Plasma metabolomics profiles and breast cancer risk

Hui-Chen Wu[1,2*], Yunjia Lai[1], Yuyan Liao[3], Maya Deyssenroth[1], Gary W. Miller[1], Regina M. Santella[1,2] and Mary Beth Terry[2,3]

Abstract

Background Breast cancer (BC) is the most common cancer in women and incidence rates are increasing; metabolomics may be a promising approach for identifying the drivers of the increasing trends that cannot be explained by changes in known BC risk factors.

Methods We conducted a nested case–control study (median followup 6.3 years) within the New York site of the Breast Cancer Family Registry (BCFR) (n = 40 cases and 70 age-matched controls). We conducted a metabolome-wide association study using untargeted metabolomics coupling hydrophilic interaction liquid chromatography (HILIC) and C_{18} chromatography with high-resolution mass spectrometry (LC-HRMS) to identify BC-related metabolic features.

Results We found eight metabolic features associated with BC risk. For the four metabolites negatively associated with risk, the adjusted odds ratios (ORs) ranged from 0.31 (95% confidence interval (CI): 0.14, 0.66) (L-Histidine) to 0.65 (95% CI: 0.43, 0.98) (N-Acetylgalactosamine), and for the four metabolites positively associated with risk, ORs ranged from 1.61 (95% CI: 1.04, 2.51, (m/z: 101.5813, RT: 90.4, 1,3-dibutyl-1-nitrosourea, a potential carcinogen)) to 2.20 (95% CI: 1.15, 4.23) (11-cis-Eicosenic acid). These results were no longer statistically significant after adjusting for multiple comparisons. Adding the BC-related metabolic features to a model, including age, the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) risk score improved the accuracy of BC prediction from an area under the curve (AUC) of 66% to 83%.

Conclusions If replicated in larger prospective cohorts, these findings offer promising new ways to identify exposures related to BC and improve BC risk prediction.

Keywords Breast cancer epidemiology, BOADICEA, Metabolome, Metabolomics, Nested case–control study, Partial least-squares discriminant analysis (PLS-DA)

*Correspondence:
Hui-Chen Wu
hw2057@cumc.columbia.edu
[1] Department of Environmental Health Sciences, Mailman School of Public Health of Columbia University, New York, NY 10032, USA
[2] Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY, USA
[3] Department of Epidemiology, Mailman School of Public Health of Columbia University, New York, NY, USA

Introduction

Breast cancer (BC), the most common cancer and the leading cause of cancer death in women worldwide [1], is increasing over time, and established risk factors cannot account for this increase [2]. Metabolic phenotype represents the metabolite profile, influenced by genetic and environmental factors [3]. Characterization of metabolic processes may provide new insights into risk factors for breast carcinogenesis [4]. Thus, a comprehensive readout of the chemical body burden and the resulting endogenous response with the fast-evolving technologies of

Wu *et al. Breast Cancer Research*   (2024) 26:141

Page 2 of 14

high-resolution mass spectrometry (HRMS) in the recent decade, metabolomics is one promising approach to gaining comprehensive insight into the etiological pathways leading to BC. [4–6] The small molecule profile of blood untargeted metabolomics provides an integrated readout of the body's chemical burden and its endogenous metabolic response.

At least 10 prospective metabolomic studies of BC risk using pre-diagnostic plasma (n = 88–1691 cases) with the mean follow-up ranging from 4 to 21 years have been carried out [7–16]. Most prior studies [7, 8, 11, 12, 15, 16], however, focused on postmenopausal women and none of these studies focused on women at high risk due to their family history. These studies reported that metabolites such as sex steroid-related metabolites, glycerolipids, and cholesteryl esters were altered several years prior to BC diagnosis; suggested that metabolomics is potentially a powerful approach to identify metabolomic biomarkers that are altered during BC development and before clinical symptoms. In addition, studies also found several BC-associated metabolic features were correlated with diet [15] or body mass index (BMI) [16]. For example, a nested case–control study identified 113 nutritional metabolites and found 3 metabolic features, including saturated fatty acids (from fats/oils), vitamin E derivatives (from desserts or vitamin supplements), and androgens (from alcohol), were associated with BC, with odds ratios (ORs) ranging from 0.6 to 2.2 [15]. These studies highlighted the associations between baseline plasma metabolomic signatures and BC risk and suggested potential metabolic pathways as a promising avenue for discovering therapeutic targets for prevention.

Women with a family history of BC are two to four times more likely to develop the disease compared to women with no family history [17]. BC risk associated with family history varies with the age of the individual, number of affected relatives and age at which the relatives were diagnosed with BC [18, 19]. Our prior study estimated lifetime risk based on the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) for women enrolled in the Breast Cancer Family Registry (BCFR) and found there was substantial variation in absolute risk among participants [17]. Therefore, the BCFR is a unique cohort to identify biomarkers for women across the risk continuum. The goal of this pilot study, which employed a prospective nested case–control study design, was to interrogate the relationship between metabolomic features with breast cancer risk in pre-diagnostic plasma of women enrolled in the New York site of the BCFR, a registry of individuals within families with breast and/or ovarian cancer [17, 20].

Materials and methods
Study design
We conducted a prospective study among the women unaffected with BC at enrollment within the New York site of the BCFR (for details see [21]). At recruitment, eligible participants completed a questionnaire that included information on demographics, lifestyles, environmental factors, and family history of cancer [20]. All BCFR participants were requested to provide a 30 ml blood sample at the time of the baseline recruitment. Biospecimens were processed according to a common standardized protocol and stored at −80 °C till metabolomic analysis.

We actively follow participants for subsequent information on cancer incidence and vital status and attempt to verify cancers through pathology reviews and medical records. For the present nested case–control study, we analyzed data for 40 prospectively ascertained BC cases and 70 age- (± 5yrs) matched controls. Of these 40 cases, 17 were diagnosed with BC within 5 years, 17 cases were diagnosed with BC between 5 and 10 years and six cases were diagnosed with BC more than 10 years after blood draw. This study was approved by Columbia University's Institutional Review Board. All methods were performed in accordance with the relevant guidelines and regulations.

Liquid chromatography-high resolution mass spectrometry (LC-HRMS) analysis
To interrogate circulating metabolic differences, we conducted global metabolomics of blood plasma samples using a liquid chromatography-high resolution mass spectrometry (LC-HRMS)-based metabolomics workflow [22]. For sample pretreatment, blood plasma samples were thawed on ice; 50 µL was aliquoted and extracted with 100 µL ice-cold acetonitrile (ACN) pre-spiked with the internal standard mix (final ACN: sample, 2:1, *v/v*). After centrifugation, supernatants were collected, of which 10 µL was injected for LC-MS analysis. The analytes were chromatographically separated, ionized and analyzed on a Thermo Fisher Scientific Vanquish dual chromatograph coupled to a high-resolution accurate-mass (HRAM) quadrupole-Orbitrap Q-Exactive HF-X mass spectrometer (Waltham, MA, USA) under two complementary modes: hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry in positive ion mode (HILIC) and C_{18}-electrospray ionization mass spectrometry in negative ion mode (C_{18}). The HILIC column uses a polar stationary phase that retains well polar species (e.g., primary metabolites including many organic acids and amino acids), while the C_{18} column uses a more nonpolar stationary phase that separates more nonpolar species well; using both

Wu *et al. Breast Cancer Research* (2024) 26:141

Page 3 of 14

columns allows us to cover a broad range of metabolites. For both modes, we operated the instrument in full scan mode at 120,000 mass resolution (full width at half maximum, fwhm) scanning a mass-to-charge (*m/z*) range of 85–1,275. For quality assurance and quality control (*QA/QC*), extracts of NIST1953 plasma (Gaithersburg, MD, USA) and BioIVT plasma (New Cassel, NY, USA) were injected intermittently with sample extracts. Other *QA/QC* procedures were implemented, spanning timely mass calibration, sample randomization, blinding technicians from case–control status of samples, method blanks, and triplicate sample injection. We further performed stringent post-data procedures, including triplicate sample filtering (keeping only features with $\geq 2/3$ occurrences), replicate median summarization, and combat correction (accounting for batch effects). The resultant dataset after *QA/QC* reached a CV of 7.92% of total ion chromatogram (TIC) intensity using all features for QC samples, and the pairwise Pearson Correlation within QC samples (averaged) has a mean of 0.95 and %CV range of 0.78–1.

