Plasma Metabolomic Signatures Associated with Long-term Breast Cancer Risk in the SU.VI.MAX Prospective Cohort

Lucie Lécuyer1, Céline Dalle2, Bernard Lyan2, Aicha Demidem3, Adrien Rossary3, Marie-Paule Vasson1,4, Mélanie Petera2, Marie Lagree5, Thomas Ferreira3, Delphine Centeno2, Pilar Galan1, Serge Hercberg1,6, Mélanie Deschasaux1, Valentin Partula1, Bernard Sour1, Paule Latino-Martel1, Emmanuelle Kesse-Guyot1, Nathalie Druesne-Pecollo1, Stéphanie Durand2, Estelle Pujos-Guillot2, and Mathilde Touvier1

Abstract

Background: Breast cancer is a major cause of death in occidental women. The role of metabolism in breast cancer etiology remains unclear. Metabolomics may help to elucidate novel biological pathways and identify new biomarkers to predict breast cancer long before symptoms appear. The aim of this study was to investigate whether untargeted metabolomic signatures from blood draws of healthy women could contribute to better understand and predict the long-term risk of developing breast cancer.

Methods: A nested case–control study was conducted within the SU.VI.MAX prospective cohort (13 years of follow-up) to analyze baseline plasma samples of 211 incident breast cancer cases and 211 matched controls by LC/MS. Multivariable conditional logistic regression models were computed.

Results: A total of 3,565 ions were detected and 1,221 were retained for statistical analysis. A total of 73 ions were associated with breast cancer risk (P < 0.01; FDR ≤ 0.2). Notably, we observed that a lower plasma level of O-succinylhomoserine (OR = 0.70, 95% CI = [0.55–0.89]) and higher plasma levels of valine/norvaline [1.45 (1.15–1.83)], glutamine/isoglutamine [1.33 (1.07–1.66)], 5-aminolevulinic acid [1.46 (1.14–1.87)], phenylalanine [1.43 (1.14–1.78)], tryptophan [1.40 (1.10–1.79)], γ-glutamyl-threonine [1.39 (1.09–1.77)], ATBC [1.41 (1.10–1.79)], and pregnene-triol sulfate [1.38 (1.08–1.77)] were associated with an increased risk of developing breast cancer during follow-up.

Conclusion: Several prediagnostic plasmatic metabolites were associated with long-term breast cancer risk and suggested a role of microbiota metabolism and environmental exposure.

Impact: After confirmation in other independent cohort studies, these results could help to identify healthy women at higher risk of developing breast cancer in the subsequent decade and to propose a better understanding of the complex mechanisms involved in its etiology.

Introduction

Breast cancer is the most common female cancer in the world (www.dietandcancerreport.org) and ranks as the fifth cause of death from cancer overall (626,679 deaths worldwide in 2018; ref. 1). Despite some well-known risk factors such as adult height, reproductive life, and hormonal factors, and, for postmenopausal women, alcoholic drinks, body fatness, and adult weight gain (www.wcrf.org/breast-cancer-2017), the role of metabolism in breast cancer etiology remains unclear and hypothesis-driven research has not yet allowed to entirely elucidate it.

Recent advances in modern explorative technologies, such as metabolomics, increase the capacity to identify additional risk factors and better understand mechanisms involved (2). Indeed, mass spectrometry (MS) and NMR allows detecting a wide range of low molecular weight biochemicals in biofluids, cells, or tissues (3) and thus can provide a signature of both exogenous exposures and endogenous metabolism. Therefore, metabolomics allows identifying biomarkers relevant to cancer etiology and elucidating novel biological pathways (4–6). Several cross-sectional epidemiologic studies highlighted differences between metabolomic profiles and patients with breast cancer and...
controls (7, 8). An NMR-based prospective study (5 years of follow-up) recently showed the benefit of metabolomic signatures of breast cancer risk (9). However, metabolomic signatures may be altered by the presence of a breast tumor itself (10), arguing for similar investigations with a long follow-up. Improving the ability to discriminate profiles at higher risk of developing breast cancer in the subsequent decades through the identification of prediagnostic biomarkers, to establish preventive approaches, remains an important focus in breast cancer research.

