Reproductive Sciences (2022) 29:2001–2012
https://doi.org/10.1007/s43032-022-00870-w

PREGNANCY: ORIGINAL ARTICLE

A Data Mining Approach Reveals Chemicals Detected at Higher Levels in Non-Hispanic Black Women Target Preterm Birth Genes and Pathways

Sean M. Harris[1] · Justin Colacino[1,2,3] · Miatta Buxton[4] · Lauren Croxton[5] · Vy Nguyen[1,6] · Rita Loch-Caruso[1] · Kelly M. Bakulski[4]

Received: 14 December 2021 / Accepted: 21 January 2022 / Published online: 2 February 2022

Abstract

Preterm birth occurs disproportionately in the USA non-Hispanic Black population. Black women also face disproportionate exposure to certain environmental chemicals. The goal of this study was to use publicly available toxicogenomic data to identify chemical exposures that may contribute to preterm birth disparities. We tested 19 chemicals observed at higher levels in the blood or urine of non-Hispanic Black women compared to non-Hispanic White women. We obtained chemical-gene interactions from the Comparative Toxicogenomics Database and a list of genes involved in preterm birth from the Preterm Birth Database. We tested chemicals for enrichment with preterm birth genes using chi-squared tests. We then conducted pathway enrichment analysis for the preterm birth genes using DAVID software and identified chemical impacts on genes involved in these pathways. Genes annotated to all 19 chemicals were enriched with preterm birth genes (FDR-adjusted p value < 0.05). Preterm birth enriched chemicals that were detected at the highest levels in non-Hispanic Black women included methyl mercury, methylparaben, propylparaben, diethyl phthalate, dichlorodiphenyldichloroethylene, and bisphenol S. The preterm birth genes were enriched for pathways including "inflammatory response" (FDR-adjusted p value $= 3 \times 10^{-19}$), "aging" (FDR-adjusted p value $= 4 \times 10^{-8}$) and "response to estradiol" (FDR-adjusted p value $= 2 \times 10^{-4}$). Chemicals enriched with preterm birth genes impacted genes in all three pathways. This study adds to the body of knowledge suggesting that exposures to environmental chemicals contribute to racial disparities in preterm birth and that multiple chemicals drive these effects. These chemicals affect genes involved in biological processes relevant to preterm birth such as inflammation, aging, and estradiol pathways.

Keywords Preterm birth · Racial disparities · Computational toxicology · Toxicogenomics · Data mining

✉ Sean M. Harris
harrsemi@umich.edu

[1] Department of Environmental Health Sciences, School of Public Health, University of Michigan, Ann Arbor, MI, USA

[2] Department of Nutritional Sciences, School of Public Health, University of Michigan, Ann Arbor, MI, USA

[3] Center for Computational Medicine and Bioinformatics, Medical School, University of Michigan, Ann Arbor, MI, USA

[4] Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA

[5] College of Literature, Science and the Arts, University of Michigan, Ann Arbor, MI, USA

[6] Department of Computational Medicine and Bioinformatics, Medical School, University of Michigan, Ann Arbor, MI, USA

Introduction

Preterm birth is defined as birth prior to 37 weeks of gestation. Preterm birth affects approximately 5–18% of pregnancies worldwide and is a major contributor to infant mortality. In the USA, the overall preterm birth rate for 2019 is estimated at 10.23%, increased 2.1% from the prior year, and increased for the fifth consecutive year [1]. For surviving offspring of preterm birth, complications include higher risks for health problems later in life including respiratory [2] and neurological complications [3]. Health complications associated with preterm birth in the USA are associated with between $6 and 14 billion in medical costs per year [4]. Racial differences exist for preterm birth rates. In 2019, the preterm birth rate for non-Hispanic Black women was 14.4%, whereas the rate for non-Hispanic White women was

9.3% [1]. Non-Hispanic Black and Hispanic women also face disproportionately high levels of exposure to certain environmental toxicants [5, 6]. Thus, the potential contribution of environmental toxicants to preterm birth disparities is a significant public health and equity issue [7].

The physiology of term and preterm birth is a complex process, and the various factors that contribute to the timing of labor are poorly understood [8]. Dynamic changes in hormone levels [9], inflammatory pathways [10], and redox status (i.e., the balance between oxidants and antioxidants) [11] in the gestational compartment all have a role in stimulating the onset labor and/or the rupture of the fetal membranes. Due to the inherent physiological complexity of labor, in vitro and in vivo models present challenges for studying chemical exposure contributions to preterm birth. For example, in vitro models cannot integrate all relevant tissues involved in the activation of term or preterm labor pathways. Moreover, in vivo laboratory animal models of pregnancy are limited by significant differences among species in uterine and placental structure [12], cytokine and prostaglandin production [13, 14], and the process of placentation [15]. In addition, pregnant women are not exposed to individual environmental chemicals in isolation but rather a complex mixture of various chemicals across an array of classes including polychlorinated biphenyls (PCBs), heavy metals, perfluorinated compounds (PFCs), volatile organic compounds, phthalates, and more [16–18]. Thus, identifying individual chemicals and chemical mixtures that may contribute to preterm birth disparities remains a daunting challenge. In a recent evaluation of chemical biomarker data for US women from the National Health and Nutrition Examination Survey (NHANES) 1999–2014, racial disparities in exposures to multiple classes of chemicals including phthalates, metals, parabens, volatile organic compounds, and perflouroalkyl substances were identified, in which these chemicals were observed at higher levels in non-Hispanic Black women compared to non-Hispanic White women. The observed differences across racial categories were independent of factors such as socioeconomic status, parity, and iron deficiency [6].

The Comparative Toxicogenomics Database (CTD) is a publicly available database that curates millions of interactions between chemicals, genes, and diseases reported throughout the toxicology and broader scientific literature [19]. The CTD offers a promising resource for researchers to test hypotheses that would otherwise not be feasible due to time or resource constraints. The Preterm Birth Database is a web-based tool that contains curated gene and pathway data related to preterm birth aggregated and reviewed by a team of trained curators [20]. Because non-Hispanic Black women face higher levels of exposures to some environmental toxicants [6] and higher preterm birth rates compared to non-Hispanic White women [1], we sought to use these unique databases to investigate the molecular basis for toxicant exposures contributing to preterm birth disparities. We used the CTD and Preterm Birth Database to test the hypothesis that chemicals observed at higher levels in non-Hispanic Black women are enriched with preterm birth gene interactions. In addition, we analyzed chemical impacts on specific biological pathways and genes relevant to preterm birth to identify potential mechanistic explanations for the link between preterm birth and chemical exposures. By assessing a diverse array of relevant chemicals, our goal was to identify both toxicants of concern and potential chemical mixtures of concern for preterm birth. We stratified our analysis by three species (human, mouse, and rat) to compare and integrate results across the mammalian species that are most commonly assessed in epidemiology and toxicology studies.

Methods

Datasets

Background Gene Lists for Human, Mouse, and Rat Species

We downloaded all genes (unique gene symbols) associated with the species *Homo sapiens* (human), *Mus musculus* (mouse), and *Rattus norvegicus* (rat) from the CTD (date of download 5/19/20). Genes that were discontinued or replaced in the PubMed gene database were discarded. The resulting number of genes was 43,485 for human, 34,529 for mouse, and 24,848 for rat.

Preterm Birth Gene List

Preterm birth genes were obtained from the Preterm Birth Database. The Preterm Birth Database contains a list of genes related to preterm birth that have been aggregated using semantic data mining from published articles, archives of transcriptomic data and pathway based interpolation, curated by researchers trained in the biology of preterm birth [20]. This list includes 640 curated human genes. Of these, there are 599 homologs in mouse and 600 homologs in rat. All statistical testing was conducted within each species separately.

Chemical Selection

Our list of chemicals was informed by an analysis of US NHANES urine or blood biomarker concentrations for 143 chemicals in 38,080 women from 1999 to 2014, stratified by Centers for Disease Control and Prevention-defined racial/ethnic categories [6]. This analysis identified 40 chemicals at significantly higher concentrations in non-Hispanic Black

versus non-Hispanic White women [6]. Due to observed disparities in preterm birth rates between non-Hispanic Black versus non-Hispanic White women [1], these chemicals were selected for preterm birth enrichment tests.