Data processing and analysis

We converted the acquired RAW format data to mzXML format in ProteoWizard msConvert, and extracted mass spectral features and aligned separately for each mode using apLCMS [23] with modifications by xMSanalyzer [24]. We used ComBat [25] for batch correction. The resultants feature table consist of 5,992 HILIC and 5,780 C_{18} features, respectively, containing accurate *m/z,* retention time (RT), and peak intensity (i.e., peak area, as a semi-quantitative measure for statistics) for individual ion features in each sample, which are referred to as *m/z* features hereafter. For QC purpose, we filtered the feature tables to remove peaks that were detected in fewer than 20% of study samples (i.e., consistently detected in analytical replicates of at least one participant's sample). We did not observe statistically significant differences in the number of missing features between cases and controls. We retained a total of 2,264 metabolic features for HILIC and 2,988 metabolic features for C_{18} for data analysis; the remaining metabolic features with values below the detection limit were imputed with half the minimum of the non-missing values. Prior to statistical analysis, we $\log_{10}$ transformed and Pareto-scaled peak intensities [26]. To annotate compound structures for these detectable metabolic features, we used a multi-layered approach, and assigned confidence of annotation according to the Schymanski Scale [27] by the guidelines of the Metabolomics Standard Initiatives (MSI) [28]. Briefly, we referenced an in-house m/z-RT library that was established from over 900 authentic chemical standards (level 1) and applied de novo annotation (level 4) through matching accurate *m/z* against annotations from Mummichog

pathway analysis (10 ppm) and filtered out unlikely annotations by (1) focusing exclusively on ESI adduct species $[M+H]^+$, $[M+H-H_2O]^+$, $[M]^+$, $[M-H]^-$, and $[M-H-H_2 0]^-$, and (2) filtering based on machine learning predicted RT, using bidirectional recurrent neural network (BRNN) for HILIC RT and random forest for C_{18} RT, respectively.

Absolute risk of BC

We assessed the 1-year risk of breast cancer by leveraging familial pedigree and vital status data, encompassing cancer diagnoses, age at diagnoses, and information on BRCA1 and BRCA2 mutations. Our analysis employed the BOADICEA model [29], utilizing the obtained probability as a continuous risk score in subsequent regression analyses. Variables included in the BOADICEA algorithm include age at baseline, year of birth, first-, second, and third-degree relatives with BC, identical twin with BC, age at cancer diagnosis, bilateral BC, ovarian cancer, pancreatic cancer, prostate cancer, molecular subtype of breast tumors, vital status of family members, *BRCA1* and *BRCA2* mutation status and Ashkenazi Jewish heritage [30].

Statistical methods

We used the Wilcoxon rank test to compare the metabolic feature levels between cases and controls. We used the original $p < 0.05$ to select the candidate metabolite features for further multivariate logistic regression analysis. We also conducted partial least squares-discriminant analysis (PLS-DA) to examine the metabolic features by case and control groups while adjusting for confounding factors. Specifically, for data pretreatment, we removed potential batch effects by *combat* [31] normalization (using *xMSanalyzer*) [24], imputed zero values with half of the minimum within-sample peak intensity, and conducted log-transformation and Pareto scaling of the alignment datasets (HIL and C_{18} separately). We then performed linear regression adjusting for potential confounding variables including age (continuous years), BMI (continuous kg/m^2), smoking, alcohol drinking, and menopausal status; the fit of the model was checked, and the resultant residuals were retrieved for PLS-DA using mixOmics. The variable importance in the projection (VIP) in the PLS-DA was retrieved in R using PLSDA.VIP() function of the mixOmics [32], and the VIP scores were plotted to assist in the sorting of the top candidate metabolic ion features contributing the most to the PLS-DA classification. We performed pathway enrichment analysis in MetaboAnalyst 5.0 using the mixed mode (combining data of HILIC and C_{18}) as input, and applied the Mummichog [33] algorithm (p-value cutoff 0.05) to identify the most enriched metabolic pathways referencing

Wu *et al. Breast Cancer Research*   (2024) 26:141

Page 4 of 14

against MFN (Homo sapiens), a human genome-scale metabolic model from the original mummichog package that has been manually curated from various sources including KEGG, BiGG and Edinburgh model.

We used logistic regression adjusting to calculate odds ratios (OR) and 95% confidence intervals (CI) for individual metabolic features with BC diagnosis. Model 1 adjusted for age at blood (continuous). Model 2 adjusted for age and BOADICEA breast cancer 1 year of risk score (continuous). Model 3 included variables in Model 2, BMI (continuous), race and ethnicity, alcohol, and smoking status (never, former and current), menopausal status (Pre and post-menopausal status). For ROC analysis, we conducted three logistic regression modes: Model 1, including age (continuous years); Model 2, including age and BOADICEA 1-year risk score; and Model 3, including age, BOADICEA risk score, and six metabolic features. All the variables were modeled as continuous rather than categorical, which makes it less likely that the model was over-fitted given the small sample size. We also conducted a sensitivity analysis by excluding the 4 cases diagnosed with breast cancer within 1 year after blood collection. Analyses were done in SAS (v. 9.4).

Results

Table 1 presents the baseline characteristics for cases and controls. The mean ages were 45.2 ± 11.4 years for cases and 46.4 ± 13.4 years for controls. The average age at BC diagnosis was 51.6 ± 12.5 years. Twenty-four cases (60%) and 49 controls (71.0%) were pre-menopausal at baseline.

We detected and aligned 11,772 m/z features (5,992 HILIC and 5,780 C_{18}) for the untargeted plasma metabolomic profiling. Of these, 5,252 (2,264 HILIC and 2,988 C_{18}) were detected in at least 80% of samples. A non-parametric test found 289 metabolic features (135 HILIC and 154 C_{18}) (Fig. 1A and B) were statistically significantly different between cases and controls if the original *p-value* was less than 0.05. Thirty-two metabolic features (17 HILIC and 15 C_{18}) had fold changes (FCs) above 1.5 (case > control) or below 0.667 (cases < control) (Fig. 2A and B). Table 2 presents the changes in annotated metabolic features with significant fold changes (original $p < 0.05$) between cases and controls [34]. In this study, among HILIC features, we observed 4 positively associated and 13 negatively associated features in cases compared to controls. Among C_{18} features, we observed 12 positively associated and 3 negatively associated features in cases compared to controls.

Table 1 Characteristics of study subjects, New York site of the BCFR

	Total (N = 110) N (%) or Mean ± SD	Prospective Cases (N = 40) N (%) or Mean ± SD	Unaffected (N = 70) N (%) or Mean ± SD
Age at blood draw	46.0 ± 12.7	45.2 ± 11.4	46.4 ± 13.4
Body mass index (BMI)	24.3 ± 4.9	24.6 ± 4.9	24.1 ± 5.0
< 25	66 (60.0)	23 (57.5)	43 (61.4)
≥ 25	44 (40.0)	17 (42.5)	27 (38.6)
Menopausal status			
Pre-menopausal	73 (66.4)	24 (60.0)	49 (70.0)
Post-menopausal	36 (32.7)	16 (40.0)	20 (28.6)
Unknown	1 (0.9)		1 (1.4)
Alcohol drinking			
Never	67 (60.9)	22 (55.0)	45 (64.3)
Former	15 (13.6)	7 (17.5)	8 (11.4)
Current	27 (24.5)	11 (27.5)	16 (22.9)
Unknown	1 (0.9)		1 (1.4)
Smoking			
Never	72 (65.5)	26 (65.0)	46 (65.7)
Former	23 (20.9)	8 (20.0)	15 (21.4)
Current	13 (11.8)	5 (12.5)	8 (11.4)
Unknown	2 (1.8)	1 (2.5)	1 (1.4)
Race/Ethnicity			
Non-hispanic white	81 (73.6)	32 (80.0)	49 (70.0)
Hispanics	27 (24.5)	8 (20.0)	19 (27.1)
Non-Hispanic Black/Asian	2 (1.9)		2 (2.9)

