Environmental Science & Technology

pubs.acs.org/est

Article

Metabolomic Signatures of Exposure to Nitrate and Trihalomethanes in Drinking Water and Colorectal Cancer Risk in a Spanish Multicentric Study (MCC-Spain)

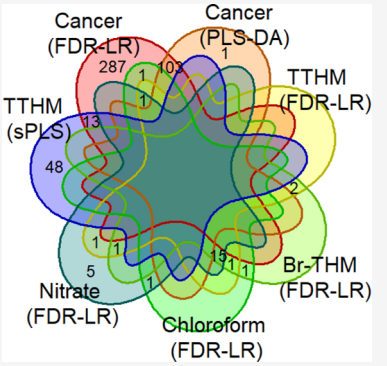
Published as part of Environmental Science & Technology virtual special issue "The Exposome and Human Health".

Jose A. Alcolea, Carolina Donat-Vargas, Anastasia Chrysovalantou Chatziioannou, Pekka Keski-Rahkonen, Nivonirina Robinot, Antonio José Molina, Pilar Amiano, Inés Gómez-Acebo, Gemma Castaño-Vinyals, Lea Maitre, Marc Chadeau-Hyam, Sonia Dagnino, Sibo Lucas Cheng, Augustin Scalbert, Paolo Vineis, Manolis Kogevinas, and Cristina M. Villanueva*

Cite This: *Environ. Sci. Technol.* 2023, 57, 19316−19329

Read Online

ABSTRACT: We investigated the metabolomic profile associated with exposure to trihalomethanes (THMs) and nitrate in drinking water and with colorectal cancer risk in 296 cases and 295 controls from the Multi Case-Control Spain project. Untargeted metabolomic analysis was conducted in blood samples using ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry. A variety of univariate and multivariate association analyses were conducted after data quality control, normalization, and imputation. Linear regression and partial least-squares analyses were conducted for chloroform, brominated THMs, total THMs, and nitrate among controls and for case-control status, together with a N-integration model discriminating colorectal cancer cases from controls through interrogation of correlations between the exposure variables and the metabolomic features. Results revealed a total of 568 metabolomic features associated with at least one water contaminant or colorectal cancer. Annotated metabolites and pathway analysis suggest a number of pathways as potentially involved in the link between exposure to these water contaminants and colorectal cancer, including nicotinamide, cytochrome P-450, and tyrosine metabolism. These findings provide insights into the underlying biological mechanisms and potential biomarkers associated with water contaminant exposure and colorectal cancer risk. Further research in this area is needed to better understand the causal relationship and the public health implications.

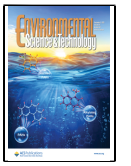


KEYWORDS: *blood, colorectal cancer, disinfection byproducts, drinking water, environment, epidemiology, exposure, exposome, metabolomics, -omics, nitrate, trihalomethanes, untargeted*

INTRODUCTION

Colorectal cancer was the top third most frequent cancer site worldwide in 2020, with 1.2 million new cases, accounting for 10% of all cancer sites.[1] Diet and lifestyle are recognized risk factors,[2] and some environmental exposures including carcinogens in drinking water, have been associated with colorectal cancer although evidence is not conclusive. Disinfection byproducts (DBPs) constitute a heterogeneous mixture including more than 700 chemicals generated during drinking water treatment. DBPs are formed from the reaction of disinfectants such as chlorine with organic matter occurring naturally in raw water.[3] Trihalomethanes (THMs) are the most common DBPs and have been employed as indicators of DBP exposure in epidemiological studies. THMs comprise chloroform, dibromochloromethane, bromodichloromethane, and bromoform, and can be incorporated through ingestion, inhalation, and dermal contact.[4] Chloroform and bromodichloromethane are possible human carcinogens.[5] Bromodichloromethane and bromomethane are animal carcinogens although human carcinogenicity evidence is insufficient.[5] Epidemiological studies suggest an increased risk of colorectal cancer associated with THM exposure, albeit with variations between THM components.[6−8]

Received: July 20, 2023
Revised: October 25, 2023
Accepted: October 25, 2023
Published: November 14, 2023



The global presence of nitrate in the water cycle is increasing, owing to the widespread use of nitrogen fertilizers and intensive farming. Nitrate exposure primarily occurs through the ingestion of food and drinking water.[9] Ingested nitrate undergoes endogenous nitrosation to form N-nitroso compounds such as nitrosamines, classified as probable human carcinogens.[10] Long-term exposure to nitrate in drinking water has been linked to colorectal cancer at exposure levels below the regulatory limits.[11,12]

Although evidence suggests that exposure to these widespread contaminants in drinking water may increase colorectal cancer risk, causal inference cannot be drawn due to the lack of understanding of the underlying biological mechanisms. Metabolomics offers a promising approach to gain insights into the relevant metabolic and molecular pathways involved. Some metabolites may be intermediate biomarkers, which directly reflect the underlying biochemical activity.[13] Previous studies have linked THM exposure in swimming pools with changes in serum metabolomic signatures.[14] However, the metabolomic profile associated with THMs and nitrate exposure in drinking water has not been evaluated so far. In light of this knowledge gap, we conducted an untargeted metabolomic study in the framework of the Multi Case-Control Spain project (MCC-Spain, www.mccspain.org).[15] We aimed to identify circulating metabolites associated with THMs and nitrate exposure, colorectal cancer, and the pathway between exposure and colorectal cancer.

MATERIALS AND METHODS

Study Design and Participants. The present study is based on a subset of the MCC-Spain project, conducted in Spain from September 2007 to November 2013.[15] Cases were diagnosed with incident colorectal cancer confirmed through histological analysis and defined following the International Statistical Classification of Diseases and Related Health Problems (10th revision) (ICD-10): C18, C19, C20, D01.0, D01.1, D01.2. Cases were identified through regular visits to hospital departments, including gastroenterology, oncology, general surgery, radiotherapy, and pathology. Controls were frequency matched to cases by sex, age (±5 years), and area of residence and were selected from the general population using lists of randomly selected family practitioners from primary care centers sharing the same catchment area as the participating hospitals. Selection criteria included age (20−85 years old), residence in the hospital catchment area for a minimum of six months before recruitment, and ability to respond to the epidemiological questionnaire. Response rates varied from 54% (Cantabria) to 80% (Barcelona, León) among cases and from 58% (Barcelona, Gipuzkoa) to 68% (Leon) among controls. The study protocol and the metabolomics study were approved by the ethics review boards of the participating centers, and all participants provided informed consent before recruitment. For the present analysis we selected a random subset of noncurrent smoking participants with most complete exposure assessment to THMs and nitrate in drinking water (≥70% years with known exposure), enrolled in 5 provinces (Barcelona, Cantabria, Gipuzkoa, León, and Navarra). The present analysis included a total of 591 participants (296 cases and 295 controls).

Individual Information. Cases were interviewed at the hospital as soon as possible after diagnosis (median of 58 days), and controls were interviewed in primary care centers. Trained interviewers administered a computer assisted personal questionnaire to study participants to gather information on sociodemographics, lifestyle (including smoking, alcohol consumption, physical activity, etc.), anthropometrics (height, weight), occupational history, medical and drug history, and family history of cancer. Residence addresses where participants lived for at least 12 months were ascertained, including the start and stop year and the type of water consumed (municipal, bottled, private well, other). Participants were asked to report the number of glasses of bottled water, tap water, and other sources of water consumed per day on average as an adult at home, workplace, and other places separately. Chemotherapy and radiotherapy treatment before the interview were ascertained among cases.

Exposure Assessment. We obtained information on water source and treatment history and THMs and nitrate concentrations in public drinking water through questionnaires to water utilities and local authorities in the study areas. Routine monitoring levels from 2004 to 2010 were also provided by the Sistema de Información Nacional en Aguas de Consumo (SINAC). We estimated the annual average concentrations of these contaminants in the study municipalities. For the present analysis, we estimated exposure to THMs and nitrate for a recent period consisting of the 3 years prior to the interview, excluding the last 2 years. Residential levels were calculated by combining the concentration in drinking water supply by year and municipality of residence of the study subjects. We focused on residential THM concentrations and waterborne ingested nitrate. For those who drank tap water, nitrate residential levels were assigned. For those who consumed bottled water, we assigned a value of 6.1 mg/L nitrate, which is the average level in the most consumed bottled water brands in Spain weighted by the sale frequency.[16] In the Leon region, where private well water was mainly consumed, we conducted a sampling campaign to measure nitrate (range 0.5−93 mg/L), the values of which were assigned to well water consumers in this area. Well water consumption was very infrequent in the other areas, where nitrate values were treated as missing values. The details on the exposure assessment have been published elsewhere.[7,11]

Blood Sample Preparation. Blood samples were collected by venipuncture after the interview, processed, and stored at −80 °C (average storage time 5 years and 10 months). Prior to the laboratory analyses, serum samples were randomized by case-control status, area, sex, and age, in order to minimize potential batch effects. Samples were prepared by mixing 30 μL of the mixture with 200 μL of acetonitrile and filtering the precipitate with 0.2 μm Captiva ND plates (Agilent Technologies). The filtrate was collected into a polypropylene well plate that was sealed and kept refrigerated until analysis. Quality control (QC) samples (N = 52) were prepared from a pooled sample made by mixing small aliquots of 90 randomly selected study samples. Blank samples (N = 7) were prepared in identical manner, leaving only serum from the process. Each well plate included four independently extracted QC samples and one blank.

Metabolomic Laboratory Analysis. Samples were analyzed as a single uninterrupted batch with an ultrahigh-performance liquid-chromatography (UHPLC)-quadrupole time-of-flight (QTOF)-mass spectrometry (MS) system (Agilent Technologies) consisting of a 1290 Binary liquid chromatography (LC) system, a Jet Stream electrospray ionization source, and a 6550 QTOF-MS. Autosampler tray was kept refrigerated at 4 °C and 2 μL of the sample solution

https://doi.org/10.1021/acs.est.3c05814
Environ. Sci. Technol. 2023, 57, 19316−19329

was injected on an ACQUITY UPLC HSS T3 column (2.1 × 100 mm, 1.8 μm; Waters). Column temperature was 45 °C and mobile phase flow rate was 0.4 mL/min, consisting of ultrapure water and LC-MS grade methanol, both containing 0.05% (v/v) of formic acid. The gradient profile was as follows: 0−6 min: 5% → 100% methanol, 6−10.5 min: 100% methanol, 10.5−13 min: 5% methanol. The MS was operated in positive polarity using the following conditions: drying gas (nitrogen) temperature 175 °C and flow 12 L/min, sheath gas temperature 350 °C and flow 11 L/min, nebulizer pressure 45 psi, capillary voltage 3500 V, nozzle voltage 300 V, and fragmentor voltage 175 V. Data acquisition was performed using the 2 GHz extended dynamic range mode across a mass range of 50−1000 Da. Scan rate was 1.67 Hz and data acquisition was in centroid mode. Continuous mass axis calibration was performed by monitoring two reference ions throughout the runs (m/z 121.050873 and m/z 922.009798). Data were acquired in full scan mode using MassHunter Acquisition B.05.01 (Agilent Technologies). The analytical run was initiated with priming injections of a QC sample to achieve stable instrument response, followed by study samples, which intervened after every 12 injections with a QC sample to monitor instrument performance and sample stability.

