



Published in final edited form as:

Reprod Toxicol. 2025 April ; 133: 108866. doi:10.1016/j.reprotox.2025.108866.

Environmental basis for early onset breast cancer

Sami Teeny^a, Zachery R. Jarrell^a, Nickilou Y. Krigbaum^b, Piera M. Cirillo^b, Young-Mi Go^a, Barbara A. Cohn^{b,**}, Dean P. Jones^{a,*}

^aClinical Biomarkers Laboratory, Division of Pulmonary, Allergy, Critical Care, and Sleep Medicine, Department of Medicine, Emory University, Atlanta, GA 30322

^bChild Health and Development Studies, Public Health Institute, Berkeley, CA 94709

Abstract

Pregnancy provokes a heightened amino acid requirement, especially in the third trimester. Alterations to late pregnancy amino acid metabolism have been associated with environmental breast carcinogen exposures, including DDT and PFAS. This project examined whether maternal serum amino acids in late pregnancy are associated with subsequent breast cancer risk. Archival third trimester serum samples from 172 women who were later diagnosed with breast cancer were compared to samples from 351 women without known breast cancer. A prospective metabolome-wide association study (MWAS) for breast cancer cases showed that associated amino acid pathways included lysine, arginine, proline, aspartate, asparagine, alanine, tyrosine, tryptophan, histidine and branched-chain amino acids. Lower mean concentrations of individual amino acids, including histidine, threonine, lysine and proline, were associated with increased risk of breast cancer, and network analyses showed that these amino acids were negatively associated with protective breast cancer risk factors. Prospective MWAS for breast cancer cases diagnosed within 15 years of sample collection showed pathway associations for tryptophan, histidine, lysine and methionine and cysteine pathways. Nutrient stresses caused by low amino acid levels impair immunosurveillance and activate oncogenic mechanisms of cell survival, thereby providing mechanisms by which environmental exposures in late pregnancy can contribute to breast cancer risk.

*Corresponding author at Whitehead Biomedical Research Building, 615 Michael St, Room 205P, Atlanta, GA, 30322. Fax: 404-727-2974. **Corresponding author at 1683 Shattuck Avenue, Suite B, Berkeley, CA 94709. dpjones@emory.edu (D.P. Jones), bcohn@chdstudies.org (B.A. Cohn).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Ethics approval and consent to participation:

The institutional review board of the Public Health Institute approved the present study, and we complied with all federal guidelines governing the use of human participants.

CRedit authorship contribution statement

ST, NK, PC, YMG, BC and DJ contributed to study design. ST, ZJ, NK, PC, YMG, BC and DJ contributed to analysis and interpretation of the data. ST, ZJ, YMG, PC, BC and DJ contributed to writing the manuscript.

Declaration of Competing Interest

The authors declare no actual or potential competing financial interests.

Keywords

Amino acids; High-resolution metabolomics; Nutrient sensing; Pregnancy

1. Introduction

Early-onset breast cancer, defined as breast cancer occurring prior to 50 years old, increased 42% worldwide between 1990 and 2019 [1]. Although improved screening can contribute to earlier diagnosis, available evidence supports the interpretation that increases in early onset cancer risks are largely driven by environmental exposures [2]. For instance, a review of monozygotic twin studies showed that genetic factors accounted for less than 20% of multiple types of cancer (breast, bladder, colorectal, leukemia, lung, ovarian, pancreatic, prostate) [3]. Known exposures of risk include persistent pollutants such as DDT and some perfluorinated alkyl substances (PFAS) [4, 5] which are relatively inert and not readily eliminated from biologic systems, i.e., can have biologic half-life of several years. Biomonitoring studies show that levels of some environmental chemicals are decreasing over time, presumably due to policy changes resulting in decreased exposures, while others are increasing over time, presumably due to new introductions and replacement of discontinued chemicals [6]. Of considerable concern, cancer diagnosis can occur decades after chemical carcinogen exposure, motivating research to identify underlying mechanisms connecting chemical exposures to health outcomes.

Prior plasma metabolomics research showed that the breast carcinogens, DDT and some PFOS, are associated with altered amino acid metabolism in pregnancy [7, 8]. Nutrition research shows that an increased amino acid requirement occurs in late pregnancy [9]. This applies to both essential and non-essential amino acids due to demand for the growing fetus as well as use of amino acids to support increased energy needs. A potentially important implication of associations of breast carcinogens with amino acid metabolism in late pregnancy is that nutrient sensing of amino acids [10] is an important oncogenic mechanism of cell survival in cancer [11, 12]. Amino acids can be a limiting factor in the proliferation and survival of cancer cells, and cancer cells rewire key amino acid pathways, such as proline and aspartate pathways, to meet requirements for energy, protein synthesis, and nucleotide synthesis [8, 9]. Because prior research established association of environmental chemicals in late pregnancy serum with 1) breast cancer outcome and 2) changes in serum amino acids in late pregnancy, the current project was designed to complement these data with 3) a test for association of serum amino acids in late pregnancy with subsequent breast cancer. In principle, these three independent studies would constitute a “meet-in-the-middle” approach in which environmental chemical associated changes in amino acid metabolism are mechanistically linked to breast cancer outcome.

Breast cancer is the most common cancer in woman, with over 2.26 million new cases worldwide in 2020, and a leading cause of early mortality [13]. In the mid-20th century, researchers reported an association of young age (≤ 20 years old) at first pregnancy with lifetime protection against breast cancer [14, 15]. This observation has since been qualified by study of a transient increase in risk of breast cancer after pregnancy, which shows that

the odds ratio 5 years after delivery among women who were 35 year old women at first delivery compared to women without childbirth was 1.26 (95% confidence interval, 1.10 to 1.44) [16]. This transient risk has multiple potential mechanisms, including changes in hormones such as estrogen and growth hormone that promote cell proliferation, immune suppression, and the involution process that returns the breasts to their pre-mammary state [17]. The magnitude and interval of increased risk depend upon a woman's age and parity at the time of her pregnancy [16, 18]. The lifetime protection of pregnancy has been linked to pregnancy-associated changes in breast cells that increase resistance to malignant transformation, specifically development of otherwise undifferentiated cells in the breasts' ductal structures where breast cancer most frequently forms [19], while the transient increase in risk in the short term has been linked to a possible promotion of an existing lesion by pregnancy hormones or to inflammation during involution of ductal tissue postpartum [18]. Results also show that ER/PR+ cancers are more common in menopausal women; whereas the more dangerous HER2+ cancers occur more frequently in pre-menopausal women [20]. Such cancer subtypes are potentially affected by age, parity and other risk factors, and currently there is no consensus regarding underlying molecular mechanisms for pregnancy associations with breast cancer [21].