A. Manhattan plot of metabolic features detected by HILIC

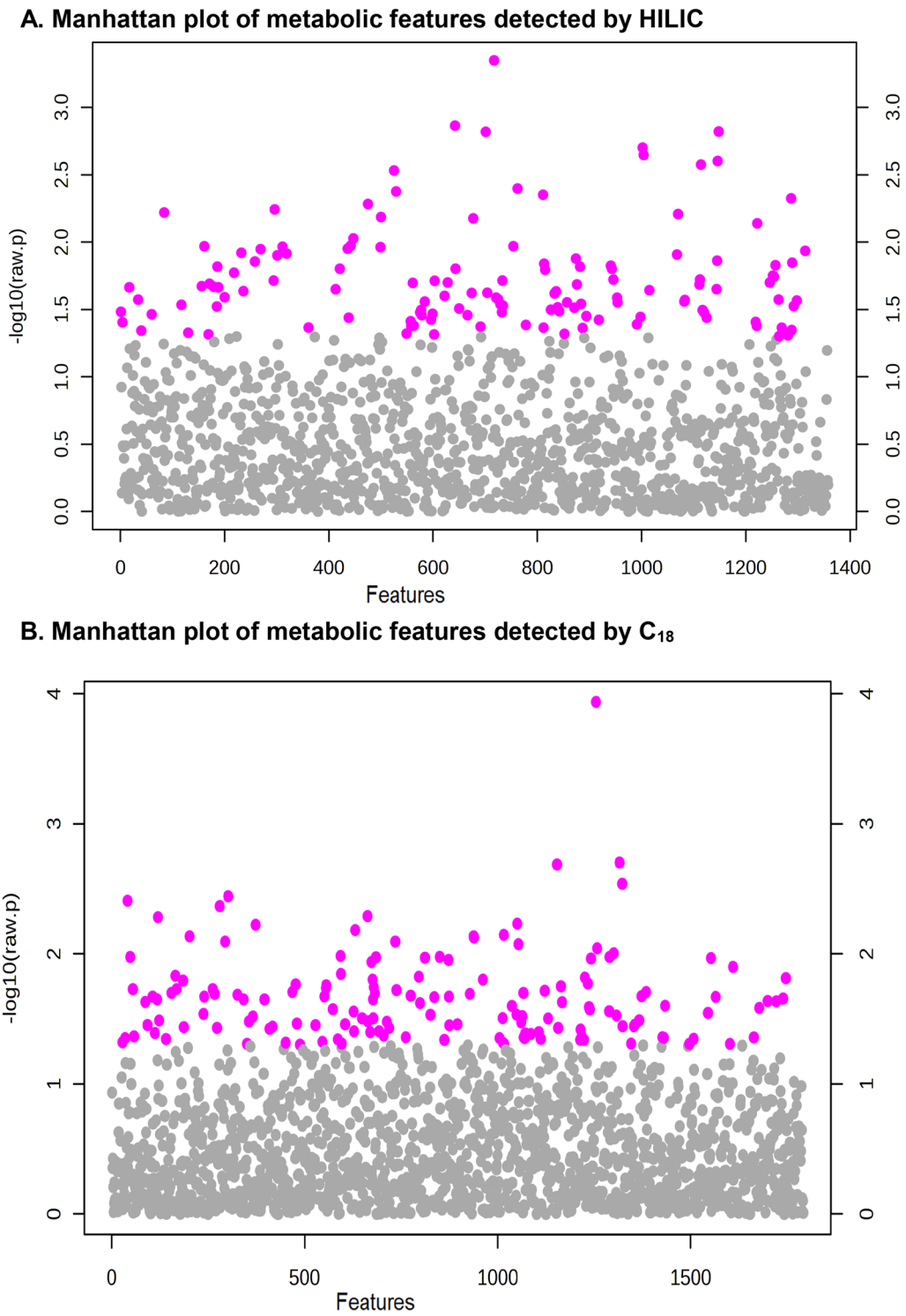## B. Manhattan plot of metabolic features detected by C$_{18}$

Fig. 1 Manhattan plots of metabolome-wide association study. Features heighted in purple indicate original *p* < 0.05 in Wilcoxon Rank Test.

A. Volcano plot of metabolites/features detected by HILIC

B. Volcano plot of metabolites/features detected by C₁₈

Fig. 2 Volcano plot of metabolites/features.

Table 3 presents the odds ratio values of BC risk for the annotated metabolic features. The ORs of log-metabolic features range from 0.31 to 2.20 in Model 3. For the metabolites negatively associated with risk, the ORs range from 0.31 (95% CI: 0.14, 0.66) for HILIC feature (m/z: 138.066, RT: 25.4 s, L-Histidine) to 0.65 (95% CI: 0.43, 0.98) for HILIC feature (m/z: 222.0984, RT: 27.5 s,

N-Acetylgalactosamine). For the metabolites positively associated with risk, ORs ranged from 1.61 (95% CI: 1.04, 2.51) for HILIC feature (m/z:101.58, RT:90.4 s, 1,3-Dobutyl-1-nitrosourea) to 2.20 (95% CI: 1.15, 4.23) for C₁₈ feature (m/z:346.246, RT:126 s, 11-cis-Eicosenoic acid). These results were no longer statistically significant after adjusting for multiple comparisons.

Table 2 Fold change of the significant metabolic features

m/z	RT (s)	Metabolites	MSI level	FC	log2(FC)	raw.pval	FDR	BC/CTRL
HILIC								
222.0984	27.5	N-Acetylgalactosamine	1	0.55	-0.86	0.001	0.38	Down
690.9543	53.7			1.57	0.65	0.002	0.38	Up
181.0721	33.0	Glucose	1	0.6	-0.74	0.003	0.43	Down
287.98	198.8			0.63	-0.66	0.004	0.46	Down
182.0754	34.7			0.6	-0.73	0.004	0.46	Down
101.5813	90.4	1,3-Dibutyl-1-nitrosourea	4	1.81	0.85	0.006	0.46	Up
236.9971	62.8			0.66	-0.6	0.007	0.46	Down
162.9844	179.9			0.38	-1.38	0.009	0.46	Down
161.1283	89.1	N(6)-Methyllysine	1	1.64	0.71	0.011	0.46	Up
113.0597	26.0	Sorbate	1	0.66	-0.59	0.011	0.46	Down
124.0507	43.2			0.62	-0.69	0.012	0.46	Down
138.0662	25.4	L-Histidine	4	0.67	-0.59	0.012	0.46	Down
129.0182	34.0	2-Oxoglutarate	4	0.46	-1.11	0.014	0.46	Down
195.0878	31.6	Caffeine	1	0.66	-0.6	0.033	0.46	Down
196.0911	30.4			0.66	-0.59	0.035	0.46	Down
190.9794	158.2			0.49	-1.02	0.042	0.48	Down
367.991	210.4			1.61	0.69	0.048	0.51	Up
C₁₈								
359.2203	128.3			1.7	0.77	0.0001	0.34	Up
389.2902	164.6			1.52	0.6	0.0029	0.63	Up
147.0491	25.6			1.65	0.73	0.0036	0.63	Up
301.2173	145.9	Eicosapentaenoic acid	1	1.5	0.59	0.0059	0.63	Up
295.1372	147.7			0.58	-0.79	0.0071	0.63	Down
302.2195	144.0			1.5	0.59	0.0084	0.63	Up
378.864	81.9			1.53	0.62	0.0099	0.63	Up
552.31	133.8	LysoPE(0:0/24:6(6Z,9Z, 12Z,15Z, 18Z, 21Z)	4	1.59	0.66	0.0108	0.63	Up
199.1634	132.1			1.96	0.97	0.0143	0.63	Up
346.2462	126.0	11-cis-Eicosenoic acid	4	1.54	0.62	0.0152	0.63	Up
555.9508	103.5			0.64	-0.65	0.0214	0.63	Down
456.4133	150.3			1.68	0.75	0.0438	0.68	Up
610.4909	156.0			1.71	0.78	0.0439	0.68	Up
95.0062	29.8			0.61	-0.72	0.0444	0.68	Down
200.173	132.9	Dodecanoic acid	4	1.62	0.7	0.0489	0.68	Up

MSI metabolomics standard initiative, *FC* Fold change: ratio of metabolic features level between cases and controls, *m/z* mass-to-charge, *RT* retention time, *FDR* false discovery rate

We calculated the Area under the Receiver Operating Curve (AUC) to evaluate the performance of our classifier. The AUC of the model that included age, and BOADICEA 1-year risk score) improved from 0.66 to 0.83 once our six candidate metabolites were incorporated into the model (Fig. 3). We did not include two metabolic features, glucose (m/z:181.0721, RT 33 s) and caffeine (m/z:195.0878, RT 31.6 s), because both metabolite features were highly correlated with L-Histidine (m/z:138.0662, RT 25.4 s), with correlation coefficients

of 0.83 ($p < 0.0001$) and 0.96 ($p < 0.0001$), respectively. We also conducted a sensitivity analysis by excluding four cases diagnosed with breast cancer within 1 year after blood collection. The results were similar (data not shown).

