hepatol* 2016;13(12):691–706 doi 10.1038/nrgastro.2016.165. [PubMed: 27848961]
31. Vippera K, O’Keefe SJ. Diet, microbiota, and dysbiosis: a ‘recipe’ for colorectal cancer. *Food & function* 2016;7(4):1731–40 doi 10.1039/c5fo01276g. [PubMed: 26840037]
32. Brenchley JM, Douek DC. Microbial translocation across the GI tract. *Annu Rev Immunol* 2012;30:149–73 doi 10.1146/annurev-immunol-020711-075001. [PubMed: 22224779]
33. Epeldegui M, Hussain SK. The Role of Microbial Translocation and Immune Activation in AIDS-Associated Non-Hodgkin Lymphoma Pathogenesis: What Have We Learned? *Crit Rev Immunol* 2020;40(1):41–51 doi 10.1615/CritRevImmunol.2020033319. [PubMed: 32421978]
34. Paramasivan S, Adav SS, Ngan SC, Dalan R, Leow MK, Ho HH, et al. Serum albumin cysteine trioxidation is a potential oxidative stress biomarker of type 2 diabetes mellitus. *Sci Rep* 2020;10(1):6475 doi 10.1038/s41598-020-62341-z. [PubMed: 32296090]
35. Kobayashi Y, Suzuki R, Yasukawa K, Oba K, Yamauchi T, Yatomi Y, et al. Oxidized albumin in blood reflects the severity of multiple vascular complications in diabetes mellitus. *Metabol Open* 2020;6:100032 doi 10.1016/j.metop.2020.100032. [PubMed: 32812907]
36. Figueroa SM, Araos P, Reyes J, Gravez B, Barrera-Chimal J, Amador CA. Oxidized Albumin as a Mediator of Kidney Disease. *Antioxidants (Basel)* 2021;10(3) doi 10.3390/antiox10030404.
37. Alcaraz-Quiles J, Casulleras M, Oettl K, Titos E, Flores-Costa R, Duran-Guell M, et al. Oxidized Albumin Triggers a Cytokine Storm in Leukocytes Through P38 Mitogen-Activated Protein Kinase: Role in Systemic Inflammation in Decompensated Cirrhosis. *Hepatology* 2018;68(5):1937–52 doi 10.1002/hep.30135. [PubMed: 30070728]
38. Oettl K, Birner-Gruenberger R, Spindelboeck W, Stueger HP, Dorn L, Stadlbauer V, et al. Oxidative albumin damage in chronic liver failure: relation to albumin binding capacity, liver dysfunction and survival. *J Hepatol* 2013;59(5):978–83 doi 10.1016/j.jhep.2013.06.013. [PubMed: 23811308]
39. Nasif WA, Mukhtar MH, El-Emshaty HM, Alwazna AH. Redox State of Human Serum Albumin and Inflammatory Biomarkers in Hemodialysis Patients with Secondary Hyperparathyroidism During Oral Calcitriol Supplementation for Vitamin D. *Open Med Chem J* 2018;12:98–110 doi 10.2174/1874104501812010098. [PubMed: 30450134]
40. Lu SC. Glutathione synthesis. *Biochimica et biophysica acta* 2013;1830(5):3143–53 doi 10.1016/j.bbagen.2012.09.008. [PubMed: 22995213]
41. Adamski J, Suhre K. Metabolomics platforms for genome wide association studies--linking the genome to the metabolome. *Curr Opin Biotechnol* 2013;24(1):39–47 doi 10.1016/j.copbio.2012.10.003. [PubMed: 23102864]
42. Dominici S, Paolicchi A, Lorenzini E, Maellaro E, Comporti M, Pieri L, et al. Gamma-glutamyltransferase-dependent prooxidant reactions: a factor in multiple processes. *Biofactors* 2003;17(1–4):187–98. [PubMed: 12897440]
43. Lu SS, Grigoryan H, Edmands WM, Hu W, Iavarone AT, Hubbard A, et al. Profiling the Serum Albumin Cys34 Adductome of Solid Fuel Users in Xuanwei and Fuyuan, China. *Environmental science & technology* 2017;51(1):46–57 doi 10.1021/acs.est.6b03955. [PubMed: 27936627]
44. Aldini G, Dalle-Donne I, Facino RM, Milzani A, Carini M. Intervention strategies to inhibit protein carbonylation by lipoxidation-derived reactive carbonyls. *Med Res Rev* 2007;27(6):817–68 doi 10.1002/med.20073. [PubMed: 17044003]
45. Grigoryan H, Edmands WMB, Lan Q, Carlsson H, Vermeulen R, Zhang L, et al. Adductomic signatures of benzene exposure provide insights into cancer induction. *Carcinogenesis* 2018;39(5):661–8 doi 10.1093/carcin/bgy042. [PubMed: 29538615]