In prior multigenerational research on maternal environmental chemical exposures, liquid chromatography-high-resolution mass spectrometry of selected archival serum samples of women in the Child Health and Development Studies (CHDS, Public Health Institute, Oakland California) were used with metabolome-wide association study (MWAS) and pathway enrichment analysis to find metabolic associations with concentrations of DDT [8] and perfluorinated alkyl substances (PFAS) [7], chemicals previously shown to influence breast cancer risk in daughters of the women and exposed in utero. Results suggested that altered amino acid levels in pregnancy could contribute to subsequent breast cancer risk, but direct tests for associations of amino acids with subsequent risk are not available. Such associations could have important implications for causal mechanisms in cancer because amino acids are used for protein synthesis and growth and also have critical non-protein functions in nutrient sensing [10], inflammation [22], blood pressure regulation [23, 24] and neurotransmission [25].

To test whether amino acid levels during pregnancy are associated with subsequent breast cancer risk, we focused on a 1960's cohort in which serum estrogens during pregnancy were found to be associated with breast cancer. In that study, the ratio of estriol (E3) to the sum of estrone (E1) and estradiol (E2) in third trimester serum samples was protective from subsequent breast cancer for the 38-year study period [26]. The association was stronger for cancers diagnosed within the first 15 y after pregnancy and for primiparas. Motivated by the association of amino acids with environmental breast carcinogens, in the present research we used archival samples from this cohort and applied an untargeted metabolomics workflow to obtain a test of amino acid pathway associations with breast cancer outcome. We selected top discriminatory features and performed pathway enrichment analysis for associations with amino acid pathways. Metabolic network analyses along with targeted measures of amino acids and available demographic and phenotypic data were then used to evaluate relationships of amino acids to prior measures of estrogens, age, parity, and interval between pregnancy and cancer diagnosis.

2. Materials and methods

2.1. Study population

Archival serum samples were from the Child Health and Development Studies (CHDS, Public Health Institute, Oakland, California). The CHDS recruited women residing in the Oakland, California area who were members of the Kaiser Foundation Health Plan and received obstetric care for pregnancies between 1959 and 1967 [27]. Breast cancer cases were identified by linkage to the California Cancer Registry and the California Vital Status Records for cases diagnosed through 1997 as previously described [28]. Record abstraction is based primarily on pathology reports, and case identification is considered to be >99% complete after a 2-year lag for cancer diagnoses [28]. Cases were defined as women with incident invasive or noninvasive breast cancer diagnosis with available prenatal serum. The median age at diagnosis was 54 years (interquartile range, 13 years). Third trimester samples of 172 women who subsequently developed breast cancer were analyzed along with samples from 351 women from the cohort who did not develop breast cancer [29]. Of women with breast cancer, 34 had cancer diagnosis within 15 years of pregnancy. From the age distributions, approximately half of these women had breast cancer diagnosis before age 50, the usual age used as a cutoff for early-onset breast cancer. The institutional review board of the Public Health Institute approved the present study (IRB# I20-002), and we complied with all federal guidelines governing the use of human participants. Blood samples were collected with informed consent from all subjects. This was an observational study and had no experiments on humans.

2.2. High-resolution metabolomics

Serum samples were randomized into batches of 20 samples for extraction and analysis using liquid chromatography-high resolution mass spectrometry (LC-MS) as described previously [30, 31]. HPLC grade acetonitrile and methanol, LC-MS water and 98% formic acid were obtained from Sigma-Aldrich (St. Louis, MO). A mixture of 14 stable isotopic chemicals were used as internal standards [32], including [$^{13}\text{C}_6$]-D-glucose, [^{15}N]-indole, [2- ^{15}N]-L-lysine dihydrochloride, [$^{13}\text{C}_5$]-L-glutamic acid, [$^{13}\text{C}_7$]-benzoic acid, [3,4- $^{13}\text{C}_2$]-cholesterol, [^{15}N]-L-tyrosine, [trimethyl- $^{13}\text{C}_3$]-caffeine, [$^{15}\text{N}_2$]-uracil, [3,3- $^{13}\text{C}_2$]-cystine, [1,2- $^{13}\text{C}_2$]-palmitic acid, [^{15}N , $^{13}\text{C}_5$]-L-methionine, [^{15}N]-choline chloride and 2'-deoxyguanosine- $^{15}\text{N}_2$, $^{13}\text{C}_{10}$ -5'-monophosphate from Cambridge Isotope Laboratories, Inc (Andover, PA).

For processing, 50 μL serum sample was treated 2:1 (v/v) with acetonitrile, and 2.5 μL internal standard of a mixture of the 14 stable isotope standards. Following mixing, extracts were incubated at 4°C for 30 min, and protein precipitates were removed by centrifugation for 10 min at 21,000 $\times$ g at 4°C. Supernatants were placed in autosampler vials and maintained at 4°C in an autosampler prior to analysis. A pooled human reference material [Qstd, pooled plasma from 2 lots (Equitech-Bio, Inc, Kerrville, Texas)] was analyzed at the beginning and end of each batch of 20 samples. A second pooled human reference material (NIST SRM1950) was analyzed at the beginning and end of the study. Samples were analyzed in triplicate with Q-Exactive HF mass spectrometer (Thermo Fisher) using two complementary analytic platforms: reversed phase C18 chromatography and negative

electrospray ionization (ESI) to measure more lipophilic compounds and hydrophilic interaction chromatography (HILIC) with positive ESI to measure more polar metabolites. Data collection occurred continuously throughout 5 min of chromatographic separation from m/z 85 to 1,275.

Peak detection, noise removal, and alignment of retention times and features were performed using the adaptive processing for LC-MS data (apLCMS) v.6.6.8 R package [29–31, 33]. Additional filtering of low-quality samples and features and batch correction were conducted with xMSanalyzer v.1.3.2 [34]. Mass spectral features for each sample consisted of mass to charge (m/z), retention time (rt), and associated integrated ion intensity; interpretations are limited to features with annotation or identification as metabolites and are therefore termed metabolites. Only samples with Pearson correlation $\rho > 0.7$ among 3 replicates were retained. Metabolic features were filtered to retain those with median coefficient of variation among technical replicates $> 70\%$. Intensities below the detection threshold were not retained by the extraction procedures, and detected intensities were median summarized across triplicates for statistical analyses. The final data included 9,042 metabolic features from C18 chromatography and 10,324 from HILIC.

Identities of amino acids and other selected metabolites were confirmed by accurate m/z match, MS/MS spectra, and co-elution with authentic standards, which is Level 1 identification by criteria of Schymanski et al [35]. Other features were annotated with a multistage clustering algorithm, xMSannotator [36] using the Human Metabolome Database [37] at 5 ppm tolerance. Annotations by xMSannotator in combination with *Mummichog* pathway enrichment analyses are considered equivalent to Level 3 identification criteria of Schymanski et al. [35].