Two recent semitargeted MS studies investigated the associations between diet- (11) or body mass index (BMI)-related metabolites (12) and long-term breast cancer risk, and found promising results supporting the role of diet or BMI in postmenopausal breast cancer etiology. We previously set up a prospective nested case–control study within the SILVIMAX cohort to investigate whether baseline metabolomic profiles from plasma of apparently healthy women could contribute to identify women who would develop breast cancer during the subsequent years, and better understand the etiology of this complex disease. Using an NMR metabolomics analysis, higher fasting plasma levels of valine, lysine, glutamine, and glucose and lower plasma levels of lipoproteins and glycerol-derived compounds were notably found to be associated with a higher risk of developing breast cancer (13). To increase the number of metabolites detected both in terms of sensitivity and metabolome coverage, this study applied for the first time a fully untargeted and nonoriented MS analysis, with the objective to highlight new endogenous or exogenous metabolites associated with long-term risk of breast cancer.

Materials and Methods

Population study

This analysis was based on a case–control study nestled within the SILVIMAX prospective cohort (clinicaltrials.gov; NCT00272428). The latter was initially designed to investigate the influence of a daily supplementation with nutritional doses of antioxidants on the incidence of cardiovascular diseases and cancers. The study design and methods have been previously detailed (14, 15). Briefly, this study is a population-based, double-blind, placebo-controlled, randomized trial enrolling a total of 13,017 participants recruited in 1994–1995. The intervention study lasted 8 years, and observational follow-up of health events was maintained until September 2007. Written informed consent was obtained from all participants. The trial was approved by the Ethics Committee for Studies with Human Subjects of Paris-Cochin Hospital (CCPPRB 706/2364) and the ‘Commission Nationale de l’Informatique et des Liberté’ (CNIL 334641/907094) and was conducted according to the Declaration of Helsinki guidelines.

Baseline data collection

At inclusion, self-administered questionnaires about sociodemographic characteristics, smoking status, medication use, health status, and family history of cancer were filled-in by participants. During a baseline clinical examination, all participants underwent anthropometric measurements (height and weight) and a blood draw (occurring after a 12-hour fasting period) by study nurses and physicians. These 35 mL venous blood samples were collected in sodium heparin Vacutainer Tubes (Becton Dickin-son). After centrifugation at +4°C, plasma aliquots were immediately prepared and stored frozen at −20°C during less than 2 days and then stored in liquid nitrogen.

Case ascertainment

Health events that occurred during follow-up (from 1994 to 2007) were self-reported by participants. Medical data were gathered through participants, physicians, and hospitals and reviewed by an independent physician expert committee. Pathologic reports were used to validate cases and extract cancer characteristics. Cancers were classified by using the International Classification of Diseases, 10th Revision, Clinical Modification (16).

Nested case–control study

This analysis initially included 215 participants with a first incident invasive breast cancer matched 1:1 with controls for baseline age (35–39 years/40–44 years/45–49 years/50–54 years/55–59 years/>60 years), menopausal status (pre/post-menopause status at baseline and at diagnosis), BMI (<18.5 kg/m², underweight; 18.5–<25 kg/m², normal weight; and ≥25 kg/m², overweight/obese), arm assignment to the initial SILVIMAX clinical trial (placebo/supplemented), smoking status (current smokers and non/former smokers at baseline), and season of blood draw (a priori–defined periods: October–November/December–January–February/March–April–May). Controls selection was based on the classical density sampling method (17), that is, every time a case was diagnosed, one control was selected from other participants from the cohort with no breast cancer at that time.

Metabolomic analyses

Metabolomic profiling was conducted using ultrahigh performance liquid chromatography/mass spectroscopy with the Metabolic Profiler Platform (Bruker Daltonique). Details on chemicals and reagents, biological samples preparation, metabolic profiling, raw data extraction, and metabolite identification are available in Supplementary Materials and Methods and Supplementary Fig. 51. Samples were randomized within the analytic sequence based on a Williams Latin Square strategy defined according to the main factors of the study. The stability of the analytic system was monitored using pooled samples as quality control and analyzed in the same sequence than participant samples (injected one time at the beginning of each sequence and then after each set of 12 samples). This metabolomic discovery approach allows semiquantitative measurements, representing relative ion intensities in arbitrary unit. However, even if it is not absolute quantitation, there are linear relations between ion intensity for each metabolite and its concentration level in the samples.

