Title

Gaussian graphical modeling of the serum exposome and metabolome reveals interactions between environmental chemicals and endogenous metabolites

Permalink

https://escholarship.org/uc/item/9f78x235

Journal

Scientific Reports, 11(1)

ISSN

2045-2322

Authors

Bessonneau, Vincent
Gerona, Roy R
Trowbridge, Jessica
et al.

Publication Date

2021

DOI

10.1038/s41598-021-87070-9

Peer reviewed

scientific reports

OPEN

Gaussian graphical modeling of the serum exposome and metabolome reveals interactions between environmental chemicals and endogenous metabolites

Vincent Bessonneau[1,2], Roy R. Gerona[3], Jessica Trowbridge[4], Rachel Grashow[5], Thomas Lin[3], Heather Buren[6], Rachel Morello-Frosch[4,7✉] & Ruthann A. Rudel[1✉]

Given the complex exposures from both exogenous and endogenous sources that an individual experiences during life, exposome-wide association studies that interrogate levels of small molecules in biospecimens have been proposed for discovering causes of chronic diseases. We conducted a study to explore associations between environmental chemicals and endogenous molecules using Gaussian graphical models (GGMs) of non-targeted metabolomics data measured in a cohort of California women firefighters and office workers. GGMs revealed many exposure-metabolite associations, including that exposures to mono-hydroxyisononyl phthalate, ethyl paraben and 4-ethylbenzoic acid were associated with metabolites involved in steroid hormone biosynthesis, and perfluoroalkyl substances were linked to bile acids—hormones that regulate cholesterol and glucose metabolism—and inflammatory signaling molecules. Some hypotheses generated from these findings were confirmed by analysis of data from the National Health and Nutrition Examination Survey. Taken together, our findings demonstrate a novel approach to discovering associations between chemical exposures and biological processes of potential relevance for disease causation.

Increasing evidence suggests that environmental, rather than genetic factors are the major causes of most chronic diseases. A recent study from Western European of monozygotic twins estimated that disease risk attributable to genetic plus shared environmental exposures ranged from 3.4% for leukemia to 48.6% for asthma with a median value of 18.5%[1]. Therefore, exploring associations between myriad exposures—originating from diet, lifestyle factors, consumer products and other sources of chemicals—received during the life course [i.e. the "exposome"[2]] and chronic diseases, may elucidate new disease risk factors. Metabolomics is recognized as a powerful approach for characterizing the exposome since it can measure thousands of small molecules in biospecimens[3–5]. These small molecules can be either substrates or end products of cellular metabolism and can originate from exogenous sources via inhalation, ingestion and dermal absorption, or from endogenous processes including human and microbial metabolism. This approach has been used to discover the joint microbial/human metabolism of the nutrient choline as a potential major cause of coronary heart disease[6–8] as well as novel risk factors for type 2 diabetes[9], hypertension[10], and all-cause mortality[11] in the U.S. population.

[1]Silent Spring Institute, 320 Nevada Street, Newton, MA 02460, USA. [2]Inserm, EHESP, Irset (Institut de Recherche en Santé, Environnement Et Travail) – UMR_S 1085, Univ. Rennes, 35000 Rennes, France. [3]Clinical Toxicology and Environmental Biomonitoring Lab, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco, CA, USA. [4]School of Public Health, University of California, 130 Mulford, Hall #3144, Berkeley, CA 94720, USA. [5]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [6]United Fire Service Women, San Francisco, CA, USA. [7]Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, USA. ✉email: rmf@berkeley.edu; rudel@silentspring.org

We recently demonstrated that the application of non-targeted metabolomics using high-resolution mass spectrometry can detect more than 600 environmental chemicals in serum samples, including phthalate metabolites, phosphate flame-retardant metabolites, phenols, pesticides, nitro and nitroso compounds, and per- and polyfluoroalkyl substances[12]. This approach can detect both exogenous and endogenous molecules in biospecimens[13–16], making metabolomics data ideal for exploring associations between environmental exposures and metabolic changes that then generate hypotheses for targeted follow-up studies. One way to achieve this is to evaluate correlation networks between exogenous and endogenous small molecules. However, a major challenge of this method is to identify direct associations (i.e. associations representing the underlying biochemical reactions), because Pearson correlations are generally high in omics data, and so it can generate indirect associations (i.e. associations between metabolites that are distantly connected in biological processes) that are not biologically relevant. One approach to circumvent the selection of indirect associations is to use Gaussian graphical models (GGMs). A GGM is an undirected probabilistic graphical model based on partial correlation coefficients between two variables (i.e. pairwise Pearson coefficients conditioned against all remaining variables)[17]. Therefore, GGMs provide an estimate of the conditional dependencies between variables (i.e. metabolites). Previous studies using GGMs have demonstrated that high partial correlations between small molecules correspond to known metabolic reactions[17–19].

In this study, we employed GGMs to identify direct associations between environmental chemicals and endogenous metabolites measured in serum samples from 69 California women firefighters (FF) and 74 office workers (OW) enrolled in the Women Firefighters Biomonitoring Collaborative (WFBC) study[12,20] using non-targeted liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS). Then, we used National Health and Nutrition Examination Survey (NHANES) data to test several hypotheses generated from exposure-metabolite associations and evaluate whether these exposures to environmental chemicals are associated with phenotypic responses in adult women from the general U.S. population.

Results

Non-targeted metabolomics analysis via LC-HRMS in negative ionization mode tentatively identified 145 small molecules in serum samples of women FF and OW. Most of these molecules were annotated as level 2 or 3 according to the metabolomics standard initiative[21]. Eight environmental chemicals were validated as level 1 (i.e. chemical identity confirmed using authentic standards).

GGM networks identify direct associations between small molecules. For this analysis, GGMs were built by combining the serum chemical exposome (i.e. environmental chemicals) data and metabolome (i.e. endogenous small molecules) data. The LC-HRMS data matrix contained 69 samples × 145 molecules (55 environmental chemicals and 90 endogenous metabolites), 74 samples × 139 molecules (49 environmental chemicals and 90 endogenous metabolites), and 143 samples × 142 molecules (52 environmental chemicals and 90 endogenous metabolites) for women FF, OW and the whole cohort, respectively. After applying GGM of the serum exposome and metabolome for the whole cohort, we found 74 significant partial correlations [partial correlation coefficients (PCC) ranged from − 0.21 to 0.49, p values (p) ranged from 2.2×10^{-16} to 1.9×10^{-4}] at a False Discovery Rate (FDR) threshold of 0.1 (Fig. 1A). When stratifying by occupational group, we observed 54 (PCC ranged from − 0.17 to 0.30, p ranged from 1.1×10^{-15} to 1.4×10^{-4}) and 85 (PCC ranged from − 0.18 to 0.29, p ranged from 2.2×10^{-16} to 3.3×10^{-4}) significant partial correlations for women OW (Fig. 1B) and FF (Fig. 1C), respectively (Supplementary Table S1).

GGM reconstructs biochemical reactions. GGMs displayed numerous significant partial correlations between endogenous metabolites which were organized into sub-networks. For example, we observed strong partial correlations between several fatty acids, including linoleic acid, stearidonic acid, eicosapentaenoic acid, docosapentaenoic acid, and docosahexaenoic acid (Fig. 2A) as well as between unconjugated and conjugated bile acids (Fig. 2C) in all three GGMs (i.e. FF, OW and the whole cohort) (Supplementary Information). Using metabolic pathway databases (i.e. KEGG PATHWAY[22] and SMPDB[23]), we found that these molecules are in close proximity in the metabolic network related to their biosynthesis and degradation. As shown in Fig. 2B, the significant association between alpha-linoleic acid and stearidonic acid (PCC = 0.20, 2×10^{-6}) can be attributed to a desaturation step regulated by the FADS2 gene, while eicosapentaenoic acid is synthesized from stearidonic acid (PCC = 0.19, 6×10^{-6}) via ELOVL5-dependent elongation and FADS1-dependent desaturation. Similarly, we found that the significant associations between bile acids represented known biochemical reactions involved in the biosynthesis of primary and secondary bile acids via cytochrome P450-mediated oxidation of cholesterol in multi-step process[24] (Fig. 2D). Taurodeoxycholic acid and deoxycholic acid glycine are secondary bile acids resulting from bacterial dehydroxylation in the colon of taurocholic acid (PCC = 0.27, 5×10^{-11}) and cholic acid glycine (PCC = 0.18, 9×10^{-6}), respectively[24].