Chemical Gene Lists

Thirty-four of the selected chemicals were annotated with chemical-gene interactions in the CTD. The CTD is a publicly available database for which trained scientists curate the scientific literature for gene, chemical, phenotype, exposure, and disease relationships. The CTD contains data on over 16,300 chemicals and 51,300 genes across hundreds of species including human, mouse, and rat [21]. The list of 34 chemicals with curated data in the CTD included dioxins, furans, metals, polychlorinated biphenyls, pesticides, personal care products, polyfluoroalkyl substances (PFAS), and volatile organic compounds. Gene lists for each of these chemicals were downloaded from the CTD website (http://ctdbase.org, accessed 2/4/20). The full list of chemicals and number of genes annotated in the CTD for each chemical is shown in Supplementary Table 1.

Statistical Tests

Enrichment Testing of Chemical Gene Lists and Preterm Birth Genes

We tested chemicals for enrichment with preterm birth genes, i.e., a significantly higher number of preterm birth genes represented in chemical gene lists than would be expected by chance. All analyses were conducted using R statistical software. The R markdown code used to conduct these analyses and generate figures is available at https://github.com/bakulskilab. For individual chemicals in each species, we generated a 2×2 descriptive table for the number of observed genes annotated to each chemical and the number of genes annotated to the preterm birth gene list. We calculated the expected number of genes annotated for each chemical gene set size. Based on these frequencies, statistical tests for enrichment with preterm birth genes for each chemical were conducted as described previously [22]. Briefly, chemicals with an expected gene number frequency less than one in any cell were excluded from the analysis for that species. When the expected gene frequencies were all greater than five, the standard chi-squared test was used to test for enrichment. When an expected gene frequency was less than five and greater than one, we used the "N-1" chi-squared test [23]. Chemicals significantly enriched with preterm birth genes were identified by $p < 0.05$ after correction for multiple comparisons within a species (Bonferroni correction; $n = 12$ tests, $p < 0.004$ for human and mouse, $n = 15$ tests, $p < 0.003$ for rat). Heatmaps for this and all subsequent tests were constructed using MultiExperiment Viewer, version 4.9 [24].

Sensitivity Analysis for Enrichment Tests

As a sensitivity test for the robustness of our preterm birth enrichment analysis, we used two additional enrichment tests to confirm our initial results. First, we used Fisher's exact test to further verify initial results obtained from chi-squared tests. We then used proportional reporting ratios to quantify the magnitude of enrichment for each chemical. A proportional reporting ratio > 1 indicates enrichment, while a proportional reporting ratio < 1 would indicate depletion (i.e., fewer overlapping between chemical and preterm birth gene lists than would be expected by chance).

Examining Differences in Exposure Levels to Preterm Birth-Enriched Chemicals Between Racial Categories

To put our enrichment testing in the context of racial disparities in chemical exposure levels, we incorporated the average fold-difference in chemical biomarker levels between non-Hispanic Black and non-Hispanic White women from Nguyen et al. [6]. For each chemical, we plotted on the x-axis the average fold difference in exposure levels between these two groups and on the y-axis a measure of the magnitude of enrichment in preterm birth genes (the proportional reporting ratio). This allowed us to visualize the degree of chemical exposure disparity relative to the degree of enrichment in preterm birth genes for each chemical. Chemicals with the highest enrichment in preterm birth genes (proportional reporting ratio > 2) and highest fold change difference in exposure between groups (> 1.5) were prioritized.

Negative Control Enrichment Testing

To assess the potential susceptibility our enrichment tests to false-positive results, we performed a negative control test [25]. We randomly generated sets of 1,000 pseudo-chemical gene lists, each containing either 100, 500, 1,000, 2,500, 5,000, or 10,000 genes. The sizes of the pseudo-chemical gene lists were selected to be inclusive of the sizes of the actual annotated chemical gene lists. We then tested each of the pseudo-chemical gene lists for enrichment with preterm birth genes. We calculated the number of tests meeting our significance criteria and visualized the resulting p values using a density plot. The results from these negative control tests can inform on the probability of observing enrichment by chance and provide a distribution to compare our observed enrichment test statistics. This procedure was conducted as described previously [22].

Testing Preterm Birth and Chemical Gene Lists for Enrichment with Gene Ontology Categories

To identify molecular mechanisms that could underlie chemical contributions to preterm birth, we conducted pathway analyses for both the preterm birth genes and chemical gene lists. Because our primary focus was on preterm birth in humans, we conducted the following analysis in the human data only. First, we used DAVID pathway enrichment software (version 6.8) [26] to perform pathway analysis for the preterm birth gene list. Of the significantly enriched pathways ($n = 77$), three were prioritized for downstream analysis because activation of one or more of these pathways represent potential mechanisms for toxicant contributions to preterm birth. These were "inflammatory response" (GO:0,006,954) [27], "aging" (GO:0,007,568) [28, 29], and "response to estradiol" (GO:0,032,355) [30]. A total of 90 preterm birth genes were involved in one or more of these pathways. We visualized chemical interactions with these genes using a heatmap. We then quantified the number genes impacted by each chemical in each of the three pathways.

Finally, to provide further evidence for chemical impacts on the three prioritized pathways (i.e., aging, inflammatory response, and response to estradiol), we used DAVID to conduct pathway enrichment analyses for each chemical gene list. We visualized the -$\log_{10}$(enrichment p value), for these prioritized pathways using heatmaps.

Results

Gene-Chemical Associations

For the 34 chemicals detected at higher levels in non-Hispanic Black women, the mean number of unique chemical-gene annotations was 535 for human, 445 for mouse, and 240 for rat (see Supplementary Table 1). The proportion of chemical-gene annotations from each species varied by chemical. For example, for methylmercury, 100% of the annotated genes came from rat data while for two chemicals (dibutyl phosphate and PCB187) 100% of the annotated genes came from human data. All other chemicals were annotated with genes from human, mouse, and rat (Supplementary Fig. 1).

Enrichment Testing for Chemical Associations with Preterm Birth

Chemicals with an expected frequency of less than one in the 2×2 descriptive tables were excluded from analysis. This resulted in 12 chemicals tested in human and mouse species and 15 tested in rat with 19 chemicals tested in at least one species. Figure 1 shows all chemicals tested. Figure 2 shows -$\log_{10}$(p values) for all

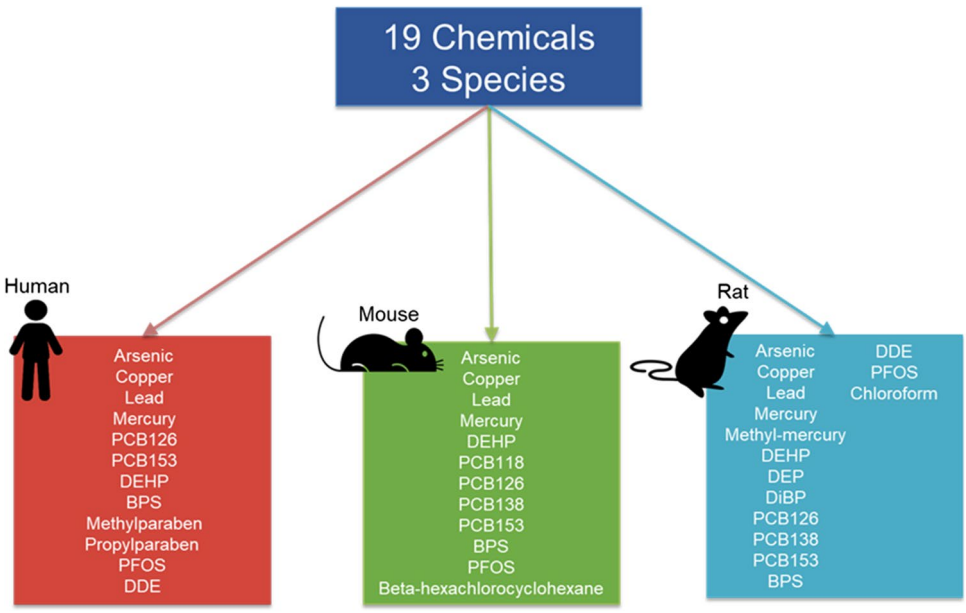


Fig. 1 Chemicals tested for enrichment with preterm birth genes. Based on disparities in average exposure levels between non-Hispanic Black vs. non-Hispanic White women (as indicated by NHANES data), we initially identified 40 chemicals for potential preterm birth enrichment testing. Nineteen of these chemicals had suffiently annotated data in the CTD to proceed with preterm birth enrichment tests in at least one species (human, mouse or rat). Chemical gene lists for all three species were obtainedfrom the CTD. Abbreviations-PCB: polychlorinated biphenyl, DEH:diethylhexyl phthalate, DEP:diethyl phthalate, DiBP: diisobutyl phthalate, BPS: bisphenol S, PFOS: perfluorooctane sulfonic acid, DDE; dichlorodiphenyldichloroethylene

