Plasma Metabolomics and Breast Cancer Risk over 20 Years of Follow-up among Postmenopausal Women in the Nurses' Health Study

Kristen D. Brantley[1], Oana A. Zeleznik[2], Bernard Rosner[2,3], Rulla M. Tamimi[4], Julian Avila-Pacheco[5], Clary B. Clish[5], and A. Heather Eliassen[1,2,6]

ABSTRACT

Background: Metabolite profiles provide insight into biologic mechanisms contributing to breast cancer development. We explored the association between prediagnostic plasma metabolites ($N = 307$) and invasive breast cancer among postmenopausal women in a nested case–control study within the Nurses' Health Study ($N = 1{,}531$ matched pairs).

Methods: Plasma metabolites were profiled via LC/MS-MS using samples taken ≥10 years (distant, $N = 939$ cases) and <10 years (proximate, $N = 592$ cases) before diagnosis. Multivariable conditional logistic regression was used to estimate ORs and 95% confidence intervals (CI) comparing the 90th to 10th percentile of individual metabolite level, using the number of effective tests (NEF) to account for testing multiple correlated hypotheses. Associations of metabolite groups with breast cancer were evaluated using metabolite set enrichment analysis (MSEA) and weighted gene coexpression network analysis (WGCNA), with adjustment for the FDR.

Results: No individual metabolites were significantly associated with breast cancer risk. MSEA showed negative enrichment of cholesteryl esters at the distant timepoint [normalized enrichment score (NES) $= -2.26$; $P_{adj} = 0.02$]. Positive enrichment of triacylglycerols (TAG) with <3 double bonds was observed at both timepoints. TAGs with ≥3 double bonds were inversely associated with breast cancer at the proximate timepoint (NES $= -2.91$, $P_{adj} = 0.03$).

Conclusions: Cholesteryl esters measured earlier in disease etiology were inversely associated with breast cancer. TAGs with many double bonds measured closer to diagnosis were inversely associated with breast cancer risk.

Impact: The discovered associations between metabolite subclasses and breast cancer risk can expand our understanding of biochemical processes involved in cancer etiology.

Introduction

Metabolite profiles reflect the integrated impact of the genome and exogenous exposures on the metabolic state and may provide insight into biologic mechanisms contributing to disease development. Breast cancer is the most common cancer among women worldwide (1). Although key sex hormone-related metabolic pathways are well-established in breast cancer etiology, knowledge on metabolic pathways in aggregate may reveal additional targets for prevention.

A handful of recent studies have explored metabolite associations with breast cancer incidence (2–9), although only a few have taken an agnostic approach to explore the metabolomics of breast cancer (3, 8, 9), instead focusing on weight-associated or nutritional metabolites. Among the studies that have explored metabolites overall with respect to breast cancer risk, one had a very small sample size ($N = 84$ cases; ref. 8), and all used different metabolomic platforms for measurement. All studies thus far have only captured metabolite profiles at a single point in time. Previous studies suggest inverse associations between carnitines (3, 9) and phosphatidylcholines (9) and breast cancer risk, and positive associations between amino acids and breast cancer risk (5), though importance of individual metabolites varied by study.

Here we used an agnostic approach in the Nurses' Health Study to investigate associations between metabolite levels, measured prior to diagnosis, and future breast cancer risk. We also examined how these measures changed over time, using measures from two different blood collections, approximately 10 years apart.

Materials and Methods

Cohort

We conducted a nested case–control study within the Nurses' Health Study (NHS), a prospective cohort of 121,700 female nurses started in 1976. Biennial follow-up questionnaires collect risk factor information as well as new disease diagnoses. Blood samples were collected in 1989 to 1990 from 32,826 cohort members, ages 43 to 69 years at blood collection. A subset of these women ($N = 18{,}743$) provided a second blood sample between 2000 and 2002. The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required.

Breast cancer cases were identified by self-report and confirmed by medical record review. Deaths were captured by next of kin, postal service, or review of the National Death Index. Cases were all women diagnosed with invasive or *in situ* breast cancer between 2000 and 2010

[1]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. [2]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts. [3]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. [4]Department of Population Health Sciences, Weill Cornell Medicine, New York, New York. [5]Metabolomics Platform, Broad Institute of MIT and Harvard, Cambridge, Massachusetts. [6]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Corresponding Author: Kristen D. Brantley, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115. E-mail: kbrantley@g.harvard.edu

Cancer Epidemiol Biomarkers Prev 2022;31:839–50

doi: 10.1158/1055-9965.EPI-21-1023

who provided a blood sample ($N = 939$ for distant 1989–1990 blood collection; $N = 592$ for proximate 2000–2002 blood collection) and had no prior reported cancer (other than nonmelanoma skin). All those with proximate blood samples also had distant blood measures. Controls were matched to cases on factors at each blood draw, including age (± 1 year), month (±1 month), time of day (±2 hours), fasting status (≥10 hours since a meal vs. <10 hours or unknown), and combined menopausal status and postmenopausal hormone use (premenopausal/postmenopausal, not on hormones//postmenopausal, on hormones, unknown).

Metabolite profiling

Plasma metabolites were profiled at the Broad Institute of MIT and Harvard. Two LC/MS-MS platforms were used for identification of metabolites, designed to measure polar metabolites and lipids, and free fatty acids, described elsewhere (10–13). Specifics on measurement procedures are described in a previous publication (14). Briefly, matched case–control pairs were distributed randomly within batch, pooled reference samples were included every 20 samples, and 64 quality controls were distributed randomly. Measures were standardized using the ratio of the value of the sample to the value of the nearest pooled reference multiplied by the median of all reference values for the metabolite. For metabolites measured with multiple metabolomics platforms, the assay laboratory provided a list of the preferred measurement platform. For metabolites measured multiple times with the same platform, the metabolite with the lowest CV was used for analysis. Metabolites that had poor stability due to delay in processing were excluded ($N = 51$; ref. 13). Following this initial data cleaning, a total of 307 known metabolites were successfully measured and included in the study. Metabolites were annotated by superclass, class, and subclass distinctions.

Covariates

Identified risk factors for breast cancer were included as covariates in the analyses: BMI at age 18 (kg/m^2), weight change since age 18 (kg), age at first birth and parity (nulliparous, 1–2 kids <25 years, 1–2 kids 25+ years, 3+ kids <25 years, 3+ kids 25+ years), age at menarche (years), breastfeeding history (yes/no), history of benign breast disease (yes/no), family history of breast cancer (yes/no), physical activity (MET-hours/week), and alcohol intake (g/day).

Statistical analysis

Metabolites with <10% missing were imputed with half the minimum value ($N = 39$ at distant blood collection, $N = 0$ at proximate blood collection). Metabolites with ≥10% missing ($N = 15$ at distant blood collection, $N = 16$ at proximate blood collection) were not imputed. We used probit transformation for all metabolites. We used multivariable conditional logistic regression (CLR) to calculate OR and 95% confidence intervals (CI) for individual metabolites with breast cancer at both distant and proximate blood collections. Unconditional logistic regression (UCLR) with adjustment for matching factors was used for estrogen receptor positive (ER+) and negative (ER−) breast cancers due to limited ER− cases. ORs represent a 2.5 SD increase in metabolites, equivalent to the comparison for 90th to 10th percentile of metabolite value under the assumption of a normal distribution.

We accounted for testing for multiple correlated hypotheses by calculating the number of effective tests by performing a principal components (PC) analysis of all metabolites among controls and calculating the number of PCs that explained 99.5% of the total variance (15). For this method, $P_{adj} = P_{unadjusted}$/number of effective tests (P_{adj} distant = 0.0003, P_{adj} proximate = 0.0002).

In a separate analysis, we explored the association of presence versus absence of metabolites with ≥10% missingness with breast cancer risk.

Correlations between metabolite measurements at distant and proximate timepoints were assessed using unadjusted Spearman correlations, and adjusted for fasting, age at blood draw, and weight change since age 18. We used unconditional logistic regression models including metabolite measures at both timepoints in the same regression along with an interaction term; the *P* value for the interaction term was used to determine potential interest in the difference measures.

The difference of metabolite levels was analyzed via unconditional logistic regression, with ORs representing comparison of the 90th to 10th percentile metabolite level from distant to proximate blood, adjusted for distant blood. For average and difference analyses, fasting status and menopausal status were assessed as a combination of the two timepoints. The remaining covariates were from the proximate blood collection.

Metabolites were grouped on the basis of structural similarities by subclasses; triacylglycerols (TAG) were further divided as TAGs with ≥3 versus TAGs with <3 double bonds. Metabolite set enrichment analysis (MSEA) combines the effect estimates from logistic regressions performed on individual metabolites by defined groups, to determine a summary enrichment score (ES) and normalized enrichment score (NES) adjusted for group size (16). The ES represents the degree to which the metabolite set is overrepresented compared to other sets; where a positive ES represents a significant positive enrichment in breast cancer, whereas a significant negative score indicates a group that is negatively enriched in breast cancer. *P* values were adjusted using the FDR to account for multiple comparisons (17).

Weighted gene (metabolite) coexpression network analysis (WGCNA) was used to identify metabolite modules associated with breast cancer risk. This analysis process is described in detail elsewhere (18). Briefly, a coexpression network is constructed using the absolute values of the correlation coefficients between metabolites to identify interconnected "nodes" based on a threshold value of similarity. Hierarchical clustering based on scale-free topology identifies densely interconnected metabolites from the network, and modules are grouped by using a Dynamic Tree Cut method (19). Within each analysis, all metabolites were assigned a module score, derived in control subjects at each timepoint separately, based on their loading on the first principal component of each module. Module scores were then included in UCLR models for breast cancer risk. The resulting OR represents the association of a particular module with breast cancer risk. The loading status of individual metabolites into each module was examined to determine the influence of individual metabolites on the resultant module association (18).