2.3. Bioinformatics and biostatistics

To test for amino acid pathway associations with subsequent breast cancer diagnosis, we used log2 transformed feature intensities to decrease heteroscedasticity and a two-step metabolome-wide association study (MWAS) design. This approach has been previously used to balance Type 1 (false positives) and Type 2 (false negatives) statistical error [38] and was considered important in the present study due to the expectation that the population may not be in an active disease state at time of blood collection, and effect sizes may be small. The first step in the MWAS was to use the ANOVA-like program *Limma* to select metabolites differing at raw $p < 0.05$ (protecting against Type 2 error) [39]. This was performed using an in-house xmsPANDA R package with the R v.4.0.3. The Benjamini-Hochberg procedure was used to test for false discovery rate (FDR) [40]. Selected features at raw $p < 0.05$ were used for the second step in which pathway enrichment analysis identified metabolic pathways associated with breast cancer. This analysis uses permutation testing and protects against Type 1 statistical error in identification of pathways associated with breast cancer outcome. In these analyses, metabolites for both columns were analyzed together using *Mummichog* version 2 (MetaboAnalyst, www.metaboanalyst.ca) [41]. Partial Least Squares-Discriminant Analysis (PLS-DA) was used to identify features that best distinguished the control and breast cancer groups. Variable Importance in Projection (VIP) values from the PLS-DA were used as selection criteria for network analysis.

To investigate relationships of amino acids with other metabolites associated with breast cancer risk, we used a data-dependent community detection tool, xMWAS [39], with all metabolites selected at $p < 0.05$ and $VIP > 1.5$. This tool enables visualization of networks of communities, which are groupings of network members based on a distance metric, for which we used PLS correlation with a threshold of $r > 0.4$. xMWAS includes an Eigenvector Centrality Score (ECS) which provides a quantitative measure of centrality in network structures. For this analysis, in addition to breast cancer status (BC) and age, we included breast cancer diagnosis within 15 y (BC15) and other factors previously associated with breast cancer risk, i.e., primiparous status, >3 live births, and log-transformed estrogen concentrations for E1, E2, E3, and the ratio of $E3/(E1+E2)$. We also plotted Pearson correlation matrices of amino acids with hierarchical clustering using the heatmap library in R [31]; to simplify the interpretation of the heat map, amino acid adducts were only included if they possessed at least one correlation with another network node where $|r| > 0.4$.

3. Results

3.1. Characteristics of the study population

Characteristics of the subjects studied are summarized in Table 1 (detailed data are available in [26]). Briefly, breast cancer cases were approximately 4 years older than non-cases, and women who were diagnosed with breast cancer within 15 years of blood sampling were about 9 years older than non-cases and 5 years older than all cases. Non-cases had differences in parity compared to cases and cases diagnosed within 15 years of sampling. Most of the women were white, with 16 to 21% African American and $<10\%$ Hispanic, Asian or Mixed race.

3.2. Pregnancy serum metabolite associations with subsequent breast cancer

Limma showed that 275 metabolites in C18 analyses and 261 metabolites in HILIC analyses differed at raw $p < 0.05$ between women who subsequently were diagnosed with breast cancer and those without known breast cancer (Fig 1A; Supplement 1). For both analyses, more metabolites were negatively associated with breast cancer than were positively associated. No metabolites were significant after FDR correction, indicating that none is likely to be useful individually as a biomarker of risk.

3.3. Pregnancy serum pathway enrichment analysis of associations with breast cancer

Pathway enrichment analysis of the metabolites that differed between women with breast cancer diagnosis and women without breast cancer diagnosis showed statistically significant enrichment in nine amino acid pathways (tyrosine; histidine; arginine and proline; alanine and aspartate; aspartate and asparagine; lysine; methionine and cysteine; valine, leucine and isoleucine; and tryptophan) and the urea cycle (Fig. 1B) along with other pathways (Supplement 2). Identified amino acid and amino acid metabolites included proline, histidine, urocanate, lysine, threonine, phenylalanine, phenylpyruvate, 5-oxoproline (pyroglutamate). Additional supportive evidence for amino acid pathway identifications was obtained from correlation analyses (Supplement 3). Although no individual amino acids could be confidently associated with breast cancer risk after FDR adjustment, the results showed that multiple amino acid pathways are associated with risk.

3.4. Metabolites separating non-cancer and breast cancer cases

We used PLS-DA to test whether metabolites that differed at raw $p < 0.05$ separated individuals with subsequent breast cancer (BC) from those without known breast cancer (NonBC). Partial but incomplete separation was observed for both C18 analyses (Fig. 1C) and HILIC (not shown). Loadings from the PLS-DA were used to select metabolites contributing to separation, and 136 metabolites from C18 data had $VIP > 1$ (Fig. 1D, Supplement 4). Of these, 33 were positively and 103 negatively associated with breast cancer. Analyses for HILIC showed that 128 metabolites contributed to separation, 33 positively and 95 negatively (Supplement 4). We used VIP, which measures the importance a given metabolite in the PLS-DA model, as a selection filter. The top 32 metabolites with $VIP > 1.5$ included threonine, histidine and amino acid metabolites, pyroglutamate and phenylpyruvate (Supplement 5). These 32 metabolites with greatest contribution to separation had relatively low fold-changes (0.78 to 1.28) for breast cancer relative to non-breast cancer. Thus, from this approach examining thousands of mass spectral signals, the results show that amino acid pathways are associated with breast cancer risk and, although no individual amino acid metabolite appears to have useful prognostic value, amino acids are among the metabolites that contribute most to metabolic differences between women subsequently diagnosed with breast cancer and those who were not.

3.5. Network analysis of amino acid associations with breast cancer

Visualization of amino acid-associated networks showed three communities (Fig 2), with a simplified version provided in Fig 2A to highlight the most the dominant characteristic of pregnancy serum amino acids associated with breast cancer outcome. Four amino acids (threonine, histidine, lysine, proline) are negatively associated (red edges) with pregnancy-related measures [>3 live births, primiparous status, $E3/(E1+E2)$] previously found to be protective against breast cancer. Notably, age did not have associations that were strong enough to be included in this network. Thus, this network analysis shows that amino acids are among the most central metabolic features in the organizational structure linked to subsequent breast cancer diagnosis, and these are negatively associated with protective phenotypic measures.

In the complete community structure (Fig 2B), an unidentified metabolite in the largest community, Community 1, was most central (Fig 2B). The accurate mass m/z 210.0499 of this metabolite matches the $M+2Na^+-H^+$ adduct of phenylalanine (Eigenvector Centrality Score, ECS = 1.0); however, in correlation analyses (Supplement 3), this metabolite did not correlate with other confirmed forms of phenylalanine. Additionally, this metabolite was unlike other amino acids in having positive correlations with breast cancer and associated phenotypic markers; therefore, despite the accurate mass match to a common adduct form of phenylalanine, this metabolite is considered unidentified.

Of the phenotypic measures in Community 1, >3 live births had the highest Eigenvector Centrality Score (0.667) followed by primiparous status (0.559) (Fig 2B). Community 1 also included BC15, with a relatively high Eigenvector Centrality Score (0.324) despite only having correlations at this threshold with histidine, threonine and the unidentified central metabolite (m/z 210.0499).

The next largest community, Community 2, contained lysine as a hub connected to 5-oxoproline, and three unidentified metabolites, one of which could be an adduct form of histidine (Fig 2B). Lysine was connected to multiple nodes in Community 1 and to the phenotypic measure, E3/(E1+E2), in Community 3 (Fig 2B). Community 3 also included proline, with edges connected to nodes in Community 1 (Fig 2B). Collectively, the evidence provided by this network analysis shows that risks of breast cancer and breast cancer within 15 years of pregnancy are negatively associated with third trimester serum amino acids. The amino acids are similarly negatively associated with pregnancy measures previously found to be protective against breast cancer risk. Notably, at the correlation threshold used, age was not present in any association, and breast cancer diagnosis within 15 years was selectively associated with threonine and histidine.

