

Exposome Epidemiology for Suspect Environmental Chemical Exposures during Pregnancy Linked to Subsequent Breast Cancer Diagnosis

Young-Mi Go¹, Jaclyn Weinberg¹, Sami Teeny¹, Piera Cirillo², Nickilou Krigbaum², Grant
Singer¹, ViLinh Ly¹, Barbara Cohn^{2†} and Dean P. Jones^{1†}

¹ Clinical Biomarkers Laboratory, Division of Pulmonary, Allergy, Critical Care, and Sleep
Medicine, Department of Medicine, Emory University, Atlanta, GA 30322

² The Center for Research on Women and Children's Health, Child Health and Development
Studies, Public Health Institute, Berkeley, CA 94709

Running Head: Exposome epidemiology of breast cancer

† Corresponding authors:

DPJ: Whitehead Biomedical Research Building, 615 Michael St, Room 205P

Atlanta, GA, 30322. Phone: 404-727-5970. Fax: 404-727-2974. E-mail: dpjones@emory.edu

BC: 1683 Shattuck Avenue, Suite B, Berkeley, CA 94709. E-mail: bcohn@chdstudies.org

1 **Author Contributions**

- 2 • Designed research – YMG, PC, DPJ, BC
- 3 • Performed research – YMG, NK, VL, GS
- 4 • Analyzed data – YMG, JW, ST, PC, NK, GS, DPJ, BC
- 5 • Wrote the paper – YMG, ST, NK, PC, DPJ, BC

6

7 **Declaration of Competing Financial Interests**

8 The authors declare they have no actual or potential competing financial interests.

ABSTRACT

9 Breast cancer is now the most common cancer globally, accounting for 12% of all new annual
10 cancer cases worldwide. Despite epidemiologic studies having established a number of risk
11 factors, knowledge of chemical exposure risks is limited to a relatively small number of
12 chemicals. In this exposome research study, we used non-targeted, high-resolution mass
13 spectrometry (HRMS) of pregnancy cohort biospecimens in the Child Health and Development
14 Studies (CHDS) to test for associations with breast cancer identified via the California Cancer
15 Registry. Second (T2) and third (T3) trimester archival samples were analyzed from 182 women
16 who subsequently developed breast cancer and 384 randomly selected women who did not
17 develop breast cancer. Environmental chemicals were annotated with the Toxin and Toxin-
18 Target Database (T3DB) for chemical signals that were higher in breast cancer cases and used
19 with an exposome epidemiology analytic framework to identify suspect chemicals and associated
20 metabolic networks. Network and pathway enrichment analyses showed consistent linkage in
21 both T2 and T3 to inflammation pathways, including linoleate, arachidonic acid and
22 prostaglandins, and identified new suspect environmental chemicals associated with breast
23 cancer, i.e., an N-substituted piperidine insecticide and a common commercial product, 2,4-
24 dinitrophenol (DNP), linked to variations in amino acid and nucleotide pathways in T2 and
25 benzo[a]carbazole and a benzoate derivative linked to glycan and amino sugar metabolism in T3.
26 The results identify new suspect environmental chemical risk factors for breast cancer and
27 provide an exposome epidemiology framework for discovery of suspect environmental
28 chemicals and potential mechanistic associations with breast cancer.

29

30 **Keywords:** Cancer, exposome, metabolomics, pregnancy

31 **Abbreviations**

32 BC, breast cancer; FDR, false discovery rate; HRM, high-resolution metabolomics; HRMS,
33 high-resolution mass spectrometry; LC-HRMS, liquid chromatography-high resolution mass
34 spectrometry; MWAS, metabolome-wide association study; m/z , mass to charge; RT, retention
35 time; T2, second trimester; T3, third trimester, T3DB, the toxin and toxin-targeted database.

36 **1. Introduction**

37 High-resolution metabolomics (HRM) uses advanced mass spectrometry and data science for
38 non-targeted analysis of biological samples to obtain insight into an individual's metabolism
39 along with the individual's exposome, including chemicals derived from diet, intestinal
40 microbiome, dietary supplements, pharmaceuticals, personal use products and environmental
41 exposures (Jones *et al.* 2012). The metabolome is a functional readout of the interactions of a
42 person's genes with exposures from diet and environment (Johnson *et al.* 2017; Jones *et al.* 2016;
43 Niedzwiecki *et al.* 2019) and has been used extensively in combination with exposure
44 measurements, such as air pollution (Liang *et al.* 2018; Ritz *et al.* 2022) and targeted
45 environmental chemicals, such as DDT, PCBs and PBB (Hu *et al.* 2020; Walker *et al.* 2019), to
46 gain understanding of functional associations with exposures. Use in a non-targeted manner for
47 discovery of new chemical associations with disease, however, are limited. A vision for
48 development of exposome epidemiology, i.e., using omics scale biomonitoring for disease
49 epidemiology, is available (Jones and Cohn 2020). The current study was undertaken to apply
50 exposome epidemiology concepts for discovery of potential new environmental factors
51 contributing to breast cancer. The long-term goal for complex diseases such as breast cancer,
52 which involves multiple genetic, behavioral and environmental factors, is to integrate omics
53 scale biomonitoring with other factors to ultimately improve prediction and intervention
54 strategies to decrease disease burden.

55 To perform a non-targeted exposome-wide association study of breast cancer, we used
56 archival blood samples from the, Child Health and Development Studies (CHDS) cohort
57 collected during pregnancy from 1959-1967. Second (T2) and third (T3) trimester archival

58 samples, collected from 1959-1967, of 182 women who subsequently developed breast cancer,
59 were compared to those from 384 women who did not develop breast cancer.

60 We developed an exposome epidemiology approach (**Fig 1**) in which we first performed
61 non-targeted statistical analysis to select HRM features that were higher in cases than in non-
62 cases. In this selection, we use raw $p < 0.05$ as a cutoff for selection based upon the expectation
63 that multiple environmental factors with small effect size may contribute to breast cancer, and
64 this more liberal selection criteria is more appropriate for a small-population discovery study
65 because it provides a better balance between Type 1 and Type 2 statistical error than more
66 rigorous False Discovery Rate (FDR) criteria. We selected HRM features that were higher in
67 cases than non-cases to search for potentially causative factors by matching to a xenobiotic
68 database, Toxin and Toxin-Target Database (T3DB) (Lim *et al.* 2010; Wishart *et al.* 2015),
69 recognizing that protective factors that are decreased in association with breast cancer would
70 prioritize use of HRM features that were lower in cases and an alternative database, such as Food
71 Database (FoodDB) (Scalbert *et al.* 2011). HRM features that were increased with breast cancer
72 and annotated as environmental chemicals were then used for network analysis with the data-
73 dependent community detection tool, xMWAS (Uppal *et al.* 2018). Pathway enrichment
74 analysis of the most central communities was then performed with *mummichog* (Li *et al.* 2013),
75 and targeted mass spectrometry was used to improve understanding of the annotated chemicals
76 associated with breast cancer outcome. The results provide evidence for new candidate
77 environmental breast carcinogens and information concerning possible mechanisms by which
78 these agents contribute to breast cancer.

79

80 **2. Materials and methods**

81 *2.1 Samples and Assays*

82 The CHDS recruited women residing in the Oakland and East Bay, California area who
83 were members of the Kaiser Foundation Health Plan and received obstetric care for pregnancies
84 between 1959 and 1966 with deliveries extending into 1967 (van den Berg *et al.* 1988). More
85 than 98% of all eligible women enrolled.