Identification of exposure-metabolite associations and hypothesis generation. Among the significant partial correlations observed, we discovered many exposure-metabolite associations (Table 1). In women FF, we found that exposures to mono-hydroxyisononyl phthalate (PCC = − 0.13, $p = 5.7 \times 10^{-5}$) and 4-ethylbenzoic acid (PCC = 0.14, $p = 4.9 \times 10^{-5}$) were associated with 11β-hydroxyprogesterone, and ethyl paraben was negatively associated with 11β-hydroxyandrosterone-3-glucuronide (PCC = − 0.12, $p = 3.2 \times 10^{-4}$)—metabolites involved in steroid hormones biosynthesis (Fig. 3A). We also observed that 4-hydroxyacetophenone was negatively associated with chenodeoxycholic acid (PCC = − 0.13, $p = 1.0 \times 10^{-4}$) and positively associated with lithocholic acid glycine (PCC = 0.15, $p = 1.1 \times 10^{-5}$)—bile acids involved in cholesterol homeostasis and energy balance (Fig. 3B). Exposures to perfluorohexanesulfonic acid (PFHxS) was also positively associated with one
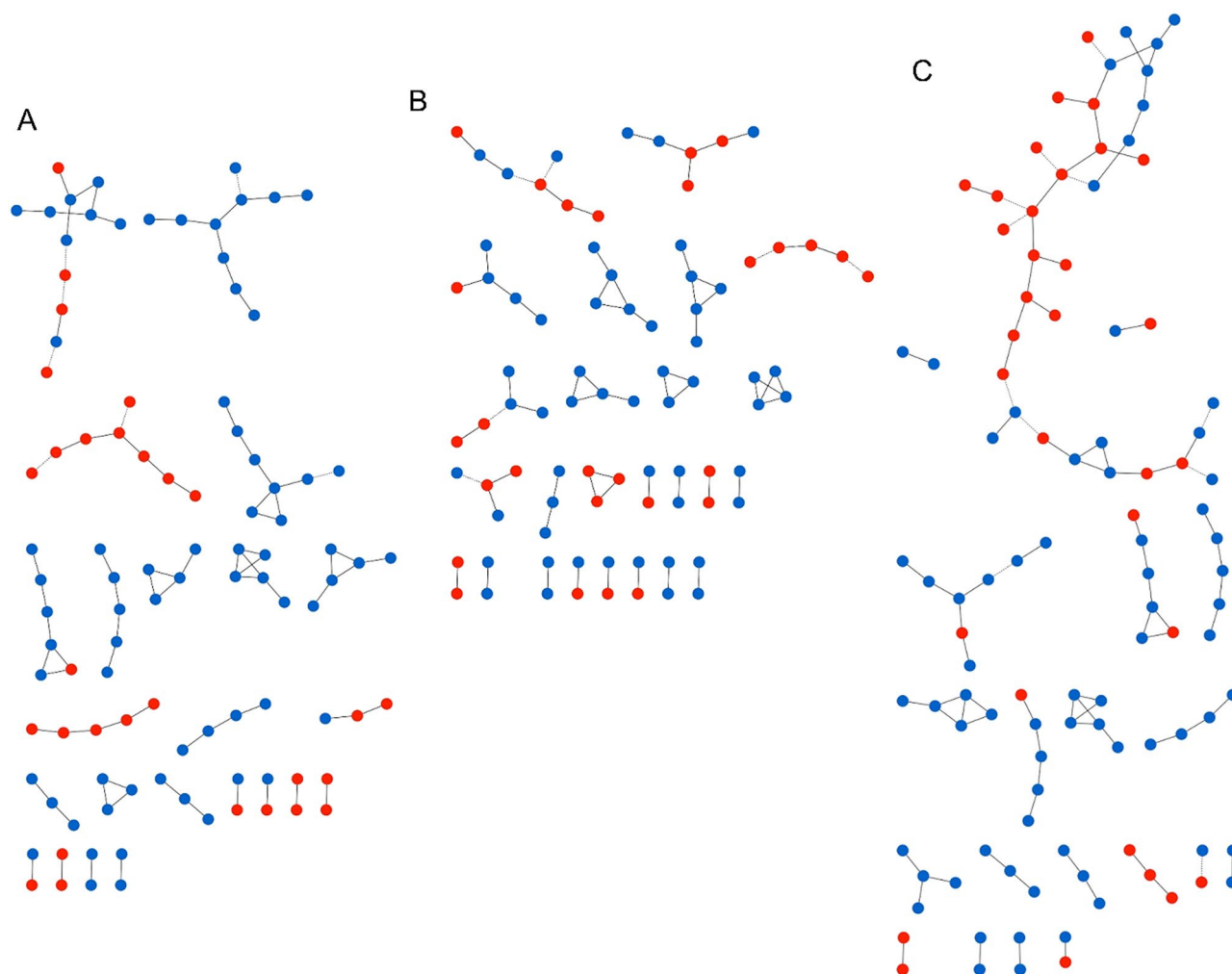
Figure 1. Network representation of Gaussian graphical models (GGM) of the serum exposome and metabolome measured in the whole cohort (**A**), in women office workers (**B**), and in women firefighters (**C**). Blue and red nodes represent endogenous metabolites and environmental chemicals, respectively. Edges connecting nodes represent significant partial correlations at FDR < 10%.

microbial-derived secondary bile acid (sulfolithocholylglycine, PCC = 0.12, $p = 2.6 \times 10^{-5}$), while perfluorooctane sulfonic acid (PFOS) was correlated with one inflammatory signaling molecule (15d PGD2, PCC = −0.13, $p = 2.4 \times 10^{-4}$) and calcitriol, the biologically active form of vitamin D (PCC = 0.14, $p = 4.4 \times 10^{-5}$) in women FF (Fig. 3B). In women OW, we observed a lower number of exposure-metabolite interactions (Table 1). We found that several phenols, including 4-hexyloxyphenol, butyl paraben, octylphenol diethoxylate and pentachlorophenol were significantly associated with molecules involved in inflammation (12-HETE, eicosapentaenoic acid, alpha-linolenic acid and 12-HTT). We also observed that mono-isobutyl phthalate was negatively associated with two metabolites of linoleic acid metabolism and lipid peroxidation (DiHODE and 9-HODE). When combining FF and OW exposome and metabolome data, we discovered only three associations between chemical exposures and metabolites (Table 1).

Next, we sought to further support exposure-metabolite associations observed from the GGM networks and evaluate whether exposure-induced metabolic perturbations are associated with phenotypic responses in adult women from the general U.S. population using NHANES. Among the significant exposure-metabolite associations, twelve were not tested because those chemicals were not measured in any NHANES cycles.

Exposures to PFHxS and metabolic syndrome (MetS).   We observed that exposures to PFHxS were associated with changes in bile acid metabolism in women FF, and so we assessed whether exposure to this environmental chemical was linked to changes in clinical measures related to cholesterol homeostasis and energy balance in NHANES. We tested for associations between serum PFHxS concentrations and MetS and individual components of MetS (Fig. 4) in women 20–79 years of age enrolled in NHANES 2003–2014.

Of the adult women in the NHANES sample, 28.3% with measured PFHxS met NCEP/ATPIII criteria for MetS. No clear associations were observed for PFHxS and MetS. Since we found that exposures to this chemical were associated with a microbial-derived bile acid in our GGMs, we also tested whether the differences in gut
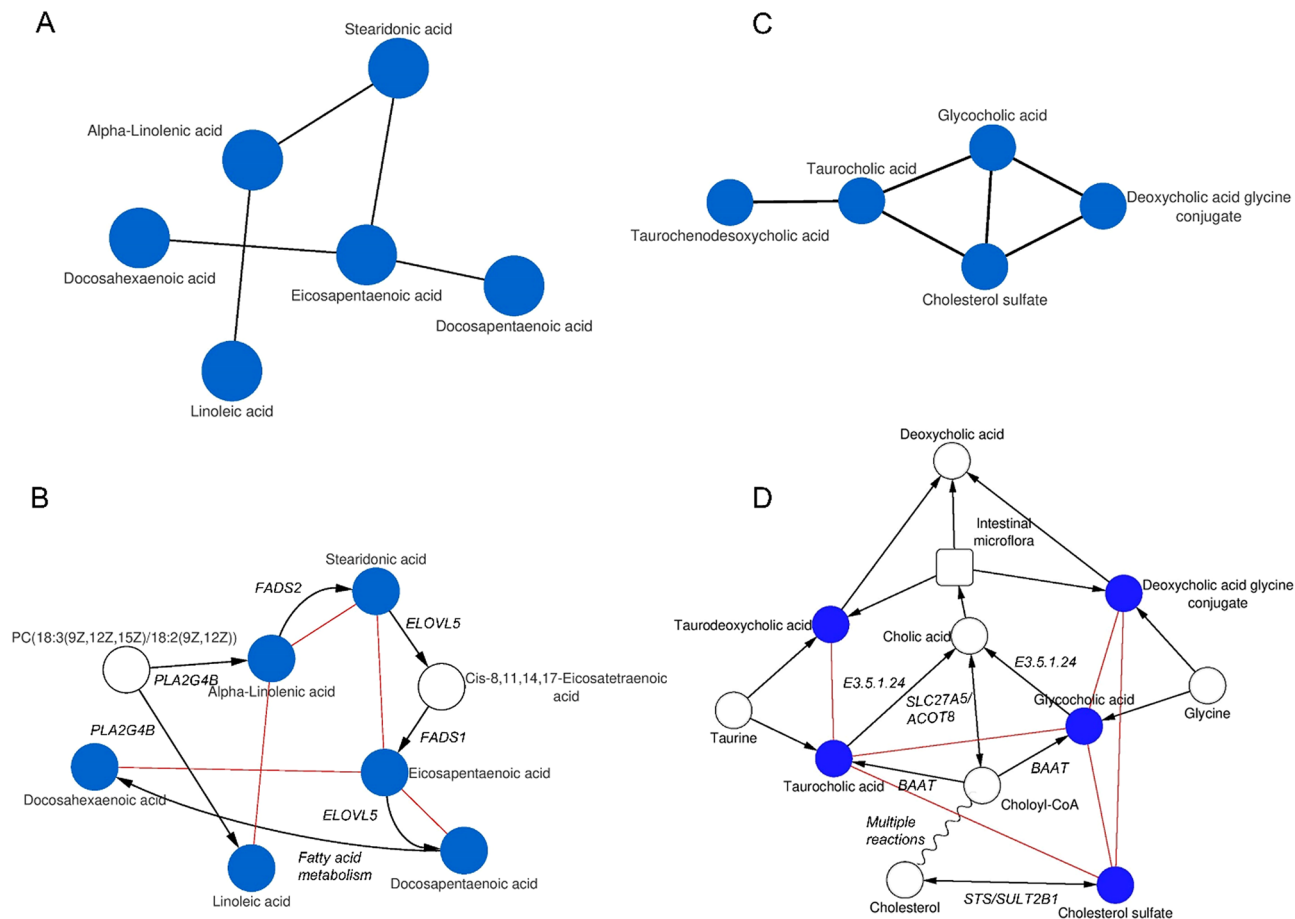
Figure 2. Linoleic acid (**A,B**) and bile acid metabolism (**C,D**) models inferred from GGMs of women FF and OW. (**A**) and (**C**) represent the sub-networks obtained from GGMs. (**B**) and (**D**) represent overlapping sub-networks of known biochemical reactions (black edges; labels correspond to regulating genes) obtained from KEGG pathway and from GGMs (red edges).

microbiota can confound associations with MetS. However, further adjusting for recent use of antibiotics as a proxy of microbial composition had no effect on the odds ratios (ORs) for MetS (data not shown).

Weak associations were found for PFHxS and individual components of MetS, with only the association with low HDL cholesterol reaching statistical significance (adj. OR for Q4 versus Q1: 0.43; 95% CI 0.21, 0.88). Further adjusting for recent use of antibiotics had no effect on the odds ratios (ORs) for MetS components and did not change the effect estimates (data not shown).

Exposures to PFAS and parabens and inflammatory responses.
Because we observed significant partial correlations between PFOS, PFHxS, ethyl paraben and butyl paraben and inflammatory signaling molecules in women FF and OW, we tested whether these exposures are associated with clinical markers of inflammatory responses in adult women in NHANES by exploring associations between serum or urinary concentration of these chemicals and clinical markers of inflammation and immune system function (Table 2).