Metabolomic Data Processing. Preprocessing of the acquired data was performed using Qualitative Analysis B.06.00, DA Reprocessor, and Mass Profiler Professional 12.1 software (Agilent Technologies). Recursive feature finding was employed to find compounds as singly charged proton adducts [M + H]+, using data from all study samples. The initial processing of the data was performed using Qualitative Analysis with the molecular feature extraction (MFE) algorithm set to small molecules. Threshold values for mass and chromatographic peak heights were 1500 and 10 000 counts, respectively. Peak spacing tolerance for isotope peaks was 0.0025 m/z plus 7 ppm, with the isotope model set to common organic molecules. After the initial feature finding, the compounds that existed in at least 2% of all the samples were combined into a target list, using windows of 0.06 min for retention time and 15 ppm +2 mDa for mass for alignment. These features were used as targets for the recursive feature extraction of all the data (samples, QCs, and blanks), which was performed using Agilent's Find by Formula (FBF) algorithm, with match tolerance for the compound mass and retention time set at ±10 ppm and ±0.03 min and ion species were limited to [M + H]+, without thresholds for the number of ions associated with a feature. Chromatographic peak area was used as a measure of intensity.

Quality Control, Normalization, and Imputation. We compared the geometric mean of feature intensities between the blank samples and study samples to identify and exclude background features. Features present in every blank sample were excluded unless their average intensity in the study samples was at least 5-fold greater. Additionally, features with intensity variation coefficient higher than 30% among QC samples were excluded. The feature-wise exclusion of missing values was performed to assess data quality. Features with >30% missing values either in cases or controls were excluded. Additionally, one participant with exceptionally high missing values (32.1%) was excluded from the analysis. Feature intensities were log-transformed to correct their skewed distribution. Data normalization based on the experimental plates was employed to minimize plate-to-plate variation. Specifically, we calculated a correction value for each feature by subtracting the overall average intensity of that feature across all plates from the plate-specific average intensity. This correction value was then subtracted from each intensity value of the corresponding feature within each plate. This process helped align the feature intensities across different experimental plates, ensuring that plate-related variations were accounted for in our analysis. Finally, a quantile regression approach was used to impute left-censored missing data separately for the control and case data sets, using the R function 'impute.QRILC' from package imputeLCMD[17] with the tune.sigma parameter set to 1.

Univariate Association Analysis. Linear regression models were fitted for each metabolomic feature as the response variable and colorectal cancer status as predictor variable, treated as a dichotomous categorical variable with a dummy encoding (0 for "Control" and 1 for "Case") and adjusting for covariates (area, sex, age, education, body mass index, smoking status, and chemotherapy and radiotherapy treatment of cases). For each model, we obtain a regression coefficient (β) for each covariate and the p-value of a two-tailed Wald test (β significantly different from 0). P-values were corrected for multiple tests using the Benjamini-Yekutieli procedure to control the false discovery rate (FDR). Finally, features significantly associated with colorectal cancer (significance level of $\alpha = 0.05$) were selected. The same procedure was used to identify metabolomic features associated with nitrate and trihalomethane exposure among controls. These analyses were performed in Python 3.6.15. Linear regressions were performed using the statsmodels package (version 0.12.1) with statsmodels.formula.api. We controlled the false discovery rate (FDR) using the multiple-tests function from statsmodels.stats.multitest. Additionally, data manipulation and analysis were performed with numpy (version 1.16.3) and pandas (version 0.24.2).

Metabolome-Colorectal Cancer Multivariate Association Analyses. We employed Unit-Variance (UV) scaling to scale the features and fitted a partial least square-discriminant analysis (PLS-DA) with two components. The Hotelling T^2 statistic at 95% confidence, a multivariate generalization of the Student's t-distribution, was used for outlier detection. The optimal number of components was selected using 10-fold cross validation with a criterion of selecting the number of components that results in a Q^2Y value, defined over the test set as 1-predictive residual error sum of squares (PRESS)/total sum of squares (TSS), with a difference <5% compared to the previous number of components. Moreover, repeated cross-validation (with 100 repetitions), where the rows in the metabolomic matrix are shuffled each time, was used to check the distribution of Q^2Y values per component, which gives a comprehensive overview of each component's robustness, since the Q^2Y value obtained with K-Fold cross validation may be sensitive to row permutation of the explanatory matrix. The model was then refitted with the optimal number of components, and cross-validation was used to estimate validation metrics such as ROC curves and area under the curve (AUC). A permutation randomization test was also performed to validate the model and obtain empirical p-values (with 1000 permutation randomizations). Finally, features were selected based on the permuted p-value of the regression coefficients, β, and of the weights of the first component, w, with a significance level of $\alpha = 0.05$. This model was performed using pyChemometrics (version 0.1) in Python 3.6.15.

https://doi.org/10.1021/acs.est.3c05814
Environ. Sci. Technol. 2023, 57, 19316−19329

Metabolome-Water Exposures Multivariate Association Analyses. Features were UV-scaled and a PLS model for regression with 4 components was fitted. To select the optimal number of components, repeated cross-validation with 10-fold and 100 repeats was performed, maximizing the Q^2Y measure. To determine the optimal number of metabolomic features for each component, a sparse Partial Least Squares (sPLS) model was fitted using the previously determined optimal number of components. The model's performance was evaluated through repeated cross-validation (10-fold, 100 repeats) using Mean Absolute Error (MAE) as the metric. The number of features that minimized the MAE for each component was selected. Subsequently, a final sPLS model was fitted with the optimal number of components and the determined number of metabolomic features for each component. Repeated cross-validation was performed to compute the Mean Squared Error of Prediction (MSEP) metric for model validation. Additionally, a PLS model with the same number of components was fitted to assess the performance of the sparse model. To ensure feature stability, the selected features from all components of the final sPLS model were analyzed. Only features that were selected for a given component in at least 60% of the cross-validation folds (repeated 100 times) were retained for further analysis. This entire process was repeated for each water contaminant variable. These analyses were done using the R package mixOmics.

N-Integration Analysis. The DIABLO method from package mixOmics was used to build an integrated multivariate model using all the exposure variables, the metabolomic profile, and the colorectal cancer status as the response variable. To avoid overfitting, 80% of the data was randomly selected for training the model, and the remaining data was held out as a test set to evaluate the performance of the model on unseen data. The features were UV-scaled, the model was built using PLS regression, and the optimal number of components was selected using 10-fold cross-validation repeated 100 times. Sparse PLS regression was used to identify the most relevant variables for each data set that contribute to the joint variation between the data sets. Finally, the model was validated by using cross-validation techniques and predictions over the test set. A feature was considered stable if it was consistently selected across different cross validation folds for a given component. Only features selected for a given component in at least 60% of the cross-validation folds (repeated 100 times) were retained.

Annotation of Metabolic Features. In order to prioritize the annotation of features, we employed a ranking approach based on the following criteria. For the FDR-corrected linear regression results in relation to cancer, we determined thresholds using the median values of significant p-values (0.00051) and the median of the absolute values of regression coefficients (0.241). Regarding the FDR-corrected linear regression results for water contaminants, due to the limited number of selected features, most of which were common across all variables, we did not apply additional criteria. For the multivariate and N-integration models, we employed a high stability threshold of 90%. Applying these criteria we retained244 out of the initial 568 features, further clustered based on retention time proximity (0.05 min) and correlation of intensities across the samples >0.8 (Pearson) to assist in finding related ions. The m/z values of the features were searched against IARC's in-house metabolite databases and the human metabolome database (HMDB, https://hmdb.ca/, accessed on June fifth 2023) using $[M + H]^+$, and $[M + Na]^+$ ions, with 10 ppm molecular weight tolerance. Candidate metabolites matching their accurate mass were confirmed by comparing the MS/MS spectra and retention times against those of pure chemical reference standards whenever available. The level of identification was based on the recommendations of the Chemical Analysis Working Group of Metabolomics Standards Initiative.[18]

Enrichment and Pathway Analysis. For identified metabolites in common between exposure and outcome, we used the MetaboAnalyst, a widely used platform dedicated to metabolomics data analysis (https://new.metaboanalyst.ca/home.xhtml), to conduct pathway analysis (integrating enrichment analysis and pathway topology analysis). Additionally, we performed an analysis using the software Mummichog version 2.6.1 in Python version 3.6,[19] focusing on all the significant features identified in univariate analysis for colorectal cancer, nitrate, chloroform, and Br-THM. This program allows analyzing significantly enriched pathways directly from feature tables, bypassing metabolite identification. We supplied all 1629 features as the reference list and a significance cutoff p-value of 0.05. These analyses allow us to explore the metabolic pathways in which these metabolites are involved and potential links to the biological mechanisms involved in colorectal cancer pathogenesis.