86 Blood samples were collected from these mothers during pregnancy in each trimester and
87 in the early post-partum period, processed to isolate serum, and stored since then at -20° Celsius.
88 Second and third trimester archival samples available were analyzed for HRM and
89 environmental chemicals. The samples used for the present study include second (T2, n=182))
90 and third (T3, n=172)) trimester archival samples of 201 women who subsequently developed
91 breast cancer, compared to second (T2, n=384) and third (T3, n=351) trimester archival samples
92 from 413 women who did not develop breast cancer. Breast cancer cases were identified by
93 linkage to the California Cancer Registry for cases diagnosed through 1997. Record abstraction
94 for cancer diagnoses to the California Cancer Registry is based primarily on pathology reports,
95 and case identification is considered to be >99% complete after a 2-year lag (Perkins *et al.*).
96 Cases were defined as mothers with incident invasive or noninvasive breast cancer diagnosed at
97 a median age at diagnosis of 54 years (interquartile range, 13 years) with available prenatal
98 serum and a standardized gross placental exam. Non-cases were an 8% sample of women not
99 known to have breast cancer randomly selected among mothers with available prenatal serum
100 and a standardized gross placental exam (Cohn *et al.* 2017). Breast cancer rates for included and
101 excluded subsets in the CHDS cohort were highly comparable suggesting selection did not
102 impose significant bias: 1.89 per 1,000 person-years (95% Confidence Interval ([95% CI] = 1.60,

103 2.23) for included vs. 1.88 per 1,000 (95% CI=1.59, 2.20) for excluded. It is possible, however,
104 that we missed some cases of breast cancer, including among women who are identified in this
105 study as “non-cases”. In this case, failure to identify cases among the non-case group would
106 result in underestimating differences between cases and non-cases in these analyses and would
107 not be expected to impact findings. The CHDS founding mothers voluntarily participated in an
108 in-person interview and gave permission to access their own medical records and those of their
109 children to researchers. The institutional review board of the Public Health Institute approved the
110 present study (IRB#120-002), and we complied with all federal guidelines governing the use of
111 human participants. Forty-six percent of cases and forty-five percent of the non-cases had
112 available data and serum.

113

114 2.2 Chemicals

115 HPLC grade acetonitrile and methanol, LC-MS water and 98% formic acid were obtained from
116 Sigma-Aldrich (St. Louis, MO). A mixture of 14 stable isotopic chemicals were used as an
117 internal standard (Go *et al.* 2015) included [$^{13}\text{C}_6$]-D-glucose, [^{15}N]-indole, [2- ^{15}N]-L-lysine
118 dihydrochloride, [$^{13}\text{C}_5$]-L-glutamic acid, [$^{13}\text{C}_7$]-benzoic acid, [3,4- $^{13}\text{C}_2$]-cholesterol, [^{15}N]-L-
119 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,
120 [^{15}N , $^{13}\text{C}_5$]-L-methionine, [^{15}N]-choline chloride and 2'-deoxyguanosine- $^{15}\text{N}_2$, $^{13}\text{C}_{10}$ -5'-
121 monophosphate from Cambridge Isotope Laboratories, Inc (Andover, PA).

122

123 2.3 High-resolution mass spectrometry. Serum samples were analyzed with liquid
124 chromatography-high resolution mass spectrometry (LC-HRMS) as described previously (Jarrell
125 *et al.* 2021; Jarrell *et al.* 2020). Briefly, 50 μL of serum was treated 2:1 (v/v) with acetonitrile,

126 and 2.5 μL of the stable isotope standard mixture was added. Proteins were precipitated by
127 incubation at 4°C for 30 min and removed by centrifugation for 10 min at 21000 x g at 4°C.
128 Supernatants were placed in autosampler vials and maintained at 4°C in an autosampler until
129 analysis. Two pooled human reference samples including NIST SRM1950 and Qstd [pooled
130 plasma from 2 separate lots from Equitech-Bio, Inc (Kerrville, Texas)] were included. NIST
131 SRM1950 was run at the beginning and end of the full sample set, and Qstd was included at the
132 beginning and end of each batch of 20 samples.

133 Samples and reference materials were analyzed with three technical replicates using a
134 High-Field Q-Exactive mass spectrometer (Thermo Fisher) with C18 chromatography and
135 electrospray ionization (ESI) in negative mode. Data collection occurred continuously
136 throughout 5 min of chromatographic separation from 85 to 1,275 mass-to-charge ratio (m/z).
137 Data extraction was performed using *apLCMS* and *xMSanalyzer*, generating mass spectral
138 features consisting of m/z , retention time (RT) and peak intensity. Feature and sample filtering
139 retained features with a median CV of 50% or less, a minimum mean Pearson correlation
140 coefficient of 0.7 between technical replicates of each sample, and presence in at least 30% of
141 samples.

142
143 *2.4 Selection of features associated with subsequent breast cancer.* A Metabolome-Wide
144 Association Study (MWAS) was used to identify m/z features associated with breast cancer
145 outcome. Intensity values were \log_2 transformed, and LIMMA was performed using the R
146 package, *xmsPANDA* (<https://github.com/kuppall2/xmsPANDA>), with retention of features at raw
147 $p < 0.05$.

148

149 *2.5 Annotation of Mass Spectral Features and Metabolite Identification.* To select mass spectral
150 features of environmental interest, features positively associated with subsequent breast cancer
151 diagnosis were subjected to a multistage clustering algorithm and annotated with xMSannotator
152 (Uppal *et al.* 2017) using T3DB (Lim *et al.* 2010; Wishart 2015) at 5 ppm tolerance. For
153 selected chemicals, identities were confirmed when possible by accurate m/z match, co-elution
154 with authentic standards and ion dissociation mass spectrometry (MS/MS); Level 1 identification
155 by criteria of Schymanski *et al.* (Schymanski *et al.* 2014)].

156
157 *2.6 Targeted mass spectrometry analysis in identification of features.* Annotated chemicals at the
158 center of the *xMWAS* network structures underwent additional investigation by ion fragmentation,
159 structure analysis, and spectral matching to publicly available spectral libraries. Plasma samples
160 with the highest intensities of the chemicals of interest were selected and analyzed with C18
161 liquid chromatography (Dionex Ultimate 3000) and MS² ion fragmentation (Thermo Scientific
162 Fusion) with negative ESI. The mass spectrometer was set to scan a minimal range of m/z 's at
163 60k resolution for MS¹. Then, an inclusion list isolated the m/z of interest. MS² fragmentation
164 was acquired in HCD mode with a normalized collision energy of 30% and analyzed at 30,000
165 resolution using the dual-pressure linear ion trap. Raw MS¹ and MS² spectra were analyzed in
166 xCalibur QualBrowser (Thermo). Fragmentation patterns were analyzed with ChemDraw
167 (PerkinElmer Informatics) and compared against experimental spectra in the MassBank of North
168 America (MoNA) library. MoNA MS² spectra were selected for negative ESI and similar
169 collision energy. When MoNA spectra were not available, experimental spectra were searched
170 against the NIST Tandem Mass Spectral Library, which provides a spectral match percentage but
171 no information on collision energy or ionization polarity. To align with the goal of high-

172 throughput analyses, chromatographic gradients and fragmentation settings were maintained to
173 be consistent with the original analysis and not extensively optimized for each chemical. Under
174 these conditions, the complexity of the plasma sample matrix and the low abundance of
175 environmental chemicals does not allow for the isolation of one precursor and its fragments.
176 Therefore, in some spectra, the precursor mass is easily visible within 10 ppm error, but the
177 fragments are not clearly visible due to interference from other ions. Spectral peaks
178 corresponding to fragments were identified by their theoretical structures in “ChemDraw” with
179 unit resolution in the dual-pressure linear ion trap detector.

180

181 *2.7 Network and Pathway Analyses.* Annotated environmental chemicals were tested for
182 associations with m/z features using *xMWAS* based on partial least-squares regression (Uppal *et*
183 *al.* 2018). Thresholds for inclusion in the networks were $|r| > 0.30$ and $p < 0.05$. Pairwise results
184 from MWAS of the metabolites used for generation of network structures were used for pathway
185 enrichment analysis using *mummichog (v3)* (Li *et al.* 2013). Enriched pathways were filtered for
186 those that included at least 3 significantly associated metabolites at $p < 0.05$.

187

188

189 **3. RESULTS**

190

191 *3.1 Study population demographics*

192 T2 and T3 samples were available from 201 individuals who subsequently developed breast
193 cancer. These individuals had a median age of 31 y at time of blood collection, and the
194 individuals without breast cancer had a median age of 36 ($p < 0.0001$). Serum collection occurred
195 between 1960 and 1964, with T2 blood collection occurring at a median gestational age of 161
196 and 160 days in cases and non-cases, respectively. For T3, blood collections were at a median
197 gestational age of 249 and 250 days, respectively. Characteristics of cases and non-cases are
198 provided in **Table 1**.

199

200 *3.2 Metabolic feature profiling on T2 and T3 serum*

201 To select m/z features which were positively associated with subsequent breast cancer diagnosis,
202 we retained 9042 m/z features present in at least 80% of samples for statistical analyses and
203 selection of features for annotation. For T2 and T3 measurements, no features had $FDR < 0.05$
204 and therefore none are likely to be useful biomarkers for breast cancer risk. Any of the m/z
205 features selected with raw $p < 0.05$ could contribute to breast cancer risk, and 521 features had
206 $p < 0.05$ for T2, and 557 features had $p < 0.05$ for T3. Of these m/z features, 188 were higher in
207 breast cancer than non-cases in T2 samples while 151 were higher in breast cancer than non-
208 cases in T3 samples (**Table 2**). These positively associated m/z features were selected to search
209 for possible matches to environmental chemicals.