For PFAS, we found that each 100% increase in serum PFOS concentration was associated with 13% increase in CRP (95% CI 4, 23) and 4.5% increase in absolute lymphocyte count (95% CI 2.1, 6.9), whereas we observed a negative association between each 100% increase in serum PFHxS concentration and absolute neutrophil count (3.5% decrease; 95% CI: -5.9, -0.9) after adjusting for age, race/ethnicity, poverty, BMI and serum cotinine. For parabens, we found negative associations between CRP and each 100% increase in urinary ethyl paraben (7.5% decrease; 95% CI − 11.7, − 3.2) and butyl paraben (6.7% decrease; 95% CI − 11.3, − 2.2) concentrations, after adjusting for age, race/ethnicity, poverty, urinary creatinine and serum cotinine. Urinary butyl paraben was also associated with 1% decrease in white blood cell count (95% CI − 1.8, − 0.3). The associations between urinary parabens and CRP disappeared after further adjusting for BMI. For PFHxS, further adjusting for recent use of antibiotics did not change effect estimates.

Discussion
To our knowledge, this is the first study to combine the serum exposome and metabolome using GGMs. The main goal of this study was to explore the links between the chemical exposome and the metabolome and generate hypotheses about possible health effects of exposures to a complex mixture of environmental chemicals. We

Environmental chemicals (MSI level[a])	Endogenous metabolites (MSI level[a])	Biological pathway	Biological role[b]	PCC[c]	p value
Women FF (n = 69)					
Phthalates					
Mono-hydroxyisononyl phthalate (3)	11β-hydroxyprogesterone (2)	Steroid hormone biosynthesis	Female menstrual cycle, pregnancy and embryogenesis	−0.13	5.7×10^{-5}
Mono-isotridecyl phthalate (3)	Myristic acid (3)	Fatty acid biosynthesis	Membrane integrity	0.13	2.4×10^{-4}
Mono-(2-ethylhexyl) phthalate (1)	12,13- -DHOME (2)	Linoleic acid metabolism	Lipid peroxidation	0.14	6.1×10^{-5}
Phenols					
Ethyl paraben (1)	11β-hydroxyandrosterone-3-glucuronide (2)			−0.12	3.2×10^{-4}
4-Hydroxyacetophenone (2)	Chenodeoxycholic acid (3)	Bile acid metabolism	Cholesterol homeostasis and energy balance	−0.13	1.0×10^{-4}
	Lithocholic acid glycine (3)	Bile acid metabolism	Cholesterol homeostasis and energy balance	0.15	1.1×10^{-5}
Eugenol (1)	Bilirubin (3)	Porphyrin metabolism	Cellular anti-oxidant	0.12	3.2×10^{-4}
Phenol (3)	Indoxyl sulfate (3)	Tryptophan metabolism	Oxidative stress	0.16	1.7×10^{-6}
Perfluorinated compounds					
Pefluorohexanesulfonic acid (PFHxS) (1)	Sulfolithocholylglycine (2)	Bile acid metabolism	Cholesterol homeostasis and energy balance	0.12	2.6×10^{-5}
Perfluorooctane sulfonic acid (PFOS) (1)	15d PGD2 (2)	Arachidonic metabolism	Inflammatory signaling	−0.13	2.4×10^{-4}
	Calcitriol (3)	Vitamin D metabolism	Maintenance of blood calcium and phosphorus	0.14	4.4×10^{-5}
Pesticide					
Pyrethrin II (3)	9,10,13-TriHOME (3)	Linoleic acid metabolism	Lipid peroxidation	0.13	1.7×10^{-4}
Women OW (n = 71)					
Pesticide					
Pyrethrin II (3)	Leucine (3)	Valine, leucine and isoleucine degradation	Protein biosynthesis	0.15	5.9×10^{-5}
Phenols					
Butyl paraben (1)	Eicosapentaenoic acid (2)	Biosynthesis of unsaturated fatty acids	Inflammatory signaling	−0.15	4.5×10^{-5}
Octyphenol diethoxylate (3)	Alpha-linolenic acid (3)	Linoleic acid metabolism	Inflammatory signaling	0.17	2.7×10^{-6}
Pentachlorophenol (1)	Calcitriol (3)	Vitamin D metabolism	Maintenance of blood calcium and phosphorus	−0.14	1.3×10^{-4}
	12-HHT (2)	Arachidonic metabolism	Inflammatory signaling	0.15	9.8×10^{-5}
Phthalates					
Mono-isobutyl phthalate (2)	9,10-DiHODE (2)	Linoleic acid metabolism	Lipid peroxidation	−0.14	1.2×10^{-4}
	9- or 13-HODE (2)	Linoleic acid metabolism	Lipid peroxidation	−0.15	6.0×10^{-5}
Mono-isotridecyl phthalate (3)	5a-Tetrahydrocorticosterone (3)	Steroid hormone biosynthesis	Signaling molecule	0.15	4.2×10^{-5}
Women FF + OW (n = 140)					
Phenols					
Ethyl paraben (1)	Linoleic acid (2)	Linoleic acid metabolism	Inflammatory signaling	−0.16	1.3×10^{-4}
Octyphenol diethoxylate (3)	Alpha-linolenic acid (3)	Linoleic acid metabolism	Inflammatory signaling	0.18	2.2×10^{-5}
Phthalates					
Mono-isotridecyl phthalate (3)	5a-Tetrahydrocorticosterone (3)	Steroid hormone biosynthesis	Signaling molecule	0.22	1.3×10^{-7}

Table 1. Summary of chemical-metabolite associations, biological changes and biological role of metabolites. [a]Metabolomics Standard Initiative (MSI) level of identification; level 1: match based on accurate mass (± 10 ppm), fragmentation pattern and retention time with authentic standards; level 2: match based on accurate mass (± 10 ppm) and fragmentation pattern using mass spectra from public metabolomics libraries or in-silico fragmentation software; level 3: match based on accurate mass (± 10 ppm) only. [b]Biological role of metabolite was inferred from the Metab2MeSH web application that annotates compounds with MeSH terms and HMDB database. [c]Partial correlation coefficient.

computed GGMs of non-targeted LC-HRMS data to map direct associations between small molecules. After controlling for multiple testing (FDR < 0.1), we observed many direct associations, including metabolite-metabolite, chemical–metabolite and chemical–chemical associations. We found that most of the significant associations between metabolites corresponded to known biochemical reactions, supporting previous studies on the ability of GGMs to reconstruct metabolic pathways from MS-based metabolomics data[17–19].

Some metabolite-metabolite associations seem to also reflect molecules originating from the same source of exposures. For example, tryptophan was significantly correlated with arachidonic acid and myristic acid.
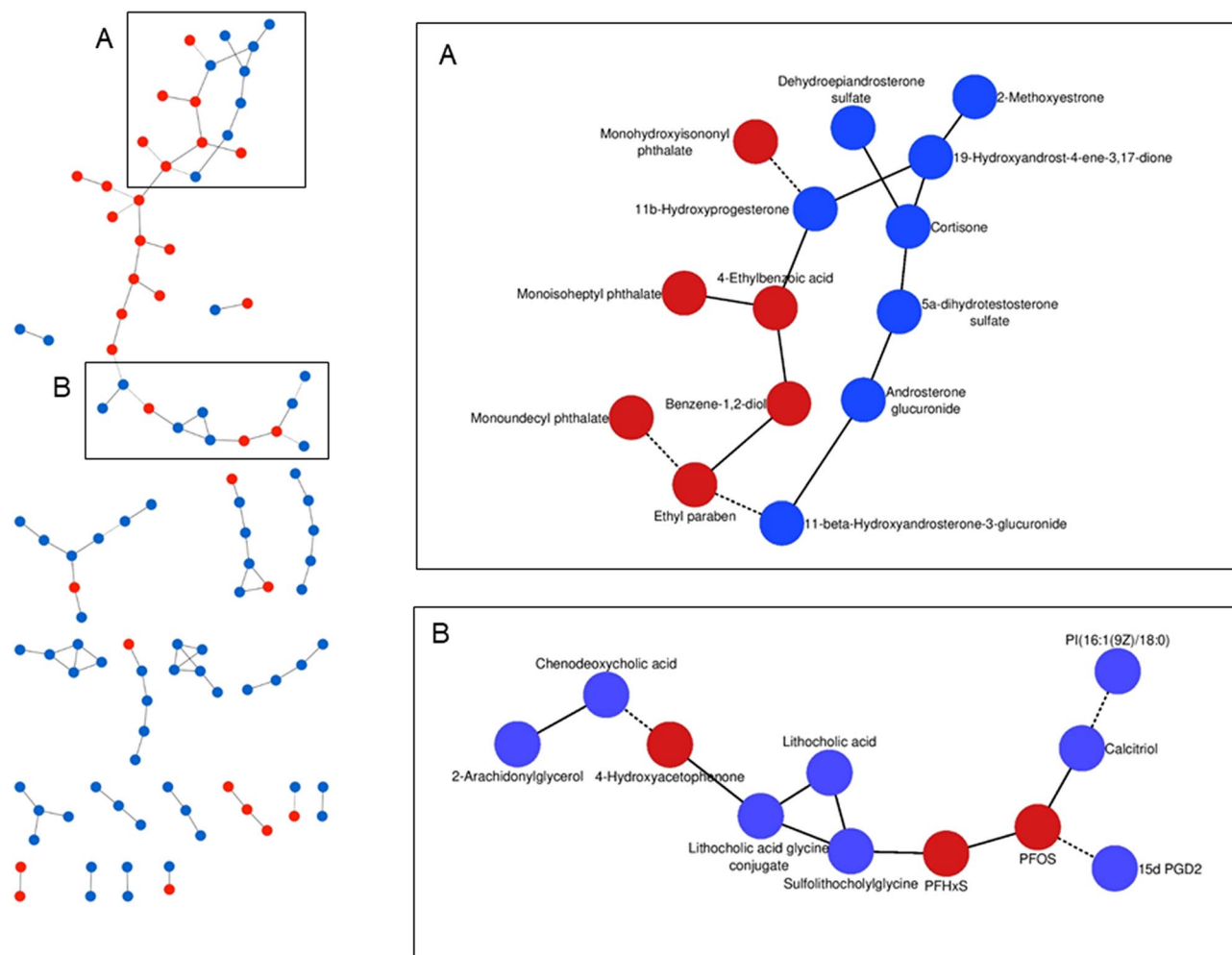
Figure 3. Exposure-metabolites subnetworks identified by the GGMs inferred from the women FF metabolomics dataset. The subnetwork A represents associations of phenols and phthalates with metabolites involved in steroid hormone biosynthesis. The subnetwork B shows associations between 4-hydroxyacetophenone, PFOS and PFHxS and metabolites involved bile acids, arachidonic and vitamin D metabolism. Blue and red nodes represent endogenous metabolites and environmental chemicals, respectively. Plain and dashed edges connecting nodes represent positive partial correlations and negative partial correlations, respectively. Edges connecting nodes represent significant partial correlations at FDR < 10%.