In addition to the metabolome-wide association analysis, we also conducted a supervised classification approach PLS-DA to differentiate cases and controls based on the metabolic profiles. Figure 4 presents the PLS-DA score plots showing the separation of the two groups and shows both HILIC and C18 features with two

Wu *et al. Breast Cancer Research*   (2024) 26:141

Page 8 of 14

Table 3 Breast cancer risk for 12 metabolic features, nested case–control study within the New York Site of the BCFR

	m/z	RT (s)	Model 1		Model 2		Model 3	
			OR (95% CI)	*p*-value	OR (95% CI)	*p*-value	OR (95% CI)	*p*-value
HILIC								
1,3-Dibutyl-1-nitrosourea	101.5813	90.4	**1.62 (1.13–2.32)**	0.01	**1.66 (1.14–2.42)**	0.01	**1.61 (1.04–2.51)**	0.03
Sorbate	113.0597	26	0.71 (0.40–1.28)	0.26	0.72 (0.40–1.30)	0.27	0.70 (0.36–1.39)	0.31
2-Oxoglutarate	129.0182	34	0.70 (0.45–1.10)	0.13	0.70 (0.44–1.11)	0.13	0.74 (0.44–1.25)	0.26
L-Histidine	138.0662	25.4	**0.45 (0.24–0.85)**	0.01	**0.45 (0.24–0.87)**	0.02	**0.31 (0.14–0.66)**	0.003
N(6)-Methyllysine	161.1283	89.1	**1.98 (1.13–3.46)**	0.02	**2.20 (1.21–4.00)**	0.01	**2.00 (1.01–3.98)**	0.047
Glucose	181.0721	33	**0.60 (0.37–0.95)**	0.03	**0.58 (0.36–0.92)**	0.02	**0.54 (0.32–0.90)**	0.02
Caffeine	195.0878	31.6	**0.61 (0.38–1.00)**	0.049	**0.61 (0.37–0.99)**	0.047	**0.49 (0.28–0.87)**	0.01
N-Acetylgalactosamine	222.0984	27.5	**0.66 (0.45–0.96)**	0.03	**0.65 (0.44–0.95)**	0.03	**0.65 (0.43–0.98)**	0.04
C18								
Dodecanoic acid	200.173	132.9	1.57 (0.90–2.75)	0.11	1.48 (0.85–2.57)	0.17	1.59 (0.87–2.91)	0.14
Eicosapentaenoic acid	301.2173	145.9	1.81 (0.89–3.69)	0.10	1.78 (0.84–3.75)	0.13	1.86 (0.87–3.96)	0.11
11-cis-Eicosenoic acid	346.2462	126	**1.99 (1.10–3.60)**	0.02	**2.07 (1.12–3.81)**	0.02	**2.20 (1.15–4.23)**	0.02
LysoPE(0:0/24:6(6Z,9Z, 12Z,15Z, 18Z, 21Z)	552.31	133.8	**1.90 (1.14–3.15)**	0.01	**1.88 (1.11–3.16)**	0.02	**1.77 (1.04–3.01)**	0.04

Significant metabolites features are indicated in bold

m/z mass-to-charge, *RT* retention time

Model 1: adjusted for age at blood (continuous). Model 2: model 1 + BOADICEA 1-year risk. Model 3: Model 2 + race and ethnicity + BMI (continuous) + alcohol (never, former and current) + smoking (never, former and current) + menopausal status

clusters by case–control status with some overlap. Figure 5 and Supplement Table 1 present the results from the pathway enrichment analysis based on the Mummichog algorithm [33]. The main pathways associated with BC include arginine and proline metabolism and urea cycle/amino group metabolism.

Discussion

We conducted a metabolome-wide association study based on an untargeted metabolomics workflow and identified eight BC related-metabolic features that were statistically significantly different between cases and controls. One of the identified features is amino acid and another feature belongs to lipids. In addition, we identified metabolic features related to diet as well as potential carcinogens. Pathway enrichment analysis identified a realm of pathways linked to both amino acid metabolism (e.g., arginine and proline metabolism) and lipid metabolism (e.g., glycerophospholipid metabolism). Our findings suggest that those metabolites and associated pathways are worthy of further evaluation using targeted, quantitative metabolomics analyses for BC risk. However, we recognized that these differences were not statistically significant after adjusted for multiple comparisons; thus, these preliminary findings thus need to be further tested and validated in larger prospective studies of BC.

1,3-Dibutyl-1-nitrosourea has demonstrated carcinogenic potential in animal models [35–38]. Specially,

exposure of rats to different doses of 1,3-dibutyl-1-nitrosourea via drinking water, resulted in a dose–response relationship with mammary tumors [36]. Other cancers such as leukemia and vaginal tumors were also observed in rats with high exposure to it [39]. However, additional data is needed in order for the International Agency for Research on Cancer (IARC) to determine whether a probable or possible carcinogen is carcinogenic in humans. To our best knowledge, this study is the first human data on an association of 1,3-dibutyl-1-nitrosourea with BC demonstrating the utility of the approach in identifying potential environmental exposures associated with the disease.

Dietary polyunsaturated fatty acids have been postulated as a modifiable factor that could influence cancer risk [40]. However, evidence for the effects of polyunsaturated fats such as omega-3 and omega-6 fatty acids on risk of cancer is conflicting [41–43]. Dietary intake of trans fatty acids was found to be associated with a slightly increased risk of BC (HR = 1.09, 95% CI: 1.01, 1.17) in the European Prospective Investigation into Cancer and Nutrition (EPIC) [44]. A systematic review and meta-analysis of randomized trials on omega-3, omega-6 and total dietary polyunsaturated fat on cancer incidence concluded that increasing omega-3 has little or no effect on BC incidence (RR = 1.03, 95% CI:0.89, 1,20) [45]. Through measuring serum phospholipid fatty acid composition among women in the

Wu *et al. Breast Cancer Research*   (2024) 26:141

Page 9 of 14

ROC Curves For Comparisons
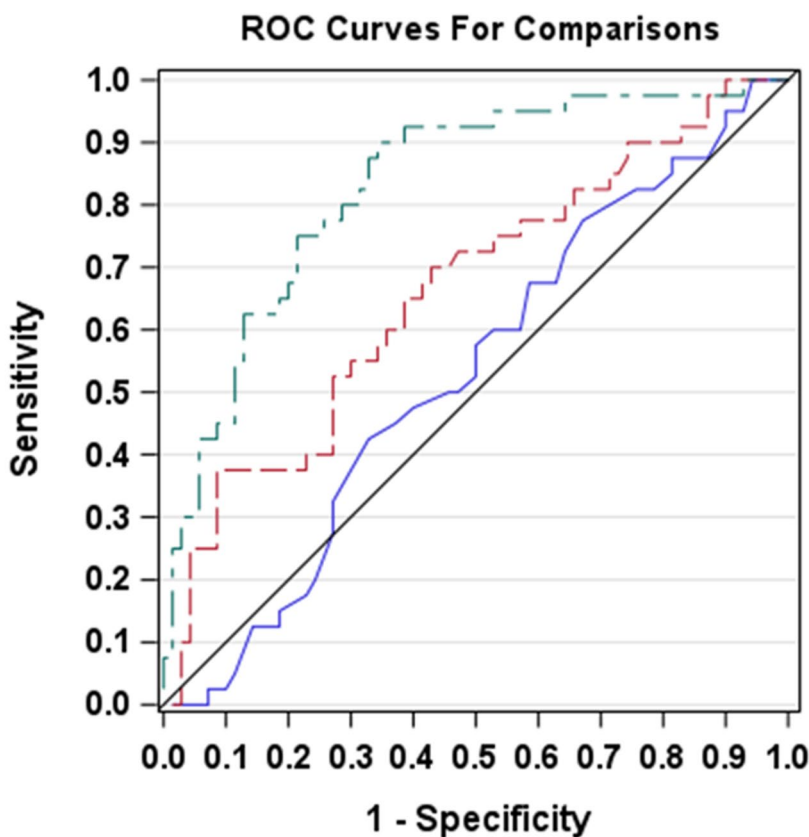
Model	AUC (95% CI)
Age	0.53 (0.42 - 0.64)
Age+BOADICEA 1-yr	0.66 (0.55 - 0.77)
Age+BOADICEA 1-yr+Metabolomics	0.83 (0.75 - 0.91)