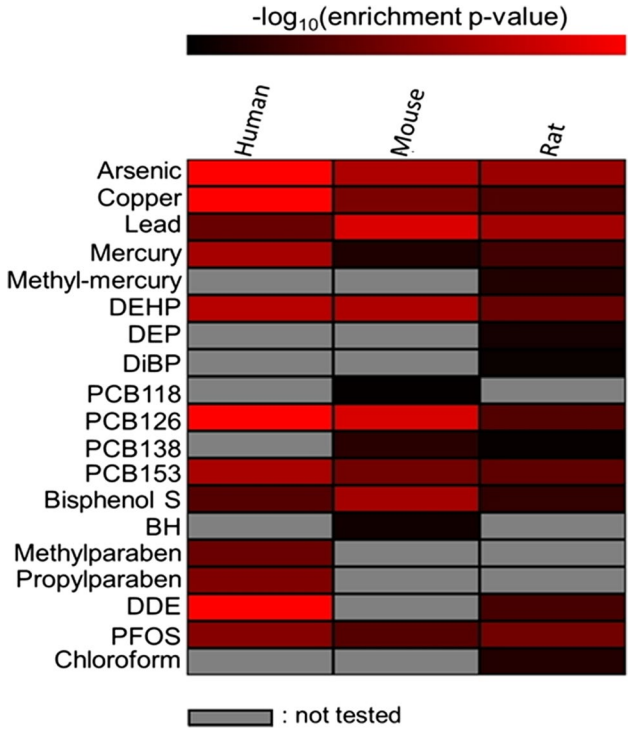


Fig. 2 Results for preterm birth enrichments tests, stratified by species. Chemical gene lists were downloaded from the Comparative Toxicogenomics Database and tested for enrichment with genes involved in preterm birth (genes obtained from the Preterm Birth Database) using Chi-squared tests. Heatmap shows-$\log_{10}$(p-value) for enrichment with preterm birth genes for each chemical. Enrichment testing was conducted in data obtained from three species: human, mouse and rat. Abbreviations: BH: beta-hexachlorocyclohexane, DDE:*p,p'*-dichlorodiphenyldichloroethylene, DEP: diethyl phthalate DEHP: diethylhexyl phthalate, PCB: polychlorinated biphenyl, PFOS: perfluorooctane sulfonate

chi-squared enrichment tests. In human data, 12 out of 12 chemicals tested (100%) were significantly enriched with preterm birth genes (adjusted chi-square p value < 0.05). In mouse data, 11 out of 12 chemicals tested (92%) were significantly enriched (PCB118 was not enriched). In rat data, 14 out of 15 chemicals tested (93%) were significantly enriched (PCB138 was not enriched). Nine chemicals were enriched in all three species: arsenic, copper, lead, mercury, DEHP, BPS, PCB126, PCB153, and PFOS.

Results of Sensitivity Analyses

We conducted Fisher's exact tests to confirm our primary chemical-disease enrichment analysis. For human and mouse data, all chemicals that were significantly enriched in the chi-squared tests were also enriched using Fisher's exact tests. For rat data, 13 out of the 15 chemicals were enriched using Fisher's exact test (PCB138 and DiBP were not significantly enriched), largely consistent with the chi-squared results.

To identify the magnitude of enrichment for significantly enriched chemicals, we calculated proportional reporting ratios. Chemicals with the highest proportional reporting ratios in human data included DDE (16), PCB126 (14.1), and propylparaben (14.1). In mouse data, the highest ratios were observed for lead (16), PCB126 (9.7), and arsenic (7.3). In rat data, the highest ratios were observed for arsenic (11), lead (9.4), and BPS (8.7). The results for all enrichment tests (chi-squared p values, Fisher's p-values, and proportional reporting ratios) are shown for all three species in Supplementary Tables 2–4.

Proportional Reporting Ratios Plotted by Relative Levels of Exposure in Non-Hispanic Black/Non-Hispanic White women

Proportional reporting ratios for all chemicals significantly enriched with preterm birth genes were plotted against the average fold difference in chemical biomarker concentrations in non-Hispanic Black vs. non-Hispanic White women. These results are shown for human data in Fig. 3A, mouse data in Fig. 3B, and rat data in Fig. 3C. As shown in Fig. 3A-C, proportional reporting ratios for all chemicals were > 1, indicating enrichment with preterm birth genes, none of the chemicals tested had proportional reporting ratios < 1 (which would indicate depletion). We noted that several chemicals were both observed at relatively high levels in non-Hispanic Black women (> 1.5-fold higher compared to non-Hispanic White women) and a relatively high degree of enrichment with preterm birth genes in at least one of the species tested. These were DDE, propylparaben, methylparaben, BPS, methyl mercury, and DEP.

Negative Control Testing of Preterm Birth Enrichment

We randomly permuted human, mouse, and rat pseudo-chemical gene sets (n = 100, 500, 1000, 2500, 5000, or 10,000) and tested for enrichment with preterm birth genes. For all gene set sizes and all species tested, less than 1% of permutation tests achieved a chi-square or Fisher's test corrected p value < 0.05, demonstrating that the random, pseudo-chemical gene sets were not enriched in preterm birth genes and suggesting that the observed enrichments with actual chemical gene lists were robust and specific. Supplementary Fig. 2 shows p value density plots for chi-square and Fisher's tests for human, mouse, and rat genes.