46. Rana I, Dahlberg S, Steinmaus C, Zhang L. Benzene exposure and non-Hodgkin lymphoma: a systematic review and meta-analysis of human studies. *Lancet Planet Health* 2021;5(9):e633–e43 doi 10.1016/S2542-5196(21)00149-2. [PubMed: 34450064]
47. Vineis P, Miligi L, Costantini AS. Exposure to solvents and risk of non-Hodgkin lymphoma: clues on putative mechanisms. *Cancer Epidemiol Biomarkers Prev* 2007;16(3):381–4 doi 10.1158/1055-9965.EPI-07-0124. [PubMed: 17337640]
48. Cust AE, Smith BJ, Chau J, van der Ploeg HP, Friedenreich CM, Armstrong BK, et al. Validity and repeatability of the EPIC physical activity questionnaire: a validation study using accelerometers as an objective measure. *Int J Behav Nutr Phys Act* 2008;5:33 doi 10.1186/1479-5868-5-33. [PubMed: 18513450]

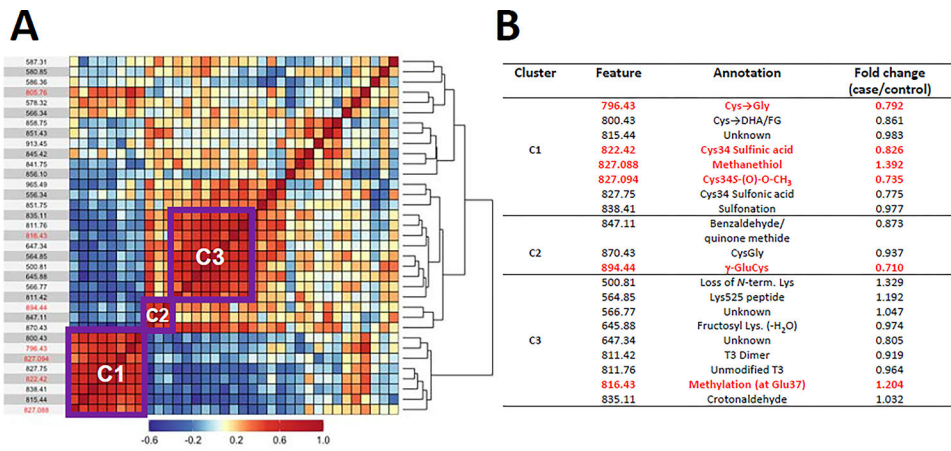


Figure 1. Heatmap showing all features in MALE subjects (A) with clusters identified in (B). Features selected for association with NHL are highlighted in red.