Datasets for analysis were created in SAS version 9 (SAS Institute Inc.). All analyses were conducted using R programming language, version 4.0.3.

Results

A total of 939 cases and 939 matched controls were included for distant blood collection analysis, and 592 cases and 592 controls were included for the proximate blood collection. At first blood collection, mean age was 55 years (SD = 6.9); 25% of women were premenopausal (**Table 1**). At the second blood draw, 98% of women were postmenopausal. Family history of breast cancer, particularly at second collection, was higher among cases (23%) compared with controls (15%). As expected, weight gain since age 18 was higher at the second blood draw; at both timepoints, cases tended to have approximately 2 kg more weight gain compared with controls.

Table 1. Descriptive characteristics of participants in NHS who provided blood samples at distant and proximate dates.[a]

Characteristic	Distant blood		Proximate blood	
	Case (*N* = 939)	Control (*N* = 939)	Case (*N* = 592)	Control (*N* = 592)
Age at blood draw [mean (SD)]	55.5 (6.9)	55.6 (6.9)	66.4 (6.9)	66.5 (6.8)
Fasting at blood draw, *N* (%)	626 (67%)	683 (73%)	515 (87%)	547 (92%)
Menopausal status and PMH use at blood draw, *N* (%)				
Premenopausal	239 (26%)	240 (26%)	3 (1%)	5 (1%)
Postmenopausal, no PMH use	288 (31%)	289 (31%)	188 (32%)	186 (31%)
Postmenopausal, PMH use	293 (31%)	292 (31%)	393 (66%)	395 (67%)
Unknown	0	0	8 (1%)	6 (1%)
Age at menarche [mean (SD)]	12.5 (1.4)	12.6 (1.4)	12.5 (1.4)	12.6 (1.4)
Nulliparous, *N* (%)	90 (10%)	75 (8%)	51 (9%)	35 (6%)
Parity [mean (SD)][b]	3.1 (1.4)	3.2 (1.6)	3.1 (1.3)	3.2 (1.6)
Age at first birth [mean (SD)][b]	25 (3.1)	25 (3.1)	24.9 (3.1)	24.7 (3.0)
Breastfeeding history, *N* (%)[b]	604 (64%)	583 (62%)	399 (67%)	381 (64%)
History of benign breast disease, *N* (%)	492 (52%)	430 (46%)	383 (65%)	346 (56%)
Family history of breast cancer, *N* (%)	136 (15%)	101 (11%)	135 (23%)	87 (15%)
Weight change from age 18 to blood draw in kg [mean (SD)]	12.3 (10.9)	10.6 (11.2)	15.1 (12.8)	13.5 (12.8)
BMI at blood draw in kg/m^2 [mean (SD)]	25.7 (4.3)	25.2 (4.7)	26.7 (5.0)	26.4 (5.1)
Average alcohol consumption at blood draw in g/day [mean (SD)]	7.0 (9.9)	5.9 (8.2)	6.7 (9.2)	5.8 (7.7)
Activity level at blood draw in MET-hours/week [mean (SD)]	15.4 (18.8)	15.9 (17.6)	25.7 (42.0)	23.4 (31.7)

[a]Distant blood draw was >10 years before diagnosis date for cases. Proximate blood draw was ≤10 years before diagnosis date for cases.
[b]Among parous women.

No individual metabolites at either distant or proximate timepoints were significantly associated with breast cancer risk after adjusting for the number of effective tests (NEF distant = 193, proximate = 186, P_{adj} distant = 0.0003, proximate = 0.0002). Despite the lack of significance at this level, several metabolites and metabolite classes stood out as nominally significant (**Table 2**; Supplementary Tables S1A and S1B). The amino acid phenylalanine was positively associated with breast cancer risk at both distant (OR = 1.41; 95% CI = 1.08–1.85; nominal *P*-value = 0.01), and proximate timepoints (OR = 1.76; 95% CI = 1.25–2.48; nominal *P*-value = 0.001). Similar positive associations at both timepoints were observed for the amino acid proline. We observed strong positive associations for TAGs with <3 double bonds at the distant timepoint, (e.g., C51:0 TAG OR = 1.30; 95% CI = 1.01–1.68; nominal *P*-value = 0.04). At the proximate timepoint, several TAGs with high numbers of double bonds were inversely associated with breast cancer risk (e.g., C54:9 TAG OR = 0.64; 95% CI = 0.47–0.87; nominal *P*-value = 0.005).

The majority of metabolites with ≥10% missingness were drug related, and none of these metabolites were associated with breast cancer risk in presence versus absence assessment (Supplementary Tables S2A and S2B).

Although most associations were consistent between ER+ and ER− breast cancer, some metabolites were associated in opposite directions for ER+ versus ER− breast cancers [**Table 3**; Supplementary Tables S3A and S3B (ER+), Supplementary Tables S4A and S4B (ER−)], although most were not significantly heterogeneous. For example, at the proximate timepoint TAGs with <3 double bonds were strongly positively associated with ER+ breast cancers, but inversely associated with ER− breast cancers (e.g., C52:0 TAG ER+ OR = 1.49; 95% CI = 1.04–2.15; nominal *P*-value = 0.03; ER− OR = 0.86; 95% CI = 0.42–0.74; nominal *P*-value = 0.668, nominal *P*-value = 0.09, P_{het} = 0.25).

Because of the differences in TAG associations by number of double bonds, we further explored TAGs by number of carbon atoms and number of double bonds (**Fig. 1**). We observed a strong inverse association for TAGs with increasing carbon atoms and double bonds at the proximate timepoint. This inverse association was not notable for the distant timepoint, though we observed a trend of more positive associations with lower number of carbon atoms and double bonds at the distant timepoint.

MSEA results mirrored individual metabolite analyses and revealed several subclasses of metabolites significantly associated with breast cancer risk after FDR correction (**Fig. 2**; Supplementary Tables S5A and S5B). TAGs with <3 double bonds at the distant timepoint were strongly positively associated with risk of overall (P_{adj} = 0.02), ER+, and ER− breast cancers. This trend remained at the proximate blood draw for ER+ breast cancers, but no association was observed for ER- breast cancers. TAGs with ≥3 double bonds were strongly inversely associated with breast cancer risk at the proximate blood draw for overall (P_{adj} = 0.03), ER+, and ER− breast cancers; however, at the distant blood draw this group was significantly positively associated with ER+ breast cancer.

Cholesteryl esters were strongly inversely associated with risk at the distant timepoint (P_{adj} overall BC = 0.02), and less strongly, although still inverse, at the proximate timepoint. Glycerophospholipids, glycerophosphoethanolamines, and glycerophosphocholines were inversely associated with risk at the distant timepoint for ER+ breast cancers, although associations were weaker and not significant at the proximate timepoint. Similarly, diacylglycerols (DAG) were strongly positively associated with risk at the distant timepoint, but less strongly associated at the proximate timepoint. Further, the group amino acids, peptides, and analogues was positively associated with overall (P_{adj} = 0.02), ER+ and ER− breast cancer at the proximate blood, although the result was stronger for ER− than ER+ breast cancers. This group was not significantly associated with breast cancer risk at the distant timepoint.

WGCNA defined 12 metabolite modules at distant collection, and 11 at proximate (Supplementary Figs. S1 and S2; Supplementary Table S6). Module 1, the grey module, represents those metabolites that remained after correlation analyses determined other metabolite groupings. Although modules were not defined by one particular

Table 2. ORs for breast cancer risk comparing 90th to 10th percentiles of selected metabolites[a] measured at distant or proximate blood.