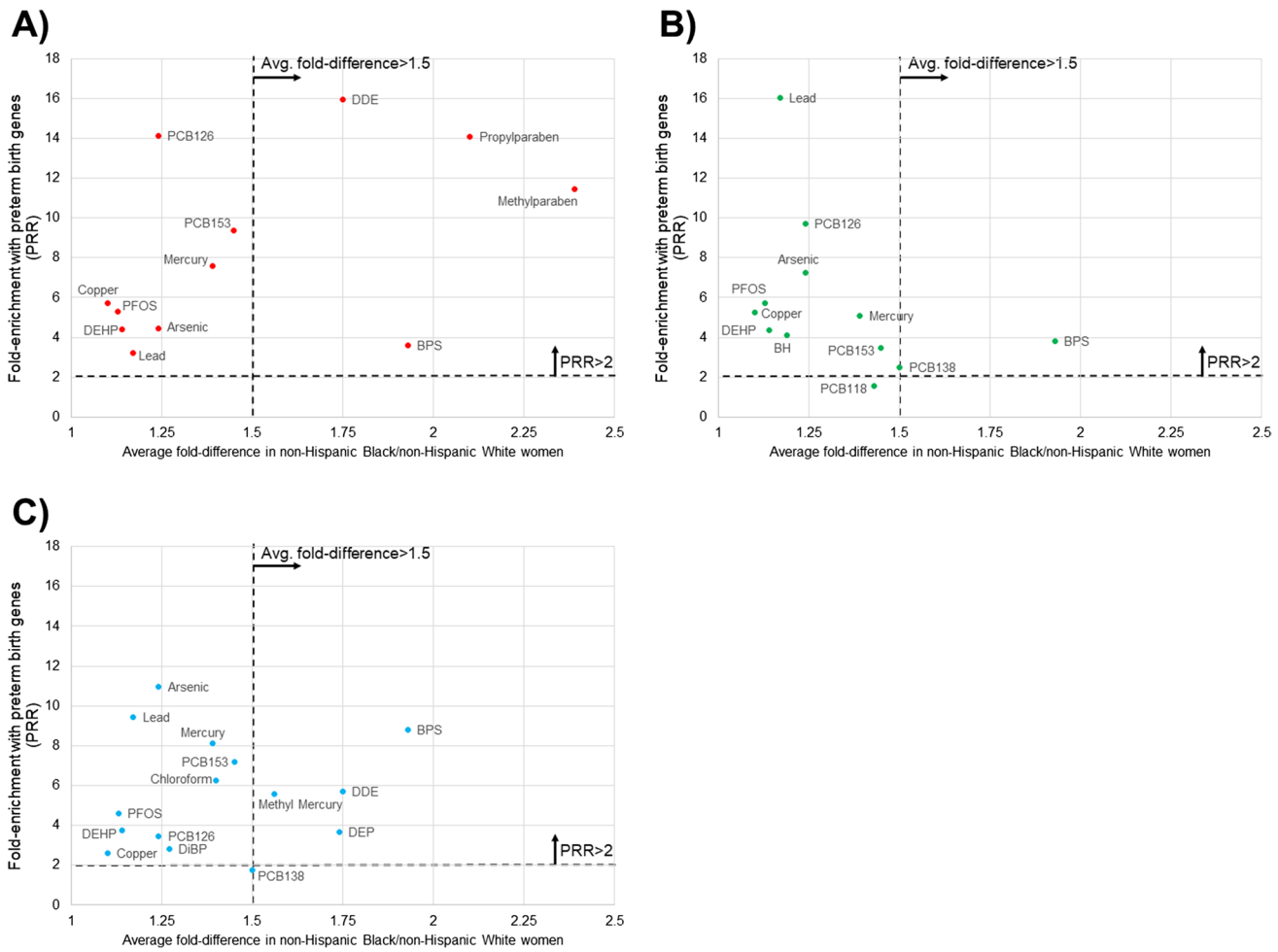


Fig. 3 Chemical enrichment with preterm birth genes plotted by relative difference in levels observed in non-Hispanic Black women vs. non-Hispanic White women. Chemical biomarker data for non-Hispanic Black and non-Hispanic White women in the United States were obtained from (Nguyen, et al. [6]). The degree of enrichment with preterm birth genes (as indicated by proportional reporting ratios) for all chemicals was plotted against the $\log_2$(fold-difference in levels observed in non-Hispanic Black/White women). Enrichment testing was conducted in data obtained from three species: **(a)** human, **(b)** mouse and **(c)** rat. Abbreviations: PRR: proportional reporting ratio, BH: beta-hexachlorocyclohexane, BPS: bisphenol S, DDE:*p,p'*-dichlorodiphenyldichloroethylene, DEP: diethyl phthalate, DEHP: diethylhexyl phthalate, PCB: polychlorinated biphenyl, PFOS: perfluorooctane sulfonate

Functional Analysis of Preterm Birth Gene List

Among the 640 preterm birth genes curated in the Database for Preterm Birth (dbPTB) [20], 417 (65%) were impacted by at least one chemical in the human data, but no genes were found to have interactions among all of the chemicals studied. The 417 genes affected by at least one chemical included collagen encoding genes (COL1A1, COL1A2, COL23A1, COL3A1, COL5A1, COL5A2), chemokines (e.g., CXCL8, CCL2, CCL3, CCL4), cytochrome P450s (CYP1A1, CYP17A1, CYP2D6, CYP2E1, and CYP19A1), coagulation factors (F2, F3, F5, F8, and F10), cytokines (e.g., IL1A, IL1B, IL2, IL4, IL5, IL6, IL7 IL10, IL13, IL15, and IL18), MAPs/MAPKs (MAPK1, MAPK2K1, MAPK2K2, MAPK2K3, MAPK2K6, MAPK14, MAPK3, MAPK8, and MAPK9), matrix metalloproteases (MMP1, MMP2, MMP3, MMP9, and MMP17), and toll-like receptors (TLR3, TLR4, TLR7, TLR8, and TLR10). Supplementary Table 5 shows gene-chemical interactions for all 640 preterm birth genes.

Pathway enrichment analysis of the 640 preterm birth genes identified 77 enriched biological pathways (gene ontology-defined biological processes) (FDR < 0.05). Several of the enriched pathways were consistent with the known etiology of preterm birth and the activation of labor. These processes included "inflammatory response" (GO:0,006,954, FDR = 3×10^{-19}), "aging" (GO:0,007,568, FDR = 4×10^{-8}), and "response to estradiol" (GO:0,032,355, FDR = 0.0002). We prioritized these

three pathways for downstream analysis. The full list of enriched pathways is contained in Supplementary Table 6.

Functional Analysis of Preterm Birth Genes Targeted by Enriched Chemicals

We identified chemical impacts on specific genes in our prioritized pathways ("inflammatory response," "aging," and "response to estradiol"). Figure 4A shows chemical interactions for preterm birth genes in each of the three pathways. Figure 4B shows the total number of preterm birth genes impacted by each chemical for the three pathways. We observed that most chemicals impact genes involved in all three pathways and that chemicals impact more genes involved in"inflammatory response" followed by "aging." We observed the fewest number of gene-chemical interactions for "response to estradiol."

Functional Analysis of Chemical Gene Lists

The number of enriched biological processes ($p < 0.05$) for each chemical in descending order were as follows: copper (498), arsenic (489), DEHP (449), DDE (356), PCB126 (335), mercury (248), PCB153 (234), propylparaben (233), PFOS (223), lead (217), BPS (207), and methylparaben (140). The full list of enriched biological pathways for each chemical is shown in Supplementary Table 7.

Figure 5 shows $-\log_{10}$(enrichment p values) for our prioritized pathways ("inflammatory response," "aging," and "response to estradiol") across all 12 chemicals. We observed two relatively distinct patterns of enrichment, with one set of chemicals primarily enriched with genes in the inflammatory response pathway (e.g., arsenic, PCB126) and another set primarily enriched with estradiol response genes (e.g., PFOS, propylparaben).

Discussion

Disparities in preterm birth rates between non-Hispanic Black and non-Hispanic White women are well-documented [31]. Due to the significant adverse maternal and child health effects associated with preterm birth, identifying environmental factors that contribute to disparities in preterm birth rates is crucial to achieving the goals of health equity and environmental justice. In this study, we used publicly available toxicogenomic data to evaluate the biological effects of exposure to a diverse set of chemicals found at higher