Statistical analysis

Participants’ baseline characteristics were compared between cases and controls using χ² tests for categorical variables and Student t test for continuous variables. Less than 5% of values were missing for all covariates and missing values were replaced by the mode.

Prospective associations between each ion’s signal intensity and breast cancer risk were characterized using multivariable conditional logistic regression models. OR for an increment of 1 SD for each metabolite and their 95% confidence intervals (95%
Cancer Epidemiol Biomarkers Prev; 28(8) August 2019
Cancer Epidemiology, Biomarkers & Prevention

Results

The baseline characteristics of breast cancer cases and controls in the study population are summarized in Table 1. The cases were more likely to be taller with more frequent family history of breast cancer than controls. A total of 25% of cases were diagnosed before 3.5 years, 50% before 5.8 years, and 75% before 9.4 years. The maximum length of follow-up was 12.8 years.

A total of 3,565 ions were detected and 1,221 were retained for statistical analysis. From logistic regression analyses, 53 ions were positively associated with breast cancer risk and 20 ions were inversely associated \( (P \leq 0.01; \text{FDR} \leq 0.2) \). Results of all associated ions are shown in Supplementary Table S1. Among these ions, some have been identified using available databases (in-house database, Metlin; https://metlin.scripps.edu), Human Metabolome Database (www.hmdb.ca), Kyoto Encyclopedia of Genes and Genomes database, and analytic data (as MS2); the associations between these ions \( (N = 9) \) and breast cancer risk are displayed in Table 2. A lower plasma level of O-succinylhomoserine \( \left[ \text{OR} = 0.70, 95\% \text{CI} = [0.55-0.89] \right] \) and higher plasma levels of valine/norvaline \( \left[ 1.45 \left( 1.15-1.83 \right) \right] \), glutamine/isoglutamline \( \left[ 1.33 \left( 1.07-1.66 \right) \right] \), 5-aminovaleric acid \( \left[ 1.46 \left( 1.14-1.87 \right) \right] \), phenylalanine \( \left[ 1.43 \left( 1.14-1.78 \right) \right] \), tryptophan \( \left[ 1.40 \left( 1.10-1.77 \right) \right] \), acetyl tributyl citrate ATBC \( \left[ 1.41 \left( 1.10-1.79 \right) \right] \), and the putative pregnene-triol sulfate \( \left[ 1.38 \left( 1.08-1.77 \right) \right] \) were associated with an increased risk of developing breast cancer during follow-up. Linearity assessments, as well as analyses performed with ions categorized in tertiles were generated using R software (v3.5.2). All of the statistical tests were two-sided.

All ions associated with breast cancer risk in logistic regressions were then analyzed by PCA. The four first PCA factors (metabolomic patterns) explained more than 36% of total variability. Loading values are displayed in Supplementary Table S7. The first factor was notably characterized by higher plasma levels of valine and 5-aminovaleric acid, the second factor was in particular characterized by higher plasma levels of phenylalanine, valine, 5-aminovaleric acid, and tryptophan, the third by lower plasma level of pregnene-triol sulfate, and the fourth by higher plasma levels of phenylalanine, valine 5-aminovaleric acid, GGT, and glutamine (loading values \( \geq 0.20 \)). Moreover, several unidentified ions also characterized these four factors. These data allowed distinguishing three metabolomic patterns (factors 1, 2, and 4).
Table 1. Baseline characteristics of breast cancer cases and controls, SU.VI.MAX cohort, France (1994–2007)