Metabolite name	HMDB ID	Class	Subclass	Unadjusted		Multivariate adjusted[b]	
				OR (95% CI)	*P* value	OR (95% CI)	*P* value
Distant blood							
Phenylalanine	HMDB0000159	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.50 (1.17–1.94)	0.002	1.41 (1.08–1.85)	0.012
Proline	HMDB0000162	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.37 (1.07–1.75)	0.012	1.33 (1.03–1.72)	0.032
Homoarginine	HMDB0000670[c]	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.41 (1.11–1.80)	0.005	1.3 (1.01–1.68)	0.039
Lysine	HMDB0000182	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.38 (1.08–1.77)	0.011	1.31 (1.01–1.69)	0.040
C5:1 carnitine	HMDB0002366	Fatty acyls	Fatty acid esters	0.80 (0.64–1.01)	0.064	0.73 (0.57–0.93)	0.010
C5-DC carnitine	HMDB0013130	Fatty acyls	Fatty acid esters	0.72 (0.57–0.92)	0.007	0.73 (0.57–0.93)	0.012
C51:0 TAG	HMDB0031106[c]	Glycerolipids	TAGs	1.46 (1.15–1.86)	0.002	1.30 (1.01–1.68)	0.044
C22:5 LPC	HMDB0010403[c]	Glycerophospholipids	Glycerophosphocholines	0.78 (0.61–0.99)	0.041	0.78 (0.60–1.00)	0.047
C22:0 LPE	HMDB0011520	Glycerophospholipids	Glycerophosphoethanolamines	0.69 (0.54–0.89)	0.004	0.75 (0.58–0.98)	0.035
C38:6 PE plasmalogen	HMDB0011387[c]	Glycerophospholipids	Glycerophosphoethanolamines	0.82 (0.65–1.03)	0.088	0.78 (0.61–0.99)	0.039
Thyroxine	HMDB0000248	NA	NA	1.50 (1.16–1.95)	0.002	1.56 (1.19–2.05)	0.001
Acetyl-galactosamine	HMDB0000212	Organooxygen compounds	Carbohydrates and carbohydrate conjugates	1.42 (1.1–1.84)	0.008	1.35 (1.02–1.77)	0.035
2-Methylguanosine	HMDB0005862	Purine nucleosides	NA	1.38 (1.07–1.77)	0.014	1.32 (1.01–1.72)	0.039
Guanosine	HMDB0000133	Purine nucleosides	NA	0.77 (0.61–0.97)	0.027	0.78 (0.61–0.99)	0.041
C22:5 CE	HMDB0010375[c]	Steroids and steroid derivatives	Cholesterol esters	0.61 (0.48–0.77)	<0.001	0.67 (0.52–0.86)	0.002
C18:3 CE	HMDB0010370[c]	Steroids and steroid derivatives	Cholesterol esters	0.70 (0.55–0.88)	0.003	0.69 (0.54–0.89)	0.004
C20:5 CE	HMDB0006731	Steroids and steroid derivatives	Cholesterol esters	0.75 (0.60–0.95)	0.016	0.74 (0.58–0.95)	0.017
Proximate blood							
2-Aminohippuric acid	NA[d]	Benzene and substituted derivatives	Benzoic acids and derivatives	1.39 (1.01–1.93)	0.046	1.45 (1.02–2.06)	0.038
N1,N12-diacetylspermine	HMDB0002172	Carboximidic acids and derivatives	Carboximidic acids	1.38 (1.02–1.85)	0.034	1.41 (1.03–1.94)	0.032
Phenylalanine	HMDB0000159	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.77 (1.29–2.42)	<0.001	1.76 (1.25–2.48)	0.001
Proline	HMDB0000162	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.52 (1.12–2.07)	0.007	1.59 (1.13–2.22)	0.007
Isoleucine	HMDB0000172	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.55 (1.15–2.08)	0.004	1.56 (1.12–2.17)	0.009
Leucine	HMDB0000687	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.50 (1.12–2.02)	0.007	1.48 (1.06–2.06)	0.02
N-alpha-acetylarginine	HMDB0004620[c]	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.39 (1.03–1.89)	0.033	1.45 (1.06–2.00)	0.022
Serine	HMDB0000187	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.35 (0.99–1.85)	0.058	1.46 (1.05–2.02)	0.023
N-acetylornithine	HMDB0003357	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	0.78 (0.58–1.03)	0.081	0.71 (0.53–0.96)	0.026
Betaine	HMDB0000043	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.27 (0.91–1.79)	0.164	1.47 (1.03–2.12)	0.035
C5-DC carnitine	HMDB0013130	Fatty acyls	Fatty acid esters	0.67 (0.5–0.91)	0.010	0.71 (0.52–0.97)	0.030
Myristoleic acid	HMDB0002000	Fatty acyls	Fatty acids and conjugates	1.5 (1.07–2.1)	0.018	1.58 (1.11–2.24)	0.012
C58:7 TAG	HMDB0005471[c]	Glycerolipids	TAGs	0.60 (0.44–0.82)	0.001	0.59 (0.42–0.82)	0.002
C56:9 TAG	HMDB0005448[c]	Glycerolipids	TAGs	0.68 (0.51–0.91)	0.010	0.64 (0.46–0.87)	0.004
C56:10 TAG	HMDB0010513[c]	Glycerolipids	TAGs	0.69 (0.52–0.93)	0.013	0.63 (0.46–0.86)	0.004
C54:9 TAG	HMDB0010498[c]	Glycerolipids	TAGs	0.70 (0.52–0.94)	0.017	0.64 (0.47–0.87)	0.005
C54:8 TAG	HMDB0010518[c]	Glycerolipids	TAGs	0.70 (0.52–0.94)	0.017	0.65 (0.47–0.88)	0.006
C58:11 TAG	HMDB0010531[c]	Glycerolipids	TAGs	0.70 (0.52–0.94)	0.017	0.64 (0.47–0.88)	0.006
C56:8 TAG	HMDB0005392[c]	Glycerolipids	TAGs	0.69 (0.51–0.93)	0.015	0.66 (0.48–0.90)	0.008
C58:9 TAG	HMDB0005463[c]	Glycerolipids	TAGs	0.67 (0.50–0.91)	0.010	0.66 (0.47–0.90)	0.010
C58:10 TAG	HMDB0005476[c]	Glycerolipids	TAGs	0.69 (0.51–0.93)	0.014	0.66 (0.48–0.90)	0.010
C56:7 TAG	HMDB0005462[c]	Glycerolipids	TAGs	0.74 (0.55–0.99)	0.044	0.68 (0.50–0.94)	0.017
C58:6 TAG	HMDB0005458[c]	Glycerolipids	TAGs	0.68 (0.49–0.92)	0.013	0.68 (0.49–0.94)	0.018
C52:7 TAG	HMDB0010517[c]	Glycerolipids	TAGs	0.76 (0.57–1.02)	0.065	0.69 (0.50–0.94)	0.019
C54:7 TAG	HMDB0005447[c]	Glycerolipids	TAGs	0.73 (0.54–0.97)	0.031	0.70 (0.51–0.95)	0.021

(Continued on the following page)

Table 2. ORs for breast cancer risk comparing 90th to 10th percentiles of selected metabolites[a] measured at distant or proximate blood. (Cont'd)

				Unadjusted		Multivariate adjusted[b]	
Metabolite name	**HMDB ID**	**Class**	**Subclass**	**OR (95% CI)**	***P* value**	**OR (95% CI)**	***P* value**
C60:12 TAG	HMDB0005478[c]	Glycerolipids	TAGs	0.76 (0.56–1.02)	0.071	0.71 (0.51–0.98)	0.035
C52:6 TAG	HMDB0005436[c]	Glycerolipids	TAGs	0.79 (0.59–1.06)	0.114	0.73 (0.53–0.99)	0.046
C18:3 LPC	HMDB0010387[c]	Glycerophospholipids	Glycerophosphocholines	1.40 (1.04–1.90)	0.026	1.40 (1.02–1.93)	0.035
C16:1 LPC	HMDB0010383[c]	Glycerophospholipids	Glycerophosphocholines	1.39 (1.04–1.87)	0.028	1.39 (1.02–1.89)	0.038
C16:0 LPC	HMDB0010382	Glycerophospholipids	Glycerophosphocholines	1.40 (1.04–1.89)	0.026	1.38 (1.01–1.89)	0.042
C18:1 LPC	HMDB0002815[c]	Glycerophospholipids	Glycerophosphocholines	1.32 (0.97–1.79)	0.072	1.39 (1.01–1.93)	0.046
C38:6 PE	HMDB0009102[c]	Glycerophospholipids	Glycerophosphoethanolamines	0.76 (0.55–1.05)	0.091	0.69 (0.49–0.97)	0.035
Tryptophan	HMDB0000929	Indoles and derivatives	Indolyl carboxylic acids and derivatives	1.39 (1.04–1.87)	0.028	1.40 (1.03–1.9)	0.030
C16:0 Ceramide (d18:1)	HMDB0004949	Sphingolipids	Ceramides	1.62 (1.18–2.22)	0.003	1.72 (1.23–2.40)	0.002
C24:1 Ceramide (d18:1)	HMDB0004953[c]	Sphingolipids	Ceramides	1.46 (1.08–1.98)	0.014	1.42 (1.04–1.94)	0.028
C22:0 Ceramide (d18:1)	HMDB0004952	Sphingolipids	Ceramides	1.43 (1.06–1.94)	0.020	1.39 (1.01–1.92)	0.044

[a]Selected metabolites are those with nominal *P* value <0.05 in fully adjusted models among metabolites with <10% missingness. Missing values were imputed with 1/2 the minimum value. Results sorted by class, subclass, and *P* value for fully adjusted model. Significant *P* value with NEF adjustment: distant blood *P* value = 0.0003, proximate blood *P* value = 0.0002.
[b]Multivariate conditional logistic regression model adjusted for BMI at age 18, weight change since age 18, age at menarche, combined age at first birth and parity, breastfeeding history, history of benign breast disease, family history of breast cancer, alcohol use (g/day), and activity level (MET-hours/week). *P* values are nominal *P* values before correction for multiple testing.
[c]Representative HMD ID.
[d]No HMD ID.

subclass, most had a majority of one subclass, or a split between two subclass distinctions.

At the distant timepoint, no modules were significantly associated with overall breast cancer risk (Supplementary Table S7A). One module, defined by several glycerophospholipids and TAGs with high numbers of double bonds, was suggestively inversely associated with ER+ breast cancer (M7 OR = 0.66; 95% CI, 0.49–0.89; nominal *P* value = 0.01; FDR adjusted *P* value = 0.08). TAGs with high numbers of double bonds were negatively weighted in this module, whereas glycerophospholipids were mainly positively weighted (Supplementary Fig. S3A). This finding, with higher glycerophospholipids and lower TAGs with ≥3 double bonds, corresponds with MSEA results (**Fig. 2**). Although glycerophospholipids were not significantly associated with ER+ breast cancer in MSEA, our results from the WGCNA highlight the importance of a few key glycerophospholipids including C20:4 LPC (OR comparing 90th to 10th percentile = 0.66; 95% CI, 0.50–0.88; *P* = 0.004) and C18:2 LPC (OR = 0.64; 95% CI, 0.47–0.87; *P* = 0.005; Supplementary Table S3A). At the proximate timepoint, no modules were associated with breast cancer risk (Supplementary Table S7B). Despite this, associative patterns that arose in module groupings aligned with MSEA results (Supplementary Fig. S3B).