3.6. Quantification of amino acids in central communities

Because proline, threonine, and histidine have previously been associated with mechanisms relevant to oncogenesis, we quantified amino acids present in the central communities in Fig. 2, including threonine, histidine, proline and lysine, to allow direct comparisons of differences between individuals who subsequently developed breast cancer and those who did not. As is evident for these amino acids (Fig. 3), even though the amino acids were lower in breast cancer at $p < 0.05$, the populations had substantial overlap so that individual measures are not likely to be useful as biomarkers of risk.

3.7. Metabolome-wide association study of breast cancer occurring within 15 y of sample collection

A relatively high Eigenvector Centrality Score for BC15 in the network analysis (Fig 2) suggested important metabolic relationships for individuals with breast cancer diagnosis within 15 years of sample collection. The top associations with threonine, histidine and the unidentified m/z 210.0499, occurred without BC15 having similar associations with primiparous status, >3 live births, or E3/E1+E2. These findings indicate that BC15 may have different metabolic antecedents than breast cancer diagnosed after 15 y. To investigate this, we repeated statistical analysis using individuals with breast cancer within 15 years of pregnancy as our target and the remaining subjects (breast cancer after 15 y and non-cases) for reference. *Limma* was used to select discriminatory metabolites at $p < 0.05$ (Supplement 6), and these were used with PLS-DA Score Plots to visualize separation of BC15 from non-breast cancer and other BC cases. Results using these selected metabolites showed that overall group-wise separation was superior for BC15 compared to non-breast cancer (Fig. 4A) compared to all breast cancer compared to non-breast cancer (Fig 1C). Results also showed relatively good separation for BC15 compared to BC cases diagnosed after 15 years (BCother, Fig. 4B). Pathway enrichment analysis of the selected metabolites showed a smaller number of pathways (Fig. 4C), including lysine, histidine, tryptophan and methionine and cysteine. Consequently, the data show that increased risk of breast cancer within 15 y is associated with a modest but important interplay of multiple amino acid pathways in late pregnancy.

4. Discussion

Prior research shows that environmental chemical exposures in pregnancy are associated with perturbed amino acid metabolism. This prospective MWAS of third trimester pregnancy provides independent evidence that a generalized lower abundance of essential and non-essential amino acids in third trimester serum is associated with subsequent breast cancer diagnosis, and this effect is greater in individuals who had breast cancer within 15 years of sample collection. Amino acids share common membrane transport systems, and this tends to keep amino acids in balance with each other to support protein synthesis. Thus, the association of breast cancer risk with decreased levels of multiple amino acids may reflect a composite effect of multiple mechanisms or a single mechanism with multiple amino acids having bystander associations.

Among the amino acids decreased with breast cancer risk, tryptophan and its metabolites function in control of inflammation in cancer [42, 43], and the tryptophan pathway association with BC15 is consistent with risk of early-onset breast cancer due to impaired immunosurveillance or inflammation [18]. Methionine and leucine have well-established roles in nutrient sensing through mechanistic target of rapamycin, which regulates metabolism and can contribute to breast cancer through support of cell survival and growth. Proline functions in cellular adaptability [12], and low proline activates a stress response in conjunction with the Warburg effect to reprogram metabolism to support tumor cell energy demands [11]. The oncogene Myc also regulates expression of key enzymes in proline synthesis, which correlate to worse outcomes in cancer, and enzymes in proline metabolism, which can behave as oncogenes [12]. Thus, the collective decrease in serum tryptophan, methionine, and proline in late pregnancy appears to reflect interactive nutrient sensing responses that could activate cell survival in premalignant/malignant cells and enhance breast cancer risk.

Serum threonine and histidine were lower for BC versus non-cancer, and both had relatively strong network connections to breast cancer diagnosis within 15 years (BC15). Prior research has emphasized an important role for threonine as a precursor for glycine to support folate-dependent methylation reactions controlling the epigenome [44]. Histidine was among the top 20 metabolites associated with breast cancer outcome for all BC, as judged by p-value and VIP score, and had a relatively strong, negative association with BC15. In addition to use for protein synthesis, histidine is deaminated to urocanic acid and decarboxylated to histamine. Urocanic acid was negatively associated with breast cancer outcome, and although no prior associations with breast cancer are available, urocanic acid induced apoptosis in melanoma, cervical carcinoma and fibrosarcoma cells [45] and therefore could have unrecognized effects in breast cancer prevention. Additionally, histamine is an immune modulator implicated in regulation of breast cancer growth [46]. Histidine supplementation makes cancer cells more susceptible to chemotherapy [47], and the tumor suppressor carnosine also contains histidine [48]. Thus, in addition to nutrient-sensing effects for tryptophan, methionine and proline, epigenetic or altered tumoricidal activities could occur due to decreased threonine and histidine.

Generalized effects on multiple amino acid pathways suggest that environmental chemical exposures disrupt nutrient sensing and homeostatic regulation in women who subsequently develop breast cancer. Nutrition research shows that serum amino acids can be lower in late pregnancy due to well-studied physiological, biochemical and structural effects of pregnancy [7]. Among these, renal glomerular filtration rate is increased, and known environmental toxicant effect on renal function could further increase urinary loss of amino acids. When combined with increased needs for protein synthesis and use of amino acid oxidation to support 20% of the energy needs for the growing fetus, the overall need for maternal protein intake during late pregnancy can be increased by 30% or more [7]. Environmental chemicals impair mitochondrial energy metabolism and thereby could potentiate this need [49]. Consequently, in addition to fetal and placental demand and non-compensated protein supply from food and excessive urinary amino acid loss, low amino acid levels could activate cell survival pathways in glandular and ductal tissues and increase subsequent breast cancer risk. Importantly, these mechanisms may not be limited to pregnancy but may be impacted by environmental exposures during any stressful, energy-limiting or energy demanding period.

A summary of amino acid-dependent mechanisms by which persistent environmental chemicals could increase breast cancer risk is provided in Fig 5. In this, selected amino acids with low concentrations in pregnancy are shown (left) along with respective mechanisms functioning to assure breast maturation for lactation in breast tissues (middle). These mechanisms can increase breast cancer risk due to altered efficacy of tumor suppressor, immune surveillance or long-term metabolic reprogramming that subsequently promotes oncogenesis (right).