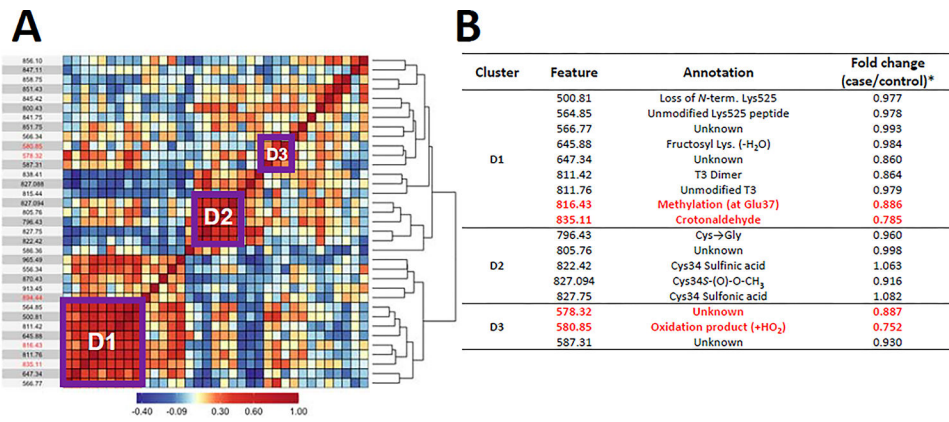


Figure 2. Heatmap showing all features in FEMALE subjects (A) with clusters identified in (B). Features selected for association with NHL are highlighted in red.

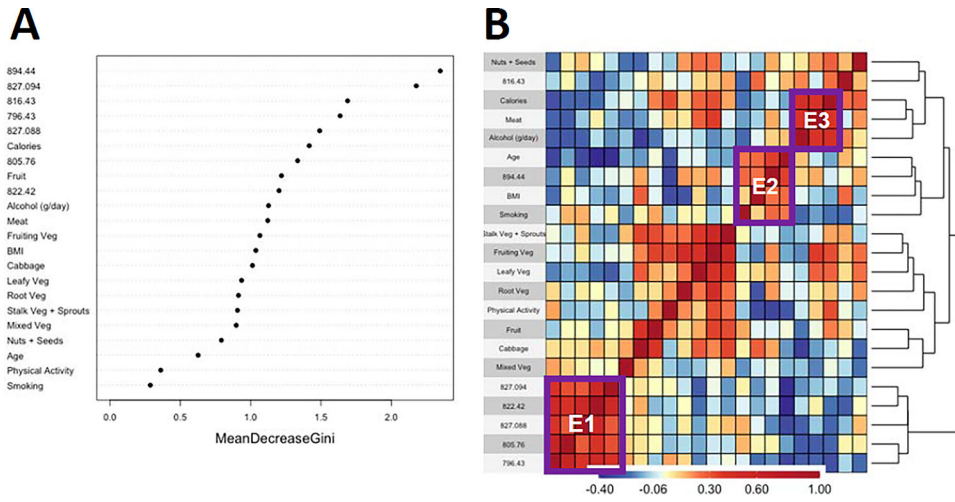


Figure 3. Variable importance of covariates and selected features as ranked by Random Forest in MALE subjects (A) and heatmap showing the corresponding correlations (B).

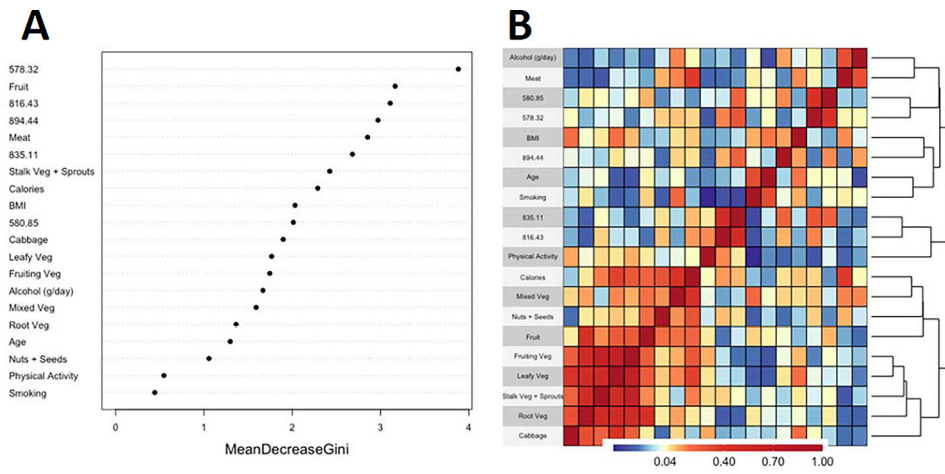


Figure 4. Variable importance of covariates and selected features as ranked by Random Forest in FEMALE subjects (A) and heatmap showing the corresponding correlations (B).