210

211 *3.3 Environmental chemical annotation of suspect chemicals in breast cancer group*

212 Most of the selected m/z features are likely to represent endogenous metabolites, but
213 environmental chemicals contributing to subsequent breast cancer could also be present. To
214 search for suspect environmental chemicals associated with subsequent breast cancer diagnosis,
215 we used xMSannotator, a network-based computational tool, with the Toxin and Toxin-Target
216 Database, T3DB (Wishart 2015; Wishart *et al.* 2007), to further characterize the mass spectral
217 features which were higher in their intensities in women who went on to develop breast cancer.
218 xMSannotator uses a multistage clustering algorithm in which intensity profiles, retention time
219 characteristics, mass defect, and isotope/adduct patterns are used to assign confidence levels to
220 annotation results relative to publicly available databases, such as T3DB. The annotated
221 environmental chemicals for T2 and T3 are shown in **Table 3** and **Table 4**, respectively. More
222 environmental chemicals were annotated in T2 than T3 serum, yielding 17 and 7, respectively,
223 and these included matches to chemicals widely used for anti-inflammatory, anticonvulsant and
224 antipsychotic drugs, quaternary ammonium salt, pesticides, herbicides, fungicides, plasticizers,
225 preservatives, cleaning materials, and flavorings. Importantly, at this level of investigation, these
226 annotations are only suspect chemicals and require further investigation and verification (see
227 below). The magnitude of difference was between 10% and 30% higher in women who went on
228 to develop breast cancer compared to those who did not ($p \leq 0.05$, **Tables 3 and 4**).

229

230 *3.4 Metabolome-Wide Association Study (MWAS) of T2 Environmental Chemicals*

231 Prior studies have identified inflammatory lipid pathways (linoleate, arachidonate,
232 prostaglandin), lipid and energy metabolism pathways (fatty acids, TCA cycle), oxidative stress
233 pathways (methionine and cysteine), and nitrogen metabolism pathways (urea cycle, pyrimidine,
234 purine) as top metabolic pathway associations with environmental chemicals and breast cancer

235 (Hu *et al.* 2019; Li *et al.* 2020; Walker *et al.* 2019). To determine whether the features annotated
236 as environmental chemicals (**Table 3**) are associated with changes in metabolites and metabolic
237 pathways, we used data-driven integration and network analysis of the T3DB annotated features
238 and all 9042 extracted metabolic signals, with xMWAS (Uppal *et al.* 2018). As shown in **Fig 2**,
239 network analysis showed three metabolic communities, Community 1 (C1, orange), Community
240 2 (C2, blue) and Community 3 (C3, green). C1 was positively associated with m/z 315.1957
241 (F1, **Fig 2, Table 3**). We performed metabolic pathway enrichment analysis (**Fig 3**) on the 716
242 metabolic features associated with C1 using the *mummichog 3* software (Li *et al.* 2013).
243 Associated pathways included linoleate, arachidonate and prostaglandins (**Fig 3A**), which are
244 closely related to inflammation and oxidative stress. Database searches showed that m/z
245 315.1957 matched multiple lipids (e.g., prostaglandins, resolvins), consistent with the strong
246 associations with related lipid species having activities in inflammation. The vitamin A pathway
247 was also associated with C1 (**Fig 3A**), and targeted mass spectrometry with MS/MS analysis
248 showed a likelihood that m/z 315.1957 was 4-hydroxyretinoic acid (**F1, Supplemental Fig. S1**).
249
250 C1 extensively overlapped with Community 2 (C2, blue), which was positively associated with
251 m/z 334.2090 (F2, **Fig 2, Table 3**). In addition to pathway associations with prostaglandins,
252 linoleic acid and arachidonic acid, a subset of C2 included phospholipids, e.g., sphingosine,
253 phosphatidyl choline (PC), and triglyceride (blue, **Fig 2**), associated with glycerophospholipid
254 metabolism. Thus, this chemical defines a separate cluster of metabolites associated with breast
255 cancer than those in C1, but this feature remains unidentified because subsequent mass
256 spectrometry analysis failed to provide support for identification. Together, results from C1 and
257 C2 support prior findings that inflammatory lipid and other lipid pathways associated with DDT,

258 PCB and PBB are associated with breast cancer, but do not reveal any new suspect
259 environmental chemicals.
260
261 A third community, C3, had metabolomic associations (green) which were separated from C1
262 and C2 and positively associated with m/z features annotated as environmental chemicals,
263 including an insecticide (F3, nitromethylene-piperidine), a wood preservation and dye production
264 chemical (F4, 2,4-dinitrophenol), oxalate (F5), and F6, m/z 156.0666 and F7, m/z 345.0777 (**Fig**
265 **2, Table 3**). Of these, oxalate was previously confirmed, and supportive targeted mass
266 spectrometry data was obtained for 2-nitromethylene-piperidine and 2,4-dinitrophenol (**F3 and**
267 **F5, Supplemental Fig. S1**). The m/z feature 156.0666 (F6) was present at high concentrations
268 in most samples and appeared likely to be an endogenous metabolite, N-acetylproline (**F6,**
269 **Supplemental Fig. S1**). No supportive mass spectrometry data could be obtained for the m/z
270 feature 345.0777 (F7). The 56 metabolic features in C3 that were positively associated with
271 these annotated environmental chemicals were related to pathways for antioxidant methionine
272 (Met) and cysteine (Cys) regulation, urea cycle, amino acids (glycine, serine, alanine, threonine,
273 aspartate, asparagine, lysine), glycerophospholipids, and purine and pyrimidine nucleotide
274 metabolism (**Fig 3C**). Representative metabolites associated with these pathways of each
275 community are indicated next to each community (**Fig 2**) and included amino acids (glutamate,
276 serine, lysine, aspartate), a pyrimidine (orotidine), a purine (uric acid), and a coenzyme for
277 mitochondrial electron transfer and the TCA cycle (dihydrolipoamide). Collectively, the results
278 for T2 serum of women who subsequently developed breast cancer show that two suspect
279 environmental chemicals, nitromethylene-piperidine and 2,4-dinitrophenol, along with oxalate,
280 are closely associated with metabolic perturbations related to amino acid and nucleotide

281 metabolism. Other communities included lipids functioning in inflammation, sphingolipids and
282 lipids, but no suspect environmental chemicals were identified for these communities.

283

284 *3.5 MWAS of T3 Environmental Chemicals and alterations in metabolic pathway*

285 Following the same experimental approach shown above, we examined the relationship of
286 annotated matches of T3 suspect environmental chemicals (**Table 4**) and changes in metabolites
287 using xMWAS (Uppal *et al.* 2018). Two major metabolic communities were identified (**Fig 4**),
288 with the first community (labeled C4, orange) including metabolites associated exclusively with
289 m/z 271.1195 (F18) and second community (C5, blue) including metabolites associated with m/z
290 214.9728 (F19), m/z 217.0866 (F20) and m/z 215.0327 (F21) (**Fig 4**). Targeted mass
291 spectrometry analysis of m/z 271.1195 (F18) showed that this feature was likely ergosterol or
292 related derivative (**F18, Supplemental Fig. S2**). Pathway enrichment analysis of the 650
293 metabolic features of C4 showed associated pathways for arachidonate, leukotriene, linoleate,
294 vitamin A, prostaglandins, vitamin E and steroid hormones (**Fig 5A**). Most of these pathways
295 are closely related to inflammation and oxidative stress, as also found in T2 associations in C1
296 and C2 (**Fig 3A, 3B**). Targeted mass spectrometry analyses showed that m/z 214.9713 (F19) was
297 likely a bromine adduct of a benzoic acid metabolite (**F19, Supplemental Fig. S2**); m/z
298 217.0866 (F20) had an MS² spectrum with 83% match to benzo[a]carbazole (**F20,**
299 **Supplemental Fig. S2**); m/z 215.0327 (F21) was a high abundance signal from glucose (**F21,**
300 **Supplemental Fig. S2**). Pathway enrichment analysis of the 155 metabolic features in C5
301 showed associations for sialic acid, pentose phosphate, N-glycan degradation, galactose,
302 phosphatidylinositol, glycolysis and gluconeogenesis, butanoate, and glycosphingolipid
303 pathways (**Fig 5B**). Representative metabolites associated with these pathways are indicated next

304 to each community (**Fig 4**). Collectively, suspect environmental chemicals in Communities 4
305 and 5 include a benzoate derivative and benzo[a]carbazole, and metabolic pathway associations
306 suggesting that breast cancer effects could be mediated through inflammatory lipids, glucose-
307 related metabolism and estrogenic signaling.