Although these metabolites are involved in different metabolic pathways, they possibly originated from dietary sources since they can co-occur within the same food component, such as eggs or meat[25].

When stratifying by occupation, most of metabolite-metabolite associations remained unchanged, while several chemical exposure-metabolite associations were only significant in FF. For example, the association between PFHxS and sulfolithocholylglycine was significant in FF (PCC = 0.12, $p = 2.6 \times 10^{-5}$), but not observed in the whole cohort (PCC = 0.05, $p = 0.01$) or in OW (PC C = − 0.03, $p = 0.02$). The instability of certain associations is possibly due to inter-occupation variability in the levels of chemical exposures as well as the small sample size. LC–MS/MS analysis of PFAS revealed that serum levels of PFHxS were higher in FF [median = 4.1 ng/mL and interquartile range (IQR) = 5.7 ng/mL] compared to OW (median = 2.5 ng/mL and IQR = 3.8 ng/mL)[20].

GGMs revealed that certain environmental chemicals were linked to endogenous metabolites involved in cholesterol homeostasis, energy metabolism, inflammation and steroid hormones biosynthesis. We found direct associations between 4-hydroxyactophenone, PFHxS and bile acids, including sulfolithocholylglycine, lithocholic acid glycine and chenodeoxycholic acid (CDCA) in women FF. PFHxS is a perfluoroalkyl substance (PFAS) used as an additive in a wide range of consumer products and food packaging due to surfactant and stain resistant properties[26]. PFAS are also components of firefighting foam and firefighter's protective clothing, so these are possible sources of exposure for women FF. A previous study of firefighters found positive associations between serum concentrations of PFOS and PFHxS and the number of years using firefighting foam[27]. As previously mentioned, we also observed higher serum PFHxS in women FF in this cohort, compared to the control group of OW[20].

Bile acids (BAs) are cholesterol-derived signaling molecules with hormonal actions that regulate lipid, glucose and energy metabolism by activating several nuclear receptors, including the constitutive androstane receptor
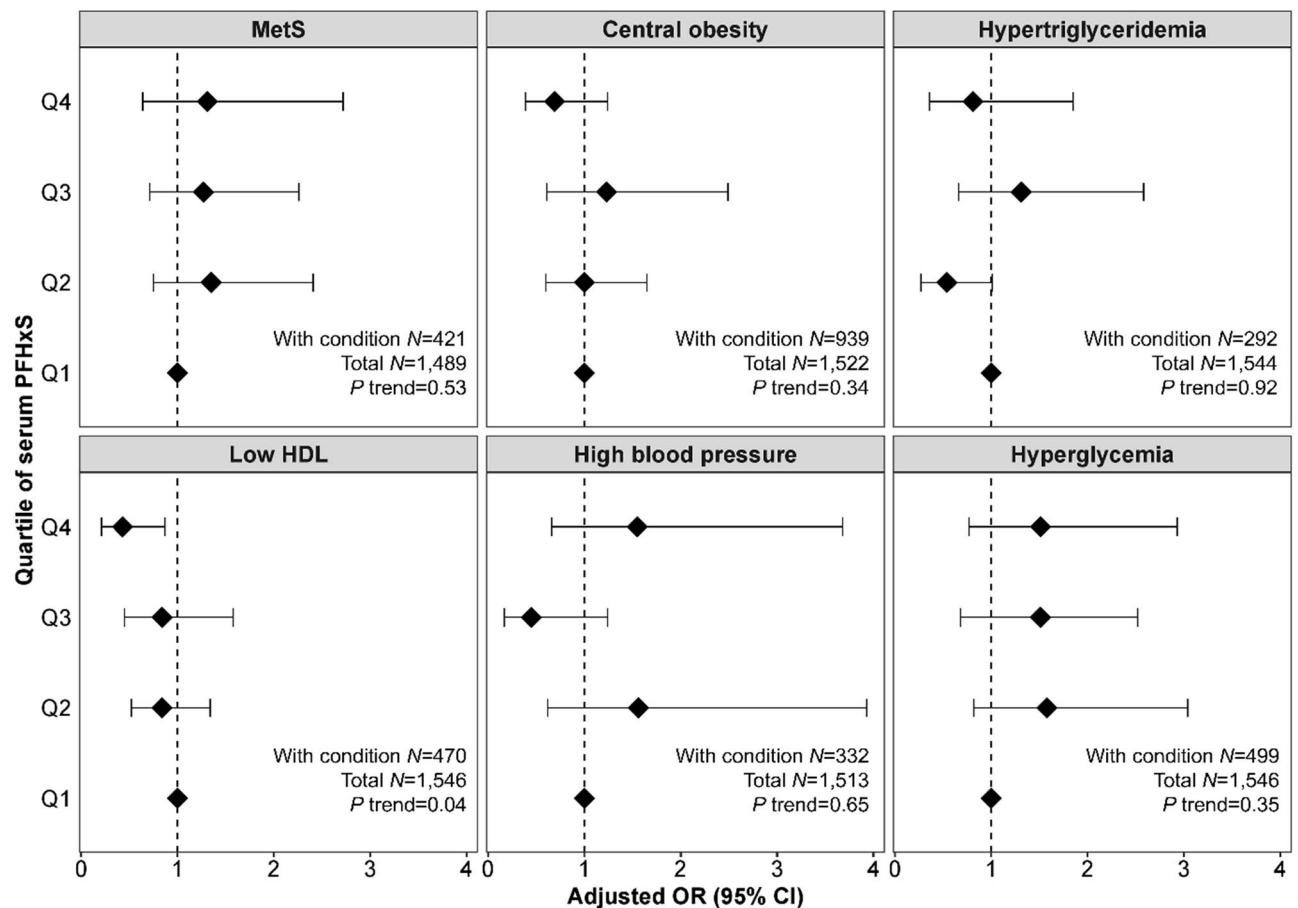
Figure 4. Odds ratio (95% CI) for metabolic syndrome (MetS) and individual components of MetS in women 20–79 years of age enrolled in NHANES 2003–2014 by quartile of serum PFHxS concentration. Models were adjusted for age, race/ethnicity, poverty, total caloric intake, physical activity, and smoking status. Central obesity: waist circumference ≥ 88 cm; hypertriglyceridemia: blood triglyceride ≥ 150 mg/dL; low HDL: high density lipid (HDL) cholesterol < 50 mg/dL; high blood pressure: blood pressure ≥ 130/85 mmHg or treatment for hypertension; hyperglycemia: fasting blood glucose ≥ 100 mg/dL or treatment for diabetes.

(CAR), pregnane X receptor (PXR), farnesoid X receptor (FXR) and vitamin D receptor (VDR), as well as G-protein coupled receptors[24,28,29]. For example, FXR activation has been found to induce expression of peroxisome proliferator activated receptor (PPAR)α and promote oxidation of lipids[30,31]. Previous research suggests that modulation of BAs levels may be an effective strategy for the treatment of components of MetS[32].

Based on the correlations between PFHxS and BAs in FF, we hypothesized that exposure to PFHxS is associated with MetS in women. When assessing components of MetS in NHANES, we found higher levels of PFHxS to be associated with decreased odds of HDL cholesterol < 50 mg/dL in adult women 20–79 years of age enrolled in NHANES. However, PFHxS had neutral effects on the prevalence of the MetS in adult women. A previous study showed that increased levels of serum PFHxS were associated with a lower prevalence of central obesity among male and female adults participants of NHANES 1999–2000 and 2003–2004[33], but we did not see that association in our NHANES analysis of female adults. Although the study by Lin et al.[33] found that increased serum PFHxS were correlated with decreased odds of HDL cholesterol < 50 mg/dL, the association did not reach statistical significance. Another study found that early-life exposure to PFHxS was inversely associated with serum concentrations of leptin—a marker of adiposity and metabolic dysfunction – among children[34]. Similar to our results, these studies observed no evidence for an adverse effect of PFHxS on the prevalence of MetS.

BAs also mediate inflammatory and immunomodulatory actions[35,36]. BAs have detergent properties, in that they dissolve fats, and they can cause membrane perturbations, resulting in production of pro-inflammatory molecules and reactive oxygen and nitrogen species which damage DNA, contributing to mutation and apoptosis[37,38]. The immune system orchestrates a network of biological processes to recognize pathogens and fight infections. The deregulation of immune functions can lead to increased susceptibility to infectious, autoimmune (e.g. rheumatoid arthritis, type 1 diabetes, lupus…) and inflammatory diseases (e.g. asthma, eczema, allergic rhinitis…), and cancer[39]. In addition to finding PFHxS correlated with BAs in our cohort, we also observed significant associations between PFOS, ethyl paraben and butyl paraben, and endogenous molecules involved in inflammatory reactions. Consistent with our GGM findings for PFOS, we found that among adult women in NHANES, serum PFOS concentrations were positively associated with inflammatory and immune reactions (positive associations