RESULTS AND DISCUSSION

Study Population and Exposures. In Table 1, we present the characteristics of the study population, along with the corresponding p-values that assess the differences between cases and controls. For numerical variables, including exposure variables, we conducted two-tailed t tests, while for categorical

Table 1. Characteristics of the Study Population (n = 585)

	Controls (n = 293)	Cases (n = 292)	p-value
Age (years), mean (SD)	67.0 (7.7)	68.5 (9.2)	0.022
Sex, N (%)			0.407
Male	182 (62.1)	191 (65.4)	
Female	111 (37.9)	101 (34.6)	
Area, N (%)			0.998
Barcelona	94 (32.1)	95 (32.6)	
Cantabria	29 (9.9)	29 (9.9)	
Gipuzkoa	30 (10.2)	30 (10.3)	
Leon	95 (32.4)	92 (31.6)	
Navarra	45 (15.4)	46 (15.8)	
Smoking status, N (%)			0.003
Never	162 (57.3)	132 (45.2)	
Former	125 (42.7)	160 (54.8)	
Body mass index (kg/m2), mean (SD)	27.3 (4.1)	27.6 (4.6)	0.510
Education, N (%)			0.228
Less than primary school	73 (24.9)	82 (28.1)	
Primary school	129 (44.0)	122 (41.8)	
Secondary school	54 (18.4)	64 (21.9)	
University	37 (12.6)	24 (8.2)	
Total THMs (μg/L), mean (SD)	37.1 (25.0)	42.9 (33.6)	<0.001
Brominated THMs (μg/L)			<0.001
Mean (SD)	17.2 (21.8)	27.5 (30.6)	
Chloroform (μg/L), mean (SD)	19.9 (13.1)	15.4 (11.0)	<0.001
Nitrate (mg/L), mean (SD)	5.4 (5.5)	7.1 (7.4)	0.105

https://doi.org/10.1021/acs.est.3c05814
Environ. Sci. Technol. 2023, 57, 19316−19329

variables, we employed Pearson's chi-square tests to evaluate the significance of these differences. No major differences were found in the covariables between cases and controls; cases were slightly older, and there were more former smokers among cases than controls. The recruitment areas contributing the largest population were Barcelona and Leon, accounting for approximately 2/3 of study subjects. Cantabria was the area with less cases and controls (Table 1). The mean concentrations of total and brominated THMs were, respectively, 5.8 and 10.3 μg/L on average lower in controls than in cases. By contrast, chloroform concentrations were 4.5 μg/L higher in the controls than in cases. Waterborne ingested nitrate did not exhibit significant differences between controls and cases (Table 1). Out of the 292 cases, 67 (22.9%) received radiotherapy, and 168 (57.5%) received chemotherapy treatment before the sample collection. We focused on a recent (3 years before the interview) rather than long-term exposure window given that in previous studies we found higher associations with effect biomarkers linked to recent exposure compared to long-term exposure.[20]

Metabolic Features Derived from Untargeted Metabolomic Analyses. A total of 5354 features were found in the 591 samples analyzed. Analysis of the QC samples revealed good reproducibility along the run, with coefficients of variation (CV) consistently below 12% for a set of 10 known compounds when screened in the raw data of all QC samples (Agilent Qual, FBF algorithm using elemental composition as target and [M + H]+ as ion): tyrosine, tryptophan, phenylalanine, leucine, lauroylcarnitine, isoleucine, indolepropionic acid, indole-3-acetic acid, hippuric acid, and decanoylcarnitine. Three samples (2 cases, 1 control) with significantly low overall feature intensity due to an analytical issue were considered outliers and excluded. Two participants were excluded due to missing DBP exposure. One control with 32.1% of missing values in the feature intensities was also excluded. After the QC process (Supporting Figure 1), a total of 1629 features and 585 study participants (292 colorectal cancer cases, 293 controls) were included in the analysis.

In Table 2, we provide a breakdown of the number of features associated with each variable as determined through various statistical methods. Collectively, considering all the different analyses, we identified a total of 568 distinct features associated with either an exposure variable or colorectal cancer. The N-integration model identified 107 features discriminating colorectal cancer cases from controls through interrogation of correlations between the exposure variables and the metabolomic features block. Supporting Table 1 presents the shared features across different statistical models.

Associations from univariate analysis. Our FDR-corrected linear regression models identified 405 features significantly associated with colorectal cancer (Table 2), of which 258 were exclusively identified through this analysis (Figure 1). A total of 259 of the 405 features (64%) were negatively associated with cancer (Figure 2). Likewise, we identified 21, 20, 20, and 24 features significantly associated, respectively, with TTHM, Br-THM, chloroform, and nitrate exposures (Table 2). Among these, 15 features were significantly associated with all of the exposure variables (Figure 3b), of which 11 were exclusively found in these FDR-linear analyses (Figure 1). The sign of the association for all metabolomic features was consistent among all exposure variables. One feature (identified as creatine) was positively associated with increased exposure levels of TTHM, chloroform, nitrate, and also with colorectal cancer status. Another feature (mass = 700.5512; retention time = 8.20 min.) was positively associated with TTHM and Br-THM levels. The rest of the selected features were negatively correlated with exposure levels to water contaminants, representing 90.5% (TTHM), 95.0% (Br-THM), 95.0% (chloroform), and 95.8% (nitrate) of the features significantly associated through this model.

Table 2. Number of Metabolomic Features Associated with Mean Residential Levels (μg/L) of Total Trihalomethanes (TTHM), Brominated THMs (Br-THMs), Chloroform ($CHCl_3$), and Waterborne Ingested Nitrate (mg/day), Colorectal Cancer, and all Variables in an N-Integration Analysis (DIABLO model) (292 cases, 293 controls)[a]

	Univariate model (FDR corrected linear regression)	Multivariate model (sparse-PLS (exposure), PLS-DA (cancer))	Union count[a]
TTHM	21	61	82
Br-THM	20	4	24
Chloroform ($CHCl_3$)	20	61	80
Nitrate	24	-	24
Colorectal cancer	405	104	406
N-integration			107
Overall profile[a]			**568**

[a]"Union count" shows the features in the union of the univariate and multivariate analysis (that may include features in common). "Overall profile" includes metabolomic features that exhibited associations with at least one exposure variable or colorectal cancer.

Colorectal Cancer Multivariate Analysis (Partial Least Squares Discriminant Analysis). The optimal number of components was determined to be 4, as the Q^2Y measure stabilized at this point. The cross-validated ROC curve showed a mean AUC of 0.87 and the Q^2Y value of 0.38, indicating a good discriminative ability of the model. Permutation randomization tests confirmed the statistical significance of the model's AUC and Q^2Y values, with p-value <0.01. By applying permutation tests, we identified 16 features with significant regression coefficients (β) and 99 features with significant weights (w) in the first component. Combining these two sets as the union set, a total of 104 features were retained for the PLS-DA model for cancer, suggesting their potential relevance in predicting the cancer status (Table 2). All of these features are shared with the features identified in the FDR-corrected model with the exception of one feature (Figure 2a), indicating the robust association across multiple analyses. Interestingly, this particular feature re-emerged in the N-integration model (Figure 3b).

Exposure Multivariate Analysis (Sparse Partial Least Squares). The sparse partial least squares (sPLS) and PLS models were evaluated with two optimal components for TTHM. The sPLS compared to the PLS model achieved lower mean squared error of prediction (MSEP) values: 0.58 and 0.56 (components 1 and 2 in sPLS), vs 0.83 and 0.71 (components 1 and 2 in PLS). This suggests that the sPLS model outperforms the PLS model in terms of prediction accuracy, potentially due to its feature selection capabilities and sparsity constraints. In the sPLS model for TTHM, a total of 85 features was initially selected. However, after ensuring stability and robustness of the model, only 61 features consistently appeared in at least 60% of the cross-validation

https://doi.org/10.1021/acs.est.3c05814
Environ. Sci. Technol. 2023, 57, 19316−19329

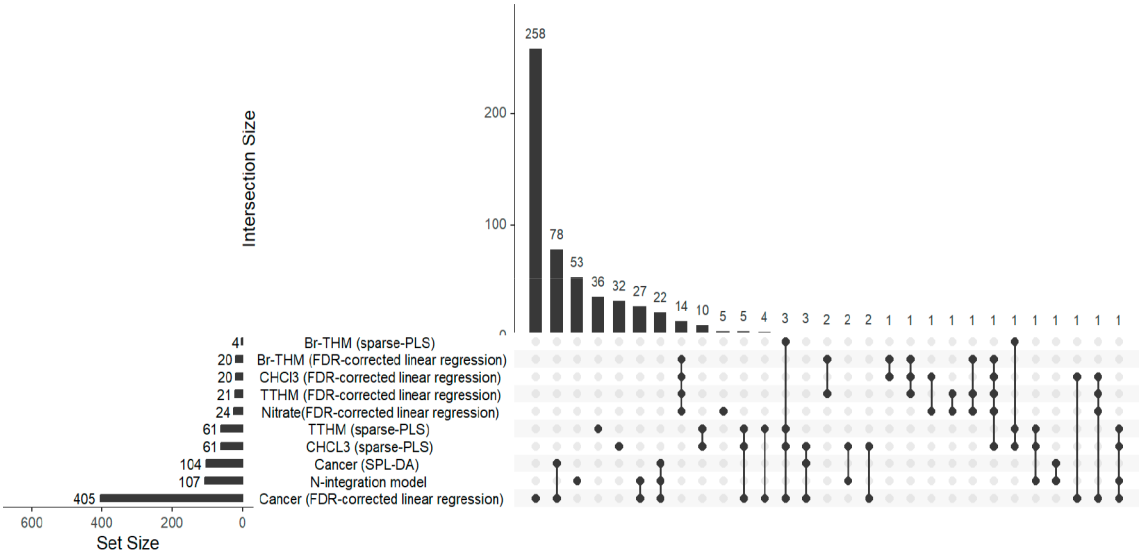


Figure 1. Upset plot showing the number of metabolomic features associated with exposure to water contaminants and colorectal cancer using different statistical models. Multivariate models are sparse-PLS (exposure) and PLS-DA (cancer). The remaining are FDR-corrected linear regression models or the N-integration model (DIABLO). The plot illustrates the intersections between the different models and the number of unique and shared features identified for each contaminant/cancer outcome. The subsets are ordered by frequency. The plot was generated using the R package UpSetR.[21]

folds, repeated 100 times (Table 2). Similarly, for chloroform, the sPLS model showed lower MSEP values (0.65 and 0.64) compared to the PLS model (0.87 and 0.78). The selected features for the sPLS model before stability were 90, of which 61 were retained after stability (Table 2). In the case of Br-THMs, both sPLS and PLS models were evaluated with one component. The sPLS model had an MSEP of 0.83, while the PLS model had an MSEP of 0.92. The selected features before stability were 5, and after stability, 4 features were retained (Table 2). Regarding waterborne ingested nitrate, the preliminary PLS model showed a very low Q^2Y measure of 0.02, indicating poor predictive ability. Therefore, a sPLS analysis was not performed for this variable. Overall, the results show the potential advantages of using the sPLS over the traditional PLS model in terms of prediction accuracy, especially when dealing with exposure variables like TTHM and chloroform. The selection of stable features based on the cross-validation process further enhances the robustness of the models, with a reduced number of features retained after stability checks. Figure 3d presents a Venn diagram illustrating the overlap among features associated with different water contaminants according to the sparse-PLS model. The total of 4 features associated with Br-THMs are shared with TTHMs and chloroform. Furthermore, it shows a substantial intersection between the features associated with TTHMs and chloroform (N = 21 features). However, unlike the features associated with cancer, the intersection between features found in the FDR-corrected linear regression models and sparse-PLS differs for water contaminants. Specifically, there is no intersection for TTHM and Br-THM, and only one feature overlaps with chloroform (Supporting Table 1). The differences between univariate and multivariate analyses can be attributed to various factors, such as independent variables (metabolite abundances) that may complement each other in the prediction of the dependent variable and the effect of consistency at large or multiple testing corrections increasing the risk of false negatives. Despite these discrepancies, seeking validation of univariate results through multivariate analysis, or vice versa, may not be appropriate. The two methods provide complementary results and offer valuable insights into the associations between metabolomic features and the studied variables.[22]