308

309 *3.6 Comparisons of T2 and T3 suspect chemical-metabolic pathway associations*

310 In comparison of findings for T3 and T2, metabolic pathways for C4 in T3 showed considerable
311 overlap with the pathways associated with C1 and C2 in T2. In contrast, the remaining
312 community, C5, in T3 differed substantially from the remaining community, C3, in T2.
313 Specifically, in T3, we identified a suspect benzoate and benzo[a]carbazole in breast cancer
314 cases that were associated with pathways of carbohydrate metabolism functioning in
315 extracellular matrix turnover and complex carbohydrate metabolism. In T2, we identified a
316 suspect insecticide, 2-nitromethylpiperidine, a widely used commercial chemical, 2,4-
317 dinitrophenol, and oxalic acid, in breast cancer cases that were associated with central pathways
318 functioning in amino acid homeostasis, urea cycle, pyrimidine and purine metabolism, and
319 defense against oxidative stress.

320

321 **4. Discussion**

322 The present study applies a new exposome epidemiology approach to discover suspect
323 environmental chemicals that may be implicated in breast cancer development. New suspect
324 chemicals from this analysis include an insecticide, nitromethylene-piperidine; a common
325 commercial chemical, 2,4-dinitrophenol; a heterocyclic amine with diverse commercial
326 applications, benzo[a]carbazole; a benzoate derivative which could be derived from natural
327 products or environmental chemicals; and oxalate, an endogenous metabolite that is also used as
328 a cleaning agent. These chemicals were at higher abundance in serum of women decades before
329 breast cancer diagnosis. Importantly, the network analysis is data-dependent, raising the
330 possibility that individual chemicals could have a causal role in breast cancer or that breast
331 cancer could occur through interaction of multiple chemicals in causal mechanisms. In the latter
332 case, individual variations in exposure could have important impact on risk.

333
334 New approaches to identify suspect carcinogens are critical because research over the last 50
335 years has not translated to a strategy for individual breast cancer prevention. The suspect
336 chemical signal F18 (m/z 271.1195) in Community 4, was consistent with identification as
337 ergosterol, which is derived from yeast and precursor of vitamin D2. Pathway mapping of this
338 community showed overlap with inflammatory lipid pathways found in Communities 1 and 2,
339 and also included vitamin A metabolism, which is consistent with feature F1 in Community 1, 4-
340 hydroxyretinoic acid, being a suspect chemical. Many diet-derived carotenoids can be converted
341 to isobaric species to 4-hydroxyretinoic acid, and these are not distinguished by the HRM
342 methods used. Thus, future research will be needed to address the potential for diet-environment
343 interactions, such as implied by the overlap of estrogen-, vitamin D- and vitamin A-linked

344 metabolites. Pregnancy impacts initiation, progression, and susceptibility to breast cancer (Troisi
345 *et al.* 2018), and therefore pregnancy provides an appropriate time frame for study. Later age at
346 pregnancy is a long-established breast cancer risk factor (Albrektsen *et al.* 2005; MacMahon *et*
347 *al.* 1970) which has become increasingly common and is still unexplained. Because first birth
348 rates have increased 6-fold for women ages 35-39 from 1973-2006 (Martin *et al.* 2015), more
349 detailed understanding of the respective contributions and interactions must be a priority to learn
350 how to mitigate risk for this population group of higher risk women.

351
352 In T2, an additional chemical cluster containing an insecticide (nitromethylene-piperidine), a
353 cleaning agent (oxalate), a chemical used in wood preservation and dye production
354 (dinitrophenol), and two unidentified chemical features (F2, F7), were associated with multiple
355 amino acid pathways and nitrogen metabolism (urea cycle, pyrimidine, purine). The amino acid
356 pathways overlap with pathways previously found to vary with persistent organic pollutants,
357 *p,p'*-DDT (Hu *et al.* 2020) and PBBs (Walker *et al.* 2019). In a multigenerational study, *p,p'*-
358 DDT exposure in women before puberty was found to be associated with breast cancer in
359 mothers (Cohn *et al.* 2007) and in utero *o',p'*-DDT exposure was associated with breast cancer in
360 daughters (Cohn *et al.* 2015). Polybrominated biphenyls (PBB) are also persistent organic
361 pollutants that cause breast cancer (IARC 2015). Potential mechanistic connections between
362 dinitrophenol, nitromethylene-piperidine, oxalate, and these POPs are not apparent, and
363 mechanisms by which POPs disrupt central amino acid and nitrogen metabolism pathways are
364 not known. In particular, amino acids are essential for growth and development, and disruption
365 of pyrimidine and purine metabolism can be expected to have long-term consequences on

366 metabolic programming which could contribute to cancer susceptibility. Thus, these findings
367 warrant further investigation into underlying mechanisms related to breast cancer development.
368
369 In T3, the large metabolic community (C4, Fig 4) containing inflammatory lipids was linked to a
370 sparse community (C5, Fig 4) containing glucose, benzo[a]carbazole and a suspect benzoate
371 derivative. Of potential importance in this network, the glucose signal was negatively associated
372 with metabolites connected to sialic acid, N-glycan and other pathways associated with
373 extracellular matrix and turnover of connective tissue. Benzoate is conjugated with glycine for
374 elimination, and the role of glycine in folate-dependent metabolism through the vitamin B₆-
375 dependent hydroxymethyltransferase raises the possibility that benzoate derivatives could have
376 unrecognized pathogenic effects. Similarly, the heterocyclic amine, benzo[a]carbazole, could be
377 bioactivated by Cyp enzymes to generate mutagenic species which have not been characterized.
378 Future hypothesis-driven research will be needed to address these possibilities.

379
380 The exposome epidemiology approach used in the present study has important assumptions, i.e.,
381 that chemical exposures which increase cancer risk can occur decades before breast cancer
382 occurrence, that these exposures are detectable and at higher abundance in serum decades prior
383 to breast cancer detection, and that network analyses of HRM data are sufficient to detect these
384 exposures and link them to biologic responses. With these assumptions, all HRM features from
385 archival blood serum of women collected decades before breast cancer diagnosis can be used in a
386 non-targeted manner to select ones that are increased ($p < 0.05$) in association with breast cancer
387 occurrence. This selection criterion is not suitable for biomarker development because of
388 multiple testing; however, this cutoff is suitable for discovery of potential chemical risk factors

389 because any of the features at this cutoff could be correct. The selected HRM features are then
390 used with a toxic exposome database, T3DB, to obtain accurate mass matches to known toxic
391 chemicals, and these are subjected to non-targeted network and pathway enrichment analyses to
392 identify environmental chemical-metabolic network associations linked to breast cancer
393 outcome.

394
395 Exposome epidemiology will benefit from populations of tens or hundreds of thousands of
396 individuals by enabling detection of exposures impacting small numbers of women and/or
397 having only small contributions individually to breast cancer risk. The present study with only
398 hundreds rather than thousands of individuals has limited capability to detect carcinogens with
399 small effect size. Additionally, selection of HRM features for statistical selection in the current
400 study required that features be present in 80% of the samples. Environmental chemicals and
401 personal use products such as hair dyes cannot be expected to be present in 80% of the women
402 and unlikely to be detectable. Application of untargeted gas chromatography-high-resolution
403 mass spectrometry, which provides capabilities to measure other hydrophobic and volatile
404 environmental chemicals (Hu *et al.* 2021), can be used to enhance coverage of environmental
405 chemicals. Thus, future studies will benefit from larger population sizes, inclusion of
406 complementary chemical analyses, and use of statistical methods which can accommodate sparse
407 environmental chemical detection.

408

409 **5. Conclusion**

410 Recent technological and statistical advances in high-resolution metabolomics (HRM) provide
411 capabilities for omics scale biomonitoring of chemicals derived from the environment along with

412 endogenous metabolism and chemicals from the diet, intestinal microbiome, dietary supplements,
413 pharmaceuticals, and personal use products. Because the metabolome is a functional readout of
414 the interactions of a person's genes with exposures, HRM of biologic samples provides one of
415 the most accessible ways to connect environmental exposures with biologic status to anticipate
416 breast cancer. As shown in the present study, these can be used with network and pathway
417 analysis to identify suspect environmental chemicals and functional communities linked to breast
418 cancer outcome. Such network approaches can be broadly applied to discover how life-long
419 exposures impact personal cancer risks. For breast cancer, we believe that this approach will
420 yield critically needed protocols to enhance protection or mediate pregnancy-associated risk.