Fig. 3 Receiver operating characteristics of a model with breast cancer risk factors and a model with breast cancer risk factors and six metabolite features, New York site of the BCFR. Metabolomic panel include: 1,3-Dibutyl-1-nitrosourea, L-Histidine, N(6)-Methyllysine, N-Acetylgalactosamine, 11-cis-Eicosenoic acid and LysoPE(0:0/24:6(6Z,9Z, 12Z,15Z, 18Z, 21Z))

E3N study, Chajes et al. found increasing levels of palmitoleic acid, a trans-monounsaturated fatty acid, was associated with an increased risk of BC (OR–1.7) [46]. We found both eicosapentaenoic acid, an omega-3 fatty acid, and 11-cis-eicosenoic acid, an omega-9-fatty acid, were associated with an increased risk of BC. Because omega-3 has been suggested as a supplement for BC prevention, a compensatory mechanistic route may occur in BC cases. Our finding needs to be validated in cohorts with a larger sample size.

LysoPE(0:0/24:6(6Z, 9Z, 12Z, 15Z, 18Z, 21Z), a lysophospholipid, is classified as a lipid mediator and elicits many biological effects such as cell proliferation, and migration [47] that are critically required for tumor formation and metastasis [47]. We found higher LysoPE(0:0/24:6(6Z, 9Z, 12Z, 15Z, 18Z, 21Z) was associated with higher BC risk. Alterations of lysoPC and lysoPE were observed in serum and plasma collected from BC patients [13, 48, 49] as well as breast tumor tissue [49]. It has been suggested that lipid oversupply enhances cancer cell proliferation by providing the raw materials needed to generate new cells [50]. Chronic lipid oversupply might increase BC risk, perhaps by supplying energy and nutrients to the growing tumors.

L-Histidine is an essential amino acid with unique roles in proton buffering, metal ion chelation, and scavenging of reactive oxygen and nitrogen species [51]. Histidine supplementation suppressed inflammation and improved insulin resistance in obese women with metabolic syndrome in a randomized controlled
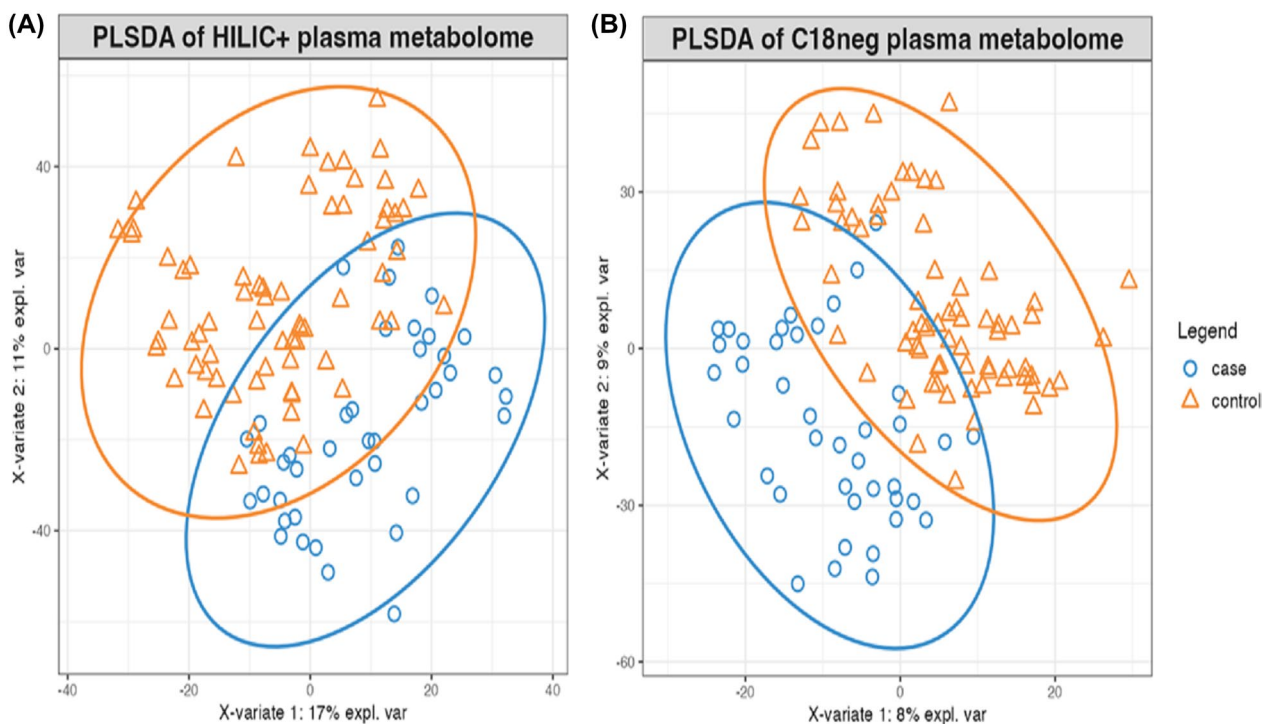
Wu *et al. Breast Cancer Research* (2024) 26:141

Page 10 of 14

Fig. 4 Partial least square discriminant analysis (PLS-DA) of plasma metabolomic data comparing breast cancer cases and unaffected controls under two complementary modes of analysis including (**A**) PLS-DA of hydrophilic interaction chromatography (HILIC) positive ESI and (**B**) C_{18} chromatography negative ESI

trial [52]. Histidine was associated with a decreased risk of BC (OR = 0.91, 95% CI, 0.84, 0.99) in a metabolome-wide association study within EPIC; however, the association was no longer statistically significant after adjustment for multiple comparisons [14]. Another metabolome-wide association study found histidine was associated with an increased risk of BC among premenopausal women in the French E3N cohort. [12]

Diabetes was associated with triple-negative breast cancer in a prospective analysis of the Sister Cohort [53]. Long term use of metformin has been associated with decreased risk of ER-positive BC [53]. Impaired glucose was associated with a non-statistically significant 40 percent higher BC risk in a cohort of 7,894 women aged 45–64 years from four US communities [54]. The inverse association between glucose and BC risk is challenging to interpret as the biospecimens were collected from non-fasting individuals in our study.

The epidemiological evidence on coffee consumption and BC risk is conflicting [55]. The EPIC study found an association between coffee intake and lower postmenopausal BC risk (HR = 0.90, 95% CI, 0.82, 0.98) [56]. While there was no evidence for an association in a cohort of 57,075 postmenopausal women [57]. Overall, current

studies of coffee consumption and BC examined coffee consumption based on self-report questionnaire. One suggestion is that possible risk differences exist with rates of caffeine metabolism [58]. Further biomarker studies measuring caffeine metabolites are needed to better characterize the preventive effect of caffeine in BC development.

In addition to the metabolome-wide association analysis to identify individual metabolic features associated with BC, pathway enrichment analysis showed that selected metabolic pathways such as arginine, proline and urea cycle might be altered in early breast tumorigenesis [59–62]. Untargeted metabolomics is a hypothesis-generating strategy to discovery early signs of metabolome-wide perturbations in BC development. Measuring metabolomic profiles may be a potential screening tool to identify higher risk individuals [63, 64]. Perturbations in fatty acid, arginine, and proline metabolism were found in plasma from BC cases at the time of cancer diagnosis [64, 65]. Our findings could provide insights for the identification of pathways for BC development.