We conducted a complementary analysis to explore the potential variations in associations across different exposure levels. Except for waterborne ingested nitrate, which lacked a multivariate analysis, sparse-PLS models were fitted for each water contaminant variable and categorized into two groups using the same methodology and criteria as the main sparse-PLS models. Controls were divided into higher and lower exposure groups based on a threshold defined as the median value of exposure among controls. We identified 4 metabolomic features associated with higher TTHM levels, 3 of which overlapped with the main analysis based on all controls (Supporting Figure 3). For lower TTHM levels, we found 42 features, 1 of which was shared with the main analysis. Notably, this feature also appeared in the high exposure category, indicating its consistent association across all levels. Among the 61 metabolomic features identified in the primary TTHM model, 58 did not exhibit differential associations across different exposure levels. In the case of chloroform, 16 metabolomic features were associated with the lower exposure group, while the higher exposure group exhibited associations with 20 features. There was no overlap between the lower and higher exposure groups. However, the lower exposure group shared 4 features with the main analysis, and the higher exposure group shared 12 features, indicating some consistency in associations. Regarding Br-THMs, the higher exposure group displayed a negative Q^2Y measure, indicating poor predictive ability. Consequently, a multivariate analysis was not conducted for this group. In the lower exposure group, we identified 18 metabolomic features associated with Br-THMs. Only 1 feature was common between the lower exposure group and the 4 features found in the main analysis. These findings suggest potential variations in the associations between water contaminants and metabolomic profiles across different exposure levels. The robustness of the main analysis findings

https://doi.org/10.1021/acs.est.3c05814
Environ. Sci. Technol. 2023, 57, 19316−19329

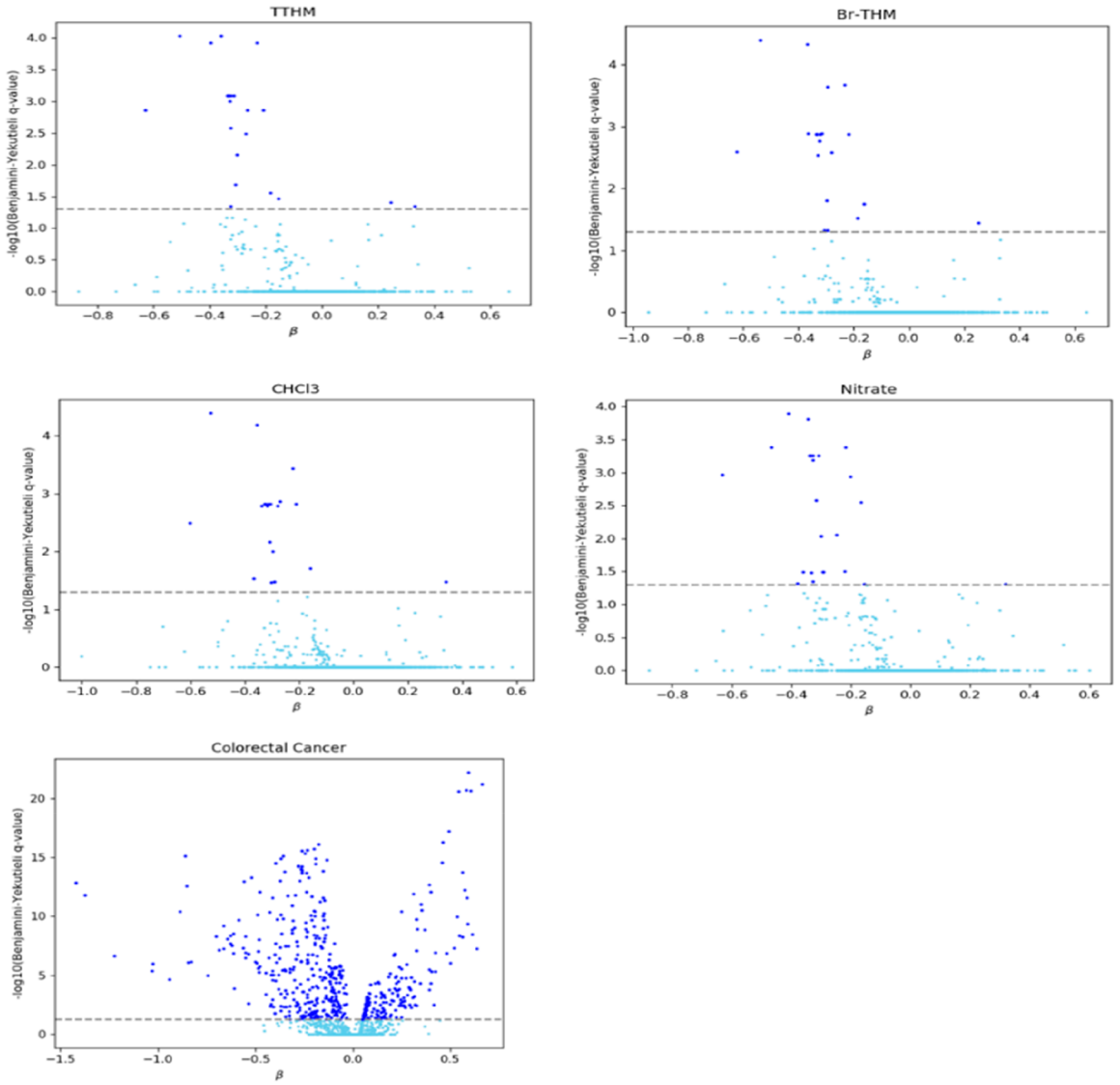


Figure 2. Results from the false discovery rate (FDR)-corrected linear regression models for colorectal cancer and water contaminants. Each dot represents a metabolomic feature and is represented by the $-\log_{10}$(Benjamini-Yekuteli q-value) against the regression coefficient (β) of the variable. Horizontal line sets the cutoff of $-\log_{10}(0.5)$, implying that the obtained signature has an estimated FDR of 5%. Features significantly associated with each variable are colored in darker blue.

is evident as several features were consistently associated with the contaminants regardless of the exposure level. In Supporting Figure 3, Venn diagrams are provided, illustrating the metabolomic features associated with each specific disinfection byproduct and their respective exposure groups.

N-Integration Analysis. In the N-integration analysis, we employed 10-fold cross-validation repeated 100 times to determine the optimal number of components, which was found to be 4. After fitting the final model, the metabolomic profile was defined as the union set of features selected across the four components. Initially, the union of these selected features comprised a total of 167 features. However, to ensure the stability and robustness of the model, only those features that consistently appeared in at least 60% of the cross-validation folds, repeated 100 times, were retained. The final metabolomic profile from the N-integration analysis model consisted of 107 features (Table 2), out of which 53 were exclusively identified through this model (Figure 1), and 50 were found to be shared with the metabolomic profile associated with cancer as determined by the FDR-corrected linear regression model. Notably, the N-integration model also identified the only feature that was associated with cancer in the PLS-DA model but not in the FDR-corrected linear regression model (Figure 3b). In Supporting Figure 2, the N-integration model reveals strong negative correlations between Br-THM and metabolomic features, while nitrate and chloroform display strong positive correlations. However, the TTHM does not exhibit a strong correlation in any of the four components of the model.

When assessing the performance of the N-integration analysis using 10-fold cross-validation repeated 100 times, we observed error rate values ranging from 0.31 to 0.24 across the four components. This indicates that the model generally performs well with the lowest error rate observed in the fourth component (0.24). Balanced error rate (BER), which considers sensitivity and specificity, ranged from 0.31 to 0.25, with the fourth component achieving the lowest balanced error rate. For the metabolomic features block with 4 components, the model

https://doi.org/10.1021/acs.est.3c05814
Environ. Sci. Technol. 2023, 57, 19316−19329

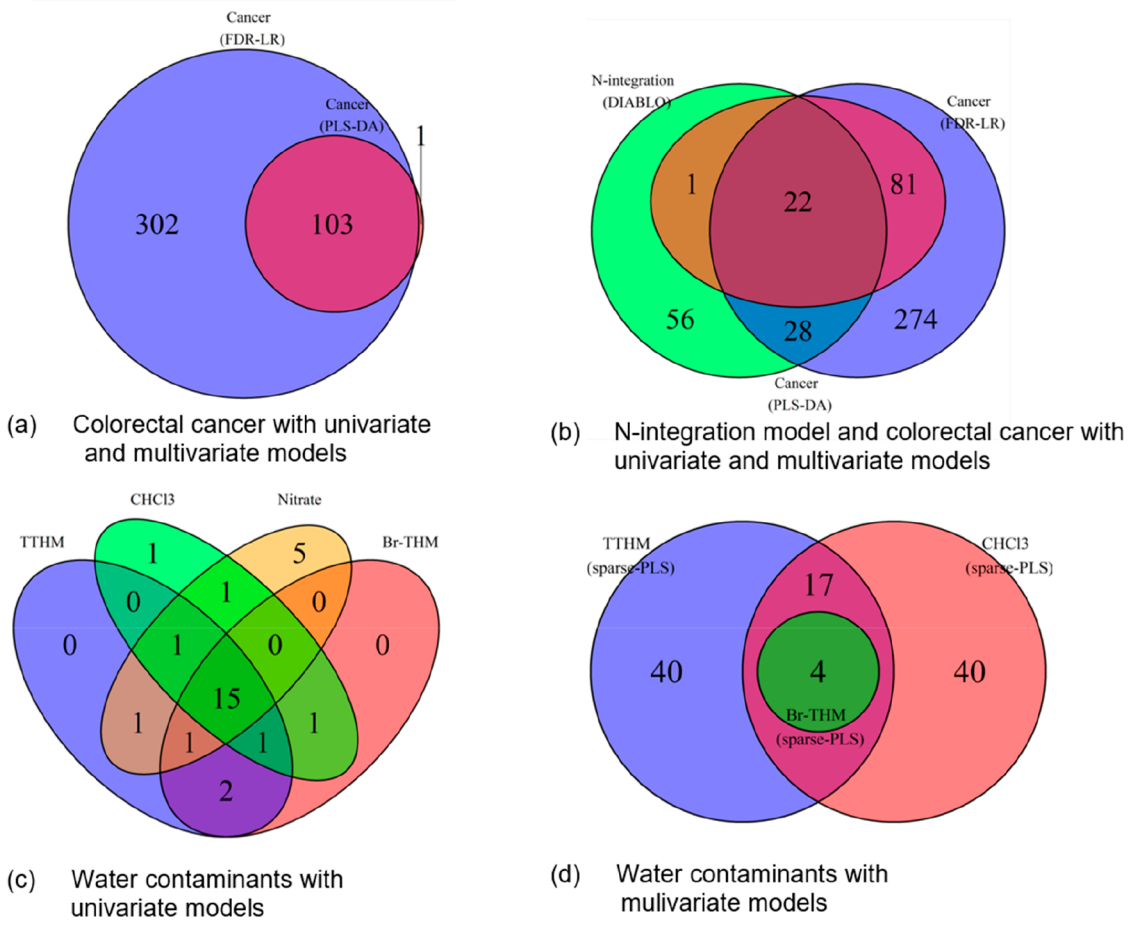


Figure 3. Venn diagrams showing the metabolomic features associated with water contaminants and colorectal cancer using different statistical models and the overlap between models.

achieved an AUC of 0.8928, indicating a high level of accuracy in discriminating between cases and controls based on the metabolite's profiles. Overall, the model demonstrated satisfactory performance.