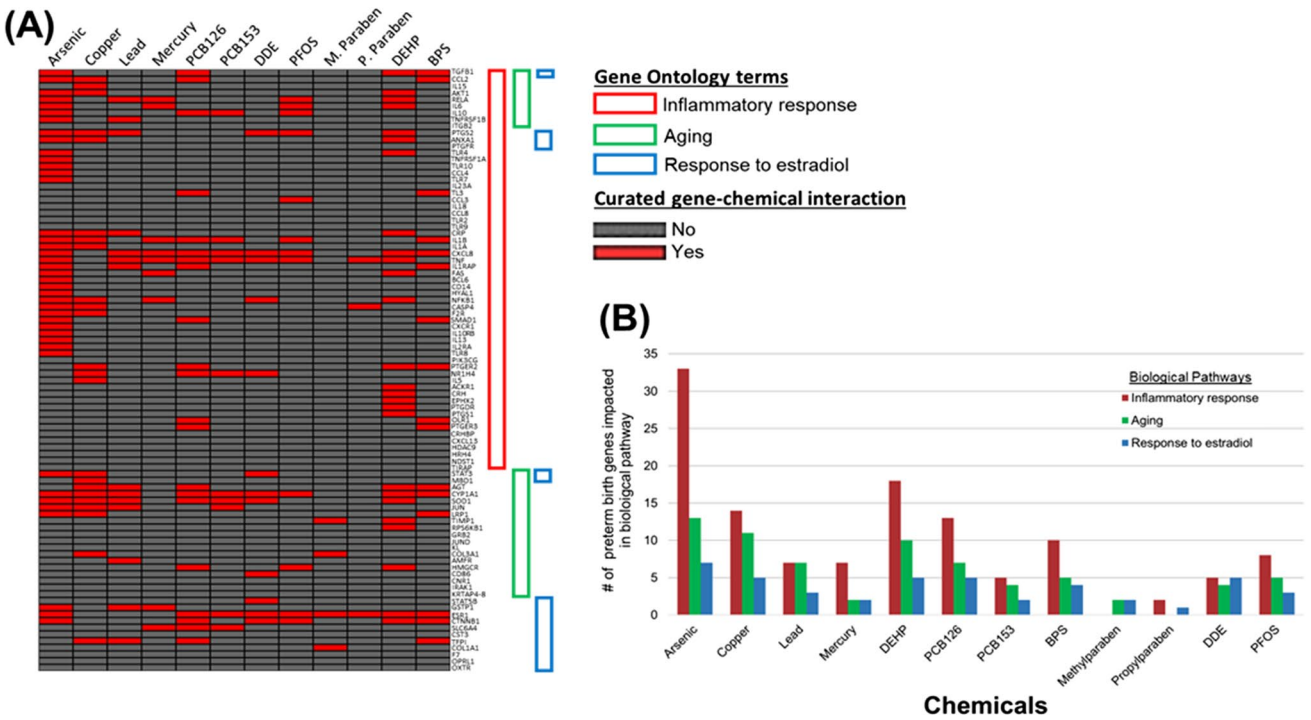


Fig. 4 Chemical-gene interactions for 90 genes in three key Gene Ontology (GO) terms for chemicals detected at higher levels in non-Hispanic Black women. We identified 77 GO terms enriched in the preterm birth gene lists. We selected three of these GO terms for downstream analysis due to their relevance to the etiology of preterm birth ("inflammatory response", "aging" and "response to estradiol") and plotted chemical-gene interactions for all preterm birth genes involved in on more of the prioritized biological pathway. Individual gene-chemical interactions are shown in **(A)** and the number of preterm birth genes in each biological pathway impacted by each chemicals are shown in **(B)**

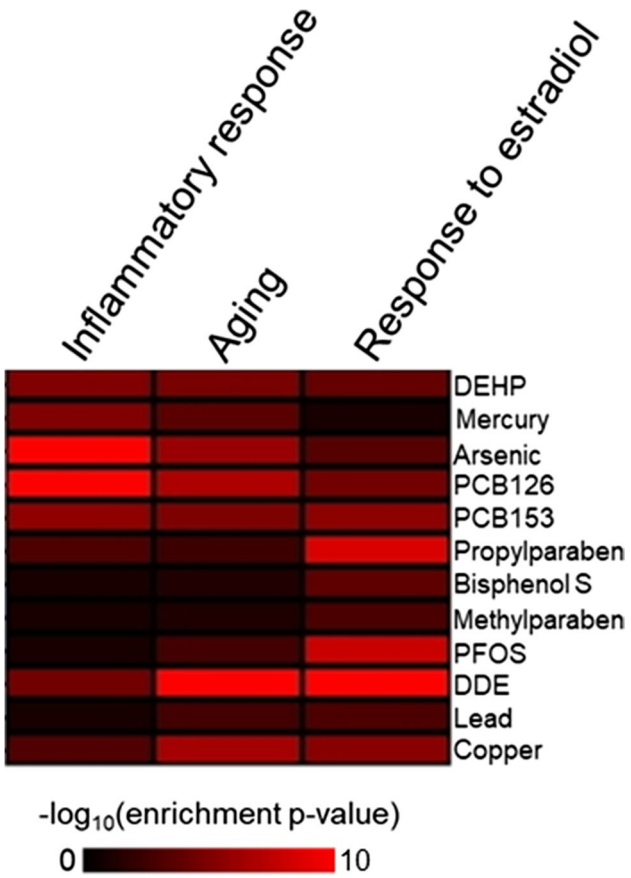


Fig. 5 Pathway analysis of chemical gene lists. Results for pathway enrichment analysis chemicals detected at higher levels in non-Hispanic Black vs. non-Hispanic White women. Gene lists for each chemical were analyzed using DAVID pathway analysis software. -$\log_{10}$(p-values) are shown for each pathway. Results are shown for the relevant following biological pathways: "Inflammatory Response", "Aging" and 'Response to Estradiol"

levels in non-Hispanic Black compared to non-Hispanic White women [6]. We found that a diverse set of chemicals detected in higher concentrations in non-Hispanic Black women in the USA have impacts on a high number of genes involved in preterm birth. These included metals, phthalates, PCBs, pesticides, personal care product components, and perfluorinated alkyl substances. Taken together, these results suggest that, as opposed to one single chemical or class of chemicals contributing to disparities in preterm birth rates, exposure to a diverse array of chemicals could play a significant role. This highlights the complex nature of studying preterm birth and a need to understand the mechanistic pathways through which toxicant exposures influence preterm birth risks.

Race and ethnicity are demographic constructs with interwoven societal, cultural, and environmental factors [32], and these factors likely play a major role in differing patterns of exposure observed between individuals of different races/ethnicities. For example, a recent study found that certain hair products used primarily by Black women contained multiple endocrine-disrupting chemicals including methylparaben, DEHP, and DEP [33]. In the current study, we observed that all three of these compounds were enriched for preterm birth genes. Moreover, our results are consistent with epidemiology studies showing links between DEHP [34], DEP [35, 36], and methylparaben [37] exposure and preterm birth or shorter pregnancy duration. Studies have identified the promotion of racialized beauty standards in advertising as well as the targeted marketing of certain beauty products which contain these chemicals to Black women (e.g., hair straighteners) as potential contributions to racial disparities in exposures [38, 39]. Importantly, Helms, et al. found that 84% of the chemicals detected in these products were not listed on product labels, presenting a key gap in information for consumers seeking to limit their exposures [33].

In addition to product usage, factors rooted in structural and environmental racism such as historical residential segregation can lead to disparities in exposures to environmental pollutants [40–42]. For example, studies have found racial disparities in the distribution of lead and arsenic in residential soil, with non-Hispanic Black mothers living near elevated soil concentrations [43]. Exposure to lead and arsenic in contaminated areas can occur through multiple pathways including dermal contact with soil, inhalation of soil/house dust particles [44–46], and incidental ingestion of soil [47, 48]. Both metals were enriched with preterm birth genes in our analysis, consistent with epidemiology studies showing links between arsenic [49–51] or lead [52, 53] exposure and preterm birth or lower gestational age. Racial disparities have also been identified in residential proximity to polluting facilities and in exposure to air pollution [54], representing another potential cause of racial disparities in exposures to air pollutants such as lead [55] and mercury [56]. Notably, both metals were enriched with preterm birth gene interactions in our study. Thus, a number of societal forces, including structural racism and patterns of product usage, potentially contribute to differences in patterns of exposure between non-Hispanic Black versus non-Hispanic White women. These factors emphasize the importance of using a multidisciplinary approach including environmental epidemiology, molecular toxicology, and exposure sciences in the fight against preterm birth. The present study demonstrates one such approach by incorporating exposure data into molecular/mechanistic evaluation of chemical impacts on preterm birth pathways.

Our preterm birth enrichment results were consistent with several epidemiology studies that showed associations between preterm birth and exposure to metals like arsenic [49], lead [57], and copper [58]. Other epidemiology studies found preterm birth associations with one or more of

the phthalates [34], phenols [59], parabens [60], PCBs [61], and pesticides [62] that were enriched with preterm birth genes. However, some epidemiology studies report mixed or inconclusive findings regarding associations between some of these chemicals and preterm birth. For example, one study showed that methylparaben was associated with decreased odds of preterm birth [63] and bisphenol S was associated with increased odds of late-term delivery when births were stratified by sex (effects were observed in female fetuses only) [64]. Overall, our results highlight the need to consider exposures to multiple diverse chemicals to assess the contribution of environmental toxicants to preterm birth, rather than exposure to one chemical in isolation.