421

422 **Funding:** This study was supported by Department of Defense Grant W81XWH2010103, and
423 National Institute of Environmental Health and Science grant R21 ES031824, R01 ES031980
424 and P30 ES019776.

425

426 **References**

- 427
- 428 Albrektsen, G.; Heuch, I.; Hansen, S.; Kvale, G. Breast cancer risk by age at birth, time since
429 birth and time intervals between births: exploring interaction effects. *British Journal of*
430 *Cancer* 2005;92:167-175
- 431 Cohn, B.A.; Cirillo, P.M.; Hopper, B.R.; Siiteri, P.K. Third Trimester Estrogens and Maternal
432 Breast Cancer: Prospective Evidence. *J Clin Endocrinol Metab* 2017;102:3739-3748
- 433 Cohn, B.A.; La Merrill, M.; Krigbaum, N.Y.; Yeh, G.; Park, J.S.; Zimmermann, L.; Cirillo, P.M.
434 DDT Exposure in Utero and Breast Cancer. *J Clin Endocrinol Metab* 2015;100:2865-
435 2872
- 436 Cohn, B.A.; Wolff, M.S.; Cirillo, P.M.; Sholtz, R.I. DDT and breast cancer in young women:
437 new data on the significance of age at exposure. *Environ Health Perspect* 2007;115:1406-
438 1414
- 439 Go, Y.M.; Kim, C.W.; Walker, D.I.; Kang, D.W.; Kumar, S.; Orr, M.; Uppal, K.; Quyyumi, A.A.;
440 Jo, H.; Jones, D.P. Disturbed flow induces systemic changes in metabolites in mouse
441 plasma: a metabolomics study using ApoE(-)/(-) mice with partial carotid ligation. *Am J*
442 *Physiol Regul Integr Comp Physiol* 2015;308:R62-72
- 443 Hu, X.; Li, S.; Cirillo, P.; Krigbaum, N.; Tran, V.; Ishikawa, T.; La Merrill, M.A.; Jones, D.P.;
444 Cohn, B. Metabolome Wide Association Study of serum DDT and DDE in Pregnancy
445 and Early Postpartum. *Reprod Toxicol* 2020;92:129-137
- 446 Hu, X.; Li, S.; Cirillo, P.M.; Krigbaum, N.Y.; Tran, V.; Jones, D.P.; Cohn, B.A. Metabolome
447 Wide Association Study of Serum Poly and Perfluoroalkyl Substances (PFASs) in
448 Pregnancy and Early Postpartum. *Reprod Toxicol* 2019;87:70-78

449 Hu, X.; Walker, D.I.; Liang, Y.; Smith, M.R.; Orr, M.L.; Juran, B.D.; Ma, C.; Uppal, K.; Koval,
450 M.; Martin, G.S.; Neujahr, D.C.; Marsit, C.J.; Go, Y.M.; Pennell, K.D.; Miller, G.W.;
451 Lazaridis, K.N.; Jones, D.P. A scalable workflow to characterize the human exposome.
452 Nat Commun 2021;12:5575

453 IARC. Polychlorinated Biphenyls and Polybrominated Biphenyls ed^eds; 2015

454 Jarrell, Z.R.; Smith, M.R.; He, X.; Orr, M.; Jones, D.P.; Go, Y.M. Firsthand and Secondhand
455 Exposure Levels of Maltol-Flavored Electronic Nicotine Delivery System Vapors Disrupt
456 Amino Acid Metabolism. Toxicol Sci 2021;182:70-81

457 Jarrell, Z.R.; Smith, M.R.; Hu, X.; Orr, M.; Liu, K.H.; Quyyumi, A.A.; Jones, D.P.; Go, Y.M.
458 Plasma acylcarnitine levels increase with healthy aging. Aging (Albany NY)
459 2020;12:13555-13570

460 Johnson, C.H.; Athersuch, T.J.; Collman, G.W.; Dhungana, S.; Grant, D.F.; Jones, D.P.; Patel,
461 C.J.; Vasiliou, V. Yale school of public health symposium on lifetime exposures and
462 human health: the exposome; summary and future reflections. Human genomics
463 2017;11:32

464 Jones, D.P.; Cohn, B.A. A vision for exposome epidemiology: The pregnancy exposome in
465 relation to breast cancer in the Child Health and Development Studies. Reprod Toxicol
466 2020;92:4-10

467 Jones, D.P.; Park, Y.; Ziegler, T.R. Nutritional metabolomics: progress in addressing complexity
468 in diet and health. Annu Rev Nutr 2012;32:183-202

469 Jones, D.P.; Walker, D.I.; Uppal, K.; Rohrbeck, P.; Mallon, C.T.; Go, Y.M. Metabolic Pathways
470 and Networks Associated With Tobacco Use in Military Personnel. Journal of
471 occupational and environmental medicine 2016;58:S111-116

- 472 Li, S.; Cirillo, P.; Hu, X.; Tran, V.; Krigbaum, N.; Yu, S.; Jones, D.P.; Cohn, B. Understanding
473 mixed environmental exposures using metabolomics via a hierarchical community
474 network model in a cohort of California women in 1960's. *Reprod Toxicol* 2020;92:57-65
- 475 Li, S.; Park, Y.; Duraisingham, S.; Strobel, F.H.; Khan, N.; Soltow, Q.A.; Jones, D.P.; Pulendran,
476 B. Predicting network activity from high throughput metabolomics. *PLoS Comput Biol*
477 2013;9:e1003123
- 478 Liang, D.; Moutinho, J.L.; Golan, R.; Yu, T.; Ladva, C.N.; Niedzwiecki, M.; Walker, D.I.;
479 Sarnat, S.E.; Chang, H.H.; Greenwald, R.; Jones, D.P.; Russell, A.G.; Sarnat, J.A. Use of
480 high-resolution metabolomics for the identification of metabolic signals associated with
481 traffic-related air pollution. *Environ Int* 2018;120:145-154
- 482 Lim, E.; Pon, A.; Djoumbou, Y.; Knox, C.; Shrivastava, S.; Guo, A.C.; Neveu, V.; Wishart, D.S.
483 T3DB: a comprehensively annotated database of common toxins and their targets.
484 *Nucleic Acids Res* 2010;38:D781-786
- 485 MacMahon, B.; Cole, P.; Lin, T.M.; Lowe, C.R.; Mirra, A.P.; Ravnihar, B.; Salber, E.J.;
486 Valaoras, V.G.; Yuasa, S. Age at first birth and breast cancer risk. *Bull World Health*
487 *Organ* 1970;43:209-221
- 488 Martin, J.; Hamilton, B.; Osterman, M.; Curtin, S.; Matthews, T. Births: final data for 2013.
489 National vital statistics reports: from the Centers for Disease Control and Prevention,
490 National Center for Health Statistics, National Vital Statistics System 2015;64:1
- 491 Niedzwiecki, M.M.; Walker, D.I.; Vermeulen, R.; Chadeau-Hyam, M.; Jones, D.P.; Miller, G.W.
492 The Exposome: Molecules to Populations. *Annu Rev Pharmacol Toxicol* 2019;59:107-
493 127

- 494 Perkins, C.; Wright, W.; Schlag, R. Cancer Incidence and Mortality in California, 1988-1994,
495 CA: California Department of Health Services, Cancer Surveillance Section.
- 496 Ritz, B.; Yan, Q.; He, D.; Wu, J.; Walker, D.I.; Uppal, K.; Jones, D.P.; Heck, J.E. Child serum
497 metabolome and traffic-related air pollution exposure in pregnancy. *Environ Res*
498 2022;203:111907
- 499 Scalbert, A.; Andres-Lacueva, C.; Arita, M.; Kroon, P.; Manach, C.; Urpi-Sarda, M.; Wishart, D.
500 Databases on food phytochemicals and their health-promoting effects. *J Agric Food*
501 *Chem* 2011;59:4331-4348
- 502 Schymanski, E.L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H.P.; Hollender, J.
503 Identifying small molecules via high resolution mass spectrometry: communicating
504 confidence. *Environ Sci Technol* 2014;48:2097-2098
- 505 Troisi, R.; Bjørge, T.; Gissler, M.; Grotmol, T.; Kitahara, C.M.; Myrtveit Saether, S.; Ording,
506 A.G.; Sköld, C.; Sørensen, H.T.; Trabert, B. The role of pregnancy, perinatal factors and
507 hormones in maternal cancer risk: a review of the evidence. *J Intern Med* 2018;283:430-
508 445
- 509 Uppal, K.; Ma, C.; Go, Y.M.; Jones, D.P.; Wren, J. xMWAS: a data-driven integration and
510 differential network analysis tool. *Bioinformatics* 2018;34:701-702
- 511 Uppal, K.; Walker, D.I.; Jones, D.P. xMSannotator: An R Package for Network-Based
512 Annotation of High-Resolution Metabolomics Data. *Anal Chem* 2017;89:1063-1067
- 513 van den Berg, B.J.; Christianson, R.E.; Oechsli, F.W. The California Child Health and
514 Development Studies of the School of Public Health, University of California at Berkeley.
515 *Paediatr Perinat Epidemiol* 1988;2:265-282