Metabolites with the most significant difference measures between blood draws included TAGs with ≥3 double bonds (**Table 4**). An increase in TAGs with ≥3 double bonds from distant to proximate measures was associated with a reduced breast cancer risk (e.g., for C56:10 TAG OR 90th–10th percentile = 0.62; 95% CI, 0.43–0.88; nominal *P* value = 0.007).

The majority of metabolites were moderately correlated between timepoints (Spearman correlation = 0.40–0.50; Supplementary Table S8; Supplementary Fig. S4). Analysis of metabolite associations with breast cancer risk taking the associations of averaged metabolites from both timepoints with breast cancer risk generally were similar to individual timepoint analyses. However, some associations were weakened due to opposing associations, whereas others were strengthened by consistent associations (Supplementary Table S9). MSEA analysis of average values highlighted the strong inverse association seen for cholesteryl esters (Supplementary Fig. S5), and the strong positive association seen for TAGS with <3 double bonds. TAGS with ≥3 double bonds showed strong inverse associations with overall and ER− BC on average, but null associations with ER+ breast cancer on average, due to opposing directions of association at distant and proximate bloods.

Discussion

In this nested case–control study examining the association between 307 plasma metabolites and breast cancer risk, we identified several metabolite groups, defined on the basis of similar biochemical structure, that were associated with risk. Individual metabolites did not reach statistical significance with correction for multiple comparisons; however, common patterns appeared for structurally similar metabolites. By subclass, cholesteryl esters were inversely associated with breast cancer risk, whereas amino acids and derivatives were associated with increased risk. The association between TAGs and breast cancer risk was dependent on the number of double bonds; TAGs with ≥3 double bonds were inversely associated, whereas TAGs with <3 double bonds were positively associated with risk. The unique ability to assess metabolite measures at two different timepoints also highlighted the potential for metabolites to influence different stages of breast cancer development, as several associations differed by time between blood draw and diagnosis.

Table 3. ORs for breast cancer risk comparing 90th to 10th percentiles of selected metabolite levels[a] measured at distant or proximate blood, by ER status of case.

Metabolite name	HMDB ID	Class	Subclass	ER+ (*N* = 585)		ER− (*N* = 91)	
				OR (95% CI)[b]	*P* value	OR (95% CI)[b]	*P* value
Distant blood							
Hippurate	HMDB0000714	Benzene and substituted derivatives	Benzoic acids and derivatives	0.67 (0.50–0.90)	0.007	1.02 (0.57–1.82)	0.952
N-alpha-acetylarginine	HMDB0004620[c]	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	0.70 (0.53–0.94)	0.016	0.69 (0.39–1.23)	0.214
Citrulline	HMDB0000904	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	0.74 (0.55–0.99)	0.045	0.82 (0.46–1.46)	0.496
C5:1 carnitine	HMDB0002366	Fatty acyls	Fatty acid esters	0.59 (0.44–0.79)	<0.001	1.12 (0.61–2.06)	0.715
C3 carnitine	HMDB0000824	Fatty acyls	Fatty acid esters	0.72 (0.54–0.97)	0.029	1.16 (0.64–2.10)	0.621
C4 carnitine	HMDB0002013	Fatty acyls	Fatty acid esters	0.69 (0.52–0.92)	0.012	0.91 (0.50–1.64)	0.748
C5-DC carnitine	HMDB0013130	Fatty acyls	Fatty acid esters	0.70 (0.53–0.93)	0.015	0.94 (0.52–1.68)	0.823
C34:2 DAG	HMDB0007103[c]	Fatty acyls	Lineolic acids and derivatives	1.50 (1.12–2.02)	0.007	1.49 (0.81–2.78)	0.203
C36:3 DAG	HMDB0007219[c]	Fatty acyls	Lineolic acids and derivatives	1.35 (1.01–1.80)	0.040	1.29 (0.70–2.39)	0.419
C32:0 DAG	HMDB0007098[c]	Glycerolipids	Diacylglycerols	1.44 (1.06–1.94)	0.018	1.49 (0.80–2.77)	0.209
C34:1 DAG	HMDB0007102[c]	Glycerolipids	Diacylglycerols	1.38 (1.03–1.87)	0.034	1.41 (0.75–2.66)	0.284
C52:4 TAG	HMDB0005363[c]	Glycerolipids	TAGs	1.43 (1.07–1.90)	0.015	1.22 (0.67–2.25)	0.518
C50:2 TAG	HMDB0005377[c]	Glycerolipids	TAGs	1.44 (1.06–1.96)	0.021	1.26 (0.67–2.38)	0.476
C50:1 TAG	HMDB0005360[c]	Glycerolipids	TAGs	1.42 (1.04–1.93)	0.027	1.24 (0.66–2.33)	0.508
C52:2 TAG	HMDB0005369[c]	Glycerolipids	TAGs	1.38 (1.02–1.87)	0.035	1.25 (0.67–2.36)	0.479
C50:3 TAG	HMDB0005433[c]	Glycerolipids	TAGs	1.38 (1.02–1.88)	0.035	1.45 (0.78–2.73)	0.245
C51:1 TAG	HMDB0042104[c]	Glycerolipids	TAGs	1.38 (1.02–1.86)	0.036	1.46 (0.79–2.68)	0.227
C43:2 TAG	HMDB0043169[c]	Glycerolipids	TAGs	1.37 (1.01–1.85)	0.041	1.46 (0.80–2.67)	0.220
C55:2 TAG	HMDB0042226[c]	Glycerolipids	TAGs	1.35 (1.00–1.82)	0.047	1.18 (0.64–2.19)	0.591
C22:5 LPC	HMDB0010403[c]	Glycerophospholipids	Glycerophosphocholines	0.58 (0.43–0.77)	<0.001	0.71 (0.39–1.27)	0.248
C18:2 LPC	HMDB0010386[c]	Glycerophospholipids	Glycerophosphocholines	0.64 (0.47–0.87)	0.005	0.78 (0.41–1.49)	0.453
C20:5 LPC	HMDB0010397	Glycerophospholipids	Glycerophosphocholines	0.65 (0.48–0.88)	0.006	0.82 (0.43–1.54)	0.535
C18:1 LPC	HMDB0002815[c]	Glycerophospholipids	Glycerophosphocholines	0.68 (0.51–0.91)	0.011	0.96 (0.52–1.74)	0.889
C18:0 LPC	HMDB0010384	Glycerophospholipids	Glycerophosphocholines	0.69 (0.51–0.93)	0.014	1.09 (0.58–2.03)	0.789
C36:5 PC plasmalogen-B	HMDB0011220[c]	Glycerophospholipids	Glycerophosphocholines	0.75 (0.57–0.99)	0.043	0.84 (0.47–1.47)	0.534
C22:0 LPE	HMDB0011520	Glycerophospholipids	Glycerophosphoethanolamines	0.65 (0.48–0.88)	0.005	0.96 (0.51–1.80)	0.900
C38:6 PE plasmalogen	HMDB0011387[c]	Glycerophospholipids	Glycerophosphoethanolamines	0.72 (0.54–0.95)	0.022	0.72 (0.41–1.27)	0.254
C36:5 PE plasmalogen	HMDB0011410[c]	Glycerophospholipids	Glycerophosphoethanolamines	0.73 (0.55–0.97)	0.028	0.84 (0.48–1.47)	0.542
Serotonin	HMDB0000259	Indoles and derivatives	Tryptamines and derivatives	1.37 (1.04–1.81)	0.025	1.03 (0.59–1.80)	0.925
C20:4 LPC	HMDB0010395	NA	NA	0.66 (0.50–0.88)	0.004	0.94 (0.52–1.67)	0.823
Thyroxine	HMDB0000248	NA	NA	1.50 (1.11–2.04)	0.009	2.16 (1.15–4.13)	0.018
C20:1 LPE	HMDB0011512[c]	NA	NA	0.69 (0.52–0.92)	0.011	0.71 (0.39–1.30)	0.270
Trigonelline	HMDB0000875	NA	NA	0.71 (0.53–0.95)	0.021	0.66 (0.37–1.18)	0.161
Carnitine	HMDB0000062	Organonitrogen compounds	Quaternary ammonium salts	0.73 (0.54–0.99)	0.041	1.12 (0.60–2.07)	0.729
Acetyl-galactosamine	HMDB0000212	Organooxygen compounds	Carbohydrates and carbohydrate conjugates	1.26 (0.95–1.69)	0.110	1.85 (1.04–3.33)	0.037
2-Methylguanosine	HMDB0005862	Purine nucleosides	NA	1.34 (1.00–1.79)	0.054	2.02 (1.13–3.64)	0.019
C22:5 CE	HMDB0010375[c]	Steroids and steroid derivatives	Cholesterol esters	0.52 (0.39–0.70)	<0.001	0.65 (0.35–1.18)	0.160
C20:5 CE	HMDB0006731	Steroids and steroid derivatives	Cholesterol esters	0.61 (0.46–0.82)	0.001	0.60 (0.33–1.10)	0.103
C18:3 CE	HMDB0010370[c]	Steroids and steroid derivatives	Cholesterol esters	0.65 (0.49–0.86)	0.003	0.55 (0.30–1.01)	0.054
C20:4 CE	HMDB0006726	Steroids and steroid derivatives	Cholesterol esters	0.67 (0.50–0.89)	0.006	0.77 (0.42–1.38)	0.374
C18:0 CE	HMDB0010368	Steroids and steroid derivatives	Cholesterol esters	0.68 (0.51–0.91)	0.010	0.99 (0.53–1.83)	0.965
C20:3 CE	HMDB0006736[c]	Steroids and steroid derivatives	Cholesterol esters	0.73 (0.55–0.96)	0.026	0.69 (0.38–1.25)	0.225
C18:1 CE	HMDB0000918[c]	Steroids and steroid derivatives	Cholesterol esters	0.72 (0.54–0.97)	0.030	0.77 (0.41–1.42)	0.404