<table>
<thead>
<tr>
<th></th>
<th>Breast cancer cases (N = 211)</th>
<th>Controls (N = 211)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at baseline (y)</td>
<td>49.0 ± 6.0</td>
<td>48.9 ± 6.0</td>
<td>0.9</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>25.2 ± 4.0</td>
<td>25.6 ± 4.2</td>
<td>0.3</td>
</tr>
<tr>
<td>&lt;18.5 kg/m² (underweight)</td>
<td>9 (4.3)</td>
<td>6 (2.8)</td>
<td></td>
</tr>
<tr>
<td>≥18.5–25 kg/m² (normal weight)</td>
<td>148 (70.1)</td>
<td>151 (71.6)</td>
<td></td>
</tr>
<tr>
<td>≥25 kg/m² (overweight)</td>
<td>54 (25.6)</td>
<td>54 (25.6)</td>
<td></td>
</tr>
<tr>
<td>Height (cm)</td>
<td>163.0 ± 6.2</td>
<td>160.9 ± 5.9</td>
<td>0.0004</td>
</tr>
<tr>
<td>Intervention group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>105 (49.8)</td>
<td>105 (49.8)</td>
<td></td>
</tr>
<tr>
<td>Antioxidants</td>
<td>106 (50.2)</td>
<td>106 (50.2)</td>
<td></td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never and former</td>
<td>164 (77.7)</td>
<td>164 (77.7)</td>
<td></td>
</tr>
<tr>
<td>Current smoker</td>
<td>47 (22.3)</td>
<td>47 (22.3)</td>
<td></td>
</tr>
<tr>
<td>Physical activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irregular</td>
<td>67 (31.8)</td>
<td>60 (28.4)</td>
<td>0.8</td>
</tr>
<tr>
<td>&lt;1 h/d walking equivalent</td>
<td>69 (32.7)</td>
<td>72 (34.1)</td>
<td></td>
</tr>
<tr>
<td>≥1 h/d walking equivalent</td>
<td>75 (35.5)</td>
<td>79 (37.4)</td>
<td></td>
</tr>
<tr>
<td>Educational level</td>
<td></td>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td>Primary</td>
<td>37 (17.5)</td>
<td>44 (20.9)</td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>82 (38.9)</td>
<td>91 (43.1)</td>
<td></td>
</tr>
<tr>
<td>Superior</td>
<td>92 (43.6)</td>
<td>76 (36.0)</td>
<td></td>
</tr>
<tr>
<td>Number of biological children</td>
<td>1.8 ± 1.2</td>
<td>2 ± 1.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Hormonal treatment for menopause</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postmenopausal status at baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prenopausal</td>
<td>129 (61.1)</td>
<td>129 (61.1)</td>
<td></td>
</tr>
<tr>
<td>Postmenopausal</td>
<td>82 (38.9)</td>
<td>82 (38.9)</td>
<td></td>
</tr>
<tr>
<td>Menopausal status at diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prenopausal</td>
<td>78 (37.0)</td>
<td>78 (37.0)</td>
<td></td>
</tr>
<tr>
<td>Postmenopausal</td>
<td>133 (63.0)</td>
<td>133 (63.0)</td>
<td></td>
</tr>
<tr>
<td>Family history of breast cancer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(yes)</td>
<td>35 (16.6)</td>
<td>21 (10.0)</td>
<td>0.04</td>
</tr>
<tr>
<td>Alcohol intake (g/d)</td>
<td>11.2 ± 12.0</td>
<td>11.5 ± 13.0</td>
<td>0.8</td>
</tr>
<tr>
<td>Month of blood draw</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>March–April–May</td>
<td>77 (35.5)</td>
<td>77 (35.5)</td>
<td></td>
</tr>
<tr>
<td>October–November</td>
<td>31 (14.7)</td>
<td>32 (15.2)</td>
<td></td>
</tr>
<tr>
<td>December–January–February</td>
<td>103 (48.8)</td>
<td>102 (48.3)</td>
<td></td>
</tr>
</tbody>
</table>

*Values are means ± SD or n (%).

**P** for the comparison between breast cancer cases and controls using *χ²* tests for categorical variables or Student *t*-test (from analysis of variance models) for continuous variables.

Among first-degree female relatives.

associated with an increased risk of breast cancer in multivariable conditional regression models and one (factor 3) with a decreased risk (FDR < 0.05; Table 3).