	C-reactive protein (mg/dL)		White blood cells (×10³ cells/μL)		Lymphocytes (×10³ cells/μL)	
	n	% change[a] (95% CI)	n	% change[a] (95% CI)	n	% change[a] (95% CI)
Serum PFOS[b]	1693	**13 (4, 23)**	2042	1.2 (−0.9, 1.9)	2039	**4.5 (2.1, 6.9)**
Serum PFHxS						
Model 1[b]	1002	−6.9 (−16, 1.8)	1012	−1.6 (−3.5, 0.4)	1012	1.5 (−0.9, 3.8)
Model 2f.	1002	−6.9 (−16, 1.8)	1012	−1.6 (−3.5, 0.4)	1012	1.5 (−0.8, 3.9)
Urinary EPB						
Model 1[c]	1529	**−7.5 (−11.7, −3.2)**	1556	−0.9 (−1.9, −0.01)	1552	−1.2 (−2.3, −0.02)
Model 2[d]	1529	−1.7 (−5.4, 2.0)	1556	−0.4 (−1.4, 0.6)	1552	−0.8 (−1.9, 0.3)
Urinary BPB						
Model 1[c]	1322	**−6.7 (−11.3, −2.2)**	1472	**−1.1 (−1.8, −0.3)**	1468	−1.1 (−2.1, 0.04)
Model 2[d]	1322	−1.1 (−5.4, 3.1)	1472	−0.3 (−1.2, 0.4)	1468	−0.6 (−1.7, 0.5)
	Monocytes (×10³ cells/μL)		Neutrophils (×10³ cells/μL)		Platelets (×10³ cells/μL)	
	N	% change[a] (95% CI)	n	% change[a] (95% CI)	n	% change[a] (95% CI)
Serum PFOS[b]	2039	1.4 (−1.6, 4.5)	2039	−0.5 (−3.1, 2.1)	2042	−0.5 (−1.8, 0.7)
Serum PFHxS						
Model 1[b]	1012	−0.5 (−3.5, 2.6)	1012	**−3.5 (−5.9, −0.9)**	1012	−1.9 (−4.0, 0.17)
Model 2f.	1012	−0.6 (−3.6, 2.5)	1012	**−3.4 (−6.0, −0.9)**	1012	−1.9 (−4.1, 0.14)
Urinary EPB						
Model 1[c]	1552	−0.5 (−1.3, 0.3)	1552	−0.9 (−2.1, 0.4)	1552	−0.1 (−1.3, 1.0)
Model 2[d]	1552	−0.3 (−1.2, 0.5)	1552	−0.2 (−1.5, 1.1)	1552	−0.5 (−1.7, 0.7)
Urinary BPB						
Model 1[c]	1468	−0.2 (−1.1, 0.7)	1468	−0.9 (−1.9, 0.04)	1472	−0.01 (−0.8, 0.8)
Model 2[d]	1468	0.001 (−0.9, 0.9)	1468	−0.1 (−1.2, 1.0)	1472	0.3 (−0.5, 1.1)
	Eosinophils (×10³ cells/μL)		Basophils (×10³ cells/μL)			
	n	% change[a] (95% CI)	n	% change[a] (95% CI)		
Serum PFOS[b]	2039	−0.2 (−1.0, 0.6)	2039	0.3 (−0.15, 0.79)		
Serum PFHxS						
Model 1[b]	1012	−0.4 (−1.1, 0.3)	1012	0.1 (−0.3, 0.6)		
Model 2f.	1012	−0.4 (−1.1, 0.3)	1012	0.1 (−0.3, 0.6)		
Urinary EPB						
Model 1[c]	1556	−0.03 (−0.3, 0.3)	1556	−0.2 (−0.4, 0.03)		
Model 2[d]	1556	0.05 (−0.3, 0.4)	1556	−0.1 (−0.4, 0.06)		
Urinary BPB						
Model 1[c]	1468	−0.02 (−0.3, 0.3)	1468	−0.01 (−0.3, 0.1)		
Model 2[d]	1468	0.08 (−0.2, 0.4)	1468	−0.01 (−0.2, 0.1)		

Table 2. Associations between markers of inflammation and immune system and each 100% increase in serum PFAS (NHANES 2003–2010) and urinary parabens (NHANES 2005–2010) concentrations in women 20–79 years of age enrolled in NHANES. Results in bold represent associations that are statistically significantly (p<0.05) Association estimates were obtained with general linear models using sampling weights and accounting for the complex survey design. [a]Adjusted for age, race/ethnicity, poverty, and log-transformed serum cotinine. [b]Adjusted for the same variables in a plus BMI. [c]Adjusted for the same variables in a plus log-transformed urinary creatinine. [d]Adjusted for the same variables in c plus BMI. [e]Adjusted for the same variables in c plus recent use of antibiotics. [f]Adjusted for the same variables in b plus recent use of antibiotics.

with CRP and lymphocyte count), while PFHxS was negatively associated with the absolute neutrophil count. Also consistent with our GGM findings, ethyl paraben and butyl paraben were negatively associated with inflammation markers in NHANES. Butyl paraben was also associated with a decrease in white blood cell count, while ethyl paraben did not affect markers of immunity. In previous cross-sectional studies, no significant associations were found between serum PFOS and CRP, and conflicting results were observed for other immune-related health conditions[40]. However, a recent prospective study has observed that elevated PFHxS concentrations at age 7 decreased anti-diphtheria and anti-tetanus antibody concentrations at age 13[41]. Parabens are widely used as antimicrobial preservatives in personal care products, pharmaceuticals, and food and beverages[42,43]. Parabens can interact with hormone receptors and have been detected in breast tissue[44,45]. Some parabens have demonstrated adverse effects on male reproduction at higher doses[46]. A previous study among pregnant women found that urinary concentrations of propyl paraben were associated with lower CRP concentrations[47]. Propyl paraben and butyl paraben have also been reported to increase the prevalence of aeroallergen sensitization in children[48,49].

The underlying mechanisms that mediate the effects of PFAS on cholesterol and glucose homeostasis, inflammation and immune response are thought to be related to PPAR-α,γ activation and fatty oxidation pathways[50–52].

The sub-networks of direct associations found in the present study seem to reinforce these relationships, which may possibly be mediated by PPAR-induced disruption of BAs metabolism and transport. Studies on the association between PFAS and BAs are limited. A previous animal study has demonstrated that mice treated with PFDA resulted in increased serum BAs, probably due to PPAR-α-driven down-regulation of BA transporters involved in the enterohepatic circulation, including the Na⁺-taurocholate cotransporting polypeptide (NTCP)[53]. Recent studies have shown that PFHxS is a substrate for NTCP in both human and rat hepatocytes[54,55]. A case study involving eight subjects has found that oral treatment with BAs sequestrant increased fecal elimination of PFHxS[56]. Although no human study supporting the paraben-BAs association was found, in vitro studies have observed that paraben acts as a strong obesogenic compound through PPAR-γ-activation at high concentrations (high μM range) in murine 3T3-L1 cells[57,58].

Since the BAs associated with PFHxS are produced by intestinal bacteria[38], we hypothesized that the gut microflora composition may confound associations between these chemicals and effects on metabolic function and inflammation. But adjusting for recent use of antibiotics—as a proxy for loss of bacterial diversity—did not modify the associations. Further studies employing direct measurements of gut bacteria diversity (e.g. targeted measurements of microbial-derived metabolites or 16S rRNA sequencing to characterize microbial communities) are needed to decipher the possible role of the microbiome on the association between chemical exposures and health outcomes.

Many exposure-metabolite associations were not further tested because chemical exposures and/or health outcomes hypothesized were not measured in NHANES. This included direct associations of phthalates with steroid hormones, and metabolites involved in oxidative stress and inflammatory response (Fig. 3A and Table 1). The associations we observed are discussed below in the context of relevant studies from the literature. Phthalates are primarily used as plasticizers in automotive, building materials and consumers products[59] and known endocrine disrupting compounds that disrupt fetal testosterone and insulin-like hormone 3 (insl3) synthesis by an unknown mechanism[60]. In our study, mono(hydroxyisononyl) phthalate (MHINP)—a major oxidative metabolite of diisononyl phthalate (DINP)—was negatively associated with one steroid hormone metabolite (11β-hydroxyprogesterone). Epidemiological studies have reported that phthalate exposure is associated with decreased levels of steroid hormones in males and females[61–64]. Meeker and Ferguson observed a suggestive negative association between DINP exposure and serum testosterone in women 40–60 years of age enrolled in NHANES[63]. Several phthalates affected steroid hormone synthesis pathways in human adrenocortical carcinoma cell lines (H295R assay), and different phthalates affected different parts of the pathway[65]. Phthalate exposure has also been linked with inflammation and oxidative stress[66–70]. Among phthalate metabolites associated with lipid peroxidation in our study, urinary levels of mono-(2-ethyl)-hexyl phthalate (MEHP) [metabolite of di(2-ethylhexyl) phthalate (DEHP)] have been positively associated with increased serum levels of gamma glutamyl-transferase (GGT)—a sensitive biomarker of oxidative stress—while mono-isobutyl phthalate (MiBP) [metabolite of dibutyl phthalate (DBP)] was linked to serum levels of CRP in participants enrolled in NHANES[66].

This study has several limitations. First, results are correlative in nature and from a small sample size. Second, due to the cross-sectional study design, we are unable to determine whether higher exposures to these chemicals precede changes in levels of endogenous metabolites or are a consequence of metabolic perturbations due to changes in health conditions, diet or exposures to other environmental factors. Third, because no health outcomes were measured in the occupational cohort, exposure-metabolite-outcome associations were tested in another cohort (i.e. NHANES). These two populations probably exhibit significant differences in terms of chemical exposure levels. For example, median serum concentrations of PFHxS in both FF (4.05 ng/mL) and OW (2.55 ng/mL) were twice as high as those measured in NHANES (1.6 ng/mL for cycles 2003–2014)[20]. As such, our NHANES analysis may underestimate or overestimate the associations between chemical exposures and hypothesized health outcomes (e.g. MetS or inflammation) in the occupational cohort. Fourth, due to the LC-HRMS analysis in negative ionization mode, most of endogenous molecules tentatively identified are lipids or hormones. Also, we only identified potential chemicals in the cohort by using an in-house database of 700 environmental chemicals[12], and so we may have missed other chemical-metabolite interactions that are present in the cohort. We may have missed other important molecules that are more easily detected in positive ionization mode, such as amino acids or vitamins, and involved in other metabolic pathways that would have generated additional exposure-outcome hypotheses.

Despite these limitations, the present study has several strengths. First, this study employed a data-driven approach that combines simultaneous measurements of a wide spectrum of environmental chemicals and endogenous metabolites detected in serum to generate hypotheses related to possible health effects of chemical exposures. This approach provides a comprehensive exploration of the impact of chemicals, individually or in combination, on critical metabolic processes. Second, exposomics and metabolomics data were combined using GGMs that provide identification of direct associations between chemicals and metabolic pathways. Pearson correlation coefficients are generally high in omics data and may lead to identification of many indirect associations that are not biologically relevant. GGMs that are based on partial correlation coefficients provide an estimate of the conditional dependencies between variables and limit the selection of indirect associations. Third, some hypotheses generated from the non-targeted GGM analyses were tested using data from a representative sample of the U.S. population (i.e. NHANES) comprising adult women from diverse racial/ethnic background and socio-economic status, and some of these findings were consistent. For example, analysis of NHANES data confirmed associations found in GGMs between exposures to PFHxS and PFOS and cholesterol metabolism and inflammation. Fourth, GGMs revealed that exposures to previously unstudied chemicals (e.g. 4-hexyloxyphenol or certain phthalates metabolites) may potentially have adverse effects on critical metabolic pathways such as metabolism of arachidonic acid and linoleic acid involved in lipid peroxidation and inflammatory and immune responses as well as steroid hormone biosynthesis.