We further tested the model on an external test set using the Weighted Vote method in N-integration analysis, resulting in a confusion matrix. The matrix showed that 26 cases were correctly predicted as cases, 17 cases were incorrectly predicted as controls, 7 controls were incorrectly predicted as cases, and 37 controls were correctly predicted as controls. The calculated BER based on the confusion matrix was 0.28, which is relatively low. This suggests that the model performed reasonably well in terms of balancing sensitivity and specificity even when applied to unseen data. In summary, the N-integration analysis demonstrated good performance with consistent results across multiple evaluation metrics.

Annotated Metabolites and Metabolic Pathways. Among the 568 metabolomic features associated with at least one of the water contaminants or with colorectal cancer, 26 metabolites could be annotated (Table 3). If multiple ion species were identified, then the most intense was presented. The level of identification was based on the recommendations of the Chemical Analysis Working Group of Metabolomics Standards Initiative.[18] Three of the annotated metabolites were associated with both colorectal cancer and at least one of the water contaminants. These were creatine, positively associated; lysophosphatidylcholines (LysoPC) (20:2), inversely associated; and 1-methylnicotinamide (MNA), inversely associated with cancer and positively associated with chloroform and Br-THMs (Table 3).

Nicotinamide (NA) metabolism involving MNA was the pathway identified as having higher significance. Previous studies have also documented an inverse correlation between MNA and colorectal cancer.[23,24] NA represents a bioactive form of vitamin B3 and is a precursor of nicotinamide-adenine dinucleotide (NAD+) coenzymes. NAD+ is a key molecule participating in a wide range of intracellular events, including transcription regulation, longevity, genome stability, and response to DNA damage.[25] Nicotinamide *N*-methyltransferase (NNMT) catalytic activity significantly contributes to the regulation of NA and NAD+ intracellular levels, participating in an irreversible catabolism of NA to MNA, which is excreted through urine and NA is no longer available as a precursor for NAD+ biosynthesis.[26] NNMT is mainly expressed in the liver, belongs to phase II metabolizing enzymes and is suggested to be involved in the biotransformation and detoxification of many xenobiotics.[27] Some studies show NNMT overexpression associated with colorectal cancer[28−30] Although further research is required to confirm this hypothesis, we suggest that exposure to chloroform and Br-THM may result in the deregulation of NNMT, causing its overexpression.

Cytochrome P-450 (CYP) metabolism is a pathway identified between nitrate exposure and colorectal cancer by Mummichog et al. (Table 4). Most xenobiotics must be biotransformed to have toxic effects, which is a two-stage process carried out by phase I and phase II metabolizing

https://doi.org/10.1021/acs.est.3c05814
Environ. Sci. Technol. 2023, 57, 19316−19329

Table 3. Annotated Metabolites

Metabolite	Mass[a]	Retention Time	ID level[b]	Direction	Associated with
1-methylnicotinamide	136.0624	0.60	1	DOWN|UP|UP	Cancer, chloroform, Br-THM
2-Hydroxy-3-methylbutyric acid	202.0141	2.76	1	DOWN	TTHM, chloroform, Br-THM, nitrate
Bilirubin	582.249	7.97	1	DOWN	TTHM, nitrate
Creatine	131.0695	0.66	1	UP	Cancer, TTHM, chloroform, nitrate
Creatinine	113.0591	0.61	1	DOWN	TTHM, chloroform, Br-THM
Cyclo(prolyl-valyl)	196.1221	3.14	1	DOWN	Cancer
Docosahexaenoic acid	350.2198	7.25	1	UP	Cancer
Ethyl glucoside related peak	230.077	0.86	-	DOWN	Cancer
Hexanoylcarnitine (C6:0)	259.1778	3.34	2	UP	Cancer
Hippuric acid	179.0597	3.10	1	DOWN	Cancer
Indole-3-propionic acid	189.08	4.60	1	DOWN	Cancer
Indolelactic acid	205.0753	3.88	1	DOWN	TTHM, chloroform, Br-THM, nitrate
Inosine	268.0821	1.67	1	DOWN	Chloroform
Isatin	147.0321	3.34	1	DOWN	TTHM
L-glutamine	146.0698	0.62	1	DOWN	TTHM
LysoPC(14:0)	467.3018	6.75	2	DOWN	Cancer
LysoPC(16:1)	515.2993	6.84	2	DOWN	Cancer
LysoPC(18:0)	523.3656	7.25	2	DOWN	Cancer
LysoPC(18:2)	519.3343	6.92	2	DOWN	Cancer
LysoPC(20:2)	547.3616	7.16	2	DOWN	Cancer, chloroform, Br-THM, nitrate
LysoPC(P-16:0)	501.3201	7.14	2	DOWN	Cancer
N6,N6,N6-Trimethyl-L-lysine	188.1528	0.54	1	DOWN	TTHM, chloroform, Br-THM, nitrate
Nonanoylcarnitine (C9:0)	301.2249	4.64	2	DOWN	Cancer
Tetradecenoylcarnitine (C14:1)	369.2883	5.83	2	UP	Cancer
Theobromine	180.0651	2.39	1	DOWN	Cancer
Trigonelline	137.048	0.68	1	DOWN	Cancer

[a]Monoisotopic mass calculated from the m/z peak that best represents the metabolite. In cases where multiple features were associated with the metabolite, data from the most intense feature is presented. [b]Identification (ID) level indicates the degree of confidence in annotation (from ref 18). Level 1 (identity confirmed): retention time and MS/MS matched with an authentic chemical standard; Level 2 (putative annotation): no standard available or analyzed but mass within 5 ppm mass error and MS/MS spectra matches with those in a database.

Table 4. Metabolic Pathways Significantly Associated with Colorectal Cancer, Chloroform, Brominated THMs, and Nitrate Exposure Based on 1629 Features As a Reference List and Significant Features (p-value <0.05) Obtained from the Univariate Regression Model[a]

Pathways	Colorectal cancer	Chloroform	Br-THM	Nitrate
Arginine and Proline Metabolism		1 (7)	1 (7)	1 (7)
Aspartate and asparagine metabolism				1 (8)
D4&E4-neuroprostanes formation	2 (3)			
Drug metabolism - cytochrome P450	6 (10)			1 (10)
Methionine and cysteine metabolism		1 (3)	1 (3)	1 (3)
Tyrosine metabolism	6 (16)	1 (16)	1 (16)	1 (16)
Urea cycle/amino group metabolism		1 (8)	1 (8)	1 (8)
Valine, leucine and isoleucine degradation		1 (2)		1 (2)

[a]N = 405 for cancer, N = 24 for nitrate, N = 20 for chloroform and brominated THMs, based on Mummichog software. The numbers indicate significant features (number of features in the pathway in our reference list).

enzymes. Phase I enzymes such as those belonging to the CYP family are involved in the initial oxidation, reduction, or dealkylation of carcinogens; this phase generally leads to the production of active intermediate metabolites. The three CYP isoenzymes, CYP2E1, CYP1A2 and CYP3A4 have previously been identified as important in the THM metabolism.[31−33] There are also data suggesting that chlorinated disinfectants mixtures are able to perturb CYP-mediated reactions and induce oxidative stress.[34] Likewise, the CYP3A superfamily specifically participates in nitric oxide formation in the liver from organic nitrates.[35] Interestingly, it has been observed that chronic exposure to organic nitrates significantly decreased hepatic P450, i.e., P450-dependent drug metabolism may be drastically affected after continuous organic nitrate exposure.[36] In turn, CYP has a major role in tumor development via metabolism of many carcinogens[37] and specific CYP have also been shown to be overexpressed in colorectal cancer, such as the isoenzymes CYP3A,[38] CYP1B1,[39] CYP2S1, CYP2U1, CYP3A5, and CYP51.[40] All together, these findings may provide an explanation for the link between THMs exposure and colorectal cancer. Long-term human exposure to THMs warrants further investigations into both the possible epigenetic and genetic mechanisms of toxicity of these compounds.

Our pathway analyses by the program Mummichog also identified that the tyrosine metabolism involved was in all water contaminants as well as colorectal cancer associations (Table 4). Tyrosine kinases play crucial roles in various biological processes including growth, differentiation, metabolism, and apoptosis in response to internal and external signals. Tyrosine kinases have been linked to the development of cancer. While these enzymes are tightly controlled in healthy cells, mutations, overexpression, and autocrine/paracrine

https://doi.org/10.1021/acs.est.3c05814
Environ. Sci. Technol. 2023, 57, 19316−19329

stimulation can confer oncogenic properties contributing to malignancy.[41] Tyrosine nitration and halogenation, which consists in the addition of a nitro-(NO2) group, chloride or bromide, to the phenolic ring of tyrosine residues in proteins, is involved in carcinogenesis.[42] It has been shown both in vitro and in vivo widespread nitration of tyrosine residues of cellular proteins in response to increased intracellular nitric oxide (NO) in colon cancer cells. We, therefore, speculate that tyrosine nitration may be responsible, at least in part, for the effect of nitrate exposure (internal NO source) on cancer cell growth and that this may represent a mechanism of colorectal carcinogenesis.[43]

Creatine participates in arginine and proline metabolism, which has also been identified as a significant pathway for biological effects related to THMs and nitrate exposures (Table 4). We also detected creatine levels higher in colorectal cancer cases than in controls and those more exposed to both THMs and nitrate (Table 3). However, prior evidence is inconsistent. While two previous studies have reported elevated levels of proline in plasma of patients with colorectal cancer,[23,44] another study observed reduced levels of proline in colorectal cases compared to control subjects.[45]

The sex-determining region Y (SRY)-box (SOX) family plays a crucial role in carcinogenesis and cancer progression. While the dysregulation of SOX12 has been linked to colorectal cancer, the underlying mechanisms remain elusive.[46] It is established that SOX12 promotes asparagine synthesis by activating genes such as glutaminase (GLS), glutamic oxaloacetic transaminase 2 (GOT2), and asparagine synthetase (ASNS). Given our data's correlation between disturbances in aspartate and asparagine metabolism with nitrate exposure (Table 4), we propose that this pathway may contribute to the association between nitrate exposure and colorectal cancer.[47]

Our pathway analysis also revealed the involvement of lysophospholipid lysoPC(20:2) in glycerophospholipid metabolism as a metabolic pathway between contaminants in water and colorectal cancer. Several other lysophosphatidylcholines have been inversely related to colorectal cancer in this study (Table 3), supporting the hypothesis of dysregulated lipid metabolism in cancer.[48] In line with our findings, multiple studies have consistently demonstrated a noteworthy decrease in the levels of various lysophosphatidylcholines (LysoPCs) in colorectal cancer cases.[23,24,49−51] Considering this prior evidence and our discoveries, it is plausible that dysregulated lipid metabolism serves as a potential mechanism by which both THMs and nitrates contribute to the development of colorectal cancer.