Due to the sample size limitations, we opted not to explore the metabolite profiles by BC molecular subtypes. The results of our study also need to be interpreted with caution. The metabolomic features were only measured
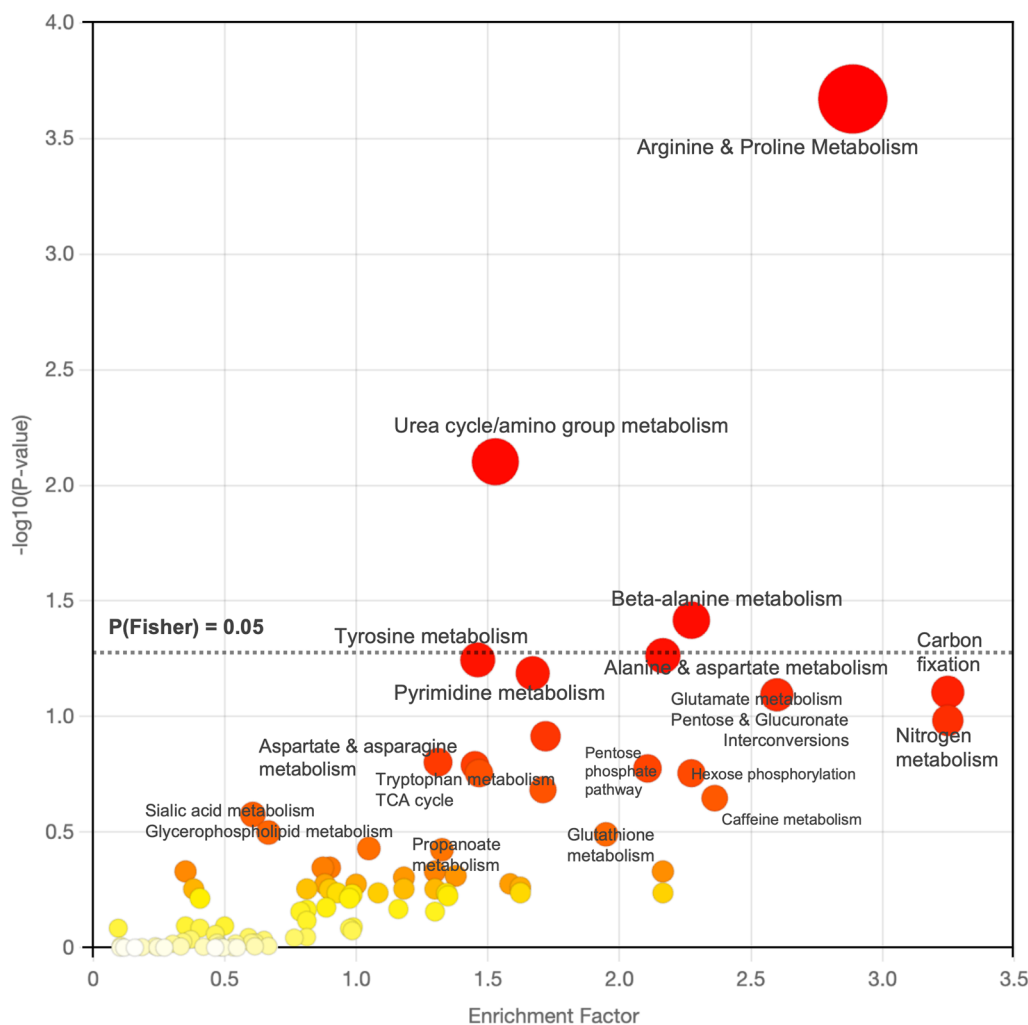
Fig. 5 Pathway analysis of the plasma metabolome comparing breast cancer cases and unaffected controls based on the Mummichog algorithm. The P-values are from Fisher's exact test applied to an enrichment test of individual metabolic features on pathways, mapping *m/z*-matched metabolites against a permutation procedure to reduce Type I error while adopting a more conservative version of Fisher's test to increase the robustness of the test

in non-fasting blood samples from a single timepoint for each participant, and we saw three metabolic features (L-Histidine, glucose, and caffeine) were positively correlated with each other. Although it is likely that most of the endogenous metabolites are biologically reproducible within a 2-year period [66], further studies are needed to examine the effect of blood collection conditions such as seasonal variation or fasting time. In addition, six metabolite features remain statistically significantly different between cases and controls after adjusting for selected risk factors; however, there might be some unadjusted confounding factors.

Accurately identifying high-risk individuals is essential for effective primary prevention (e.g., chemoprevention) [67–71], and for risk-based screening options [72, 73] which emphasize risk rather than age for optimal screening outcomes. BC risk assessment models used in the clinic only have very modest discriminatory accuracy in the range of 65% [74–77], meaning that 35% of women are misclassified. Inaccurate risk assessment means that women are either subject to over-treatment with biopsies and multiple screens or under-treatment with missed opportunities for optimal prevention, including chemoprevention. The most widely known and most commonly used model for BC risk assessment is the Breast Cancer Risk Assessment Tool (BCRAT, or Gail model) [78, 79], which although it is well-calibrated, only has modest discriminatory accuracy at the individual level (AUC ~ 0.6-0.65) [80–82]. Recently, modest improvements were achieved by incorporating

polygenetic risk score [83], epigenetic markers [84], and lifestyle factors [85]. Metabolome studies identified diet-and/or lifestyle-related metabolic features and their associations with breast cancer [16, 86, 87]. Metabolomics can detect metabolic shifts resulting from lifestyle behaviors and may provide insight on the relevance of changes to carcinogenesis. In addition, metabolomics analysis can also identify metabolic features associated with environmental exposure, such as polycyclic aromatic hydrocarbons (PAHs) [88]. Our prior study showed women with a higher risk of BC based on their genetic factors are more susceptible to PAH exposure [89]. Incorporating metabolite markers related to modifiable factors might result in substantially greater magnitudes of association with BC risk.

Strengths of our study include the collection of plasma before diagnosis (range of 1–15 years), and the use of an untargeted metabolomic approach allowing us to identify novel contributors to BC. In summary, our study identified selected metabolic pathways and potential exposure factors related to breast cancer. If replicated in larger prospective cohorts, these findings offer promising new ways to identify environmental exposures related to BC and improve BC risk prediction.

Abbreviations

AUC	The area under the curve
BC	Breast cancer
BCFR	Breast cancer family registry
BMI	Body mass index
BOADICEA model	Breast and ovarian analysis of disease incidence and carrier estimation algorithm
CI	Confidence interval
LC-MS	Liquid chromatography-mass spectrometry
LC-HRMS	Liquid chromatography-high resolution mass spectrometry
HILIC	Hydrophilic interaction liquid chromatography
MSI	Metabolomics standard initiatives (MSI)
m/z	Mass-to-charge
ORs	Odds ratios
PLS-DA	Partial least squares-discriminant analysis
RT	Retention time
VIP	Variable importance in the projection

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13058-024-01896-5.

> Additional file1

Author contributions
Wu, Lai wrote the main manuscript text. Wu, Lai and Liao prepared figures. All authors reviewed the manuscript.

Availability of data and materials
No datasets were generated or analysed during the current study.

Code availability statement
The underlying code for this study are available upon reasonable request.

Declarations

Ethics approval and consent to participate
All aspects of the BCFR were approved by the Columbia University Institutional Review Board.

Informed consent Statement
Informed consent was obtained from all subjects involved in the study.

Competing interests
Dr. Mary Beth Terry is an associated editor of the breast cancer research.