516 Walker, D.I.; Marder, M.E.; Yano, Y.; Terrell, M.; Liang, Y.; Barr, D.B.; Miller, G.W.; Jones,
517 D.P.; Marcus, M.; Pennell, K.D. Multigenerational metabolic profiling in the Michigan
518 PBB registry. *Environ Res* 2019;172:182-193

519 Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.;
520 Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; Bouwman, J.; Brookes,
521 A.J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C.T.; Finkers, R.;
522 Gonzalez-Beltran, A.; Gray, A.J.; Groth, P.; Goble, C.; Grethe, J.S.; Heringa, J.; t Hoen,
523 P.A.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S.J.; Martone, M.E.; Mons, A.;
524 Packer, A.L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.A.;
525 Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M.A.; Thompson, M.; van der
526 Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.;
527 Zhao, J.; Mons, B. The FAIR Guiding Principles for scientific data management and
528 stewardship. *Sci Data* 2016;3:160018

529 Wishart, D.; Arndt, D.; Pon, A.; Sajed, T.; Guo, A.C.; Djoumbou, Y.; Knox, C.; Wilson, M.;
530 Liang, Y.; Grant, J.; Liu, Y.; Goldansaz, S.A.; Rappaport, S.M. T3DB: the toxic
531 exposome database. *Nucleic Acids Res* 2015;43:D928-934

532 Wishart, D.S. Identifying putative drug targets and potential drug leads: starting points for virtual
533 screening and docking. *Methods Mol Biol* 2015;1215:425-444

534 Wishart, D.S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A.C.; Young, N.; Cheng, D.; Jewell, K.;
535 Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M.A.; Forsythe, I.;
536 Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D.D.;
537 Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan,
538 G.E.; Macinnis, G.D.; Weljie, A.M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner,

- 539 R.; Li, L.; Marrie, T.; Sykes, B.D.; Vogel, H.J.; Querengesser, L. HMDB: the Human
540 Metabolome Database. *Nucleic Acids Res* 2007;35:D521-526

541 **Figure Legends**

542

543 **Figure 1.** Exposome Epidemiology Workflow. The steps used in the current analysis are
544 summarized in general form. In the first step, the study population and sample types can be
545 selected for diverse purposes. In the second step, liquid chromatography, gas chromatography or
546 other separation methods can be used with high-resolution mass spectrometry to generate tens of
547 thousands of mass spectral features. For this purpose, multiple data extraction tools are
548 available, and extraction parameters and filtration criteria can be selected to increase stringency
549 or leniency in feature retention. For the third step, directionality of change is selected for
550 features expected to have a causal or protective activity. For the fourth step, selected features are
551 annotated using a relevant database. In the fifth step, selected and annotated features are used
552 with network analysis to identify associated communities of mass spectral features. The sixth
553 step applies pathway enrichment analysis to the associated communities of mass spectral features
554 to characterize the associated biologic response. Finally, targeted mass spectral analysis of
555 relevant mass spectral features provides prioritization for subsequent hypothesis-driven research
556 to verify environmental chemical associations and underlying causal mechanisms.

557

558 **Figure 2.** Metabolome-Wide Association Study (MWAS) of T3DB-annotated chemicals higher
559 in second trimester (T2) serum. Association of seventeen T3DB-annotated features (Table 3) that
560 are higher in breast cancer than non-cases with the metabolome (9,042 metabolic features) from
561 T2 serum (n=384 for non-cases, n=182 for breast cancer) are examined using xMWAS. The
562 outcome of xMWAS analysis is visualized with two separate networks at $|r| \geq 0.3$. The networks
563 include three communities (C1, C2, C3) and show central seven chemicals (F1-F7) with tight

564 connections [C1 (orange), C2 (blue), C3 (green)]. Representative metabolites of pathways
565 associated with community are shown next to each community. DHET: dihydroxyeicosatrienoic
566 acid, EETA: epoxyeicosatrienoic acid, TG: triglyceride, PC: phosphatidylcholine.

567
568 **Figure 3.** Metabolic pathway associated with second trimester (T2) chemicals. Pathway
569 enrichment analysis with metabolites of three network communities in **Fig. 2** was conducted
570 using mummichog. **A)** A total of seven pathways were found altered with m/z 315.1957 (F1) in
571 **C1**, **B)** Eight pathways were altered with m/z 334.2090 (F2) in C2, and **C)** Seven pathways were
572 altered with environmental chemicals and drug metabolites (F3-F7) ($p < 0.05$). The ratio of
573 selected metabolites mapped to the listed pathway over the number of total pathway metabolites
574 detected is provided to the right of each bar.

575
576 **Figure 4.** Metabolome-Wide Association Study (MWAS) of T3DB-annotated chemicals higher
577 in third trimester (T3) serum. Association of seven T3DB-annotated chemicals (Table 4) that are
578 higher in breast cancer than non-cases with the metabolome (9,042 metabolic features) from T3
579 serum ($n=351$ for non-cases, $n=172$ for breast cancer) are examined using xMWAS. The network
580 is visualized at $|r| \geq 0.3$ and includes two communities (C4 and C5) showing central four
581 chemicals with tight connections [C4 (orange)] than C5 (blue)]. Representative metabolites of
582 pathways associated with community are shown next to each community.

583
584 **Figure 5.** Metabolic pathway associated with T3 chemicals. Pathway enrichment analysis with
585 metabolites of two network communities in **Fig. 4** was conducted using mummichog. **A)** Seven
586 pathways were found altered with F18 chemical in C1 and **B)** eight pathways were altered with

587 chemicals (F19-F21) in C2 ($p < 0.05$). The ratio of selected metabolites mapped to the listed
588 pathway over the number of total pathway metabolites detected is provided to the right of each
589 bar.

590 **Table 1.** Characteristics of Study Population

Characteristic	Non-Cases (N=413)		Cases (N=201)	
	Median	(IQR)	Median	(IQR)
Year of mother's birth	1936	(9)	1931	(11)
Year of blood draw	1962	(2)	1962	(2)
Gestational day of T2 blood draw	160	(33)	161	(29)
Gestational day of T3 blood draw	250	(17)	249	(16)
BMI at first prenatal visit (kg/m ²) ^a	22.1	(4.3)	22.6	(4.2)
Characteristic	Percent		Percent	
Race				
non-Hispanic Caucasian	70		69	
African American	20		19	
Hispanic	2		1	
Asian	6		7	
Mixed	2		3	

591 **Table 2.** HRM profiling of T2 and T3 serum samples comparing breast case and non-case
592 groups ($P \leq 0.05$). BC, breast cancer.

	T2 (182 BC, 384 non-cases)	T3 (172 BC, 351 non-cases)
Total number of metabolic features	9042	9042
Features differing at. $p < 0.05$	521	557
Differing features with BC/non-cases ratio > 1.0	188	151

593 **Table 3.** T3DB-annotated m/z features higher in T2 serum of women who subsequently
 594 developed breast cancer (BC) compared with non-cases ($P \leq 0.05$). Targeted mass spectrometry
 595 analysis was performed to aid in identification of features F1 to F7 which were found to occur in
 596 central communities of Fig 2, indicated by asterisk (*). T3DB annotations are provided for others
 597 (F8-F17).