Among the strengths of this study, the two-step approach enabled identification of metabolic associations with small effect size measured decades before breast cancer diagnosis. In the first step of analysis using *Limma* to test for associations with breast cancer, any of the metabolites associated with outcome at $p < 0.05$ could be correct even though adjustment for false discovery showed that most not useful as biomarkers for risk. The second step with *Mummichog* pathway enrichment analysis used permutation testing to show that the set of metabolites differing at $p < 0.05$ was enriched in metabolites of amino acid pathways even though no individual amino acid differed after FDR adjustment. The study also has limitations in that there were differences in age of sampling between non-breast cancer, breast cancer and breast cancer within 15 years of sampling, which could contribute to associations observed. Additionally, based upon the age at sampling, only about half of the women who developed breast cancer within 15 years of sampling would have had breast cancer prior to 50 years old. Thus, the associations observed may underestimate true associations for early-onset breast cancer or include associations not relevant to early-onset breast cancer. A further limitation is that the proposed amino acid-dependent mechanisms linking environmental chemicals and breast cancer (Fig. 5) relies upon an assumption that decreased amino acids are at least partially due to environmental chemical exposures. Serum pregnancy DDT and PFAS associations with breast cancer have been established for other women within the CHDS, but this association has not been established for the women in the current study. Other limitations lie in the generalized meet-in-the-middle construct for environmental exposures $\rightarrow$ amino acid limitation $\rightarrow$ oncogenic cell survival (Fig. 5); many

environmental exposures other than DDT and PFAS could contribute to altered amino acid metabolism. Additionally, environmental chemicals such as DDT have aromatic structures similar to tryptophan, and PFAS have diverse structures which could interact with nutrient sensing sites and cancer surveillance mechanisms other than those discussed. Importantly, chemicals such as DDT and PFAS are poorly metabolized and eliminated, so binding to such sites could have long-acting effects. Finally, the complexity of metabolic changes in pregnancy and mechanisms of breast cancer development requires that study of specific environmental chemicals and amino acid mechanisms be verified in experimental models of breast carcinogenesis.

In summary, the present research supports a mediator role for amino acid pathway disruptions previously associated with environmental breast carcinogens. The results of pathway enrichment and network analysis of untargeted metabolomics data show that decreased amino acid concentrations in the third trimester of pregnancy are associated with subsequent breast cancer diagnosis, especially within the first 15 years after pregnancy. These amino acids are negatively associated with parity and the ratio of E3 to E1 and E2, measures which have been previously established to be protective in breast cancer risk. Prior mechanistic studies support the interpretation that low amino acid levels during late pregnancy impact nutrient sensing and metabolic reprogramming mechanisms contributing to subsequent breast cancer development.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

The collection of cancer incidence data was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute Surveillance, Epidemiology, and End Results Program under Contract HHSN261201000140C awarded to the Cancer Prevention Institute of California, Contract HHSN261201000035C awarded to the University of Southern California, and Contract HHSN261201000034C awarded to the Public Health Institute; and the Centers for Disease Control and Prevention National Program of Cancer Registries, under Agreement U58DP003862-01 awarded to the California Department of Public Health. The ideas and opinions expressed herein are those of the authors and endorsement by the State of California, Department of Public Health, National Institutes of Health, and Centers for Disease Control and Prevention or their contractors and subcontractors, or any of the funders of this research is not intended nor should be inferred. Calculation for rise of early-onset breast cancer between 1990 and 2019 using data from Global Health Data Exchange was kindly provided by Caitlin Murphy, PhD, MPH, University of Texas Health Science Center, Houston.

Funding

This study was supported in part by Department of Defense Breast Cancer Research Program Grants DAMD17-99-1-9358 and W81XWH-20-1-0102, and National Institute of Environmental Health and Science grant R01 ES031980 and P30 ES019776.

Data Availability

The data are available from the authors upon reasonable request.

Abbreviations

BC	breast cancer
FDR	false discovery rate
HRM	high-resolution metabolomics
HRMS	high-resolution mass spectrometry
LC-HRMS	liquid chromatography-high resolution mass spectrometry
MWAS	metabolome-wide association study
m/z	mass to charge
RT	retention time
T2	second trimester
T3	third trimester

References

1. Global Health Data Exchange. Available from: <https://gbd2019.healthdata.org/gbd-results/>.
2. Scott D and Singer DS, Harnessing the Power of Discovery. *Cancer Discov*, 2023.13(4): p. 819–823. [PubMed: 36884310]
3. Rappaport SM, Genetic Factors Are Not the Major Causes of Chronic Diseases. *PLoS One*, 2016. 11(4): p. e0154387. [PubMed: 27105432]
4. Cohn BA, Cirillo PM, and Terry MB, DDT and Breast Cancer: Prospective Study of Induction Time and Susceptibility Windows. *J Natl Cancer Inst*, 2019. 111(8): p. 803–810. [PubMed: 30759253]
5. Wan MLY, Co VA, and El-Nezami H, Endocrine disrupting chemicals and breast cancer: a systematic review of epidemiological studies. *Crit Rev Food Sci Nutr*, 2022. 62(24): p. 6549–6576. [PubMed: 33819127]
6. Stanfield Z, et al. Characterizing Chemical Exposure Trends from NHANES Urinary Biomonitoring Data. *Environ Health Perspect*, 2024. 132(1): p. 17009. [PubMed: 38285237]
7. Hu X, et al. Metabolome Wide Association Study of Serum Poly and Perfluoroalkyl Substances (PFASs) in Pregnancy and Early Postpartum. *Reprod Toxicol*, 2019. 87: p. 70–78. [PubMed: 31121237]
8. Hu X, et al. Metabolome Wide Association Study of serum DDT and DDE in Pregnancy and Early Postpartum. *Reprod Toxicol*, 2020. 92: p. 129–137. [PubMed: 31102720]
9. Elango R and Ball RO, Protein and Amino Acid Requirements during Pregnancy. *Adv Nutr*, 2016. 7(4): p. 839S–44S. [PubMed: 27422521]
10. Efeyan A, Comb WC, and Sabatini DM, Nutrient-sensing mechanisms and pathways. *Nature*, 2015. 517(7534): p. 302–10. [PubMed: 25592535]
11. Phang JM, Proline Metabolism in Cell Regulation and Cancer Biology: Recent Advances and Hypotheses. *Antioxid Redox Signal*, 2019. 30(4): p. 635–649. [PubMed: 28990419]
12. Burke L, et al. The Janus-like role of proline metabolism in cancer. *Cell Death Discov*, 2020. 6: p. 104. [PubMed: 33083024]
13. Wilkinson L and Gathani T, Understanding breast cancer as a global health concern. *Br J Radiol*, 2022. 95(1130): p. 20211033. [PubMed: 34905391]
14. MacMahon B and Cole P, Endocrinology and epidemiology of breast cancer. *Cancer*, 1969. 24(6): p. 1146–51. [PubMed: 5353960]