References
1. Torre LA, Siegel RL, Ward EM, Jemal A. Global cancer incidence and mortality rates and trends—an update. Cancer Epidemiol Biomark Prev. 2016;25(1):16–27.
2. Lima SM, Kehm RD, Terry MB. Global breast cancer incidence and mortality trends by region, age-groups, and fertility patterns. EClinicalMed. 2021;38:100985.
3. Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. Cell. 2008;134(5):714–7.
4. McCartney A, Vignoli A, Biganzoli L, et al. Metabolomics in breast cancer: a decade in review. Cancer Treat Rev. 2018;67:88–96.
5. Hiatt RA, Brody JG. Environmental determinants of breast cancer. Annu Rev Public Health. 2018;39(1):113–33.
6. Rudolph A, Chang-Claude J, Schmidt MK. Gene-environment interaction and risk of breast cancer. Br J Cancer. 2016;114(2):125–33.
7. Brantley KD, Zeleznik OA, Rosner B, et al. Plasma metabolomics and breast cancer risk over 20 years of follow-up among postmenopausal women in the nurses' health study. Cancer Epidemiol Biomark Prev. 2022;31(4):839–50.
8. Moore SC, Mazzilli KM, Sampson JN, et al. A metabolomics analysis of postmenopausal breast cancer risk in the cancer prevention study II. Metabolites. 2021;11(2):95.
9. Lécuyer L, Victor Bala A, Deschasaux M, et al. NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer. Int J Epidemiol. 2018;47(2):484–94.
10. Yoo HJ, Kim M, Kim M, et al. Analysis of metabolites and metabolic pathways in breast cancer in a Korean prospective cohort: the Korean cancer prevention study-II. Metabolomics. 2018;14(6):85.
11. Mrowiec K, Kurczyk A, Jelonek K, et al. Association of serum metabolome profile with the risk of breast cancer in participants of the HUNT2 study. Front Oncol. 2023;13:1116806.
12. Jobard E, Dossus L, Baglietto L, et al. Investigation of circulating metabolites associated with breast cancer risk by untargeted metabolomics: a

case-control study nested within the French E3N cohort. Br J Cancer. 2021;124(10):1734–43.

13. Stevens VL, Carter BD, Jacobs EJ, McCullough ML, Teras LR, Wang Y. A prospective case–cohort analysis of plasma metabolites and breast cancer risk. Breast Cancer Res. 2023;25(1):5.

14. His M, Viallon V, Dossus L, et al. Prospective analysis of circulating metabolites and breast cancer in EPIC. BMC Med. 2019;17(1):178.

15. Playdon MC, Ziegler RG, Sampson JN, et al. Nutritional metabolomics and breast cancer risk in a prospective study. Am J Clin Nutr. 2017;106(2):637–49.

16. Moore SC, Playdon MC, Sampson JN, et al. A metabolomics analysis of body mass index and postmenopausal breast cancer risk. J Natl Cancer Inst. 2018;110(6):588–97.

17. Terry MB, Phillips KA, Daly MB, et al. Cohort profile: the breast cancer prospective family study cohort (ProF-SC). Int J Epidemiol. 2016;45(3):683–92.

18. Pharoah P, Day N, Duffy S, Easton D, Ponder B. Family history and the risk of breast cancer: a systematic review and meta-analysis. Int J Cancer. 1997;71:800–9.

19. Braithwaite D, Miglioretti DL, Zhu W, et al. Family history and breast cancer risk among older women in the breast cancer surveillance consortium cohort. JAMA Intern Med. 2018;178(4):494–501.

20. John E, Hopper J, Beck J, et al. The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. Breast Cancer Res. 2004;6(4):R375–89.

21. Terry MB, Phillips K-A, Daly MB, et al. Cohort profile: the breast cancer prospective family study cohort (ProF-SC). Int J Epidemiol. 2015;45:1–10.

22. Soltow QA, Strobel FH, Mansfield KG, Wachtman L, Park Y, Jones DP. High-performance metabolic profiling with dual chromatography-Fourier-transform mass spectrometry (DC-FTMS) for study of the exposome. Metabolomics. 2013;9(1 Suppl):S132–43.

23. Yu T, Park Y, Li S, Jones DP. Hybrid feature detection and information accumulation using high-resolution LC–MS metabolomics data. J Proteome Res. 2013;12(3):1419–27.

24. Uppal K, Soltow QA, Strobel FH, et al. xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. BMC Bioinform. 2013;14(1):15.

25. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118–27.

26. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genom. 2006;7:142.

27. Schymanski EL, Jeon J, Gulde R, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol. 2014;48(4):2097–8.

28. Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). Metabolomics. 2007;3(3):211–21.

29. Lee A, Mavaddat N, Wilcox AN, et al. BOADICEA: a comprehensive breast cancer risk prediction modelincorporating genetic and nongenetic risk factors. Genet Med. 2019;21(8):1708–18.

30. Antoniou AC, Cunningham AP, Peto J, Evans DG, Lalloo F, Narod SA. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. Br J Cancer. 2008;98:1457–66.

31. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882–3.

32. Rohart F, Gautier B, Singh A, Lê Cao K-A. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol. 2017;13(11): e1005752.

33. Li S, Park Y, Duraisingham S, et al. Predicting network activity from high throughput metabolomics. PLoS Comput Biol. 2013;9(7): e1003123.

34. Mahieu NG, Patti GJ. Systems-level annotation of a metabolomics data set reduces 25 000 features to Fewer than 1000 unique metabolites. Anal Chem. 2017;89(19):10397–406.

35. Maekawa A, Onodera H, Tanigawa H, et al. Experimental induction of ovarian sertoli cell tumors in rats by N-Nitrosoureas. Environ Health Perspect. 1987;73:115–23.

36. Ogiu T, Kajiwara T, Furuta K, Takeuchi M, Odashima S, Tada K. Mammary tumorigenic effect of a new nitrosourea, 1,3-dibutyl-l-nitrosourea (B-BNU), in female Donryu rats. J Cancer Res Clin Oncol. 1980;96(1):35–41.

37. Rudel RA, Attfield KR, Schifano JN, Brody JG. Chemicals causing mammary gland tumors in animals signal new directions for epidemiology, chemicals testing, and risk assessment for breast cancer prevention. Cancer. 2007;109(S12):2635–66.

38. Odashima S, Hashimoto Y, Ogiu T, Maekawa A. Carcinogenic effect of 1-butyl-1-nitrosourea on female Sprague-Dawley rats. Gan. 1975;66(6):615–21.

39. Takeuchi M, Maekawa A, Tada K, Odashima S. Leukemias and vaginal tumors induced in female Donryu rats by continuous administration of 1-butyl-3,3-dimethyl-1-nitrosourea in the drinking water. J Natl Cancer Inst. 1976;56(6):1177–81.

40. Azrad M, Turgeon C, Demark-Wahnefried W. Current evidence linking polyunsaturated Fatty acids with cancer risk and progression. Front Oncol. 2013;3:224.

41. Hooper L, Thompson RL, Harrison RA, et al. Risks and benefits of omega 3 fats for mortality, cardiovascular disease, and cancer: systematic review. BMJ. 2006;332(7544):752–60.

42. Zhang Y-F, Gao H-F, Hou A-J, Zhou Y-H. Effect of omega-3 fatty acid supplementation on cancer incidence, non-vascular death, and total mortality: a meta-analysis of randomized controlled trials. BMC Public Health. 2014;14(1):204.

43. MacLean CH, Newberry SJ, Mojica WA, et al. Effects of Omega-3 fatty acids on cancer RiskA systematic review. JAMA. 2006;295(4):403–15.

44. Matta M, Huybrechts I, Biessy C, et al. Dietary intake of trans fatty acids and breast cancer risk in 9 European countries. BMC Med. 2021;19(1):81.

45. Hanson S, Thorpe G, Winstanley L, et al. Omega-3, omega-6 and total dietary polyunsaturated fat on cancer incidence: systematic review and meta-analysis of randomised trials. Br J Cancer. 2020;122(8):1260–70.

46. Chajès V, Thiébaut AC, Rotival M, et al. Association between serum trans-monounsaturated fatty acids and breast cancer risk in the E3N-EPIC Study. Am J Epidemiol. 2008;167(11):1312–20.

47. Mills GB, Moolenaar WH. The emerging role of lysophosphatidic acid in cancer. Nat Rev Cancer. 2003;3(8):582–91.

48. Amiri-Dashatan N, Yekta RF, Koushki M, et al. Metabolomic study of serum in patients with invasive ductal breast carcinoma with LC-MS/MS approach. Int J Biol Mark. 2022;37(4):349–59.

49. Díaz-Beltrán L, González-Olmedo C, Luque-Caro N, et al. Human plasma metabolomics for biomarker discovery: targeting the molecular subtypes in breast cancer. Cancers (Basel). 2021;13(1):147.

50. Carracedo A, Cantley LC, Pandolfi PP. Cancer metabolism: fatty acid oxidation in the limelight. Nat Rev Cancer. 2013;13(4):227–32.