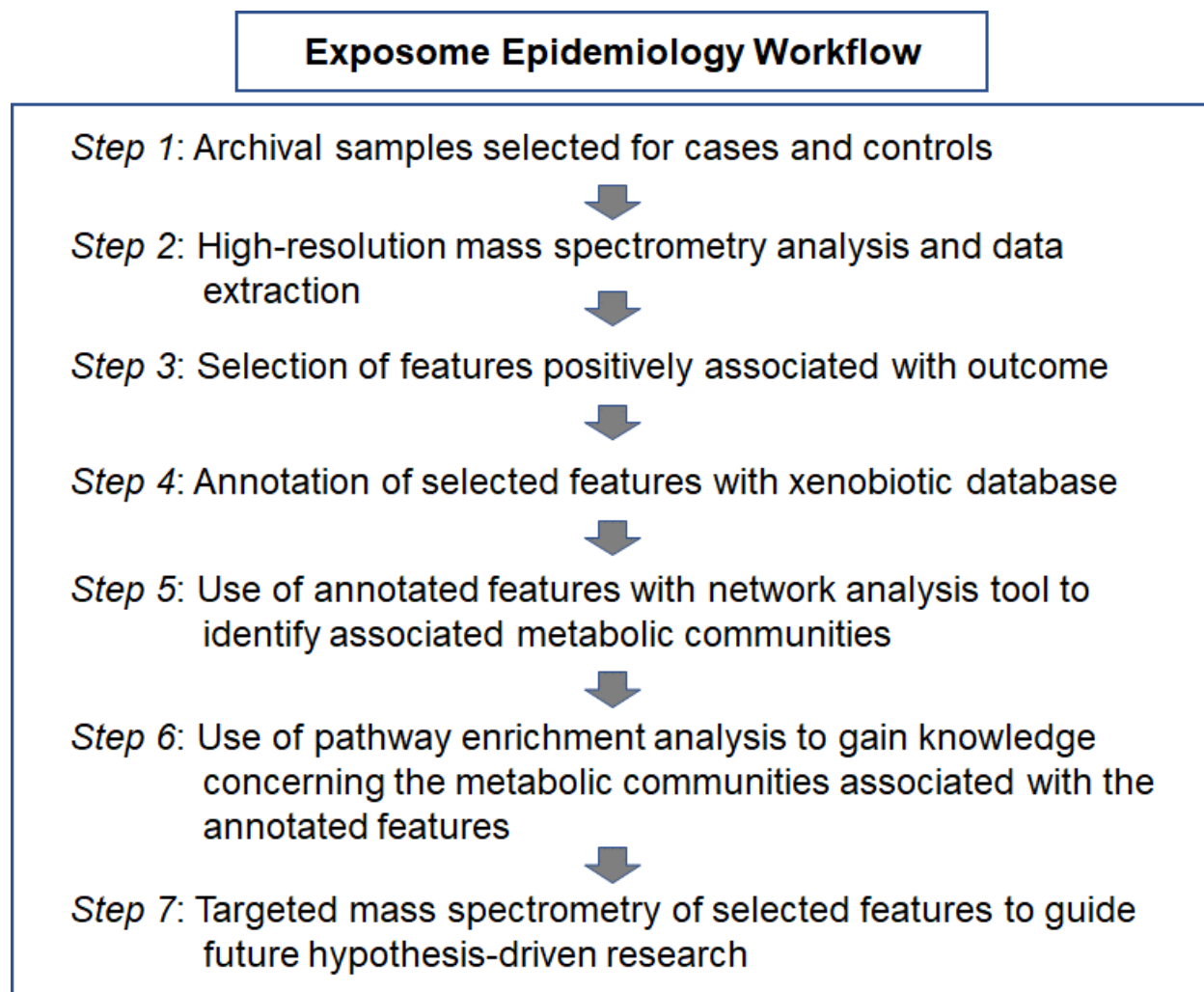
Feature	Name	m/z	RT (sec)	% Higher in BC	Category
F1*	4-Hydroxyretinoic acid	315.1957	177	20	Vitamin A metabolite
F2*	Unidentified	334.2090	125	20	Unidentified
F3*	2-Nitromethylene-piperidine	141.0670	278	10	Insecticide
F4*	2,4-Dinitrophenol	183.0046	283	10	Multiple commercial uses
F5*	Oxalic Acid	88.9883	129	10	Cleaning agent
F6*	N-acetyl-proline	156.0666	283	10	Endogenous metabolite
F7*	Unidentified	345.0777	51	20	Unidentified
F8	Flurbiprofen	225.0712	139	30	Anti-inflammatory drug

F9	Quaternium-52	610.4093	195	20	Quaternary ammonium salt
F10	Ethyl cyanoacrylate	124.0404	14	10	Cyanoacrylate glue component
F11	2-Chloro-4,5-xylol N-hydroxy-N-methylcarbamate	228.0427	284	10	Cholinesterase inhibitor
F12	Psoralen, Angelicin	167.0135	20	20	Skin nodule treatment,
F13	Hexazinone	273.1344	32	10	Herbicide
F14	MHP (Methyl hydrogen phthalate)	179.0350	138	10	Organic compound
F15	2-Hydroxyphenyl methylcarbamate	166.0508	212	10	Pesticide
F16	Levodopa	178.0510	23	20	Dopamine precursor drug
F17	Phenylmercuric acetate	337.0157	22	20	Preservative

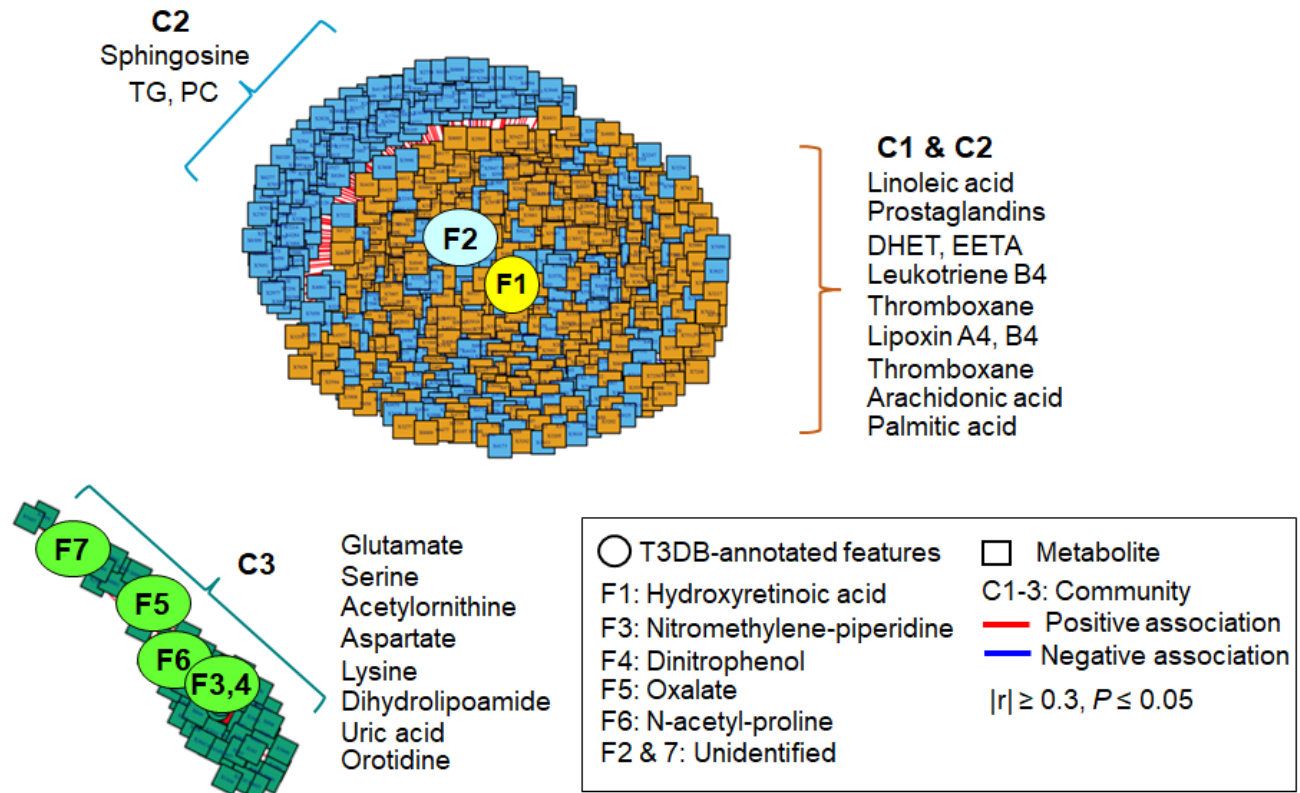
598 **Table 4.** T3DB-annotated m/z features higher in T3 serum of women who subsequently
599 developed breast cancer (BC) compared with those in the non-case women ($P \leq 0.05$). Targeted
600 mass spectrometry analysis was performed to aid in identification of features F18-F21 which
601 were found to occur in central communities of Fig 4, indicated by asterisk (*). T3DB annotation
602 is provided for others (F22-F24).