In agreement with previous research,[23] we identified many acyl carnitines as associated with colorectal cancer (Table 3). Metabolites associated with the carnitine cycle play a crucial role in modulating fatty acid metabolism and facilitating mitochondrial fatty acid transport. These metabolites can also exert an influence on the composition of the gut microbiota.[52] In the PISCINA-II study (EXPOsOMICS project), which investigated volunteers who swam for 40 min in an indoor pool, the hexanoylcarnitine was associated with bromodichloromethane, and nonanoylcarnitine was associated with bromoform in exhaled breath. However, in this data, we found these metabolites associated with colorectal cancer but not with any water contaminant variable. This discrepancy might be attributed to our grouping of all bromates together, whereas the PISCINA-II study analyzed separately each individual Br-THM.[14]

Finally, bilirubin was negatively associated with TTHM and nitrate (Table 3). Although bilirubin was not associated with cancer risk in our study, it has been found to exhibit significant antioxidant and anticancer properties in previous studies, which have proposed the bilirubin as a potentially valuable prognostic biomarker for overall survival in advanced colorectal cancer.[53] Likewise, there is robust evidence substantiating the participation of diminished levels of bilirubin in the pathogenesis of colorectal carcinogenesis.[23,24,51] Thus, an additional potential mechanism by which THMs and nitrate elevate the risk of colorectal cancer may involve the reduction of bilirubin levels (porphyrin and chlorophyll metabolism).

Strengths and Limitations. Metabolomic profiles can be influenced by various factors, including the disease status. Given that metabolomic analysis was conducted after cancer diagnosis, reverse causation cannot be ruled out. The associations observed with cancer need to be cautiously interpreted, even if we accounted for chemotherapy and radiotherapy treatment, which led to 106 features less compared to the model without radio- and chemotherapy treatment adjustment. Longitudinal studies are needed to elucidate the dynamic nature of these associations and establish causal relationships. In contrast, exposure assigned corresponded to a period before diagnosis and sample collection and excluded 2 years before the interview. This is better aligned with the hypothesized causal pathway, where exposure precedes changes in the metabolomic profile and, subsequently, the development of cancer. Thus, the associations observed between water contaminant exposure and metabolomic features among controls provide more robust evidence for a plausible causal relationship.

Finally, it is important to emphasize that pathway analyses, while valuable for generating hypotheses and estimating pathway-level differences using metabolomics data, are typically not exhaustive or definitive in nature. Conflicting findings regarding metabolite levels across different studies can often be attributed to variations in study populations, divergent approaches to sample collection and preparation, diverse analytical platforms employed, and disparities in the statistical methodologies employed.

The advantages of untargeted metabolomic analysis include the ability to identify novel metabolic features and gain insights into impacted pathways from an agnostic perspective, enhancing our understanding of the underlying mechanisms. However, independent validation in other cohorts and the confirmation of identity by targeted metabolomics would be needed to confirm the detected associations. The statistical analysis of our study encompasses univariate, multivariate, and N-integration methods. This multifaceted approach allows us to capture a broader and more nuanced understanding of the associations between metabolomic features, water contaminants, and colorectal cancer. By applying different analytical techniques, we maximize the robustness and reliability of our findings, providing a more comprehensive picture of the complex relationships within the metabolome.

In conclusion, this is, to the best of our knowledge, the first study evaluating the metabolomic profile associated with exposure to trihalomethanes and nitrate in drinking water and colorectal cancer risk that constitute widespread environmental exposures and one of the most frequent cancer sites. Our comprehensive analysis using untargeted metabolomic analysis and a variety of bioinformatic approaches suggests the involvement of various metabolic pathways including nicoti-

https://doi.org/10.1021/acs.est.3c05814
Environ. Sci. Technol. 2023, 57, 19316−19329

namide, cytochrome P-450, and tyrosine metabolism, among others. These findings provide insights into potential biological mechanisms involved and underscore the need for deeper investigation into these pathways as potential targets for future research on colorectal cancer and the evaluated exposures.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.3c05814.

Supporting Figures 1, 2, and 3; Supporting Table 1 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Cristina M. Villanueva − *ISGlobal, Barcelona 08003, Spain; CIBER Epidemiología y Salud Pública (CIBERESP), Madrid 28029, Spain; Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain; IMIM (Hospital del Mar Medical Research Institute), Barcelona 08003, Spain;* orcid.org/0000-0002-0783-1259; Email: cristina.villanueva@isglobal.org

Authors

Jose A. Alcolea − *ISGlobal, Barcelona 08003, Spain; CIBER Epidemiología y Salud Pública (CIBERESP), Madrid 28029, Spain; Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain*

Carolina Donat-Vargas − *ISGlobal, Barcelona 08003, Spain; CIBER Epidemiología y Salud Pública (CIBERESP), Madrid 28029, Spain; Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain; Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm 17177, Sweden*

Anastasia Chrysovalantou Chatziioannou − *International Agency for Research on Cancer, CS 90627 69366 Lyon, France;* orcid.org/0000-0002-1973-7542

Pekka Keski-Rahkonen − *International Agency for Research on Cancer, CS 90627 69366 Lyon, France*

Nivonirina Robinot − *International Agency for Research on Cancer, CS 90627 69366 Lyon, France*

Antonio José Molina − *Research Group in Gene - Environment and Health Interactions (GIIGAS)/Institute of Biomedicine (IBIOMED) and Faculty of Health Sciences, Department of Biomedical Sciences, Area of Preventive Medicine and Public Health, Universidad de León, León 24071, Spain*

Pilar Amiano − *CIBER Epidemiología y Salud Pública (CIBERESP), Madrid 28029, Spain; Ministry of Health of the Basque Government, Sub Directorate for Public Health and Addictions of Gipuzkoa, BioGipuzkoa (BioDonostia) Health Research Institute, San Sebastián 20013, Spain*

Inés Gómez-Acebo − *CIBER Epidemiología y Salud Pública (CIBERESP), Madrid 28029, Spain; Universidad de Cantabria-IDIVAL, Santander 39011, Spain*

Gemma Castaño-Vinyals − *ISGlobal, Barcelona 08003, Spain; CIBER Epidemiología y Salud Pública (CIBERESP), Madrid 28029, Spain; Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain; IMIM (Hospital del Mar Medical Research Institute), Barcelona 08003, Spain*

Lea Maitre − *ISGlobal, Barcelona 08003, Spain; CIBER Epidemiología y Salud Pública (CIBERESP), Madrid 28029, Spain; Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain;* orcid.org/0000-0003-3682-7117

Marc Chadeau-Hyam − *MRC Centre for Environment and Health, School of Public Health, Imperial College London, London W2 1PG, United Kingdom;* orcid.org/0000-0001-8341-5436

Sonia Dagnino − *MRC Centre for Environment and Health, School of Public Health, Imperial College London, London W2 1PG, United Kingdom; Transporters in Imaging and Radiotherapy in Oncology (TIRO), School of Medicine, Direction de la Recherche Fondamentale (DRF), Institut des Sciences du Vivant Frédéric Joliot, Commissariat à l'Energie Atomique et aux Énergies Alternatives (CEA), Université Côte d'Azur (UCA), Nice 06107, France*

Sibo Lucas Cheng − *MRC Centre for Environment and Health, School of Public Health, Imperial College London, London W2 1PG, United Kingdom*

Augustin Scalbert − *International Agency for Research on Cancer, CS 90627 69366 Lyon, France*

Paolo Vineis − *MRC Centre for Environment and Health, School of Public Health, Imperial College London, London W2 1PG, United Kingdom;* orcid.org/0000-0001-8935-4566

Manolis Kogevinas − *ISGlobal, Barcelona 08003, Spain; CIBER Epidemiología y Salud Pública (CIBERESP), Madrid 28029, Spain; Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain; IMIM (Hospital del Mar Medical Research Institute), Barcelona 08003, Spain*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.3c05814

Author Contributions

Jose A. Alcolea: conceptualization, data curation, formal data analysis, investigation, methodology, software, visualization, writing original draft, and writing review and editing; Carolina Donat-Vargas: conceptualization, investigation, methodology, software, visualization, writing original draft, and writing review and editing; Anastasia Chrysovalantou Chatziioannou: conceptualization, data curation, formal laboratory analysis, investigation, methodology, software, visualization, and writing review and editing; Pekka Keski-Rahkonen: conceptualization, data curation, formal laboratory analysis, investigation, methodology, software, visualization, and writing review and editing; Nivonirina Robinot: conceptualization, data curation, formal laboratory analysis, investigation, methodology, software, visualization, and writing review and editing; Antonio José Molina: conceptualization, resources, investigation, methodology, and writing review and editing; Pilar Amiano: conceptualization, resources, investigation, methodology, and writing review and editing; Inés Gómez-Acebo: conceptualization, resources, investigation, methodology, and writing review and editing; Gemma Castaño-Vinyals: conceptualization, investigation, methodology, and writing review and editing; Lea Maitre: conceptualization, methodology, and writing review and editing. Marc Chadeau-Hyam: conceptualization, investigation, methodology, and writing review and editing; Sonia Dagnino: conceptualization, investigation, methodology, and writing review and editing; Sibo Lucas Cheng: conceptualization, investigation, methodology, and writing review and editing; Augustin Scalbert: conceptualization, data curation, formal laboratory analysis, investigation, methodology, software, visualization, and writing review and editing;

https://doi.org/10.1021/acs.est.3c05814
Environ. Sci. Technol. 2023, 57, 19316−19329