51. Peterson JW, Boldogh I, Popov VL, Saini SS, Chopra AK. Anti-inflammatory and antisecretory potential of histidine in Salmonella-challenged mouse small intestine. Lab Invest. 1998;78(5):523–34.

52. Feng RN, Niu YC, Sun XW, et al. Histidine supplementation improves insulin resistance through suppressed inflammation in obese women with the metabolic syndrome: a randomised controlled trial. Diabetologia. 2013;56(5):985–94.

53. Park YMM, Bookwalter DB, O'Brien KM, Jackson CL, Weinberg CR, Sandler DP. A prospective study of type 2 diabetes, metformin use, and risk of breast cancer. Ann Oncol. 2021;32(3):351–9.

54. Mink PJ, Shahar E, Rosamond WD, Alberg AJ, Folsom AR. Serum insulin and glucose levels and breast cancer incidence: the atherosclerosis risk in communities study. Am J Epidemiol. 2002;156(4):349–52.

55. Pauwels EKJ, Volterrani D. Coffee consumption and cancer risk: an assessment of the health implications based on recent knowledge. Med Princ Pract. 2021;30(5):401–11.

56. Bhoo-Pathy N, Peeters PH, Uiterwaal CS, et al. Coffee and tea consumption and risk of pre- and postmenopausal breast cancer in the European prospective investigation into cancer and nutrition (EPIC) cohort study. Breast Cancer Res. 2015;17(1):15.

57. Gapstur SM, Gaudet MM, Wang Y, et al. coffee consumption and invasive breast cancer incidence among postmenopausal women in the cancer prevention study-II nutrition cohort. Cancer Epidemiol Biomark Prev. 2020;29(11):2383–6.

58. Nehlig A, Reix N, Arbogast P, Mathelin C. Coffee consumption and breast cancer risk: a narrative review in the general population and in different subtypes of breast cancer. Eur J Nutr. 2021;60(3):1197–235.

Wu *et al. Breast Cancer Research*   (2024) 26:141

Page 14 of 14

59. D'Aniello C, Patriarca EJ, Phang JM, Minchiotti G. Proline metabolism in tumor growth and metastatic progression. Front Oncol. 2020. https://doi.org/10.3389/fonc.2020.00776.
60. Chen CL, Hsu SC, Ann DK, Yen Y, Kung HJ. Arginine signaling and cancer metabolism. Cancers (Basel). 2021;13(14):3541.
61. Keshet R, Szlosarek P, Carracedo A, Erez A. Rewiring urea cycle metabolism in cancer to support anabolism. Nat Rev Cancer. 2018;18(10):634–45.
62. Kus K, Kij A, Zakrzewska A, et al. Alterations in arginine and energy metabolism, structural and signalling lipids in metastatic breast cancer in mice detected in plasma by targeted metabolomics and lipidomics. Breast Cancer Res. 2018;20(1):148.
63. Wei Y, Jasbi P, Shi X, et al. Early breast cancer detection using untargeted and targeted metabolomics. J Proteome Res. 2021;20(6):3124–33.
64. Wei Y, Jasbi P, Shi X, et al. Early breast cancer detection using untargeted and targeted metabolomics. J Proteome Res. 2021;20(6):3124–33.
65. Jasbi P, Wang D, Cheng SL, et al. Breast cancer detection using targeted plasma metabolomics. J Chromatogr B. 2019;1105:26–37.
66. Carayol M, Licaj I, Achaintre D, et al. Reliability of serum metabolites over a two-year period: a targeted metabolomic approach in fasting and non-fasting samples from EPIC. PLoS ONE. 2015;10(8): e0135437.
67. Padamsee TJ, Wills CE, Yee LD, Paskett ED. Decision making for breast cancer prevention among women at elevated risk. Breast Cancer Res. 2017;19(1):34.
68. Howell A, Anderson AS, Clarke RB, Duffy SW, Evans DG, Garcia-Closas M. Risk determination and prevention of breast cancer. Breast Cancer Res. 2014. https://doi.org/10.1186/s13058-014-0446-2.
69. Gagnon J, Lévesque E, The Clinical Advisory Committee on Breast Cancer S, et al. Recommendations on breast cancer screening and prevention in the context of implementing risk stratification: impending changes to current policies. Curr Oncol. 2016; 23(6):e615-e625.
70. Reimers LL, Sivasubramanian PS, Hershman D, et al. Breast cancer chemoprevention among high-risk women and those with ductal carcinoma in situ. Breast J. 2015;21(4):377–86.
71. Pruthi S, Heisey RE, Bevers TB. Chemoprevention for breast cancer. Ann Surg Oncol. 2015;22(10):3230–5.
72. Lee CI, Chen LE, Elmore JG. Risk-based breast cancer screening: implications of breast density. Med Clin North Am. 2017;101(4):725–41.
73. Vilaprinyo E, Forné C, Carles M, et al. Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer. PLOS One. 2014;9(2):e86858.
74. Amir E, Freedman OC, Seruga B, Evans DG. Assessing women at high risk of breast cancer: a review of risk assessment models. J Natl Cancer Inst. 2010;102(10):680–91.
75. Jacobi C, de Bock G, Siegerink B, van Asperen C. Differences and similarities in breast cancer risk assessment models in clinical practice: Which model to choose? Breast Cancer Res Treat. 2009;115(2):381–90.
76. Meads C, Ahmed I, Riley R. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. Breast Cancer Res Treat. 2012;132(2):365–77.
77. Anothaisintawee T, Teerawattananon Y, Wiratkapun C, Kasamesup V, Thakkinstian A. Risk prediction models of breast cancer: a systematic review of model performances. Breast Cancer Res Treat. 2012;133(1):1–10.
78. Gail M, Brinton L, Byar D, Corle D, Green S, Schairer C. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst. 1989;81:1879–86.
79. Constantino JP, Gail MH, Pee D, Anderson S, Redmond CK, Benichou J. Validation studies for models projecting the risk of invasive and total breast cancer incidence. J Natl Cancer Inst. 1999;91:1541–8.
80. Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA. Validation of the Gail et al. Model of breast cancer risk prediction and implications for chemoprevention. J Natl Cancer Inst. 2001;93(5):358–366
81. Banegas MP, John EM, Slattery ML, et al. Projecting individualized absolute invasive breast cancer risk in US Hispanic women. JNCI: J Natl Cancer Inst. 2017;109(2):djw215.
82. Quante AS, Whittemore AS, Shriver T, Hopper JL, Strauch K, Terry MB. Practical problems with clinical guidelines for breast cancer prevention based on remaining lifetime risk. JNCI J Natl Cancer Inst. 2015;107(7):djv124.
83. Wacholder S, Hartge P, Prentice R, et al. Performance of common genetic variants in breast-cancer risk models. Engl J Med. 2010;362(11):986–93.
84. Xu Z, Bolick SC, DeRoo LA, Weinberg CR, Sandler DP, Taylor JA. Epigenome-wide association study of breast cancer using prospectively collected sister study samples. J Natl Cancer Inst. 2013;105(10):694–700.
85. Guan Z, Raut JR, Weigl K, et al. Individual and joint performance of DNA methylation profiles, genetic risk score and environmental risk scores for predicting breast cancer risk. Mol Oncol. 2020;14(1):42–53.
86. Qi J, Spinelli JJ, Dummer TJB, et al. Metabolomics and cancer preventive behaviors in the BC Generations Project. Sci Rep. 2021;11(1):12094.
87. Lécuyer L, Dalle C, Lefevre-Arbogast S, et al. Diet-related metabolomic signature of long-term breast cancer risk using penalized regression: an exploratory study in the SU.VI.MAX. Cohort. Cancer Epidemiol Biomark Prev. 2020;29(2):396–405.
88. Wang Z, Zheng Y, Zhao B, et al. Human metabolic responses to chronic environmental polycyclic aromatic hydrocarbon exposure by a metabolomic approach. J Proteome Res. 2015;14(6):2583–93.
89. Shen J, Liao Y, Hopper JL, Goldberg M, Santella RM, Terry MB. Dependence of cancer risk from environmental exposures on underlying genetic susceptibility: an illustration with polycyclic aromatic hydrocarbons and breast cancer. Br J Cancer. 2017;116(9):1229–33.

Publisher's Note