15. MacMahon B, et al. Age at first birth and breast cancer risk. *Bull World Health Organ*, 1970. 43(2): p. 209–21. [PubMed: 5312521]
16. Lambe M, et al. Transient increase in the risk of breast cancer after giving birth. *N Engl J Med*, 1994. 331(1): p. 5–9. [PubMed: 8202106]
17. Park S, et al. The Risk Factors, Incidence and Prognosis of Postpartum Breast Cancer: A Nationwide Study by the SMARTSHIP Group. *Front Oncol*, 2022. 12: p. 889433. [PubMed: 35847886]
18. Lyons TR, Schedin PJ, and Borges VF, Pregnancy and breast cancer: when they collide. *J Mammary Gland Biol Neoplasia*, 2009. 14(2): p. 87–98. [PubMed: 19381788]
19. Russo J, et al. The protective role of pregnancy in breast cancer. *Breast Cancer Res*, 2005. 7(3): p. 131–42. [PubMed: 15987443]
20. Ulgen A, Gürkut Ö, and Li W, Potential Predictive Factors for Breast Cancer Subtypes from a North Cyprus Cohort Analysis. *Cyprus Journal of Medical Sciences*, 2020. 5(4): p. 339–349.
21. Troisi R, et al. The role of pregnancy, perinatal factors and hormones in maternal cancer risk: a review of the evidence. *J Intern Med*, 2018. 283(5): p. 430–445. [PubMed: 29476569]
22. Aquilani R, et al. Essential amino acid supplementation is associated with reduced serum C-reactive protein levels and improved circulating lymphocytes in post-acute inflamed elderly patients. *Int J Immunopathol Pharmacol*, 2021. 35: p. 20587384211036823. [PubMed: 34387512]
23. Lin C, et al. The causal associations of circulating amino acids with blood pressure: a Mendelian randomization study. *BMC Med*, 2022. 20(1): p. 414. [PubMed: 36307799]
24. Teymoori F, et al. Dietary amino acids and incidence of hypertension: A principle component analysis approach. *Sci Rep*, 2017. 7(1): p. 16838. [PubMed: 29203783]
25. Brann DW and Mahesh VB, Excitatory amino acid neurotransmission evidence for a role in neuroendocrine regulation. *Trends Endocrinol Metab*, 1992. 3(4): p. 122–6. [PubMed: 18407090]
26. Cohn BA, et al. Third Trimester Estrogens and Maternal Breast Cancer: Prospective Evidence. *J Clin Endocrinol Metab*, 2017. 102(10): p. 3739–3748. [PubMed: 28973345]
27. van den Berg BJ, Christianson RE, and Oechsli FW, The California Child Health and Development Studies of the School of Public Health, University of California at Berkeley. *Paediatr Perinat Epidemiol*, 1988. 2(3): p. 265–82. [PubMed: 3070486]
28. Perkins C, Wright W, and Schlag R, Cancer Incidence and Mortality in California, 1988–1994, CA: California Department of Health Services, Cancer Surveillance Section.
29. Go YM, et al. Exposome epidemiology for suspect environmental chemical exposures during pregnancy linked to subsequent breast cancer diagnosis. *Environ Int*, 2023. 178: p. 108112. [PubMed: 37517180]
30. Jarrell ZR, et al. Firsthand and Secondhand Exposure Levels of Maltol-Flavored Electronic Nicotine Delivery System Vapors Disrupt Amino Acid Metabolism. *Toxicol Sci*, 2021. 182(1): p. 70–81. [PubMed: 34009373]
31. Jarrell ZR, et al. Plasma acylcarnitine levels increase with healthy aging. *Aging (Albany NY)*, 2020. 12(13): p. 13555–13570. [PubMed: 32554854]
32. Go YM, et al. Disturbed flow induces systemic changes in metabolites in mouse plasma: a metabolomics study using ApoE(–)/(–) mice with partial carotid ligation. *Am J Physiol Regul Integr Comp Physiol*, 2015. 308(1): p. R62–72. [PubMed: 25377480]
33. Yu T, et al. apLCMS--adaptive processing of high-resolution LC/MS data. *Bioinformatics*, 2009. 25(15): p. 1930–6. [PubMed: 19414529]
34. Uppal K, et al. xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinformatics*, 2013. 14: p. 15. [PubMed: 23323971]
35. Schymanski EL, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ Sci Technol*, 2014. 48(4): p. 2097–8. [PubMed: 24476540]
36. Uppal K, Walker DI, and Jones DP, xMSannotator: An R Package for Network-Based Annotation of High-Resolution Metabolomics Data. *Anal Chem*, 2017. 89(2): p. 1063–1067. [PubMed: 27977166]

37. Wishart DS, et al. HMDB: the Human Metabolome Database. *Nucleic Acids Res*, 2007. 35(Database issue): p. D521–6. [PubMed: 17202168]
38. Uppal K, et al. Computational Metabolomics: A Framework for the Million Metabolome. *Chem Res Toxicol*, 2016. 29(12): p. 1956–1975. [PubMed: 27629808]
39. Uppal K, et al. xMWAS: a data-driven integration and differential network analysis tool. *Bioinformatics*, 2018. 34(4): p. 701–702. [PubMed: 29069296]
40. Benjamini Y and Hochberg Y, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B*, 1995. 57(1): p. 289–300.
41. MetaboAnalyst. MetaboAnalyst 5.0. 2023; Available from: <https://new.metaboanalyst.ca/MetaboAnalyst/upload/PeakUploadView.xhtml>.
42. Platten M, et al. Tryptophan metabolism as a common therapeutic target in cancer, neurodegeneration and beyond. *Nat Rev Drug Discov*, 2019. 18(5): p. 379–401. [PubMed: 30760888]
43. Zeitler L and Murray PJ, IL4i1 and IDO1: Oxidases that control a tryptophan metabolic nexus in cancer. *J Biol Chem*, 2023. 299(6): p. 104827. [PubMed: 37196768]
44. Locasale JW, Serine, glycine and one-carbon units: cancer metabolism in full circle. *Nat Rev Cancer*, 2013. 13(8): p. 572–83. [PubMed: 23822983]
45. Laihia JK, et al. Protodynamic intracellular acidification by cis-urocanic acid promotes apoptosis of melanoma cells in vitro and in vivo. *J Invest Dermatol*, 2010. 130(10): p. 2431–9. [PubMed: 20520626]
46. Branco A, et al. Role of Histamine in Modulating the Immune Response and Inflammation. *Mediators Inflamm*, 2018. 2018: p. 9524075. [PubMed: 30224900]
47. Frezza C, Histidine metabolism boosts cancer therapy. *Nature*, 2018. 559(7715): p.484–485. [PubMed: 30030511]
48. Holecsek M, Histidine in Health and Disease: Metabolism, Physiological Importance, and Use as a Supplement. *Nutrients*, 2020. 12(3).
49. Walker DI, Pennell KD, and Jones DP, The Mitochondrial Exosome, in *Mitochondrial Dysfunction by Drug and Environmental Toxicants*, Will Y and Dykens JA, Editors. 2018, John Wiley & Sons., p. 615–637.

Highlights

- Nutrient stresses due to low amino acid levels may activate oncogenic mechanisms
- Lower concentrations for histidine, threonine, lysin and proline are associated with an increased risk of early-onset breast cancer.
- Histidine, threonine, lysin and proline show a negative association with the ratio of E3 to E1 and E2.