### Discussion

To our knowledge, this prospective study was the first fully untargeted and nonoriented LC/MS-based analysis aiming to investigate the associations between prediagnosis metabolic profiles and long-term breast cancer risk. Untargeted metabolomics, by the detection of a large number of metabolites, allowed discriminating early phenotypes and revealing a plasma metabolic signature announcing the evolution toward breast cancer in the subsequent 13 years. This could be relevant for detecting specific phenotypic profiles of women at higher risk for this disease and improve personalized prevention. A profile (signature) of circulating levels of multiple metabolites provides insights to draw an overall picture of the different pathways involved in the etiology of this cancer (20). Even if the origin of metabolites variations (exogenous or endogenous) cannot be determined in this study, our results also illustrate that metabolomics integrates information regarding the influence of both extrinsic and intrinsic factors, such as lifestyle exposures and basal metabolism, and consequently address the complex issue of environment and health interaction.

In this study, we found that a lower plasma level of O-succinyl-homoserine and higher plasma levels of valine/norvaline, glutamine/isoglutamine, 5-aminolevulinic acid, phenylalanine, tryptophan, γ-glutamyl-threonine, ATBC, and pregnene-triol sulfate were associated with an increased risk of developing breast cancer during the following decade. Three metabolomic patterns characterized by several concurrently metabolites variations were associated with an increased risk of breast cancer, as well as another metabolomic pattern associated with a decreased risk. These univariate analysis (analysis of one by one metabolic variation) and multivariate analysis (analysis of combinations of simultaneous metabolites variations) allowed getting complementary overviews of metabolic signatures discriminating phenotypes at higher risk for this disease well before breast cancer clinical onset.

We previously investigated a higher plasma level of glutamine and valine in the future breast cancer cases with NMR analyses on the same population study (13), which was consistent with our current results. An increase in plasmatic levels of valine and three γ-glutamyl dipeptides was also recently found to be positively associated with BMI and estrogen receptor–positive breast cancer risk in postmenopausal women through a semuntargeted MS study (12). The pregnene-triol sulfate is a steroid hormone and belongs to progestative family compounds. In line with our results, two semuntargeted MS studies found a positive association between steroid hormones and breast cancer risk in postmenopausal women (11, 12). Four metabolomic studies reported that an increased level of phenylalanine was positively associated with hepatocellular carcinoma (HCC) risk (21–24) and three of these studies found that a higher level of glutamine was associated with a decreased risk of HCC (21–23). Concerning valine, findings were more contrasted; one of these studies showed a positive association with HCC risk (21), whereas another observed an inverse association (22). An increase in plasmatic valine level was also reported in two metabolic studies to be positively associated with pancreatic adenoma risk (25) and inversely associated with prostate cancer risk (26). In three metabolomic studies an inverse association was found between phenylalanine (26), tryptophan (27), γ-glutamyl-histidine (28), and γ-glutamyl-phenylalanine (26, 27) and prostate cancer risk, whereas a positive association was reported between γ-glutamyl-glutamine and nonaggressive prostate cancer risk (27). The comparisons of our results with other prospective metabolomic studies should be interpreted with caution, taking into account methodologic and technological variations (as the wide range of metabolomics platforms and types of analyses), because the etiology may vary according to cancer locations.

The findings highlighted in this study provide evidence for circulating metabolites as putative risk biomarkers for breast cancer. Given that metabolites concerned by the observed variations are not specific of one metabolic pathway and as their variations could arise from several types of exposure and endogenous reactions, precaution is needed in the mechanistic interpretation of our epidemiologic findings. Even so, we propose...
Table 2. Identified ions associated with breast cancer risk from multivariable conditional logistic regression, SU.VI.MAX cohort, France (1994–2007)