(Continued on the following page)

Table 3. ORs for breast cancer risk comparing 90th to 10th percentiles of selected metabolite levels[a] measured at distant or proximate blood, by ER status of case. (Cont'd)

Metabolite name	HMDB ID	Class	Subclass	ER+ (*N* = 585)		ER− (*N* = 91)	
				OR (95% CI)[b]	*P* value	OR (95% CI)[b]	*P* value
Proximate blood							
Hippurate	HMDB0000714	Benzene and substituted derivatives	Benzoic acids and derivatives	0.64 (0.45–0.91)	0.014	0.52 (0.26–1.03)	0.062
N1,N12-diacetylspermine	HMDB0002172	Carboximidic acids and derivatives	Carboximidic acids	1.46 (1.02–2.08)	0.038	2.33 (1.14–4.81)	0.020
Proline	HMDB0000162	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.52 (1.04–2.21)	0.029	0.89 (0.41–1.91)	0.760
Phenylalanine	HMDB0000159	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.48 (1.03–2.14)	0.035	1.67 (0.80–3.50)	0.171
Isoleucine	HMDB0000172	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.47 (1.00–2.16)	0.049	1.31 (0.62–2.79)	0.487
C5-DC carnitine	HMDB0013130	Fatty acyls	Fatty acid esters	0.60 (0.42–0.86)	0.004	1.46 (0.72–2.97)	0.289
C52:0 TAG	HMDB0005365[c]	Glycerolipids	TAGs	1.49 (1.04–2.15)	0.030	0.86 (0.42–1.74)	0.668
C54:9 TAG	HMDB0010498[c]	Glycerolipids	TAGs	0.68 (0.48–0.98)	0.037	0.60 (0.29–1.23)	0.160
C58:11 TAG	HMDB0010531[c]	Glycerolipids	TAGs	0.69 (0.48–0.98)	0.038	0.61 (0.29–1.27)	0.189
C58:9 TAG	HMDB0005463[c]	Glycerolipids	TAGs	0.69 (0.48–0.99)	0.043	0.70 (0.34–1.46)	0.345
C58:7 TAG	HMDB0005471[c]	Glycerolipids	TAGs	0.68 (0.47–0.99)	0.044	0.55 (0.27–1.10)	0.093
C56:10 TAG	HMDB0010513[c]	Glycerolipids	TAGs	0.69 (0.49–0.99)	0.045	0.55 (0.26–1.15)	0.113
C58:10 TAG	HMDB0005476[c]	Glycerolipids	TAGs	0.70 (0.49–0.99)	0.046	0.67 (0.32–1.37)	0.272
C52:1 TAG	HMDB0005367[c]	Glycerolipids	TAGs	1.45 (1.01–2.09)	0.047	0.83 (0.41–1.70)	0.614
Tryptophan	HMDB0000929	Indoles and derivatives	Indolyl carboxylic acids and derivatives	1.56 (1.09–2.23)	0.015	0.98 (0.49–1.97)	0.956
Guanosine	HMDB0000133	Purine nucleosides	NA	1.45 (1.02–2.06)	0.039	0.71 (0.35–1.42)	0.332
C22:0 Ceramide (d18:1)	HMDB0004952	Sphingolipids	Ceramides	1.50 (1.05–2.16)	0.027	1.28 (0.63–2.59)	0.488
C24:1 Ceramide (d18:1)	HMDB0004953[c]	Sphingolipids	Ceramides	1.48 (1.03–2.13)	0.035	1.26 (0.61–2.6)	0.533
C16:0 Ceramide (d18:1)	HMDB0004949	Sphingolipids	Ceramides	1.48 (1.02–2.15)	0.037	1.89 (0.95–3.82)	0.071
C22:5 CE	HMDB0010375[c]	Steroids and steroid derivatives	Cholesteryl esters	0.69 (0.49–0.98)	0.040	1.08 (0.56–2.10)	0.814
Hydroxyproline	HMDB0000725	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	0.77 (0.54–1.10)	0.157	2.03 (1.01–4.15)	0.048
C5:1 carnitine	HMDB0002366	Fatty acyls	Fatty acid esters	0.80 (0.55–1.14)	0.218	2.44 (1.19–5.07)	0.015
C45:0 TAG	HMDB0042093[c]	Glycerolipids	TAGs	1.16 (0.80–1.67)	0.465	0.67 (0.33–1.34)	0.039
C22:0 LPE	HMDB0011520	Glycerophospholipids	Glycerophosphoethanolamines	0.96 (0.66–1.39)	0.818	2.39 (1.16–5.01)	0.018
Deoxyguanosine	HMDB0000085	NA	NA	1.18 (0.83–1.68)	0.346	0.47 (0.23–0.98)	0.043
Methyl N-methylanthra-nilate	HMDB0034169	NA	NA	1.10 (0.78–1.55)	0.580	0.50 (0.25–0.98)	0.045
Kynurenic acid	HMDB0000715	Quinolines and derivatives	Quinoline carboxylic acids	0.97 (0.67–1.41)	0.888	2.14 (1.03–4.5)	0.041

[a]Selected metabolites are those with <10% missingness and a nominal $P < 0.05$ for either ER+ or ER− breast cancers (those identified as significant in ER− breast cancers are in bold). Missing values were imputed with half the minimum value. Results sorted by class, subclass, and *P* value for fully adjusted model. Metabolites in bold represent those chosen as top hits for ER− breast cancer.

[b]Multivariate unconditional logistic regression model adjusted for BMI at age 18, weight change since age 18, age at menarche, combined age at first birth and parity, breastfeeding history, history of benign breast disease, family history of breast cancer, alcohol use (g/day), and activity level (MET-hours/week). *P* values are nominal *P* values before correction for multiple testing.

[c]Representative HMDB ID.

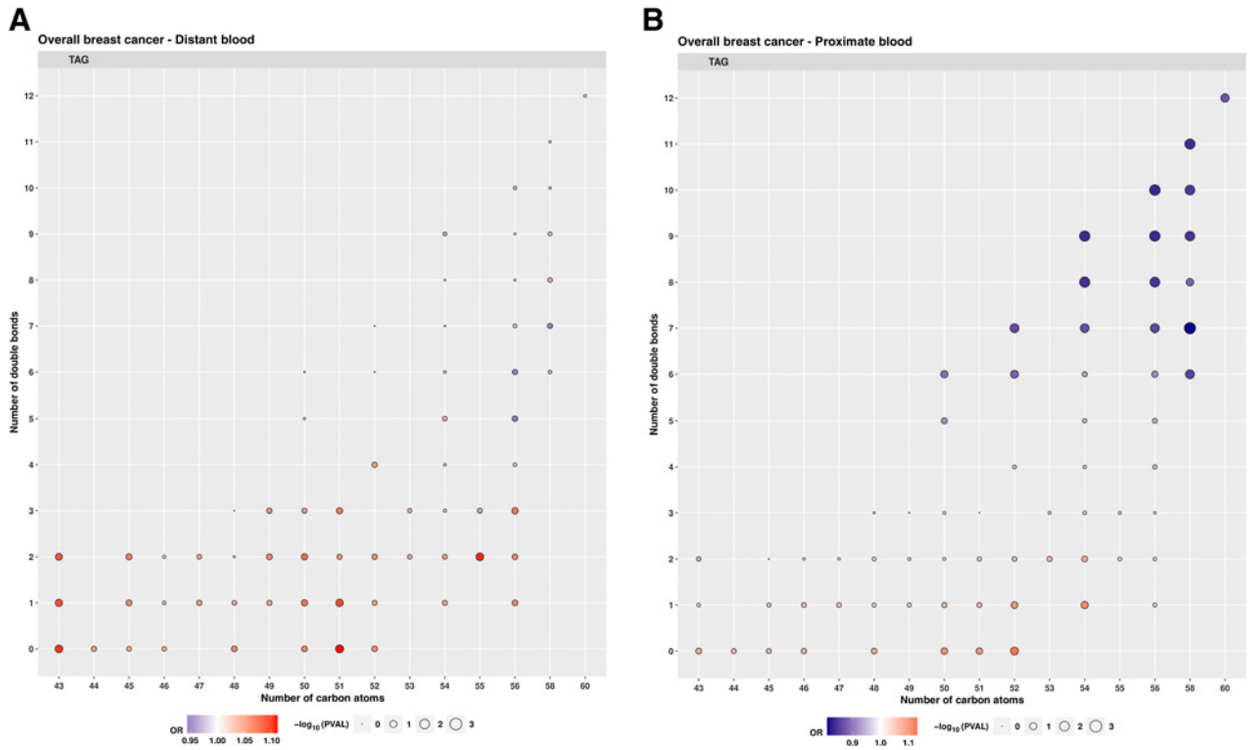

Figure 1.