Conclusion

In conclusion, we used GGMs, which have previously been shown to identify high partial correlations between endogenous small molecules that correspond to known metabolic reactions, to identify associations between tentatively identified exogenous chemicals and endogenous molecules in serum from a cohort of California women firefighters and office workers. These GGMs revealed many exposure-metabolite associations, including that exposures to mono-hydroxyisononyl phthalate, ethyl paraben and 4-ethylbenzoic acid were associated with metabolites involved in steroid hormone biosynthesis, and PFASs were linked to BAs and inflammatory signaling molecules. Some hypotheses generated from these findings were further supported by analysis of data from NHANES, specifically the cross-sectional associations between some PFASs and cholesterol metabolism and inflammation. However, many hypotheses generated from this work were not further tested because chemical exposures and/or health outcomes were not measured in NHANES. Taken together, our findings demonstrate a novel approach to discovering associations between chemical exposures and upstream biological processes potentially involved in disease. We encourage further studies to apply these techniques in other cohorts and to further evaluate associations between chemical exposures and health outcomes that we have reported here.

Methods

Study population.   This study was conducted using LC-QTOF/MS data available from an established cohort of California women workers known as the Women Firefighter Biomonitoring Collaborative (WFBC). The WFBC is a cross-sectional study designed to measure and compare exposures to potential breast carcinogens and other endocrine disrupting compounds (EDCs) in 69 San Francisco (SF) women firefighters (FF) and 74 female controls among office workers (OW) from the City of San Francisco. Detailed description of this cohort has been published elsewhere[12,20]. Demographic characteristics and breast cancer risk factor information, including menopausal status, history of hormone replacement therapy, and reproductive history were collected using questionnaires during in-person interviews. Body Mass Index (BMI; kg m^{-2}) was calculated from participants' height and weight measured at the time of the in-person interview. A 50 mL blood sample was collected by a certified phlebotomist. Written informed consent was obtained from all participants. All research was performed in accordance with relevant guidelines/regulation. The study following protocols were approved by the Institutional Review Board of the University of California, Berkeley (# 2013-07-5512).

Overall, the FF and OW cohorts were similar in terms of age, race/ethnicity, BMI, parity, and hormone use, as previously reported[12]. However, the household income for women FF was significantly higher compared to OW. There were significantly more pre-menopausal women in the FF group and women FF had a higher proportion with a 4-year college as their highest degree, while a higher proportion of OW had additional degrees beyond 4-year college degrees.

Non-targeted LC-QTOF/MS metabolomics analysis.   Non-targeted analysis of serum samples was performed as previously described[71]. Briefly, 250 µL of serum sample was spiked with 2.5 µL of 1 µg/mL of internal standard (2.5 ng BPA-d16) and centrifuged at 3000 rpm for 10 min. Analytes were extracted using solid-phase extraction (SPE; Waters Oasis HLB 10 mg, 1 cc). Extracts were dried under a stream of nitrogen gas and reconstituted in 250 µL of 10% methanol.

Extracts were analyzed on a LC-QTOF/MS system consisting of an LC 1260 autosampler, pumps and a QTOF/MS 6550 (Agilent, Santa Cruz, CA, USA). Analytes were separated by a reversed-phase method using a C18 column (Agilent Poroshell 120, 2.1 mm × 100 mm, 2.7 µm particle size) maintained at 55 °C. Mobile phase A consisted of water with 0.05% ammonium acetate (pH = 7.8) and mobile phase B consisted of methanol with 0.05% ammonium acetate (pH = 7.8). The elution gradient employed was: 0–0.5 min, 5% B; 1.5 min, 30% B; 4.5 min, 70% B; 7.5–10 min, 100% B; 10.01–14 min, 5% B. The injection volume was 50 µL. Analyses were performed with a QTOF/MS operating in negative electrospray ionization mode (ESI-). Ions were collected in the m/z 80–600 range at high resolution for eluates coming out of the LC from 1 to 12 min. Using the Auto MS/MS mode (information-dependent acquisition), a product ion scan (MS/MS) of the three most abundant peaks at high resolution was triggered each time a precursor ion with an intensity of ≥ 500 counts/second was generated in the TOF–MS scan using a collision voltage ranging from 0 to 40 V depending of ions m/z. The LC-QTOF/MS analysis produces a total ion chromatogram for each sample, which includes the following: the accurate mass of each unique compound (expressed as *m/z* of their respective anion), peak area, retention time and spectral data on the parent and fragment ions, including isotopic pattern.

Data processing.   *Exposome annotation.*   Chemical exposures were identified from non-targeted LC-QTOF/MS data as previously described[12,71]. Briefly, all detected *m/z* were matched with those from an in-house MS database of environmental chemicals with a mass tolerance value of 10 ppm. The in-house database consists of more 700 chemicals: (1) environmental organic acids including parabens and paraben metabolites, phthalates and phthalate metabolites and pesticides and pesticide metabolites; (2) chemicals that increase breast cancer risk including mammary carcinogens and mammary gland developmental disruptors[72,73]; (3) known firefighting-related occupational exposures including perfluorinated compounds found in firefighting foams, polychlorinated and polybrominated dioxins and furans and other flame retardants. A list of tentatively annotated chemicals was generated for all samples, with corresponding exact mass, retention time, mass error, peak area, chemical formula, and match scores. Then, we performed retention time correction using an in-house R script in order to align LC-QTOF/MS data. We assigned the chemicals with the same formula to be two different entities if the difference in retention time of two adjacent chemicals was greater than 0.16 min. The exposome data matrix consisted of peak area and *m/z* of 620 unique chemicals which matched to 300 chemical formulas[12].

Metabolome annotation. LC-QTOF/MS metabolomics data were pre-processed using the R package "XCMS"[74] for peak detection, retention time correction and peak alignment (the R script used for data pre-processing can be found in Supplementary Information). After pre-processing, a data matrix containing retention time, mass-to-charge ratio *(m/z)* and intensity of features was generated. Metabolomics features were annotated using the R package "xMSannotator"[75]. We used the Human Metabolome Database 3.6 (HMDB) that contains detailed molecular information about 42,632 small molecules[76] as the reference library for annotation of metabolomics features. We retained only annotated molecules with a confidence score equal to 3 (the highest score of confidence) and classified by HMDB as previously detected and quantified in biological matrices (the R script used for annotation of metabolomics features can be found in Supplementary Information). After this process, the metabolome data matrix contained *m/z* and intensity of 90 annotated molecules (Supplementary Table S5).

The identification level of environmental chemicals and endogenous metabolites was reported as proposed by the Metabolic Standards Initiative[21]: level 1: match based on accurate mass (± 10 ppm), fragmentation pattern and relative retention time with authentic standards; level 2: match based on accurate mass (± 10 ppm) and fragmentation pattern using mass spectra from public metabolomics libraries or in-silico fragmentation software[77].; level 3: match based on accurate mass (± 10 ppm) using public metabolomics libraries. Most small molecules tentatively identified in this study corresponded to level 2 or level 3 annotation. Eight environmental chemicals were validated as level 1 annotation (Supporting Information Table S1).

Statistical analysis. *Gaussian graphical modeling (GGM).* GGMs were built by combining the exposome data matrix with the metabolome data matrix. We first excluded molecules with more than 50% missing values. The filtered data matrix contained 69 samples $\times$ 145 molecules (55 environmental chemicals and 90 endogenous metabolites), 74 samples $\times$ 139 molecules (49 environmental chemicals and 90 endogenous metabolites), and 143 samples $\times$ 142 molecules (52 environmental chemicals and 90 endogenous metabolites) for women FF, OW and the whole cohort, respectively. Missing values imputation, data normalization and transformation were conducted using MetaboAnalyst 3.0[78]. Remaining missing values were imputed using the k-nearest neighbor (KNN) method[79]. Exposome data (i.e. peak area) and metabolome data (i.e. metabolite intensity) were normalized, separately, by the sum of peak area or peak intensity of each sample (i.e. sample normalization. Peak area or peak intensity of each molecule was divided by the sum of peak area or peak intensity of each sample) to reduce analytical variations. Then, normalized data were generalized log-transformed (i.e. feature normalization), and exposome data and metabolome data were merged into one dataset. . We computed GGMs using the ggm.estimator.pcor function from the R package "GeneNet"[80]. We considered partial correlations between two molecules to be significant if the resulting *p* value was below the False Discovery Rate (FDR) threshold of 0.1 (for FF $p < 1.79 \times 10^{-4}$, for OW $p < 1.21 \times 10^{-4}$ and for the whole cohort $p < 1.71 \times 10^{-4}$). The GGM networks were constructed and visualized with Cytoscape 3.8.0[81] using "organic" as layout. Edges connecting nodes represent significant partial correlations. Partial correlation coefficients were used as the edge attributes.

Testing of exposome-metabolome associations. For each significant exposure-metabolite partial correlation, we used several approaches to attempt to validate the associations by checking whether a similar type of chemical-effect relationship has been reported elsewhere, for example in National Health and Nutrition Examination Survey (NHANES) or in PubMed.

We used NHANES to further support or not the association found in our occupational cohort if (1) the chemical of interest and (2) one marker of the biological pathway perturbed were measured in at least one NHANES cycle. NHANES is an ongoing cross-sectional study of the civilian noninstitutionalized U.S. population designed to collect data on dietary and health factors (ref, see Supplementary Information for more details). NHANES study protocols were approved by the National Center for Health Statistics' Research Ethics Review Board. Since our study population consists of women workers, we selected only non-pregnant women adults (20–79 years of age) enrolled in NHANES. All NHANES data (demographics, examination data, laboratory data, and questionnaire data) were downloaded using the R package "RNHANES" [82]. Full description of exposure measurements, outcomes and covariates in NHANES can be found in Supplementary Information.

To test our hypothesis that certain environmental chemicals measured in women's serum affect the level of inflammatory signaling molecules, we explored the relationships between serum or urinary concentrations of each chemical and serum markers of inflammation [C-reactive protein (CRP), and the complete blood count which measured the number of white blood cells, lymphocytes, neutrophils, monocytes, eosinophils, basophils and platelets] in adult women 20–79 years of age. Details about measurements of environmental chemicals and markers of inflammation in NHANES can be found in Supplementary Information. Since we were interested only in the effects of chemicals on chronic inflammation, we excluded participants who reported poor health status, or acute infection at the time of examination, including head or chests colds, stomach or intestinal illness, or flu, pneumonia, or ear infection. We further excluded individuals with CRP concentrations > 10 mg/L because these extreme values likely reflect acute inflammation [83,84]. We constructed multivariable linear regression models with natural log-transformed CRP or absolute blood cells count as the dependent variable and one natural log-transformed chemical divided by 100 as a predictor (e.g. [ln(PFOS)]/100).