<table>
<thead>
<tr>
<th>Name (level of confidence for identificationa)</th>
<th>Annotation</th>
<th>Measured mass m/z</th>
<th>Retention time (minute)</th>
<th>Mode</th>
<th>Cases/controls</th>
<th>Mean ± SD</th>
<th>OR (95% CI)</th>
<th>Pcorr</th>
</tr>
</thead>
<tbody>
<tr>
<td>l-phenylalanine (+)</td>
<td>[M + H]+</td>
<td>166.08618</td>
<td>10.11</td>
<td>ESI+</td>
<td>21/211</td>
<td>227.718 ± 42.535</td>
<td>1.43 (1.14–1.78)</td>
<td>0.002</td>
</tr>
<tr>
<td>l-valine/norvaline (+)</td>
<td>[M + H]+</td>
<td>72.08069</td>
<td>3.08</td>
<td>ESI+</td>
<td>21/211</td>
<td>115.033 ± 24.435</td>
<td>1.45 (1.15–1.83)</td>
<td>0.002</td>
</tr>
<tr>
<td>5-aminolevulinic acid (+)</td>
<td>[M + H]2+</td>
<td>140.06816</td>
<td>2.29</td>
<td>ESI+</td>
<td>21/211</td>
<td>10.299 ± 5.319</td>
<td>1.49 (1.14–1.87)</td>
<td>0.003</td>
</tr>
<tr>
<td>O-succinyl-L-homoserine (+)</td>
<td>[M + Na]+</td>
<td>240.04755</td>
<td>2.86</td>
<td>ESI−</td>
<td>203/203</td>
<td>919 ± 400</td>
<td>0.70 (0.55–0.89)</td>
<td>0.004</td>
</tr>
<tr>
<td>ATBC (acetyl tributyl citrate) (-)</td>
<td>[M + Na]+</td>
<td>425.21694</td>
<td>19.31</td>
<td>ESI+</td>
<td>21/211</td>
<td>2.998 ± 964</td>
<td>1.41 (1.10–1.79)</td>
<td>0.006</td>
</tr>
<tr>
<td>L-tryptophan (+)</td>
<td>[M + Na]+</td>
<td>227.07867</td>
<td>11.09</td>
<td>ESI+</td>
<td>21/211</td>
<td>799 ± 180</td>
<td>1.40 (1.10–1.79)</td>
<td>0.007</td>
</tr>
<tr>
<td>L-lysoglutamic acid (+)</td>
<td>[M + H2O]+</td>
<td>233.1451</td>
<td>2.52</td>
<td>ESI+</td>
<td>21/211</td>
<td>1.054 ± 601</td>
<td>1.39 (1.09–1.77)</td>
<td>0.008</td>
</tr>
<tr>
<td>Delaminic acid (+)</td>
<td>[M + H]+</td>
<td>90.0551</td>
<td>2.32</td>
<td>ESI+</td>
<td>21/211</td>
<td>13.187 ± 2982</td>
<td>1.33 (1.07–1.66)</td>
<td>0.009</td>
</tr>
<tr>
<td>Pregnenetriol sulfate (-)</td>
<td>C21H35O6S</td>
<td>413.19771</td>
<td>16.11</td>
<td>ESI−</td>
<td>203/203</td>
<td>2.44 ± 1.254</td>
<td>1.38 (1.08–1.77)</td>
<td>0.01</td>
</tr>
</tbody>
</table>

*aMultivariable models were adjusted for age (continuous), BMI (continuous), smoking status (current smokers, former smokers, and nonsmokers), season of blood draw (a priori–defined periods: October–November/December–January–February/March–April–May), height (continuous), alcohol intake (continuous), physical activity (irregular/c1 h/d walking equivalent/c1 h/d walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous), and use of hormone replacement therapy for menopause (yes/no; reference model). Tests for linear trend were performed using the continuous variables. ORs were presented for an increase of 1 SD for each metabolite.

*bMetabolites were classified according to Sumner and colleagues (46) concerning the levels of confidence in the identification process: identified (level 1: confirmed by standard), putatively annotated (level 2: based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries).

*cFDR after Benjamini Hochberg correction for multiple testing = 0.1.