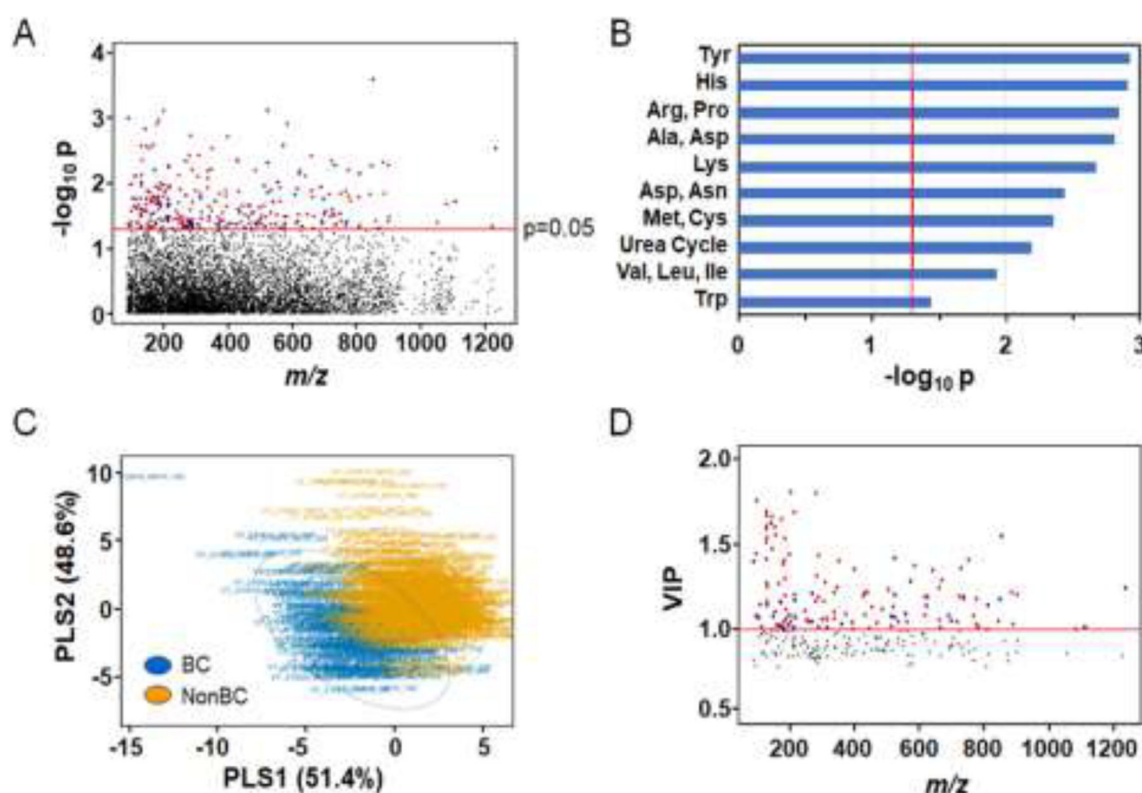


Fig. 1. Metabolome-Wide Association Study (MWAS) for metabolic prediction of breast cancer risk. A. Manhattan Plot of mass spectral features from *Limma* analysis of untargeted liquid chromatography-mass spectrometry analysis of third trimester serum samples of 172 women who subsequently developed breast cancer compared to 351 women without known breast cancer. Data are from LC-HRMS using C18 chromatography and negative electrospray ionization. Broken line represents $p<0.05$; no signals were significant after FDR adjustment. Analysis showed that 276 differed at raw $p<0.05$, listed in along with HILIC data in Supplement 1. For HILIC chromatography with positive electrospray ionization, 262 features at raw $p<0.05$ was obtained, and none were significant after FDR adjustment (not shown). B. Amino acid pathways in third trimester pregnancy serum associated with subsequent breast cancer diagnosis. Metabolites differing between non-cancer and cancer cases for both C18 and HILIC were selected at $p < 0.05$ and used with permutation testing for pathway enrichment using *mummichog*. Other associated pathways are provided in Supplement 2. C. PLS-DA Score Plot showing women who subsequently developed breast cancer (BC) in blue and those not known to develop breast cancer (NonBC) in yellow using 276 features that differed between groups at raw $p<0.05$. D. Manhattan plot of VIP values from PLS-DA analysis of selected top 276 features from *Limma* analysis of C18 in Panel A. Broken line indicates $VIP>1.0$. Features with a positive correlation to breast cancer diagnosis are shown in red, and those with a negative correlation are shown in blue. A list of these 136 features is provided in Supplement 4. A similar number of features with $VIP>1.0$ was obtained for HILIC chromatography with positive electrospray ionization (Supplement 4).

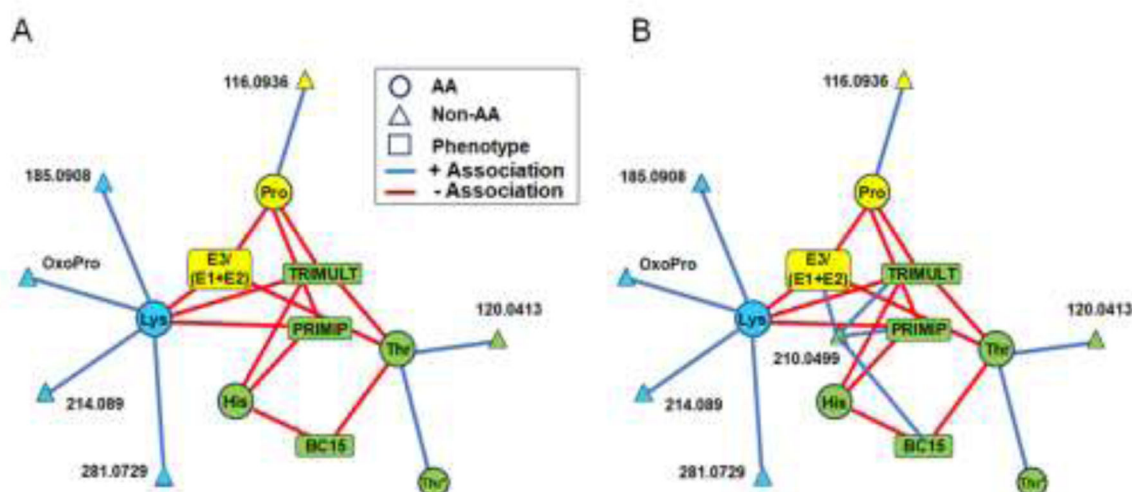


Fig. 2. Network analysis of amino acids, phenotypic measures and other top features associated with breast cancer. A. Simplified version of full network from Panel B with unidentified metabolite (*mz* 210.0499) removed to emphasize negative relationship of amino acids with phenotypic measures of pregnancy. Node color indicates community, node size indicates centrality, and edge color indicates direction of correlation where red corresponds to negative correlation and blue to positive correlation. Nodes with unknown identity are annotated with their ionization mode and accurate mass. Community 1 (green) centers on >3 live births (TRIMULT), primiparous status (PRIMIP), BC15, threonine, and histidine, Community 2 (blue) centers on lysine and 5-oxoproline and Community 3 (yellow) centers on E3/E1+E2 and proline. B. Complete network structure showing unidentified metabolite (*mz* 210.0499) with positive community associations with BC15 and other phenotypic measures.