In a 2013 review, Ferguson et al. outlined multiple challenges in assessing links between chemical exposures and preterm birth [62]. These included a need for further research into the biological pathways underpinning links between environmental contaminants and preterm birth as well as research into the effects of mixed exposures. The methods used in the current study addressed some of these challenges using an approach which allowed us to screen a structurally diverse set of chemicals for significant associations with preterm birth genes while also allowing us to identify common potential pathway targets across these diverse compounds. In addition, by integrating exposure and toxicogenomic data, we were able to identify multiple compounds with a both a relatively high level of exposure in non-Hispanic Black women and high degrees of enrichment with preterm birth genes: DDE, propylparaben, methylparaben, BPS, methyl mercury, and DEP. This diverse set of compounds include both environmental toxicants such as pesticides (DDE, a metabolite of the pesticide dichlorodiphenyltrichloroethane, i.e., DDT) [65] and heavy metals (methyl mercury) as well as contaminants found in personal care products like hair and skin lotions (methylparaben and DEP) [33, 66]. These results may help in prioritizing chemicals for further study and/or exposure reduction in the fight against preterm birth disparities.

By identifying specific preterm birth genes and pathways impacted by chemicals, we were able to gain insight into plausible mechanisms linking exposures to preterm birth. Our analysis showed that the chemicals we studied impact genes involved in multiple biological pathways relevant to the etiology of preterm birth. These included inflammatory response, aging, and hormone pathways. All of these biological processes play important roles in the process of term and preterm labor. For example, increases in pro-inflammatory cytokines play important roles in both term and preterm labor, such as stimulating fetal membrane rupture and uterine activation (i.e., increased uterine expression of contraction-associated proteins and sensitivity to utertonins) [67]. In addition, the process of aging in the fetal membranes involves telomere-dependent cellular senescence (loss of cell division potential), which is a key factor in membrane weakening and rupture at the end of pregnancy [29]. Finally, hormone pathways such as estradiol signaling are critical for the maintenance of pregnancy and in the process of labor, e.g., estradiol promotes the expression of pro-contraction proteins in the uterine myometrium [68]. As shown in our analyses, most of the chemicals tested impact genes involved in all three of these pathways ("inflammatory response," "aging," and "response to estradiol"), providing a potential mechanistic link between chemical exposures and preterm birth. Our findings could help to identify important molecular or cellular targets of toxicants and aid in the design of future toxicological studies. For example, chemicals with a relatively high number of interactions with genes involved in aging could be assessed for their capacity to induce cellular senescence in fetal membranes, either individually or in chemical mixtures. Such studies could provide a mechanistic link between chemical exposures and premature rupture of the membranes [69]. Similarly, BPS was shown to impact genes in the estradiol signaling pathway ("response to estradiol"), consistent with studies showing an ability for BPS to bind estrogen receptors and trigger downstream transcriptional activity [70, 71]. This provides a potential mechanistic explanation for BPS contributions to preterm birth, as estrogens can promote myometrial activation and upregulate uterotonin receptors [72]. The potential for BPS to act through this mechanism warrants further toxicological study, including in mixtures with other preterm birth-enriched chemicals that impact this pathway. Thus, our results could inform future studies which seek to group or prioritize chemicals based on similarity of pathways affected in order to identify chemical mixtures of concern for preterm birth.

Strengths of this study include the use of human exposure data to direct the selection of chemicals for our analysis, allowing us to focus on chemicals most likely to contribute to preterm birth disparities. In addition, our use of publicly available datasets allowed us to evaluate a range of chemicals in a relatively short amount of time compared to what is required for traditional in vitro or in vivo toxicology studies. Our methods also allowed us to compare chemical impacts across critical pathways relevant to preterm birth, providing plausible mechanistic linkages between chemical exposures and preterm birth. Finally, by validating the preterm birth gene list and conducting sensitivity analysis for our enrichment results, we implemented an improved level of statistical rigor compared to other studies using similar methods.

In addition to the strengths of the study, some important limitations should also be noted. Our methods were limited to detecting statistical associations between chemicals and preterm birth genes. Thus, the directionality of toxicant-induced gene expression changes, dose–response relationships, and specificity of tissues or cell type affected were not included in our analyses. Future experiments

using in vitro or in vivo models could be used to confirm the results we observed in this study and determine causality between toxicant exposure and activation of preterm birth pathways. In addition, future epidemiology studies that focus on links between the usage of products containing preterm birth enriched chemicals identified in this study and/or environmental exposures could help to confirm relationships between exposures and preterm birth. In addition, our chemical and preterm birth gene lists were limited to those curated in the CTD and Preterm Birth Database. Although both of these databases are curated by trained experts, biases could arise due to incomplete data curation or a lack of available data (e.g., a lack of adequate toxicogenomic data for a given toxicant). Additional studies using data from additional databases could be used to confirm the results observed in this study.

In conclusion, our study adds to the body of knowledge that environmental toxicant exposures contribute to disparities in preterm birth rates between non-Hispanic Black versus non-Hispanic White women. Our results suggest that exposures to multiple environmental toxicants may drive these effects as opposed to one chemical in isolation. Future toxicological and epidemiological studies investigating the chemicals highlighted in this study could inform important public health efforts to address racial disparities in preterm birth rates, such as changes in environmental policies or behavioral interventions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43032-022-00870-w.

Funding This research was supported by the Michigan Center on Lifestage Environmental Exposures and Disease (P30ES017885). Drs. Harris, Bakulski and Loch-Caruso were supported by the National Institute of Environmental Health Sciences of the National Institutes of Health under Award Number P42ES017198. Additional funding for Dr. Harris was provided by the National Center for Advancing Translational Sciences (UL1TR002240). Dr. Bakulski was supported by grants from the National Institutes of Health (R01ES025531; R01ES025574; R01AG055406; R01MD013299). Dr. Colacino and Dr. Nguyen were supported by grants from the National Institutes of Health (R01ES028802 and R01ES028802S1). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability Not applicable.

Code Availability The R markdown code used to conduct these analyses and generate figures is available at https://github.com/bakulskilab.

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The authors declare no competing interests.

References

1. Hamilton, B.M., JA., Osterman, MJK., *Births: provisional data for 2019*. 2020, Centers for Disease Control and Prevention.
2. Pike KC, Lucas JS. Respiratory consequences of late preterm birth. Paediatr Respir Rev. 2015;16(3):182–8.
3. Behrman, R.E., Butler, A. S., *Preterm birth: causes, consequences, and prevention*, R.E. Behrman and A.S. Butler, Editors. 2007: Washington (DC).
4. Grosse, S.D., et al., *Employer-sponsored plan expenditures for infants born preterm.* Pediatrics, 2017. **140**(4).
5. Ruiz D, et al. Disparities in environmental exposures to endocrine-disrupting chemicals and diabetes risk in vulnerable populations. Diabetes Care. 2018;41(1):193–205.
6. Nguyen VK, et al. A comprehensive analysis of racial disparities in chemical biomarker concentrations in United States women, 1999–2014. Environ Int. 2020;137:105496.
7. York TP, et al. The contribution of genetic and environmental factors to the duration of pregnancy. Am J Obstet Gynecol. 2014;210(5):398–405.
8. Norwitz, E.R., et al., *Molecular regulation of parturition: The role of the decidual clock.* Cold Spring Harb Perspect Med, 2015. **5**(11).
9. Kota SK, et al. Endocrinology of parturition. Indian J Endocrinol Metab. 2013;17(1):50–9.
10. Romero R, et al. Inflammation in preterm and term labour and delivery. Semin Fetal Neonatal Med. 2006;11(5):317–26.
11. Menon R, et al. Histological evidence of oxidative stress and premature senescence in preterm premature rupture of the human fetal membranes recapitulated in vitro. Am J Pathol. 2014;184(6):1740–51.
12. Schmidt A, et al. Only humans have human placentas: molecular differences between mice and humans. J Reprod Immunol. 2015;108:65–71.
13. Faas MM, et al. Species differences in the effect of pregnancy on lymphocyte cytokine production between human and rat. J Leukoc Biol. 2005;78(4):946–53.
14. Keirse MJ, et al. Comparison of intrauterine prostaglandin metabolism during pregnancy in man, sheep and guinea pig. Eur J Obstet Gynecol Reprod Biol. 1978;8(4):195–203.
15. Grigsby PL. Animal models to study placental development and function throughout normal and dysfunctional human pregnancy. Semin Reprod Med. 2016;34(1):11–6.
16. Woodruff TJ, Zota AR, Schwartz JM. Environmental chemicals in pregnant women in the United States: NHANES 2003–2004. Environ Health Perspect. 2011;119(6):878–85.
17. Maekawa R, et al. Evidence of exposure to chemicals and heavy metals during pregnancy in Japanese women. Reprod Med Biol. 2017;16(4):337–48.
18. Boyle EB, et al. Assessment of Exposure to VOCs among Pregnant Women in the National Children's Study. Int J Environ Res Public Health. 2016;13(4):376.
19. Davis AP, et al. The comparative toxicogenomics database: update 2019. Nucleic Acids Res. 2019;47(D1):D948–54.
20. Uzun, A., et al., *dbPTB: a database for preterm birth.* Database (Oxford), 2012. **2012**: p. bar069.
21. Davis AP, et al. Comparative toxicogenomics database (CTD): update 2021. Nucleic Acids Res. 2021;49(D1):D1138–43.
22. Harris SM, et al. Identification of environmental chemicals targeting miscarriage genes and pathways using the comparative toxicogenomics database. Environ Res. 2020;184:109259.