*dFDR after Benjamini Hochberg correction for multiple testing = 0.2.

interpretation hypotheses for some presumed pathways, summarized in Fig. 1. Among the possible interpretations, our results could suggest a perturbation of energy and amino acids metabolism, a set-up of conditions favoring cell proliferation, inflammation, and oxidative stress in women who will develop breast cancer involving environmental exposure and microbiota metabolism. Glutamine and valine (one of the branched chain amino acids; BCAA) positively correlated in our study, are two glycogenic acids; BCAA are known to interact with several signaling pathways as a ligand for the mTOR pathway; one of the main cell proliferation routes (31). BCAAs are an alternative plasticizer to phthalates (33) widely used in products such as food wrap, vinyl toys, and pharmaceutical excipients. Although this contaminant does not seem to accumulate in tissues (34), recent data revealed potential biological activity on tissue growth (35). The combination of plasminogen activator in both ATBC and in a progestative sulfate steroid may interact either with PPAR signaling involved in cell proliferation and inflammation, and/or with steroid signaling (36). Indeed, ATBC has been shown to influence steroid levels (35) and is known to interact with several signaling pathways as a ligand for steroid xenobiotic receptor or sexual hormone–binding globulin involved in steroid signaling (33). Phenylalanine and tryptophan are aromatic amino acids (AAA), correlated in our study, R = 0.27 and high level of AAs in blood have been considered as an inflammatory marker in patients with liver cirrhosis (37). Disruptions in AAs and BCAAs levels are found in several pathologies with chronic low-grade inflammation such as obesity and diabetes (38, 39); in our study valine (BCAA) is highly positively correlated with AAs (phenylalanine; R = 0.47 and tryptophan; R = 0.24). Data on the kynurenine/tryptophan ratio, unfortunately unavailable in this study, would be useful to support this hypothesis, as it reflects IDO1 gene expression, which increases in situations of chronic low-grade inflammation. 5-aminolevulinic acid is a catabolism product from lysine; its increase could witness the presence of cellular necrosis or microbiota metabolism. In mice, higher urinary levels of 5-aminolevulinic acid were associated with intestinal inflammation (40). Concerning identified metabolites, the second PCA-derived factor highlighted in our study could represent an inflammatory pattern, as it was notably characterized by higher plasma levels of phenylalanine, valine, 5-aminolevulinic acid, and tryptophan, the third by lower plasma level of pregnene-triol sulfate, and the fourth by higher plasma levels of phenylalanine, valine 5-aminolevulinic acid, and glutamine (loading values ≥0.20). Moreover, several unidentified ions also characterized these four factors. Another metabolite synthesized by gut microbiota, 2-amino-4-
cyano butanoic acid, was associated with increased breast cancer risk in this study before FDR correction only (OR = 1.28 (1.03–1.58); P = 0.02; FDR = 0.3; data not tabulated). The variation of these two microbiota-related metabolites could be interpreted as an oxidative stress response given their potential role in glutamate and glutathione synthesis. Indeed, humans’ metabolism can produce cysteine (one component of glutathione) using O-succinyl-homoserine instead of homocysteine (41). 2-amino-4-cyano butanoic acid is the substrate of nitrilase responsible for the de novo synthesis of glutamate in bacteria and fungus (42). Moreover it is also implied in heme synthesis, recognized as an oxidative biocompound (43). Likewise, the increase in plasmatic glutamine and in γ-glutamyl-threonine (positively correlated in our study) can suggest both a decrease in antioxidant defenses and an oxidative stress response (γ-glutamyl-dipeptides may indicate increased glutathione synthesis). The data available in this study do not allow to draw firm conclusions about these mechanistic interpretations, which need further investigations for their validation, such as complementary analyses of other metabolites variation involved in the same pathway (but not detected with the current analytic technic), fluxomic analyses to determine the rates of metabolic reactions, or lipidomic and proteomic analyses to cover a wide range of biggest molecules.

The major strengths of this study pertained to the fully untargeted MS metabolomics approach and the prospective design. However, several limitations should be acknowledged. First, the relatively small sample size was probably sufficient for main analyses, but it was certainly limited for secondary stratified analyses. For instance, a larger sample size could have help to investigate more precisely the relation between the metabolites and the duration before cancer development (e.g., short, medium, and long term). Second, we cannot rule out the possibility of residual or unmeasured confounding despite the large spectrum of covariates taken into account, and we cannot exclude the possibility of overfitting of the models. Third, because the size of the sample did not allow keeping apart a validation set, we did not compute predictive performance such as AUC. However the main interest of our study lied in a better understanding of breast cancer etiology rather than in an improvement in discrimination models. Then, only LC/MS and not gas chromatography–mass spectroscopy was applied. This may have limited our ability to detect some categories of metabolites which could have