ORs for breast cancer risk comparing 90th to 10th percentile of TAGs, by number of carbon atoms and double bonds at (**A**) distant and (**B**) proximate blood. Models are for CLR, fully adjusted for BMI at age 18, weight change since age 18, age at menarche, combined age at first birth and parity, breastfeeding history, history of benign breast disease, family history of breast cancer, alcohol use (g/day), and activity level (MET-hours/week). *Y* axis is number of double bonds, *X* axis is number of carbon atoms. Protective associations are shown in blue; harmful associations are shown in red.

Our results add novel knowledge and provide support for several findings from other similar agnostic metabolomic approaches. For example, a nested case–control study in EPIC (with 1,624 cases), examined 127 metabolites in prediagnostic blood samples (9). Although individual metabolite results were not consistent with our study, associations by metabolite classes were similar. For example, C2 carnitine was inversely associated with breast cancer risk in EPIC but not in our study, although we observed an inverse association between high levels of carnitines and risk in general, with C5-DC carnitine appearing inversely related to breast cancer with a nominal *P* value <0.05 at both timepoints. This finding is also reflected in a recent nested case–control study in Cancer Prevention Study 2 (CPS 2, $n =$ 782 cases), with 1,275 metabolites (3), where acyl fatty acid derivatives of carnitine were inversely associated with risk. Carnitine deficiency is associated with increased insulin sensitivity (20), suggesting that the inverse association between carnitines and breast cancer may be due to insulin-dependent signaling pathways. In fact, carnitine supplementation has been shown to improve glucose homeostasis (20). Among BMI-associated metabolites in the prostate, lung, colorectal, and ovarian cancer screening (PLCO) cohort, acylcarnitines 3-methylglutarylcarnitine and 2-methylbutyrylcarnitine were associated with increased breast cancer risk (2), contrasting with the finding of the agnostic analysis within CPS 2 (3). Although we generally observed inverse associations with breast cancer risk for carnitines and derivatives, a few carnitines were suggestively positively associated with risk, including C14 carnitine. Higher levels of acylcarnitines have been associated with increased meat consumption, and higher blood concentrations may be indicative of changes in mitochondrial function and β-oxidation (21, 22); thus, the positive association with breast cancer seen here may represent breakdown products of animal sources of protein. In addition, accumulation of long chain (C14–C20) acylcarnitines has been associated with decreased insulin sensitivity (23, 24).

Phosphatidylcholines were inversely associated with risk in EPIC; we found nominally significant inverse association between several glycerophosphocholines (derivatives of phosphatidylcholines) and risk, especially for ER+ breast cancer. Dietary choline intake from glycerophosphocholines was inversely associated with breast cancer risk in a Chinese case–control study (25). In an earlier study within the EPIC-Heidelberg cohort, higher levels of lysophosphatidylcholines (lysoPC) were associated with decreased breast cancer risk (6), which was suggestive, although not statistically significant, in our study. In contrast, a positive association between glycerophosphocholines and breast cancer risk was observed in CPS 2 (3), which was not observed in our study or in EPIC (9). The associations may be dependent on side chains of interest; because each study measured a slightly different set of metabolites, direct comparisons are not possible, although this group may be important in breast cancer development.

Although the Korean Cancer Prevention Study II was of much smaller scale ($N = 84$ cases), amino acid metabolism, fatty acid metabolism, and linoleic acid metabolism differed between cases and controls, similar to some of our findings (8). In pathway analysis, increased breast cancer risk was observed for metabolites involved in phenylalanine, tyrosine, and tryptophan biosynthesis, suggesting that

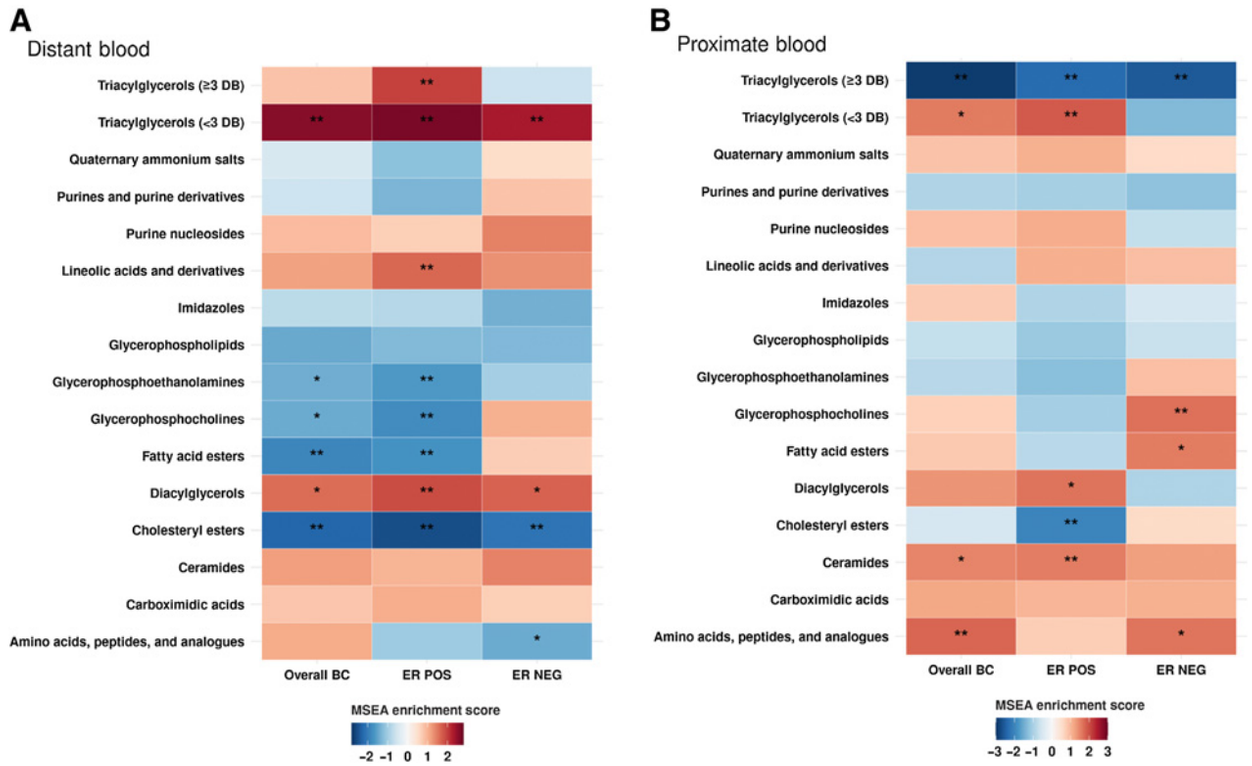

Figure 2.

Gene set enrichment analysis by subclass of metabolites for overall, ER+, and ER− breast cancer, distant (≥10 years before dx) and proximate blood (<10 years before dx). Overall breast cancer results use conditional logistic regression; ER status–specific models use unconditional logistic regression models adjusted for matched factors. Models are fully adjusted for the following: BMI at age 18, weight change from age 18 to blood draw, age at menarche, combined age at first birth and parity, breastfeeding history, history of benign breast disease, family history of breast cancer, alcohol use (g/day), activity level (MET-hours/week). Stars denote *P* values adjusted by FDR: *, $P_{adj} < 0.2$ and **, $P_{adj} < 0.05$. Darker blue is a more negative enrichment score; darker red is a more positive enrichment score.

amino acid metabolism may be an important driver in breast cancer development. Lower circulating levels of amino acid were apparent in cases at diagnosis compared to controls (26, 27), suggesting that the tumor-specific metabolic reprogramming focuses on amino acids. Perhaps a high level of amino acids many years prior to diagnosis provides a hospitable environment for tumor cells that will later use these amino acids to drive their formation. In our analysis, phenylalanine was one of the strongest hits at both the distant and proximate timepoints. The importance of this metabolite may be further highlighted by the need for cancer cells to uptake phenylalanine to survive; in fact, a recent study used nanoparticles coated with phenylalanine to target and cause cancer cells to self-destruct (28). Proline also appeared nominally significant in our analyses; this amino acid plays a key role in metabolic reprogramming important for cancer cell survival (29). Further supporting our amino acid findings, plasma levels of amino acids including valine, lysine, arginine, glutamine were associated with increased breast cancer risk in SU.VI.MAX cohort (5), which used untargeted NMR metabolomic profiles ($N = 206$ cases). In contrast to our current findings and those reported separately (14), in the Women's Health Study, branched-chain amino acids (BCAA) valine, leucine, and isoleucine were not associated with breast cancer risk.

Uniquely, in our study we observed a strong inverse association between cholesteryl esters and breast cancer risk. Cholesteryl esters form the components of cholesterol, high density lipoprotein (HDL) and low-density lipoprotein (LDL), levels of which have been associated with breast cancer risk (30), although epidemiologic evidence for these associations remain inconsistent (31). In addition, laboratory studies and *in vivo* studies suggest cholesterol metabolism as a driver for breast cancer tumor growth (32). The role of cholesteryl esters in lipid metabolism and transport is of interest, as lipid metabolic reprogramming occurs in cancer cells (33). Our finding of an inverse association of cholesteryl esters with breast cancer risk was more notable at the distant timepoint. Although the biologic basis for this difference over time is unclear, women with more prominent cholesteryl ester profiles earlier in life may be more likely to benefit from their protective effect, making cholesterol metabolism across the life course an important avenue for further research.