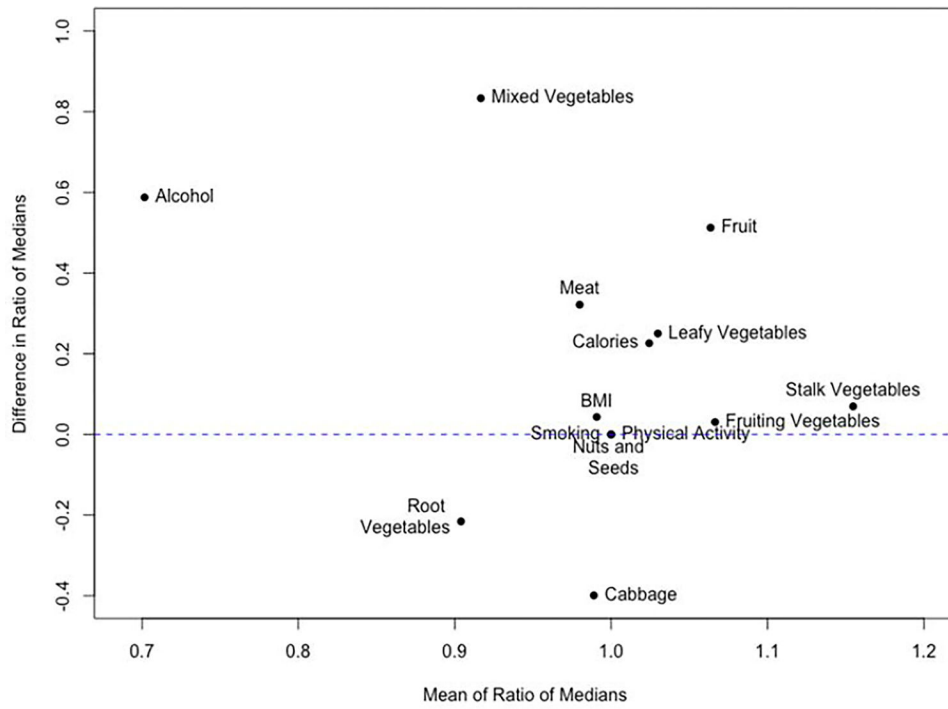


Figure 5. Mean-Difference plot of median values of case/control ratios for covariates by sex. The y-axis shows the case/control ratio of median values in males minus the case/control ratio of median values in females. The x-axis shows the mean of these two quantities. The dashed line indicates no difference in the ratio between males and females.

Table 1.

Summary statistics for covariates included with NHL cases and controls.

Variable		Cases (n=67)	Controls (n=82)	<i>p</i> -value ^I
Sex	Male	23	Male 32	0.67
	Female	44	Female 50	
BMI (kg/m ²)	Min	19.84	Min 15.53	0.63
	Mean	26.42	Mean 25.76	
	Median	25.69	Median 26.06	
	Max	36.59	Max 36.16	
Alcohol Consumption (g/d)	Min	0	Min 0	0.38
	Mean	12.45	Mean 15.35	
	Median	5.13	Median 8.78	
	Max	64.78	Max 78.98	
Smoking Status	Yes	32	Yes 41	0.34
	Ex	17	Ex 28	
	No	16	No 13	
Meat Consumption (g/d)	Min	0	Min 10.4	0.40
	Mean	108.9	Mean 114.1	
	Median	100.6	Median 105.3	
	Max	348.0	Max 350.5	
Physical Activity ²	1	17	1 22	0.67
	2	23	2 29	
	3	17	3 16	
	4	8	4 15	
Age at Recruitment (years)	Min	40	Min 39	1.00
	Mean	54.21	Mean 54.17	
	Median	54	Median 53	
	Max	74	Max 75	
Calorie Consumption (kCal)	Min	1048	Min 1021	0.64
	Mean	2269	Mean 2323	
	Median	2051	Median 2110	
	Max	4571	Max 5727	
Leafy Vegetable Consumption (g/d)	Min	0	Min 0.3	0.51
	Mean	35.22	Mean 36.06	
	Median	26.60	Median 26.75	
	Max	121.9	Max 138.3	
Fruiting Vegetable Consumption (g/d)	Min	13.5	Min 9.6	0.43