23. Campbell I. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. Stat Med. 2007;26(19):3661–75.
24. Saeed AI, et al. TM4: a free, open-source system for microarray data management and analysis. Biotechniques. 2003;34(2):374–8.
25. Camargo A, et al. Permutation - based statistical tests for multiple hypotheses. Source Code Biol Med. 2008;3:15.
26. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57.
27. Cappelletti M, et al. Inflammation and preterm birth. J Leukoc Biol. 2016;99(1):67–78.
28. Polettini J, et al. Aging of intrauterine tissues in spontaneous preterm birth and preterm premature rupture of the membranes: a systematic review of the literature. Placenta. 2015;36(9):969–73.
29. Menon R, Richardson LS, Lappas M. Fetal membrane architecture, aging and inflammation in pregnancy and parturition. Placenta. 2019;79:40–5.
30. Smith R. Parturition. N Engl J Med. 2007;356(3):271–83.
31. Martin, J.A., et al., *Births: final data for 2018*, in *National Vital Statistics Reports*. 2019.
32. James-Todd TM, Chiu YH, Zota AR. Racial/ethnic disparities in environmental endocrine disrupting chemicals and women's reproductive health outcomes: epidemiological examples across the life course. Curr Epidemiol Rep. 2016;3(2):161–80.
33. Helm JS, et al. Measurement of endocrine disrupting and asthma-associated chemicals in hair products used by Black women. Environ Res. 2018;165:448–58.
34. Ferguson KK, McElrath TF, Meeker JD. Environmental phthalate exposure and preterm birth. JAMA Pediatr. 2014;168(1):61–7.
35. Polanska K, et al. Effect of environmental phthalate exposure on pregnancy duration and birth outcomes. Int J Occup Med Environ Health. 2016;29(4):683–97.
36. Radke EG, et al. Phthalate exposure and female reproductive and developmental outcomes: a systematic review of the human epidemiological evidence. Environ Int. 2019;130:104580.
37. Baker BH, et al. Methylparaben in meconium and risk of maternal thyroid dysfunction, adverse birth outcomes, and attention-deficit hyperactivity disorder (ADHD). Environ Int. 2020;139:105716.
38. Zota, A.R. and B. Shamasunder, *The environmental injustice of beauty: framing chemical exposures from beauty products as a health disparities concern.* Am J Obstet Gynecol, 2017. **217**(4): p. 418 e1-418 e6.
39. Bristor J, Gravois R, Hunt M. Race and ideology: African-American images in television advertising. J Public Policy Mark. 1995;14(1):48–59.
40. Bravo MA, et al. Racial isolation and exposure to airborne particulate matter and ozone in understudied US populations: environmental justice applications of downscaled numerical model output. Environ Int. 2016;92–93:247–55.
41. Morello-Frosch R, Jesdale BM. Separate and unequal: residential segregation and estimated cancer risks associated with ambient air toxics in U.S. metropolitan areas. Environ Health Perspect. 2006;114(3):386–93.
42. Nardone A, et al. Redlines and greenspace: The relationship between historical redlining and 2010 greenspace across the united states. Environ Health Perspect. 2021;129(1):17006.
43. Davis HT, et al. Potential sources and racial disparities in the residential distribution of soil arsenic and lead among pregnant women. Sci Total Environ. 2016;551–552:622–30.
44. Chen H, et al. Contamination features and health risk of soil heavy metals in China. Sci Total Environ. 2015;512–513:143–53.
45. Li, S., et al., *Heavy Metal(loid)s Contamination in ground dust and associated health risks at a former indigenous zinc smelting area.* Int J Environ Res Public Health, 2021. **18**(3).
46. Hogervorst J, et al. House dust as possible route of environmental exposure to cadmium and lead in the adult general population. Environ Res. 2007;103(1):30–7.
47. Murphy BL, Toole AP, Bergstrom PD. Health risk assessment for arsenic contaminated soil. Environ Geochem Health. 1989;11(3–4):163–9.
48. Juhasz AL, Weber J, Smith E. Impact of soil particle size and bioaccessibility on children and adult lead exposure in peri-urban contaminated soils. J Hazard Mater. 2011;186(2–3):1870–9.
49. Almberg KS, et al. Arsenic in drinking water and adverse birth outcomes in Ohio. Environ Res. 2017;157:52–9.
50. Ahmad SA, et al. Arsenic in drinking water and pregnancy outcomes. Environ Health Perspect. 2001;109(6):629–31.
51. Laine JE, et al. Maternal arsenic exposure, arsenic methylation efficiency, and birth outcomes in the Biomarkers of Exposure to ARsenic (BEAR) pregnancy cohort in Mexico. Environ Health Perspect. 2015;123(2):186–92.
52. Torres-Sanchez LE, et al. Intrauterine lead exposure and preterm birth. Environ Res. 1999;81(4):297–301.
53. Zhang B, et al. Prenatal exposure to lead in relation to risk of preterm low birth weight: A matched case-control study in China. Reprod Toxicol. 2015;57:190–5.
54. Mohai P, et al. Racial and socioeconomic disparities in residential proximity to polluting industrial facilities: evidence from the Americans' Changing Lives Study. Am J Public Health. 2009;99(Suppl 3):S649–56.
55. Frank JJ, et al. Systematic review and meta-analyses of lead (Pb) concentrations in environmental media (soil, dust, water, food, and air) reported in the United States from 1996 to 2016. Sci Total Environ. 2019;694:133489.
56. ThanosBourtsalas AC, Themelis NJ. *Major sources of mercury emissions to the atmosphere: The U.S. case*. Waste Manag. 2019;85:90–4.
57. Cheng L, et al. Fetal exposure to lead during pregnancy and the risk of preterm and early-term deliveries. Int J Hyg Environ Health. 2017;220(6):984–9.
58. Kim SS, et al. Urinary trace metals individually and in mixtures in association with preterm birth. Environ Int. 2018;121(Pt 1):582–90.
59. Aung MT, et al. Preterm birth in relation to the bisphenol A replacement, bisphenol S, and other phenols and parabens. Environ Res. 2019;169:131–8.
60. Geer LA, et al. Association of birth outcomes with fetal exposure to parabens, triclosan and triclocarban in an immigrant population in Brooklyn. New York J Hazard Mater. 2017;323(Pt A):177–83.
61. Tsukimori K, et al. Long-term effects of polychlorinated biphenyls and dioxins on pregnancy outcomes in women affected by the Yusho incident. Environ Health Perspect. 2008;116(5):626–30.
62. Ferguson KK, O'Neill MS, Meeker JD. Environmental contaminant exposures and preterm birth: a comprehensive review. J Toxicol Environ Health B Crit Rev. 2013;16(2):69–113.
63. Aker AM, et al. The associations between prenatal exposure to triclocarban, phenols and parabens with gestational age and birth weight in northern Puerto Rico. Environ Res. 2019;169:41–51.
64. Wan Y, et al. Relationship between maternal exposure to bisphenol S and pregnancy duration. Environ Pollut. 2018;238:717–24.
65. Kezios KL, et al. Dichlorodiphenyltrichloroethane (DDT), DDT metabolites and pregnancy outcomes. Reprod Toxicol. 2013;35:156–64.
66. Koniecki D, et al. Phthalates in cosmetic and personal care products: concentrations and possible dermal exposure. Environ Res. 2011;111(3):329–36.
67. Romero R, Dey SK, Fisher SJ. Preterm labor: one syndrome, many causes. Science. 2014;345(6198):760–5.