Feature	Name	m/z	RT (sec)	% Higher in BC	Category
F18*	Ergosterol	271.1195	25	10	Vitamin D2 precursor
F19*	Benzoate derivative	214.9728	22.8	10	Food preservative
F20*	Benzo[a]carbazole	217.0866	21	10	Heterocyclic amine
F21*	Glucose	215.0327	23	10	Endogenous metabolite
F22	Trimethadione	142.0509	277	10	Anticonvulsant
F23	Heptabromobiphenyl	734.4246	195	20	Aromatic compound
F24	Chlorpromazine	317.0876	135	10	Antipsychotics

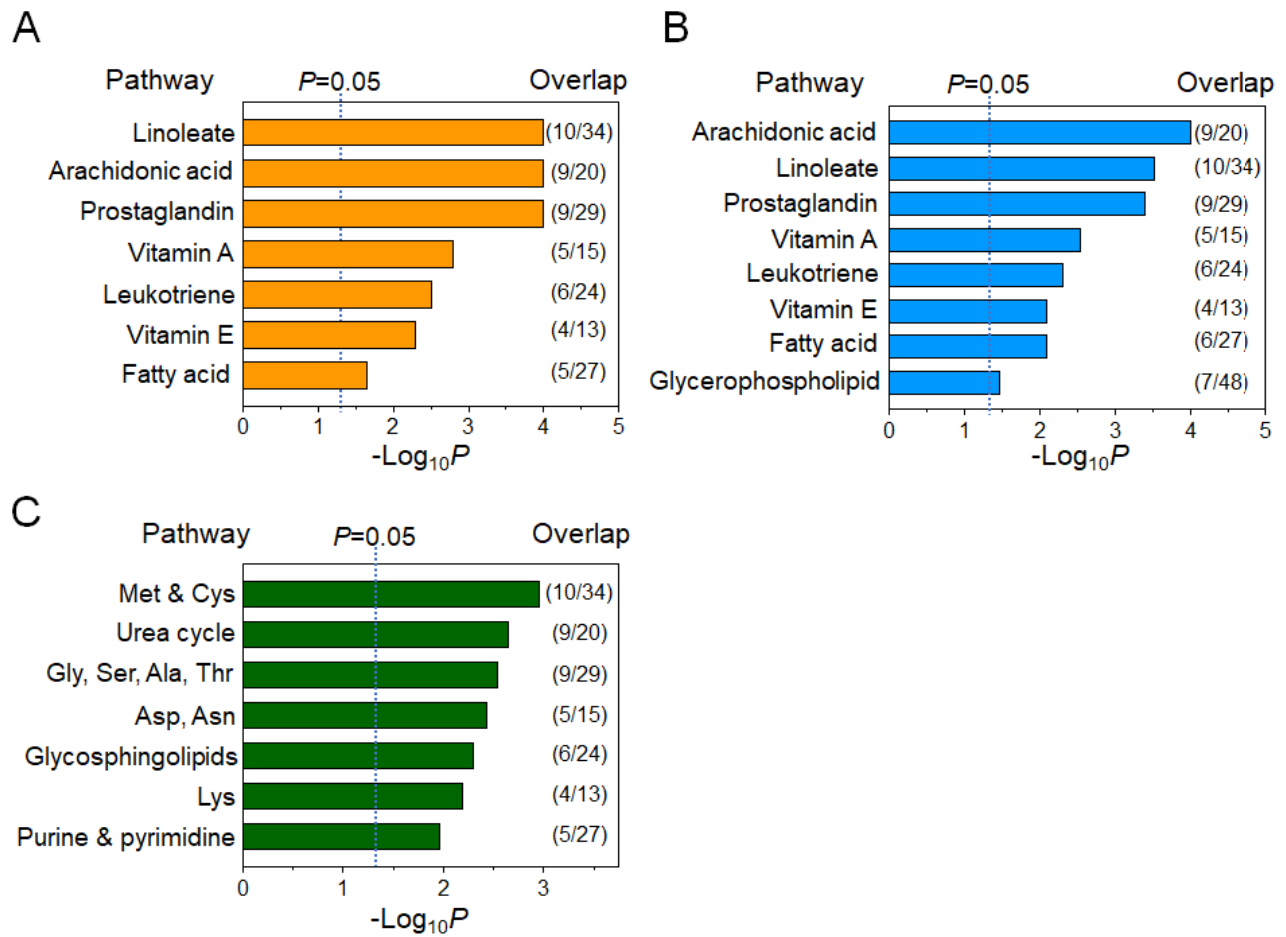
603 **Fig 1.**



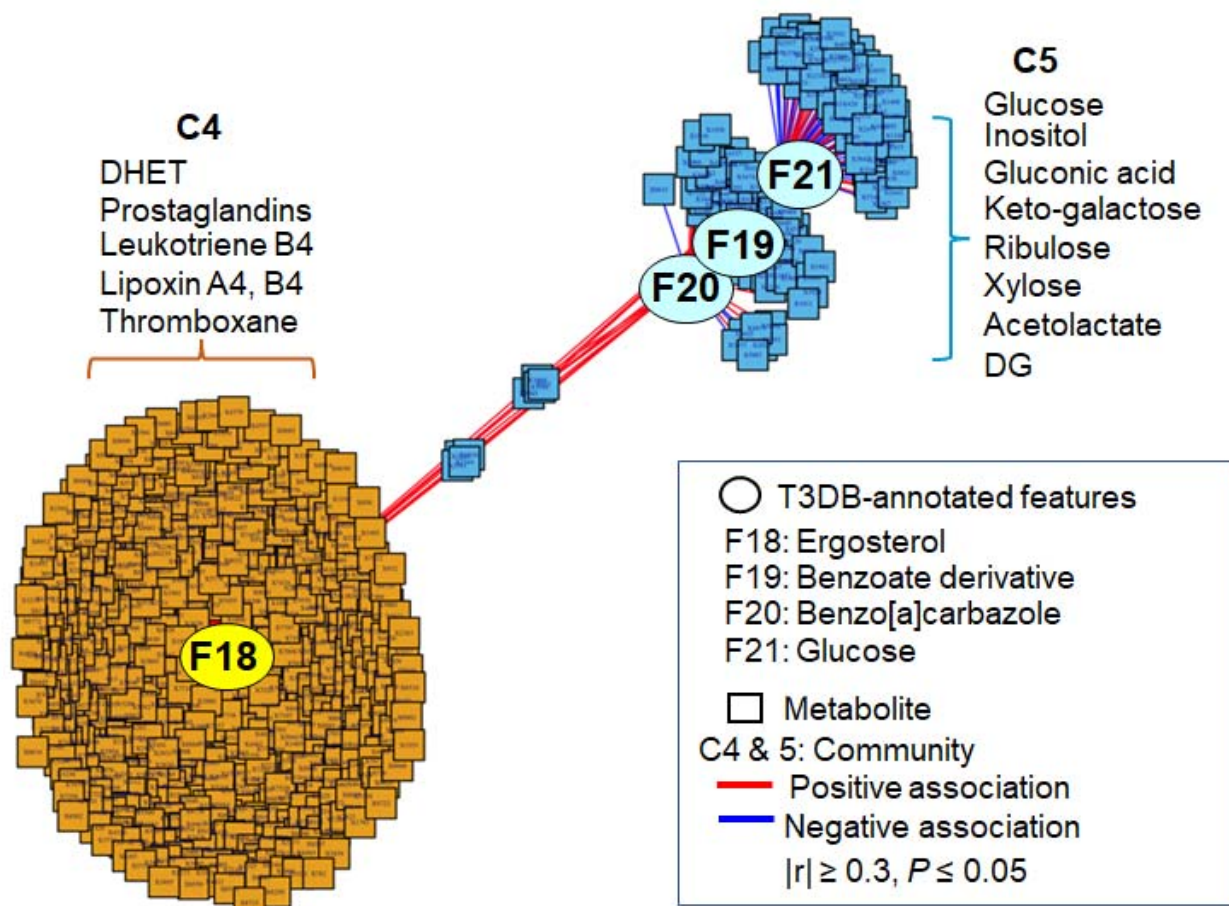
604 **Fig. 2**



605 **Fig. 3**



606 **Fig. 4**



607 **Fig. 5**