Variable	Cases (n=67)		Controls (n=82)		<i>p</i> -value ¹
	Mean	Median	Mean	Median	
	76.24	69.30	70.18	66.65	
	Max	195.0	Max	202.8	
Root Vegetable Consumption (g/d)	Min	0	Min	0	0.55
	Mean	15.09	Mean	15.75	
	Median	8.60	Median	8.85	
	Max	216.7	Max	127.8	
Cabbage Consumption (g/d)	Min	0	Min	0	0.54
	Mean	6.02	Mean	5.09	
	Median	3.80	Median	3.50	
	Max	29.6	Max	40.5	
Stalk Vegetable and Sprouts Consumption (g/d)	Min	0.3	Min	0.1	0.50
	Mean	14.61	Mean	12.49	
	Median	12.3	Median	9.8	
	Max	54.5	Max	41.0	
Mixed Vegetable Consumption (g/d)	Min	0	Min	0	0.23
	Mean	4.74	Mean	4.73	
	Median	1.1	Median	2.05	
	Max	94.8	Max	126.1	
Fruit Consumption (g/d)	Min	41.3	Min	27.5	0.26
	Mean	331.6	Mean	346.1	
	Median	260.9	Median	320.6	
	Max	1,057	Max	894.6	
Nut and Seed Consumption (g/d)	Min	0	Min	0	0.27
	Mean	1.16	Mean	1.24	
	Median	0.3	Median	0.3	
	Max	8.5	Max	14.3	
Time to Diagnosis (days)	Min	45			
	Mean	2010			
	Median	2005			
	Max	3721			

¹ *p*-values were computed with chi-squared statistics for categorical variables and with Wilcoxon-rank sum tests for continuous variables.

² For definitions and validation of physical activity see (48).

Table 2.

Features selected by the ensemble pipeline for association with NHL. The boldface font indicates selection criteria that were satisfied. [Fold-change based on normalized and imputed data (FC); nominal *p*-value; bootstrap LASSO; Random Forest (RF)].

Feature	Annotation	All subjects					Males					Females					
		FC	<i>p</i> -value	LASSO	RF	FC	<i>p</i> -value	LASSO	RF	FC	<i>p</i> -value	LASSO	RF	FC	<i>p</i> -value	LASSO	RF
578.32	Unknown	0.892	0.034	Yes	Yes	0.921	0.353	No	No	0.887	0.083	Yes	Yes				
580.85	Oxidation of Lys525 (+HO ₂)	0.798	0.133	No	No	0.913	0.713	No	No	0.752	0.112	Yes	No				
796.43	Cys34→Gly	0.893	0.054	No	No	0.792	0.017	Yes	Yes	0.960	0.581	No	No				
805.76	Cys34→OA/FG	0.943	0.211	No	No	0.822	0.022	Yes	No	0.998	0.971	No	No				
816.43	Methylation (at Glu37)	0.995	0.953	No	No	1.204	0.136	Yes	Yes	0.886	0.219	No	Yes				
822.42	Cys34 sulfonic acid	0.958	0.468	No	No	0.826	0.069	Yes	No	1.063	0.388	No	No				
827.088	Methanethiol	1.135	0.314	No	No	1.392	0.127	Yes	Yes	1.025	0.865	No	No				
827.094	Cys34S-(O)-O-CH ₃	0.844	0.005	Yes	Yes	0.735	0.005	Yes	Yes	0.916	0.240	No	No				
835.11	Crotonaldehyde	0.867	0.093	No	No	1.032	0.804	No	No	0.785	0.019	Yes	Yes				
894.44	γ-GluCys	0.855	0.047	Yes	Yes	0.710	0.029	Yes	Yes	0.922	0.380	No	Yes				

OA/FG, oxoalanine/formyl glycine.