To test the hypothesis that exposures to PFHxS are associated with metabolic syndrome (MetS) through alteration of bile acid metabolism, we evaluated the relationships between serum PFHxS and MetS in women adults 20–79 years of age. Data were pooled from NHANES 2003–2014 for PFHxS. We excluded participants with fasting time < 8 h. We used the National Cholesterol Education Program's Adult Treatment Panel III report (NCEP/ATPIII) to define MetS[85]. The NCEP/ATPIII classifies women as having MetS, if at least 3 of the following 5 criteria are met: (1) waist circumference ≥ 88 cm; (2) triglyceride ≥ 150 mg/dL; (3) high density lipid (HDL) cholesterol < 50 mg/dL; 4) blood pressure ≥ 130/85 mmHg or treatment for hypertension; (5) fasting

blood glucose ≥ 100 mg/dL or treatment for diabetes (further details can be found in Supplementary Information. To assess associations between chemical measurements and MetS, we used logistic regression models to estimate adjusted odds ratio (ORs) and corresponding 95% CIs. Serum concentrations of PFHxS were natural log-transformed to address skewness.

All models were adjusted for likely sources of confounding, including age (years, continuous), race and ethnicity (non-Hispanic white, non-Hispanic black, Mexican American, multiracial or other), and poverty/income ratio [PIR (the ratio of self-reported family income to the family's appropriate threshold value), divided into tertiles]. When urinary concentrations of chemicals were used as predictors, we further adjusted for urinary creatinine (continuous, natural log-transformed) to adjust for urine dilution. For models with inflammatory markers as dependent variables, we further adjusted for serum cotinine (natural log-transformed, continuous exposure) as a measure of exposure to tobacco smoke. For models with MetS and individual components of MetS as dependent variables, we further adjusted for self-reported physical activity (none, moderate or vigorous), smoking status (never, past or current), and total caloric intake derived from the average of 2-day 24 recalls and divided into quartiles. Association estimates for models with non-persistent chemicals (i.e. ethyl paraben or propyl paraben) as predictors were reported before and after adjustment for body mass index (BMI) since it can be argued that BMI can act as a confounder or a mediator. For chemicals associated with bile acids in GGM networks (i.e. PFHxS), ORs and association estimates were also reported before and after further adjustment for recent use of antibiotics (in the past 30 days, dichotomous) since the gut microbiota is a possible confounder of associations between exposures to these chemicals and outcome variables related to MetS and inflammation.

Data were analyzed using the R package "survey" to obtain estimates of association or ORs and 95% CIs accounting for the complex NHANES sampling design. We also used the weights to adjust for the oversampling of certain population subgroups and to account for non-response and non-coverage in NHANES. When multiple NHANES cycles were combined, we recalculated new sample weights for each participant by dividing the 2-year sample weights provided by the number of cycles combined. All tests were two sided, and $p < 0.05$ was the level of significance.

References

1. Rappaport, S. M. Genetic factors are not the major causes of chronic diseases. *PLoS ONE* **11**, e0154387 (2016).
2. Wild, C. P. Complementing the genome with an 'exposome': the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **14**, 1847–1850 (2005).
3. Bessonneau, V., Pawliszyn, J. & Rappaport, S. M. The saliva exposome for monitoring of individuals' health trajectories. *Environ. Health Perspect.* **125**, 077014 (2017).
4. German, J. B., Hammock, B. D. & Watkins, S. M. Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics Off. J. Metabolomic Soc.* **1**, 3–9 (2005).
5. Nicholson, J. K. & Wilson, I. D. Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nat. Rev. Drug Discov.* **2**, 668–676 (2003).
6. Koeth, R. A. *et al.* Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.* **19**, 576–585 (2013).
7. Tang, W. H. W. *et al.* Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMoa1109400 (2010).
8. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
9. Patel, C. J., Bhattacharya, J. & Butte, A. J. An environment-wide association study (EWAS) on type 2 diabetes mellitus. *PLoS ONE* **5**, e10746 (2010).
10. Tzoulaki, I. *et al.* A nutrient-wide association study on blood pressure. *Circulation* **126**, 2456–2464 (2012).
11. Patel, C. J. *et al.* Systematic evaluation of environmental and behavioural factors associated with all-cause mortality in the United States national health and nutrition examination survey. *Int. J. Epidemiol.* **42**, 1795–1810 (2013).
12. Grashow, R. *et al.* Integrating exposure knowledge and serum suspect screening as a new approach to biomonitoring: an application in firefighters and office workers. *Environ. Sci. Technol.* https://doi.org/10.1021/acs.est.9b04579 (2020).
13. Walker, D. I. *et al.* High-resolution metabolomics of occupational exposure to trichloroethylene. *Int. J. Epidemiol.* **45**, 1517–1527 (2016).
14. Soltow, Q. A. *et al.* High-performance metabolic profiling with dual chromatography-Fourier-transform mass spectrometry (DC-FTMS) for study of the exposome. *Metabolomics Off. J. Metabolomic Soc.* **9**, S132–S143 (2013).
15. Kingsley, S. L. *et al.* Metabolomics of childhood exposure to perfluoroalkyl substances: a cross-sectional study. *Metabolomics Off. J. Metabolomic Soc.* **15**, 95 (2019).
16. Maitre, L. *et al.* Urine metabolic signatures of multiple environmental pollutants in pregnant women: an exposome approach. *Environ. Sci. Technol.* **52**, 13469–13480 (2018).
17. Krumsiek, J., Suhre, K., Illig, T., Adamski, J. & Theis, F. J. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst. Biol.* **5**, 21 (2011).
18. Jourdan, C. *et al.* Body fat free mass is associated with the serum metabolite profile in a population-based study. *PLoS ONE* **7**, e40009 (2012).
19. Mittelstrass, K. *et al.* Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet.* **7**, e1002215 (2011).
20. Trowbridge, J. *et al.* Exposure to perfluoroalkyl substances in a cohort of women firefighters and office workers in San Francisco. *Environ. Sci. Technol.* https://doi.org/10.1021/acs.est.9b05490 (2020).
21. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics Off. J. Metabolomic Soc.* **3**, 211–221 (2007).
22. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
23. Jewison, T. *et al.* SMPDB 2.0: big improvements to the small molecule pathway database. *Nucleic Acids Res.* **42**, D478–D484 (2014).
24. Chiang, J. Y. L. Bile acids: regulation of synthesis. *J. Lipid Res.* **50**, 1955–1966 (2009).
25. USDA. *USDA Food Composition Databases* (2018).