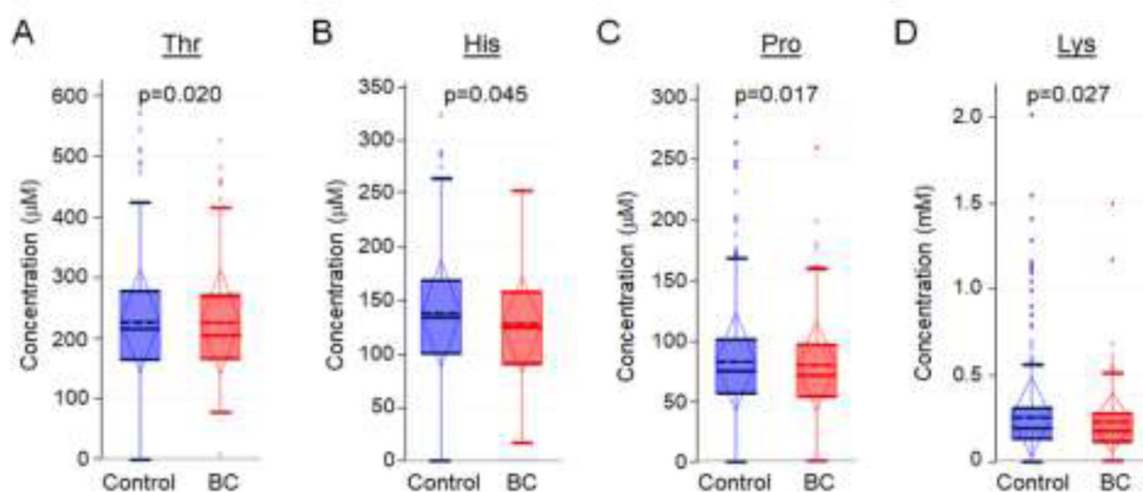


Fig. 3. Concentrations of hub amino acids in metabolic network analysis of breast cancer risk. Threonine, histidine, proline and lysine concentrations (μM) are represented with Box and Whisker Plots, where the solid lines show the median and broken lines show the mean and standard deviations of the metabolites. Concentrations for breast cancer and non-breast cancer groups showed higher median values in the non-cancer group for all selected amino acids.

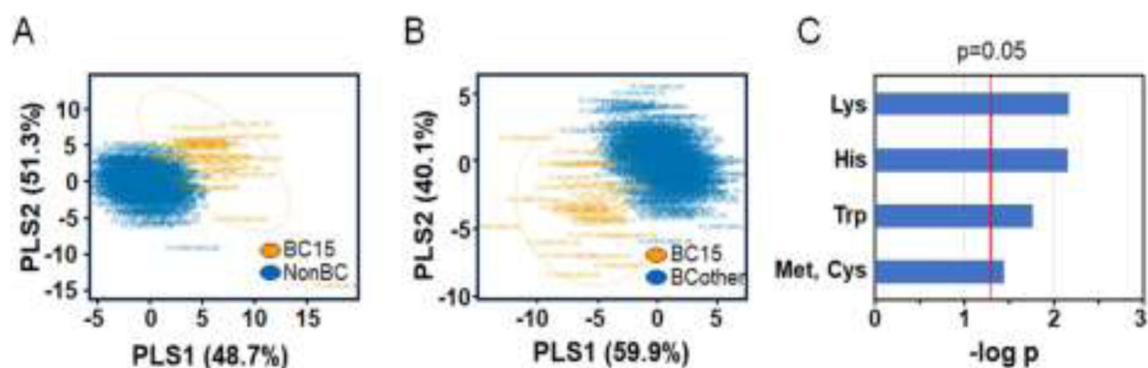


Fig. 4. PLS-DA Score Plots of BC15 and associated pathway enrichment analysis. A. Comparison of BC15 (n = 34) to non-breast cancer (NonBC) group (n = 351). B. Comparison of BC15 (n = 34) to group with breast cancer diagnosed after 15 years (BCother; n = 117). C. *Mummichog* pathways with selected metabolites differing for BC15 compared to all others.

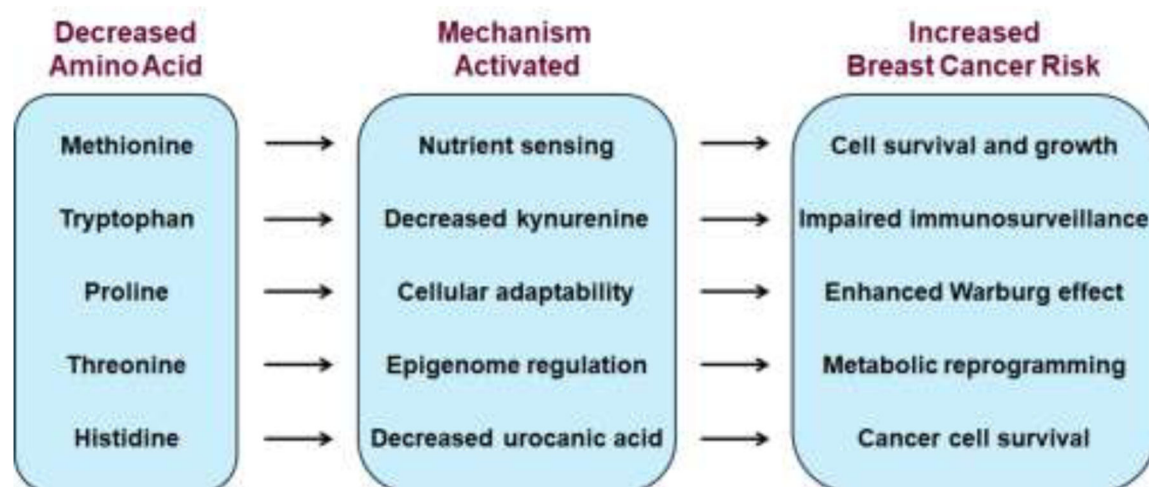


Fig. 5. Amino Acid-Associated Mechanisms of Environmental Breast Cancer. A selected list of amino acids that are decreased in third trimester serum in association with subsequent breast cancer is provided in the left column. In the middle column provides corresponding mechanisms from the literature (see text for discussion) that are activated by low concentration of amino acid. Potential consequences of these mechanisms in breast cancer risk are provided in the right column. Environmental chemicals impact many of these mechanisms, and experimental studies are needed to investigate environmental chemical exposures in pregnancy to define impact on specific amino acids and associated carcinogenicity.

Table 1.

Study population characteristics. Comparison between women diagnosed with breast cancer (BC), women diagnosed with breast cancer within 15 years of sample collection, and women who were not diagnosed with breast cancer (control; CTRL).

Characteristic	CTRL (n = 351)	Cases (n = 172)	BC within 15 y (n = 34)
Age at blood sample, y	26 (8)	30 (10)	35 (7.5)
Parity at blood sampling	%	%	%
0	38	30	26
1–3	19	47	50
>3	43	23	24
Race	%	%	%
White	73	70	62
African American	18	16	21
Hispanic	2	2	3
Asian	5	9	9
Mixed	2	3	6