Here we found TAGs, a metabolite subclass not measured in EPIC (9) and CPS2 (3), were significantly associated with breast cancer risk, but in opposite directions depending on the size and number of carbon atom double bonds. Recent studies of diabetes suggest that lipid composition is important in the association between lipids and diabetes, and may reflect insulin activity (34). TAGs with low carbon atoms and low double bonds are associated with insulin resistance and consequently with diabetes, whereas TAGs with high carbon atoms and high double bonds are higher in those with normal insulin function (34). Insulin signaling is a marker for metabolic health, which is a predictor of breast cancer risk (35, 36). Insulin signaling is responsible for activating the MAPK and PI3K/Akt pathways, which promote cancer cell proliferation and invasion (37). As noted above, the role of carnitines in the insulin-signaling pathway may also

Table 4. ORs for breast cancer risk comparing 90th to 10th percentiles in proximate-distant metabolite measures for metabolites with nominal *P* value <0.05.

Metabolite	HMDB ID	Class	Subclass	OR (95% CI)[a]	*P* value	Spearman correlation[b]
N-Alpha-acetylarginine	HMDB0004620[c]	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.86 (1.25–2.78)	0.002	0.655
2-Aminooctanoic acid	HMDB0000991[c]	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	0.67 (0.47–0.93)	0.019	0.458
Aminoisobutyric acid	HMDB0001906[c]	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	0.65 (0.45–0.93)	0.020	0.502
Isoleucine	HMDB0000172	Carboxylic acids and derivatives	Amino acids, peptides, and analogues	1.43 (1.01–2.03)	0.044	0.422
Myristoleic acid	HMDB0002000	Fatty acyls	Fatty acids and conjugates	1.47 (1.03–2.11)	0.033	0.513
C58:9 TAG	HMDB0005463[c]	Glycerolipids	TAGs	0.60 (0.42–0.85)	0.004	0.475
C56:9 TAG	HMDB0005448[c]	Glycerolipids	TAGs	0.60 (0.42–0.86)	0.005	0.489
C58:11 TAG	HMDB0010531[c]	Glycerolipids	TAGs	0.60 (0.41–0.86)	0.005	0.520
C56:10 TAG	HMDB0010513[c]	Glycerolipids	TAGs	0.62 (0.43–0.88)	0.007	0.493
C56:7 TAG	HMDB0005462[c]	Glycerolipids	TAGs	0.62 (0.43–0.88)	0.008	0.500
C58:7 TAG	HMDB0005471[c]	Glycerolipids	TAGs	0.64 (0.45–0.90)	0.010	0.423
C56:8 TAG	HMDB0005392[c]	Glycerolipids	TAGs	0.64 (0.45–0.90)	0.010	0.426
C54:8 TAG	HMDB0010518[c]	Glycerolipids	TAGs	0.65 (0.46–0.91)	0.013	0.438
C54:9 TAG	HMDB0010498[c]	Glycerolipids	TAGs	0.65 (0.46–0.92)	0.015	0.469
C58:10 TAG	HMDB0005476[c]	Glycerolipids	TAGs	0.64 (0.45–0.92)	0.015	0.488
C60:12 TAG	HMDB0005478[c]	Glycerolipids	TAGs	0.64 (0.45–0.92)	0.017	0.504
C58:8 TAG	HMDB0005413[c]	Glycerolipids	TAGs	0.67 (0.48–0.95)	0.023	0.459
C52:7 TAG	HMDB0010517[c]	Glycerolipids	TAGs	0.68 (0.48–0.97)	0.032	0.473
C58:6 TAG	HMDB0005458[c]	Glycerolipids	TAGs	0.70 (0.49–0.98)	0.037	0.434
C54:7 TAG	HMDB0005447[c]	Glycerolipids	TAGs	0.71 (0.51–0.99)	0.045	0.388
C18:1 LPC	HMDB0002815[c]	Glycerophospholipids	Glycerophosphocholines	1.47 (1.06–2.06)	0.022	0.390
C18:0 LPC	HMDB0010384	Glycerophospholipids	Glycerophosphocholines	1.45 (1.05–2.01)	0.026	0.351
C40:9 PC	HMDB0008731[c]	Glycerophospholipids	Glycerophosphocholines	0.67 (0.46–0.98)	0.039	0.552
C38:6 PC	HMDB0007991[c]	Glycerophospholipids	Glycerophosphocholines	0.67 (0.46–0.98)	0.040	0.547
C18:3 LPC	HMDB0010387[c]	Glycerophospholipids	Glycerophosphocholines	1.38 (1.00–1.90)	0.047	0.301
C38:6 PE	HMDB0009102[c]	Glycerophospholipids	Glycerophosphoethanolamines	0.64 (0.44–0.92)	0.017	0.549
C22:0 LPE	HMDB0011520	Glycerophospholipids	Glycerophosphoethanolamines	1.44 (1.03–2.03)	0.035	0.438
C20:1 LPE	HMDB0011512[c]	Glycerophospholipids	Glycerophosphoethanolamines	1.40 (1.00–1.94)	0.048	0.375
C22:6 LPE	HMDB0011526	Glycerophospholipids	Glycerophosphoethanolamines	0.70 (0.49–1.00)	0.048	0.434
Tryptophan	HMDB0000929	Indoles and derivatives	Indolyl carboxylic acids and derivatives	1.42 (1.02–1.98)	0.039	0.395
Ribothymidine	HMDB0000884	Pyrimidine nucleosides	Pyrimidine nucleosides	1.49 (1.03–2.16)	0.033	0.552
C16:0 Ceramide (d18:1)	HMDB0004949	Sphingolipids	Ceramides	1.57 (1.12–2.20)	0.009	0.417
C22:0 Ceramide (d18:1)	HMDB0004952	Sphingolipids	Ceramides	1.48 (1.04–2.12)	0.030	0.509
C24:1 Ceramide (d18:1)	HMDB0004953[c]	Sphingolipids	Ceramides	1.44 (1.01–2.04)	0.043	0.463
C18:3 CE	HMDB0010370[c]	Steroids and steroid derivatives	Cholesteryl esters	1.53 (1.07–2.19)	0.021	0.495
C18:0 CE	HMDB0010368	Steroids and steroid derivatives	Cholesteryl esters	1.44 (1.04–2.02)	0.030	0.403

[a]All models adjusted for distant blood measure. Estimate is for difference in proximate-distant blood measure. Results are sorted by class, subclass, and *P* value for Model 2. ORs are for unconditional logistic regressions adjusted for BMI at age 18, weight change from 18 to blood draw, age at menarche, combined age at first birth and parity, breastfeeding history, history of benign breast disease, family history of breast cancer, alcohol use (g/day), activity level (MET-hours/week). *P* values are nominal *P* values before correction for multiple testing.
[b]Correlations between proximate and distant time point, adjusted for fasting status and age at blood draw.
[c]Representative HMDB ID.

underlie their associations with breast cancer risk. We also found several ceramides (C16:0, C24:1, C22:0), also potential markers for insulin resistance (38), to be associated with an increased risk of breast cancer. In addition, several TAGS with many double bonds and carbon atoms are associated with higher fish intake (39), indicating a potential protective mechanism of dietary fish intake. More generally, polyunsaturated, omega-3, and omega-6 fatty acids are associated with a higher alternative healthy eating index (AHEI) score (40), further demonstrating potential of dietary intake to influence metabolite levels and future breast cancer risk. In contrast to our findings, lower plasma levels of unsaturated lipids were associated with a higher breast cancer risk in the SU.VI.MAX cohort (5). Further research is needed to fully understand our findings, including why these relationships changed over time.

There are several differences between our study and previous studies. The platforms used for metabolomic profiling differed, which may account for the inconsistencies between studies (3), and actual metabolites measured, constituting various stages of breakdown pathways, differed between studies. Moreover, the timing of blood collections and median time from blood draw to diagnosis differed across studies. This may have contributed to observed differences between studies. Although lag-time between blood draw and diagnosis was explored in the EPIC-Heidelberg cohort (6), median follow-up time was <10 years from blood draw.

There are several strengths of our study. The large sample size allowed adequate power for analyses. We had the ability to control for covariates at the time of metabolite measure, as data were collected for all pertinent covariates every 2 years with follow-up questionnaires, and on blood draw specific questionnaires. Our study assessed how metabolite associations with breast cancer change over time, with samples taken covering a period 0 to 20 years prior to cancer diagnosis.

Although the metabolomic platforms used for profiling were a strength of our study, these also represent a limitation given the inability to directly compare results to those of others. Assessment by ER status was limited by ER− cases. We were unable to investigate premenopausal breast cancer, as most women in NHS were already postmenopausal by the second blood draw.

In conclusion, we found several metabolite subclasses that may be of further interest to explore in breast cancer etiology, including cholesteryl esters, amino acids, and TAGs. Our findings clarified some previous findings, supporting the idea that carnitine metabolism and glycerophosphocholines may be involved in reduction of breast cancer risk. Notably, several metabolite–breast cancer associations we observed may be explained, at least in part, by their role in insulin-signaling pathways. Future studies are needed to determine the intricacies of the biologic mechanisms contributing to breast cancer risk.