---

**Figure 1.** Summary of proposed mechanistic interpretations for some putative pathways regarding observed metabolites variations in plasma of women who will develop breast cancer in subsequent years. Among the possible interpretations, our results could suggest a perturbation of energy and amino acids metabolism, a set-up of conditions favoring cell proliferation, inflammation, and oxidative stress in women who will develop breast cancer involving environmental exposure and microbiota metabolism.
supported certain mechanistic interpretations. Moreover, as the metabolic profiles were established a decade before the diagnosis, some small variations may not have been detected. Some of the metabolites highlighted in our study are at the crossroads of several metabolic pathways, thus these mechanistic interpretations resulting from literature should be considered with caution and need further investigations to be validated. Other interpretations could also be considered. Furthermore, this metabolic analysis was based on a single blood draw so we could not investigate the stability of the profiles. Nevertheless, several studies showed a good stability and reproducibility of metabolic measurements for most of metabolites (44, 45). Finally, as it is usually the case in a fully untargeted metabolomic approach, several associated ions could not be identified to date. However, these findings will be useful in the future in a context of strong development of metabolites databases and to compare our results with other future LC/MS studies.

In conclusion, this prospective study highlighted associations between baseline untargeted metabolomic profiles and subsequent long-term breast cancer risk. These findings generate mechanistic hypotheses regarding the etiology of this complex disease, which were consistent with those reported in our NMR analysis, and include a possible implication of gut microbiota activity and environmental contamination/exposure. If replicated in other independent prospective studies and after quantification of key metabolites to set discrimination thresholds, these results might contribute to develop screening strategies for the identification, before tumor set up, of women at higher risk of developing a breast cancer in the decade following the blood draw.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Disclaimer
The funders had no role in the design, analysis, or writing of this article.

References
11. Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): L. Lyan, E. Pujos-Guillot, M. Touvier.

Authors’ Contributions
Conception and design: L. Lyan, M. Touvier.
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): B. Iyan, M. Petera, D. Centeno, P. Galan, S. Hercberg, E. Kesse-Guyot, M. Touvier, M. Lagree.

Acknowledgments
The authors thank Yvoness Esseddk, Frederic Coffinieres, Thong Hong Van Duong, Paul Flanzey, Regis Gazibelza, Jagatjit Mohinder, and Mai Khaly Shiwyavalp (computer scientists), Rachid Mehoog and Frederique Ferrat (logistic assistants), Nathalie Arnaud, Veronique Gourlet, PhD, Fabien Szabo, PhD, Julien Allerge, and Laurent Bousin (data management/statistician) and Cedric Agaese (dietician) for their technical contribution to the SU.VI.MAX study. We also thank Nathalie Druene-Pecollo, PhD (operational coordination), as well as all participants of the SU.VI.MAX study. The authors thank Dan Chatiel for his help in R programming. This work was conducted in the framework of the French network for Nutrition And Cancer Research (NACRe network), www.inra.fr/nacre and received the NACRe Partnership Label. Metabolomic analysis was performed within the metaboHUB French infrastructure (ANR-INBS-0010). This work was supported by the French National Cancer Institute (grant no. INCA_108065 for the project, and PhD grant no. INCA_11323 for L. Lyan), the Federation Institute for Biomedical Research (I2B2 Paris 13, and the Cancéropole Ile-de-France/Region Ile de France.

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 February 5, 2019; revised April 12, 2019; accepted May 28, 2019; published first June 4, 2019.


Plasma Metabolomic Signatures Associated with Long-term Breast Cancer Risk in the SU.VI.MAX Prospective Cohort

Lucie Lécuyer, Céline Dalle, Bernard Lyan, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-19-0154

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2019/06/01/1055-9965.EPI-19-0154.DC1

Cited articles
This article cites 45 articles, 6 of which you can access for free at:
http://cebp.aacrjournals.org/content/28/8/1300.full#ref-list-1

Citing articles
This article has been cited by 1 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/28/8/1300.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link:
http://cebp.aacrjournals.org/content/28/8/1300.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.