Paolo Vineis: conceptualization, resources, investigation, methodology, and writing review and editing; Manolis Kogevinas: conceptualization, resources, investigation, methodology, and writing review and editing; Cristina M. Villanueva: conceptualization, investigation, methodology, supervision, and writing review and editing.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge support from the grant CEX2018-000806-S funded by MCIN/AEI/10.13039/501100011033, and support from the Generalitat de Catalunya through the CERCA Program. This work was funded by the seventh Framework Programme EXPOSOMICS Project (grant agreement 308610), the Acción Transversal del Cáncer del Consejo de Ministros del 11/10/2007, and the Instituto de Salud Carlos III-FEDER (PI08/1770, PI08/0533, PI08/1359, PS09/00773, PS09/01286, PS09/01903, PS09/02078, PS09/01662, PI11/01403, PI11/01889, PI11/00226) FIS grants, by the Fundación Marqués de Valdecilla (API 10/09), by the ICGC International Cancer Genome Consortium CLL (The ICGC CLL-Genome Project is funded by Spanish Ministerio de Economía y Competitividad (MINECO) through the Instituto de Salud Carlos III (ISCIII) and Red Temática de Investigación del Cáncer (RTICC) del ISCIII (RD12/0036/0036)), by the Junta de Castilla y León (LE22A10-2), by the Consejería de Salud of the Junta de Andalucía (PI-0571-2009, PI-0306-2011, salud201200057018tra), by the Conselleria de Sanitat of the Generalitat Valenciana (AP_061/10), by the Recercaixa (2010ACUP 00310), by the Regional Government of the Basque Country, by the Consejería de Sanidad de la Región de Murcia, by the European Commission grants FOOD-CT-2006-036224-HIWATE, by the Spanish Association Against Cancer (AECC) Scientific Foundation, by the Catalan Government- Agency for Management of University and Research Grants (AGAUR) grants 2017SGR723 and 2014SGR850, by the Fundación Caja de Ahorros de Asturias and by the University of Oviedo. We are grateful to Ana Espinosa (ISGlobal) for contributing to the quality control of data. Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article, and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization.

■ ABBREVIATIONS

AUC, area under the curve; BER, Balanced error rate; Br-THMs, brominated trihalomethanes; CHCl3, chloroform; CV, coefficients of variation; DBP, disinfection byproduct; FDR, false discovery rate; LC, liquid chromatography; MAE, mean absolute error; MFE, molecular feature extraction; ms, mass spectrometry; MSEP, mean squared error of prediction; PLS, partial Least Squares; PLS-DA, partial least square-discriminant analysis; QC, quality control; QTOF, quadrupole time-of-flight; sPLS, sparse partial least squares; TTHM, total trihalomethanes; UHPLC, ultrahigh performance liquid-chromatography

■ REFERENCES

(1) Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin* **2021**, *71* (3), 209−249.

(2) Bosman, F. T.; Yan, P. Molecular pathology of colorectal cancer. *Pol J. Pathol* **2014**, *65* (4), 257−266.

(3) Richardson, S. D.; Plewa, M. J.; Wagner, E. D.; Schoeny, R.; Demarini, D. M. Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: a review and roadmap for research. *Mutat. Res.* **2007**, *636* (1−3), 178−242.

(4) Villanueva, C. M.; Cantor, K. P.; Grimalt, J. O.; Malats, N.; Silverman, D.; Tardon, A.; Garcia-Closas, R.; Serra, C.; Carrato, A.; Castano-Vinyals, G. Bladder cancer and exposure to water disinfection by-products through ingestion, bathing, showering, and swimming in pools. *Am. J. Epidemiol* **2006**, *165* (2), 148−156.

(5) Guha, N.; Loomis, D.; Grosse, Y.; Lauby-Secretan, B.; Ghissassi, F. E.; Bouvard, V.; Benbrahim-Tallaa, L.; Baan, R.; Mattock, H.; Straif, K. Carcinogenicity of trichloroethylene, tetrachloroethylene, some other chlorinated solvents, and their metabolites. *Lancet Oncol* **2012**, *13* (12), 1192−1193.

(6) Jones, R. R.; DellaValle, C. T.; Weyer, P. J.; Robien, K.; Cantor, K. P.; Krasner, S.; Beane Freeman, L. E.; Ward, M. H. Ingested nitrate, disinfection by-products, and risk of colon and rectal cancers in the Iowa Women's Health Study cohort. *Environ. Int.* **2019**, *126*, 242−251.

(7) Villanueva, C. M.; Gracia-Lavedan, E.; Bosetti, C.; Righi, E.; Molina, A. J.; Martin, V.; Boldo, E.; Aragones, N.; Perez-Gomez, B.; Pollan, M.; Acebo, I. G.; Altzibar, J. M.; Zabala, A. J.; Ardanaz, E.; Peiró, R.; Tardón, A.; Chirlaque, M. D.; Tavani, A.; Polesel, J.; Serraino, D.; Pisa, F.; Castaño-Vinyals, G.; Espinosa, A.; Espejo-Herrera, N.; Palau, M.; Moreno, V.; La Vecchia, C.; Aggazzotti, G.; Nieuwenhuijsen, M. J.; Kogevinas, M. Colorectal Cancer and Long-Term Exposure to Trihalomethanes in Drinking Water: A Multicenter Case-Control Study in Spain and Italy. *Environ. Health Perspect* **2017**, *125* (1), 56−65.

(8) Rahman, M. B.; Driscoll, T.; Cowie, C.; Armstrong, B. K. Disinfection by-products in drinking water and colorectal cancer: a meta-analysis. *Int. J. Epidemiol* **2010**, *39* (3), 733−745.

(9) Ward, M. H.; Jones, R. R.; Brender, J. D.; de Kok, T. M.; Weyer, P. J.; Nolan, B. T.; Villanueva, C. M.; van Breda, S. G. Drinking Water Nitrate and Human Health: An Updated Review. *Int. J. Environ. Res. Public Health* **2018**, *15* (7), 1557.

(10) IARC. Ingested nitrate and nitrite, and cyanobacterial peptide toxins. In *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*; International Agency for Research on Cancer, 1987; Vol. *94*.

(11) Espejo-Herrera, N.; Gracia-Lavedan, E.; Boldo, E.; Aragones, N.; Perez-Gomez, B.; Pollan, M.; Molina, A. J.; Fernandez, T.; Martin, V.; La Vecchia, C.; et al. Colorectal cancer risk and nitrate exposure through drinking water and diet. *Int. J. Cancer* **2016**, *139* (2), 334−346.

(12) Picetti, R.; Deeney, M.; Pastorino, S.; Miller, M. R.; Shah, A.; Leon, D. A.; Dangour, A. D.; Green, R. Nitrate and nitrite contamination in drinking water and cancer risk: A systematic review with meta-analysis. *Environ. Res.* **2022**, *210*, No. 112988.

(13) Kim, S. J.; Kim, S. H.; Kim, J. H.; Hwang, S.; Yoo, H. J. Understanding Metabolomics in Biomedical Research. *Endocrinol Metab (Seoul)* **2016**, *31* (1), 7−16.

(14) van Veldhoven, K.; Keski-Rahkonen, P.; Barupal, D. K.; Villanueva, C. M.; Font-Ribera, L.; Scalbert, A.; Bodinier, B.; Grimalt, J. O.; Zwiener, C.; Vlaanderen, J.; Portengen, L.; Vermeulen, R.; Vineis, P.; Chadeau-Hyam, M.; Kogevinas, M. Effects of exposure to water disinfection by-products in a swimming pool: A metabolome-wide association study. *Environ. Int.* **2018**, *111*, 60−70.

(15) Castano-Vinyals, G.; Aragones, N.; Perez-Gomez, B.; Martin, V.; Llorca, J.; Moreno, V.; Altzibar, J. M.; Ardanaz, E.; de Sanjose, S.; Jimenez-Moleon, J. J.; Tardón, A.; Alguacil, J.; Peiró, R.; Marcos-

https://doi.org/10.1021/acs.est.3c05814
Environ. Sci. Technol. 2023, 57, 19316−19329