Authors' Disclosures

O.A. Zeleznik reports grants from NIH during the conduct of the study. B. Rosner reports grants from NIH during the conduct of the study. R.M. Tamimi reports grants from NIH/NCI during the conduct of the study. A. Eliassen reports grants from NIH during the conduct of the study. No disclosures were reported by the other authors.

Authors' Contributions

K.D. Brantley: Conceptualization, formal analysis, visualization, methodology, writing–original draft. **O.A. Zeleznik:** Formal analysis, supervision, visualization, methodology, writing–review and editing. **B. Rosner:** Conceptualization, supervision, methodology, writing–review and editing. **R.M. Tamimi:** Conceptualization, supervision, methodology, writing–review and editing. **J. Avila-Pacheco:** Resources, data curation. **C.B. Clish:** Resources, data curation, supervision, writing–review and editing. **A.H. Eliassen:** Conceptualization, supervision, funding acquisition, methodology, project administration, writing–review and editing.

Acknowledgments

This study was funded by the NCI. A.H. Eliassen received UM1 CA186107, P01 CA87969, and T32 CA009001 grants from NCI; Susan E. Hankinson (Channing Division of Network Medicine, Brigham and Women's Hospital and University of Massachusetts Amherst, Department of Epidemiology) received R01 CA49449 from NCI. We would like to thank the participants and staff of the Nurses' Health Study for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 27, 2021; revised December 8, 2021; accepted January 10, 2022; published first January 21, 2022.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394–424.
2. Moore SC, Playdon MC, Sampson JN, Hoover RN, Trabert B, Matthews CE, et al. A metabolomics analysis of body mass index and postmenopausal breast cancer risk. J Natl Cancer Inst 2018;110:588–97.
3. Moore SC, Mazzilli KM, Sampson JN, Matthews CE, Carter BD, Playdon MC, et al. A metabolomics analysis of postmenopausal breast cancer risk in the cancer prevention study II. Metabolites 2021;11:95.
4. Tobias DK, Hazra A, Lawler PR, Chandler PD, Chasman DI, Buring JE, et al. Circulating branched-chain amino acids and long-term risk of obesity-related cancers in women. Sci Rep 2020;10:16534.
5. Lécuyer L, Victor Bala A, Deschasaux M, Bouchemal N, Triba MN, Vasson MP, et al. NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer. Int J Epidemiol 2018;47:484–94.
6. Kühn T, Floegel A, Sookthai D, Johnson T, Rolle-Kampczyk U, Otto W, et al. Higher plasma levels of lysophosphatidylcholine 18:0 are related to a lower risk of common cancers in a prospective metabolomics study. BMC Med 2016;14:13.
7. Playdon MC, Ziegler RG, Sampson JN, Stolzenberg-Solomon R, Thompson HJ, Irwin ML, et al. Nutritional metabolomics and breast cancer risk in a prospective study. Am J Clin Nutr 2017;106:637–49.
8. Yoo HJ, Kim M, Kim M, Kang M, Jung KJ, Hwang S, et al. Analysis of metabolites and metabolic pathways in breast cancer in a Korean prospective cohort: the Korean Cancer Prevention Study-II. Metabolomics 2018;14:85.
9. His M, Viallon V, Dossus L, Gicquiau A, Achaintre D, Scalbert A, et al. Prospective analysis of circulating metabolites and breast cancer in EPIC. BMC Med 2019;17:178.
10. Mascanfroni ID, Takenaka MC, Yeste A, Patel B, Wu Y, Kenison JE, et al. Metabolic control of type 1 regulatory T cell differentiation by AHR and HIF1-α. Nat Med 2015;21:638–46.
11. O'Sullivan JF, Morningstar JE, Yang Q, Zheng B, Gao Y, Jeanfavre S, et al. Dimethylguanidino valeric acid is a marker of liver fat and predicts diabetes. J Clin Invest 2017;127:4394–402.
12. Paynter NP, Balasubramanian R, Giulianini WDD, Tinker LF, Gopal S, et al. Metabolic predictors of incident coronary heart disease in women. Circulation 2018;137:841–53.
13. Townsend MK, Clish CB, Kraft P, Wu C, Souza AL, Deik AA, et al. Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. Clin Chem 2013;59:1657–67.

14. Zeleznik OA, Eliassen AH, Kraft P, Poole EM, Rosner BA, Jeanfavre S, et al. A prospective analysis of circulating plasma metabolites associated with ovarian cancer risk. Cancer Res 2020;80:1357–67.
15. Gallois A, Mefford J, Ko A, Vaysse A, Julienne H, Ala-Korpela M, et al. A comprehensive study of metabolite genetics reveals strong pleiotropy and heterogeneity across time and context. Nat Commun 2019;10:4788.
16. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.
17. Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 2006;93:491–507.
18. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinf 2008;9:559.
19. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 2008;24:719–20.
20. Ringseis R, Keller J, Eder K. Role of carnitine in the regulation of glucose homeostasis and insulin sensitivity: evidence from *in vivo* and *in vitro* studies with carnitine supplementation and carnitine deficiency. Eur J Nutr 2012;51:1–18.
21. Rebholz CM, Zheng Z, Grams ME, Appel LJ, Sarnak MJ, Inker LA, et al. Serum metabolites associated with dietary protein intake: results from the modification of diet in renal disease (MDRD) randomized clinical trial. Am J Clin Nutr 2019;109:517–25.
22. Bouchard-Mercier A, Rudkowska I, Lemieux S, Couture P, Vohl M-C. The metabolic signature associated with the Western dietary pattern: a cross-sectional study. Nutr J 2013;12:158.
23. Vilks K, Videja M, Makrecka-Kuka M, Katkevics M, Sevostjanovs E, Grandane A, et al. Long-chain acylcarnitines decrease the phosphorylation of the insulin receptor at Tyr1151 through a PTP1B-dependent mechanism. Int J Mol Sci 2021;22:6470.
24. McCann MR, George De la Rosa MV, Rosania GR, Stringer KA. L-carnitine and acylcarnitines: mitochondrial biomarkers for precision medicine. Metabolites 2021;11:51.
25. Zhang C-X, Pan M-X, Li B, Wang L, Mo X-F, Chen Y-M, et al. Choline and betaine intake is inversely associated with breast cancer risk: a two-stage case-control study in China. Cancer Sci 2013;104:250–8.
26. Yuan B, Schafferer S, Tang Q, Scheffler M, Nees J, Jörg H, et al. A plasma metabolite panel as biomarkers for early primary breast cancer detection. Int J Cancer 2019;144:2833–42.
27. Jové M, Collado R, Quiles JL, Ramírez-Tortosa M-C, Sol J, Ruiz-Sanjuan M, et al. A plasma metabolomic signature discloses human breast cancer. Oncotarget 2017;8:19522–33.
28. Wu Z, Lim HK, Tan SJ, Gautam A, Hou HW, Ng KW, et al. Potent-by-design: amino acids mimicking porous nanotherapeutics with intrinsic anticancer targeting properties. Small 2020;16:e2003757.
29. Burke L, Guterman I, Palacios Gallego R, Britton RG, Burschowsky D, Tufarelli C, et al. The Janus-like role of proline metabolism in cancer. Cell Death Discov 2020;6:104.
30. Pelton K, Coticchia CM, Curatolo AS, Schaffner CP, Zurakowski D, Solomon KR, et al. Hypercholesterolemia induces angiogenesis and accelerates growth of breast tumors in vivo. Am J Pathol 2014;184:2099–110.
31. Cedó L, Reddy ST, Mato E, Blanco-Vaca F, Escolà-Gil JC. HDL and LDL: Potential new players in breast cancer development. J Clin Med 2019;8:853.
32. Danilo C, Frank PG. Cholesterol and breast cancer development. Curr Opin Pharmacol 2012;12:677–82.
33. Beloribi-Djefaflia S, Vasseur S, Guillaumond F. Lipid metabolic reprogramming in cancer cells. Oncogenesis 2016;5:e189.
34. Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. J Clin Invest 2011;121:1402–11.
35. Iyengar NM, Arthur R, Manson JE, Chlebowski RT, Kroenke CH, Peterson L, et al. Association of body fat and risk of breast cancer in postmenopausal women with normal body mass index: a secondary analysis of a Randomized Clinical Trial and Observational Study. JAMA Oncol 2019;5:155–63.
36. Park Y-MM, White AJ, Nichols HB, O'Brien KM, Weinberg CR, Sandler DP. The association between metabolic health, obesity phenotype and the risk of breast cancer. Int J Cancer 2017;140:2657–66.
37. Yee LD, Mortimer JE, Natarajan R, Dietze EC, Seewaldt VL. Metabolic health, insulin, and breast cancer: why oncologists should care about insulin. Front Endocrinol 2020;11:58.
38. Sokolowska E, Blachnio-Zabielska A. The role of ceramides in insulin resistance. Front Endocrinol 2019;10:577.
39. Mazzilli KM, McClain KM, Lipworth L, Playdon MC, Sampson JN, Clish CB, et al. Identification of 102 correlations between serum metabolites and habitual diet in a metabolomics study of the prostate, lung, colorectal, and ovarian cancer trial. J Nutr 2020;150:694–703.
40. Akbaraly T, Würtz P, Singh-Manoux A, Shipley MJ, Haapakoski R, Lehto M, et al. Association of circulating metabolites with healthy diet and risk of cardiovascular disease: analysis of two cohort studies. Sci Rep 2018;8:8620.