26. Björklund, J. A., Thuresson, K. & De Wit, C. A. Perfluoroalkyl compounds (PFCs) in indoor dust: concentrations, human exposure estimates, and sources. *Environ. Sci. Technol.* **43**, 2276–2281 (2009).
27. Rotander, A., Toms, L.-M.L., Aylward, L., Kay, M. & Mueller, J. F. Elevated levels of PFOS and PFHxS in firefighters exposed to aqueous film forming foam (AFFF). *Environ. Int.* **82**, 28–34 (2015).
28. Chiang, J. Y. L. Bile acid metabolism and signaling. *Compr. Physiol.* **3**, 1191–1212 (2013).
29. Kliewer, S. A. & Mangelsdorf, D. J. Bile acids as hormones: the FXR-FGF15/19 pathway. *Dig. Dis. Basel Switz.* **33**, 327–331 (2015).
30. Pineda Torra, I. *et al.* Bile acids induce the expression of the human peroxisome proliferator-activated receptor alpha gene via activation of the farnesoid X receptor. *Mol. Endocrinol. Baltim. Md* **17**, 259–272 (2003).
31. Zhang, Y., Castellani, L. W., Sinal, C. J., Gonzalez, F. J. & Edwards, P. A. Peroxisome proliferator-activated receptor-γ coactivator 1α (PGC-1α) regulates triglyceride metabolism by activation of the nuclear receptor FXR. *Genes Dev.* **18**, 157–169 (2004).
32. Ma, H. & Patti, M. E. Bile acids, obesity, and the metabolic syndrome. *Best Pract. Res. Clin. Gastroenterol.* **28**, 573–583 (2014).
33. Lin, C.-Y., Chen, P.-C., Lin, Y.-C. & Lin, L.-Y. Association among serum perfluoroalkyl chemicals, glucose homeostasis, and metabolic syndrome in adolescents and adults. *Diabetes Care* **32**, 702–707 (2009).
34. Fleisch, A. F. *et al.* Early-life exposure to perfluoroalkyl substances and childhood metabolic function. *Environ. Health Perspect.* **125**, 481–487 (2017).
35. Mencarelli, A. *et al.* The bile acid sensor farnesoid X receptor is a modulator of liver immunity in a rodent model of acute hepatitis. *J. Immunol. Baltim. Md.* **1950**(183), 6657–6666 (2009).
36. Vavassori, P., Mencarelli, A., Renga, B., Distrutti, E. & Fiorucci, S. The bile acid receptor FXR is a modulator of intestinal innate immunity. *J. Immunol. Baltim. Md.* **1950**(183), 6251–6261 (2009).
37. Ding, L., Yang, L., Wang, Z. & Huang, W. Bile acid nuclear receptor FXR and digestive system diseases. *Acta Pharm. Sin. B* **5**, 135–144 (2015).
38. Ridlon, J. M. & Bajaj, J. S. The human gut sterolbiome: bile acid-microbiome endocrine aspects and therapeutics. *Acta Pharm. Sin. B* **5**, 99–105 (2015).
39. Suzuki, T., Hidaka, T., Kumagai, Y. & Yamamoto, M. Environmental pollutants and the immune response. *Nat. Immunol.* **21**, 1486–1495 (2020).
40. Chang, E. T., Adami, H.-O., Boffetta, P., Wedner, H. J. & Mandel, J. S. A critical review of perfluorooctanoate and perfluorooctanesulfonate exposure and immunological health conditions in humans. *Crit. Rev. Toxicol.* **46**, 279–331 (2016).
41. Grandjean, P. *et al.* Serum vaccine antibody concentrations in adolescents exposed to perfluorinated compounds. *Environ. Health Perspect.* **125**, 077018 (2017).
42. Calafat, A. M., Ye, X., Wong, L.-Y., Bishop, A. M. & Needham, L. L. Urinary concentrations of four parabens in the U.S. population: NHANES 2005–2006. *Environ. Health Perspect.* **118**, 679–685 (2010).
43. Dodson, R. E. *et al.* Endocrine disruptors and asthma-associated chemicals in consumer products. *Environ. Health Perspect.* **120**, 935–943 (2012).
44. Barr, L., Metaxas, G., Harbach, C. A. J., Savoy, L. A. & Darbre, P. D. Measurement of paraben concentrations in human breast tissue at serial locations across the breast from axilla to sternum. *J. Appl. Toxicol.* **32**, 219–232 (2012).
45. Okubo, T., Yokoyama, Y., Kano, K. & Kano, I. ER-dependent estrogenic activity of parabens assessed by proliferation of human breast cancer MCF-7 cells and expression of ERα and PR. *Food Chem. Toxicol.* **39**, 1225–1232 (2001).
46. Meeker, J. D., Yang, T., Ye, X., Calafat, A. M. & Hauser, R. Urinary Concentrations of parabens and serum hormone levels, semen quality parameters, and sperm DNA damage. *Environ. Health Perspect.* **119**, 252–257 (2011).
47. Watkins, D. J. *et al.* Associations between urinary phenol and paraben concentrations and markers of oxidative stress and inflammation among pregnant women in Puerto Rico. *Int. J. Hyg. Environ. Health* **218**, 212–219 (2015).
48. Savage, J. H., Matsui, E. C., Wood, R. A. & Keet, C. A. Urinary levels of triclosan and parabens are associated with aeroallergen and food sensitization. *J. Allergy Clin. Immunol.* **130**, 453–60.e7 (2012).
49. Spanier, A. J., Fausnight, T., Camacho, T. F. & Braun, J. M. The associations of triclosan and paraben exposure with allergen sensitization and wheeze in children. *Allergy Asthma Proc.* **35**, 475–481 (2014).
50. Buhrke, T., Kibellus, A. & Lampen, A. In vitro toxicological characterization of perfluorinated carboxylic acids with different carbon chain lengths. *Toxicol. Lett.* **218**, 97–104 (2013).
51. Wolf, C. J., Schmid, J. E., Lau, C. & Abbott, B. D. Activation of mouse and human peroxisome proliferator–activated receptor-alpha (PPARα) by perfluoroalkyl acids (PFAAs): further investigation of C4–C12 compounds. *Reprod. Toxicol.* **33**, 546–551 (2012).
52. Zhang, L., Ren, X.-M., Wan, B. & Guo, L.-H. Structure-dependent binding and activation of perfluorinated compounds on human peroxisome proliferator-activated receptor γ. *Toxicol. Appl. Pharmacol.* **279**, 275–283 (2014).
53. Cheng, X. and Klaassen C. D. Critical role of PPAR-alpha in perfluorooctanoic acid- and perfluorodecanoic acid-induced down-regulation of Oatp uptake transporters in mouse livers. PubMed-NCBI. https://www.ncbi.nlm.nih.gov/pubmed/18703564.
54. Han, X., Nabb, D. L., Russell, M. H., Kennedy, G. L. & Rickard, R. W. Renal elimination of perfluorocarboxylates (PFCAs). *Chem. Res. Toxicol.* **25**, 35–46 (2012).
55. Zhao, W. *et al.* Na+/taurocholate cotransporting polypeptide and apical sodium-dependent bile acid transporter are involved in the disposition of perfluoroalkyl sulfonates in humans and rats. *Toxicol. Sci. Off. J. Soc. Toxicol.* **146**, 363–373 (2015).
56. Genuis, S. J., Curtis, L. & Birkholz, D. Gastrointestinal elimination of perfluorinated compounds using cholestyramine and *Chlorella pyrenoidosa. Int. Sch. Res. Notic.* https://doi.org/10.1155/2013/657849 (2013).
57. Hu, P. *et al.* Effects of parabens on adipocyte differentiation. *Toxicol. Sci.* **131**, 56–70 (2013).
58. Pereira-Fernandes, A. *et al.* Evaluation of a screening system for obesogenic compounds: screening of endocrine disrupting compounds and evaluation of the PPAR dependency of the effect. *PLoS ONE* **8**, e77481 (2013).
59. Giovanoulis, G. *et al.* Multi-pathway human exposure assessment of phthalate esters and DINCH. *Environ. Int.* **112**, 115–126 (2018).
60. Howdeshell, K. L., Rider, C. V., Wilson, V. S. & Gray, L. E. Mechanisms of action of phthalate esters, individually and in combination, to induce abnormal reproductive development in male laboratory rats. *Environ. Res.* **108**, 168–176 (2008).
61. Hauser, R. *et al.* Male reproductive disorders, diseases, and costs of exposure to endocrine-disrupting chemicals in the European Union. *J. Clin. Endocrinol. Metab.* **100**, 1267–1277 (2015).
62. Marsee, K., Woodruff, T. J., Axelrad, D. A., Calafat, A. M. & Swan, S. H. Estimated daily phthalate exposures in a population of mothers of male infants exhibiting reduced anogenital distance. *Environ. Health Perspect.* **114**, 805–809 (2006).
63. Meeker, J. D. & Ferguson, K. K. urinary phthalate metabolites are associated with decreased serum testosterone in men, women, and children from NHANES 2011–2012. *J. Clin. Endocrinol. Metab.* **99**, 4346–4352 (2014).
64. Mendiola, J. *et al.* Urinary concentrations of di(2-ethylhexyl) phthalate metabolites and serum reproductive hormones: pooled analysis of fertile and infertile men. *J. Androl.* **33**, 488–498 (2012).
65. Haggard, D. E. *et al.* High-throughput H295R steroidogenesis assay: utility as an alternative and a statistical approach to characterize effects on steroidogenesis. *Toxicol. Sci. Off. J. Soc. Toxicol.* **162**, 509–534 (2018).
66. Ferguson, K. K., Loch-Caruso, R. & Meeker, J. D. Urinary phthalate metabolites in relation to biomarkers of inflammation and oxidative stress: NHANES 1999–2006. *Environ. Res.* **111**, 718–726 (2011).
67. Ferguson, K. K., Loch-Caruso, R. & Meeker, J. D. Exploration of oxidative stress and inflammatory markers in relation to urinary phthalate metabolites: NHANES 1999–2006. *Environ. Sci. Technol.* **46**, 477–485 (2012).

68. Ferguson, K. K., McElrath, T. F., Chen, Y.-H., Mukherjee, B. & Meeker, J. D. Urinary phthalate metabolites and biomarkers of oxidative stress in pregnant women: a repeated measures analysis. *Environ. Health Perspect.* **123**, 210–216 (2015).
69. Ferguson, K. K. *et al.* Urinary phthalate metabolite associations with biomarkers of inflammation and oxidative stress across pregnancy in puerto rico. *Environ. Sci. Technol.* **48**, 7018–7025 (2014).
70. Guo, Y. *et al.* Urinary concentrations of phthalates in couples planning pregnancy and its association with 8-hydroxy-2′-deoxyguanosine, a biomarker of oxidative stress: longitudinal investigation of fertility and the environment study. *Environ. Sci. Technol.* **48**, 9804–9811 (2014).
71. Gerona, R. R. *et al.* Suspect screening of maternal serum to identify new environmental chemical biomonitoring targets using liquid chromatography-quadrupole time-of-flight mass spectrometry. *J. Expo. Sci. Environ. Epidemiol.* **28**, 101–108 (2018).
72. Rudel, R. A., Attfield, K. R., Schifano, J. N. & Brody, J. G. Chemicals causing mammary gland tumors in animals signal new directions for epidemiology, chemicals testing, and risk assessment for breast cancer prevention. *Cancer* **109**, 2635–2666 (2007).
73. Rudel, R. A., Ackerman, J. M., Attfield, K. R. & Brody, J. G. New exposure biomarkers as tools for breast cancer epidemiology, biomonitoring, and prevention: a systematic approach based on animal evidence. *Environ. Health Perspect.* **122**, 881–895 (2014).
74. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**, 779–787 (2006).
75. Uppal, K., Walker, D. I. & Jones, D. P. xMSannotator: an R package for network-based annotation of high-resolution metabolomics data. *Anal. Chem.* **89**, 1063–1067 (2017).
76. Wishart, D. S. *et al.* HMDB 3.0—the human metabolome database in 2013. *Nucleic Acids Res.* **41**, D801–D807 (2013).
77. Dührkop, K., Shen, H., Meusel, M., Rousu, J. & Böcker, S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12580–12585 (2015).
78. Xia, J. & Wishart, D. S. Using MetaboAnalyst 3.0 for Comprehensive metabolomics data analysis. In *Current Protocols in Bioinformatics* (Wiley, 2002). doi:https://doi.org/10.1002/cpbi.11.
79. Xia, J. & Wishart, D. S. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nat. Protoc.* **6**, 743–760 (2011).
80. Schäfer, J. & Strimmer, K. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* **21**, 754–764 (2005).
81. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
82. Susmann, H. *RNHANES: Facilitates Analysis of CDC NHANES Data* (2016).
83. Shaheen, M. *et al.* Hepatitis C, metabolic syndrome, and inflammatory markers: results from the third national health and nutrition examination survey [NHANES III]. *Diabetes Res. Clin. Pract.* **75**, 320–326 (2007).
84. Yoon, S. S., Dillon, C. F., Carroll, M., Illoh, K. & Ostchega, Y. Effects of statins on serum inflammatory markers: the U.S National Health and Nutrition Examination Survey 1999–2004. *J. Atheroscler. Thromb.* **17**, 1176–1182 (2010).
85. Grundy, S. M. *et al.* Definition of metabolic syndrome: report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. *Arterioscler. Thromb. Vasc. Biol.* **24**, e13-18 (2004).

Acknowledgements

Author contributions

R.A.R., R.M.F., R.R.G. and H.B. designed the WFBC study. R.R.G. and T.L. performed non-targeted analysis of serum samples. R.R.G., R.A.R. and R.M.F. developed the in-house MS database of environmental chemicals used to annotate the serum exposome. R.R.G., T.L., R.G. and V.B. generated the serum exposome data. V.B. generated the serum metabolome data, designed and led the data analysis to combine the exposome and metabolome data, and wrote the initial draft of the manuscript. J.T., R.M.F., R.A.R. and R.R.G. edited the submitted version. All authors critically revised the draft and approved the final manuscript.

Competing interests

RAR, RG, and VB, are employed at the Silent Spring Institute, a scientific research organization dedicated to studying environmental factors in women's health. The Institute is a 501(c)3 public charity funded by federal grants and contracts, foundation grants, and private donations, including from breast cancer organizations. HB is former president and member of United Fire Service Women, a 501(c)3 public charity dedicated to supporting the welfare of women in the San Francisco Fire Department. The authors declare they have no actual or potential competing financial interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-87070-9.

Correspondence and requests for materials should be addressed to R.M.-F. or R.A.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.