Gragera, R.; Navarro, C.; Pollán, M.; Kogevinas, M. Population-based multicase-control study in common tumors in Spain (MCC-Spain): rationale and study design. *Gac Sanit* **2015**, *29* (4), 308−315.
(16) Espejo-Herrera, N.; Kogevinas, M.; Castano-Vinyals, G.; Aragones, N.; Boldo, E.; Ardanaz, E.; Azpiroz, L.; Ulibarrena, E.; Tardon, A.; Molina, A. J.; López-Rojo, C.; Jiménez-Moleón, J. J.; Capelo, R.; Gómez-Acebo, I.; Ripoll, M.; Villanueva, C. M. Nitrate and trace elements in municipal and bottled water in Spain. *Gac Sanit* **2013**, *27* (2), 156−160.
(17) Lazar, C. imputeLCMD: a collection of methods for left-censored missing data imputation. *R package*; 2015.
(18) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.-M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; et al. Proposed minimum reporting standards for chemical analysis: chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics* **2007**, *3*, 211−221.
(19) Li, S.; Park, Y.; Duraisingham, S.; Strobel, F. H.; Khan, N.; Soltow, Q. A.; Jones, D. P.; Pulendran, B. Predicting network activity from high throughput metabolomics. *PLoS Comput. Biol.* **2013**, *9* (7), No. e1003123.
(20) Villanueva, C. M.; Espinosa, A.; Gracia-Lavedan, E.; Vlaanderen, J.; Vermeulen, R.; Molina, A. J.; Amiano, P.; Gomez-Acebo, I.; Castano-Vinyals, G.; Vineis, P.; Kogevinas, M. Exposure to widespread drinking water chemicals, blood inflammation markers, and colorectal cancer. *Environ. Int.* **2021**, *157*, No. 106873.
(21) Conway, J. R.; Lex, A.; Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **2017**, *33* (18), 2938−2940.
(22) Saccenti, E.; Hoefsloot, H. C.; Smilde, A. K.; Westerhuis, J. A.; Hendriks, M. M. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* **2014**, *10*, 361−374.
(23) Gumpenberger, T.; Brezina, S.; Keski-Rahkonen, P.; Baierl, A.; Robinot, N.; Leeb, G.; Habermann, N.; Kok, D. E. G.; Scalbert, A.; Ueland, P. M.; Ulrich, C. M.; Gsur, A. Untargeted Metabolomics Reveals Major Differences in the Plasma Metabolome between Colorectal Cancer and Colorectal Adenomas. *Metabolites* **2021**, *11* (2), 119.
(24) Geijsen, A.; Brezina, S.; Keski-Rahkonen, P.; Baierl, A.; Bachleitner-Hofmann, T.; Bergmann, M. M.; Boehm, J.; Brenner, H.; Chang-Claude, J.; van Duijnhoven, F. J. B.; Gigic, B.; Gumpenberger, T.; Hofer, P.; Hoffmeister, M.; Holowatyj, A. N.; Karner-Hanusch, J.; Kok, D. E.; Leeb, G.; Ulvik, A.; Robinot, N.; Ose, J.; Stift, A.; Schrotz-King, P.; Ulrich, A. B.; Ueland, P. M.; Kampman, E.; Scalbert, A.; Habermann, N.; Gsur, A.; Ulrich, C. M. Plasma metabolites associated with colorectal cancer: A discovery-replication strategy. *Int. J. Cancer* **2019**, *145* (5), 1221−1231.
(25) Zhang, J. Are poly(ADP-ribosyl)ation by PARP-1 and deacetylation by Sir2 linked? *Bioessays* **2003**, *25* (8), 808−814.
(26) Brachs, S.; Polack, J.; Brachs, M.; Jahn-Hofmann, K.; Elvert, R.; Pfenninger, A.; Barenz, F.; Margerie, D.; Mai, K.; Spranger, J.; Kannt, A. Genetic Nicotinamide N-Methyltransferase (Nnmt) Deficiency in Male Mice Improves Insulin Sensitivity in Diet-Induced Obesity but Does Not Affect Glucose Tolerance. *Diabetes* **2019**, *68* (3), 527−542.
(27) Rini, J.; Szumlanski, C.; Guerciolini, R.; Weinshilboum, R. M. Human liver nicotinamide N-methyltransferase: ion-pairing radiochemical assay, biochemical properties and individual variation. *Clin. Chim. Acta* **1990**, *186* (3), 359−374.
(28) Song, M.; Li, Y.; Miao, M.; Zhang, F.; Yuan, H.; Cao, F.; Chang, W.; Shi, H.; Song, C. High stromal nicotinamide N-methyltransferase (NNMT) indicates poor prognosis in colorectal cancer. *Cancer Med.* **2020**, *9* (6), 2030−2038.
(29) Xie, X.; Liu, H.; Wang, Y.; Zhou, Y.; Yu, H.; Li, G.; Ruan, Z.; Li, F.; Wang, X.; Zhang, J. Nicotinamide N-methyltransferase enhances resistance to 5-fluorouracil in colorectal cancer cells through inhibition of the ASK1-p38 MAPK pathway. *Oncotarget* **2016**, *7* (29), 45837−45848.
(30) Xie, X.; Yu, H.; Wang, Y.; Zhou, Y.; Li, G.; Ruan, Z.; Li, F.; Wang, X.; Liu, H.; Zhang, J. Nicotinamide N-methyltransferase enhances the capacity of tumorigenesis associated with the promotion of cell cycle progression in human colorectal cancer cells. *Arch. Biochem. Biophys.* **2014**, *564*, 52−66.
(31) Allis, J. W.; Brown, B. L.; Zhao, G.; Pegram, R. A. The effects of inhalation exposure to bromo-dichloromethane on specific rat CYP isoenzymes. *Toxicology* **2001**, *161* (1−2), 67−77.
(32) Zhao, G.; Allis, J. W. Kinetics of bromodichloromethane metabolism by cytochrome P450 isoenzymes in human liver microsomes. *Chem. Biol. Interact* **2002**, *140* (2), 155−168.
(33) Constan, A. A.; Sprankle, C. S.; Peters, J. M.; Kedderis, G. L.; Everitt, J. I.; Wong, B. A.; Gonzalez, F. L.; Butterworth, B. E. Metabolism of chloroform by cytochrome P450 2E1 is required for induction of toxicity in the liver, kidney, and nose of male mice. *Toxicol. Appl. Pharmacol.* **1999**, *160* (2), 120−126.
(34) Sapone, A.; Gustavino, B.; Monfrinotti, M.; Canistro, D.; Broccoli, M.; Pozzetti, L.; Affatato, A.; Valgimigli, L.; Forti, G. C.; Pedulli, G. F.; Biagi, G. L.; Abdel-Rahman, S. Z.; Paolini, M. Perturbation of cytochrome P450, generation of oxidative stress and induction of DNA damage in Cyprinus carpio exposed in situ to potable surface water. *Mutat. Res.* **2007**, *626* (1−2), 143−154.
(35) Minamiyama, Y.; Takemura, S.; Akiyama, T.; Imaoka, S.; Inoue, M.; Funae, Y.; Okada, S. Isoforms of cytochrome P450 on organic nitrate-derived nitric oxide release in human heart vessels. *FEBS Lett.* **1999**, *452* (3), 165−169.
(36) Minamiyama, Y.; Takemura, S.; Yamasaki, K.; Hai, S.; Hirohashi, K.; Funae, Y.; Okada, S. Continuous administration of organic nitrate decreases hepatic cytochrome P450. *J. Pharmacol Exp Ther* **2004**, *308* (2), 729−735.
(37) Guengerich, F. P.; Shimada, T. Activation of procarcinogens by human cytochrome P450 enzymes. *Mutat. Res.* **1998**, *400* (1−2), 201−213.
(38) Martinez, C.; Garcia-Martin, E.; Pizarro, R. M.; Garcia-Gamito, F. J.; Agundez, J. A. Expression of paclitaxel-inactivating CYP3A activity in human colorectal cancer: implications for drug therapy. *Br. J. Cancer* **2002**, *87* (6), 681−686.
(39) Gibson, P.; Gill, J. H.; Khan, P. A.; Seargent, J. M.; Martin, S. W.; Batman, P. A.; Griffith, J.; Bradley, C.; Double, J. A.; Bibby, M. C.; Loadman, P. M. Cytochrome P450 1B1 (CYP1B1) is overexpressed in human colon adenocarcinomas relative to normal colon: implications for drug development. *Mol. Cancer Ther.* **2003**, *2* (6), 527−534.
(40) Kumarakulasingham, M.; Rooney, P. H.; Dundas, S. R.; Telfer, C.; Melvin, W. T.; Curran, S.; Murray, G. I. Cytochrome p450 profile of colorectal cancer: identification of markers of prognosis. *Clin. Cancer Res.* **2005**, *11* (10), 3758−3765.
(41) Paul, M. K.; Mukhopadhyay, A. K. Tyrosine kinase - Role and significance in Cancer. *Int. J. Med. Sci.* **2004**, *1* (2), 101−115.
(42) Zhan, X.; Huang, Y.; Qian, S. Protein Tyrosine Nitration in Lung Cancer: Current Research Status and Future Perspectives. *Curr. Med. Chem.* **2018**, *25* (29), 3435−3454.
(43) Williams, J. L.; Ji, P.; Ouyang, N.; Kopelovich, L.; Rigas, B. Protein nitration and nitrosylation by NO-donating aspirin in colon cancer cells: Relevance to its mechanism of action. *Exp. Cell Res.* **2011**, *317* (10), 1359−1367.
(44) Gao, P.; Zhou, C.; Zhao, L.; Zhang, G.; Zhang, Y. Tissue amino acid profile could be used to differentiate advanced adenoma from colorectal cancer. *J. Pharm. Biomed Anal* **2016**, *118*, 349−355.
(45) Gu, J.; Xiao, Y.; Shu, D.; Liang, X.; Hu, X.; Xie, Y.; Lin, D.; Li, H. Metabolomics Analysis in Serum from Patients with Colorectal Polyp and Colorectal Cancer by (1)H-NMR Spectrometry. *Dis Markers* **2019**, *2019*, No. 3491852.
(46) Scharer, C. D.; McCabe, C. D.; Ali-Seyed, M.; Berger, M. F.; Bulyk, M. L.; Moreno, C. S. Genome-wide promoter analysis of the SOX4 transcriptional network in prostate cancer cells. *Cancer Res.* **2009**, *69* (2), 709−717.
(47) Du, F.; Chen, J.; Liu, H.; Cai, Y.; Cao, T.; Han, W.; Yi, X.; Qian, M.; Tian, D.; Nie, Y.; Wu, K.; Fan, D.; Xia, L. SOX12 promotes colorectal cancer cell proliferation and metastasis by regulating asparagine synthesis. *Cell Death Dis* **2019**, *10* (3), 239.

https://doi.org/10.1021/acs.est.3c05814
Environ. Sci. Technol. 2023, 57, 19316−19329

(48) Michalopoulou, E.; Bulusu, V.; Kamphorst, J. J. Metabolic scavenging by cancer cells: when the going gets tough, the tough keep eating. *Br. J. Cancer* **2016**, *115* (6), 635−640.
(49) Tan, B.; Qiu, Y.; Zou, X.; Chen, T.; Xie, G.; Cheng, Y.; Dong, T.; Zhao, L.; Feng, B.; Hu, X.; Xu, L. X.; Zhao, A.; Zhang, M.; Cai, G.; Cai, S.; Zhou, Z.; Zheng, M.; Zhang, Y.; Jia, W. Metabonomics identifies serum metabolite markers of colorectal cancer. *J. Proteome Res.* **2013**, *12* (6), 3000−3009.
(50) Zhao, Z.; Xiao, Y.; Elson, P.; Tan, H.; Plummer, S. J.; Berk, M.; Aung, P. P.; Lavery, I. C.; Achkar, J. P.; Li, L.; Casey, G.; Xu, Y. Plasma lysophosphatidylcholine levels: potential biomarkers for colorectal cancer. *J. Clin Oncol* **2007**, *25* (19), 2696−2701.
(51) Gold, A.; Choueiry, F.; Jin, N.; Mo, X.; Zhu, J. The Application of Metabolomics in Recent Colorectal Cancer Studies: A State-of-the-Art Review. *Cancers (Basel)* **2022**, *14* (3), 725.
(52) Ghonimy, A.; Zhang, D. M.; Farouk, M. H.; Wang, Q. The Impact of Carnitine on Dietary Fiber and Gut Bacteria Metabolism and Their Mutual Interaction in Monogastrics. *Int. J. Mol. Sci.* **2018**, *19* (4), 1008.
(53) Yang, L.; Ge, L. Y.; Yu, T.; Liang, Y.; Yin, Y.; Chen, H. The prognostic impact of serum bilirubin in stage IV colorectal cancer patients. *J. Clin Lab Anal* **2018**, *32* (2), e22272.

https://doi.org/10.1021/acs.est.3c05814
Environ. Sci. Technol. 2023, 57, 19316−19329