68. Welsh T, et al. Estrogen receptor (ER) expression and function in the pregnant human myometrium: estradiol via ERalpha activates ERK1/2 signaling in term myometrium. J Endocrinol. 2012;212(2):227–38.
69. Mercer, B.M., *Preterm premature rupture of the membranes: diagnosis and management.* Clin Perinatol, 2004. **31**(4): p. 765-82, vi.
70. Hashimoto Y, et al. Measurement of estrogenic activity of chemicals for the development of new dental polymers. Toxicol In Vitro. 2001;15(4–5):421–5.
71. Kang JS, et al. Estrogenic potency of bisphenol S, polyethersulfone and their metabolites generated by the rat liver S9 fractions on a MVLN cell using a luciferase reporter gene assay. Reprod Biol Endocrinol. 2014;12:102.
72. Wood SL, et al. Endocrine disruptors and spontaneous premature labor: a case control study. Environ Health. 2007;6:35.

Title: A Data Mining Approach Reveals Chemicals Detected at Higher Levels in Non-Hispanic Black Women Target Preterm Birth Genes and Pathways

Authors: Sean M. Harris[a], Justin Colacino[a,b,c], Miatta Buxton[e], Lauren Croxton[f], Vy Nguyen[a,d], Rita Loch-Caruso[a], and Kelly M. Bakulski[e]

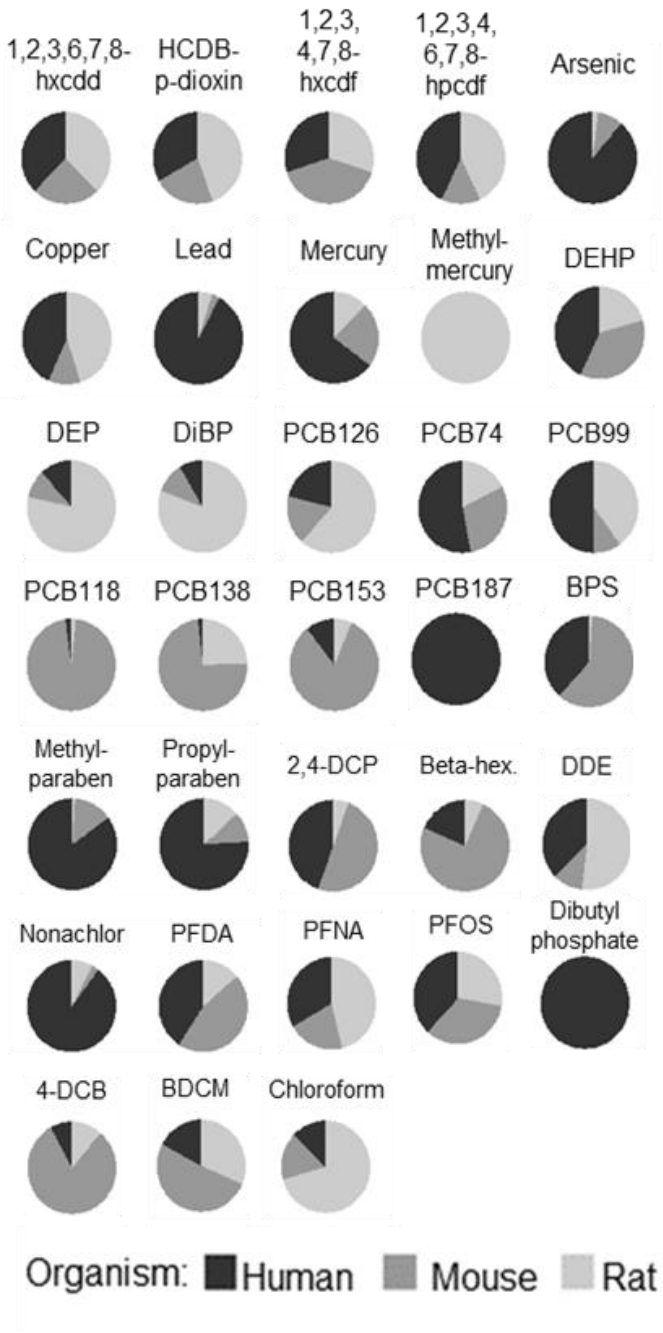
Affiliations:

[a]Department of Environmental Health Sciences, School of Public Health, University of Michigan, Ann Arbor, MI
[b]Department of Nutritional Sciences, School of Public Health, University of Michigan, Ann Arbor, MI
[c]Center for Computational Medicine and Bioinformatics, Medical School, University of Michigan, Ann Arbor, MI
[d]Department of Computational Medicine and Bioinformatics, Medical School, University of Michigan, Ann Arbor, MI
[e]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI
[f]College of Literature, Science and the Arts, University of Michigan, Ann Arbor, MI

Supplementary Material

Supplementary Table 1. Number of unique genes associated with 34 chemicals in the Comparative Toxicogenomics Database detected at significantly higher levels in non-Hispanic Black women compared to non-Hispanic White women.

	# of curated chemical-gene interactions			
Chemical	**All Species**	**Human**	**Mouse**	**Rat**
Arsenic	5,378	4,530	546	187
Copper	8,854	2,994	1,386	2,892
Lead	3,512	2,864	217	320
Mercury	980	456	297	95
Methyl-mercury	85	0	0	85
PCB74	22	11	7	3
PCB99	15	5	2	7
PCB118	1,461	82	1,302	63
PCB126	1,475	306	249	919
PCB138	2,223	41	1,610	557
PCB153	2,864	340	2,124	220
PCB187	1	1	0	0
beta-hexachlorocyclohexane	201	60	119	15
1,4-dichlorobenzene	64	5	52	7
2,4-dichlorophenol	18	8	9	1
DDE	881	241	57	224
Bromodichloromethane	75	6	42	24
Chloroform	114	14	25	72
PFNA	204	26	17	84
PFOS	2,849	834	1,007	673
Methyl Paraben	129	104	19	1
Propyl Paraben	504	93	11	12
PFDA	192	33	97	17
DEHP	7,847	2,382	3,216	1,311
DEP	220	39	19	157
DiBP	1,341	26	18	136
Bisphenol S	5,840	2,623	2,660	49
Heptachlorodibenzo-p-dioxin	9	3	2	4
1,2,3,4,7,8-hexachlorodibenzofuran	10	3	4	3
1,2,3,6,7,8-hexachlorodibenzodioxin	8	3	2	3
1,2,3,4,6,7,8-heptachlorodibenzofuran	7	3	1	3
Nonachlor	60	48	2	3
Oxychlordane	1	0	0	0
Dibutyl phosphate	2	1	0	0
Mean	**1,395**	**535**	**445**	**240**
Median	**197**	**36**	**22**	**37**



Supplementary Figure 1. Proportion of genes for each chemical by species. Lists of genes interacting with 34 chemicals in any of three species (*Homo sapiens*, *Mus musculus* or *Rattus norvegicus*) were downloaded from the Comparative Toxicogenomics Database. The relative percentage of genes annotated to each species is shown for each chemical. One chemical (oxychlordane) had no gene interactions for any of the three species and is not shown. Abbreviations: 1,2,3,6,7,8-hxcdd (1,2,3,6,7,8-hexachlorodibenzodioxin), HCDB-p-dioxin (1,2,3,6,7,8-hexachlorodibenzodioxin), 1,2,3,4,7,8-hxcdf (1,2,3,4,7,8-hexachlorodibenzofuran), 1,2,3,4,6,7,8-hpcdf (1,2,3,4,6,7,8-hepatochlorodibenzofuran), DEHP (diethylhexyl phthalate), DEP (diethyl phthalate), DiBP (diisobutyl phthalate), PCB (polychlorinated biphenyl), BPS (bisphenol S), 2,4-DCP (2,4-dichlorophenol), Beta-hex (beta-hexachlorocyclohexane), DDE (dichlorodiphenyl dichloroethylene), PFDA (perfluorodecanoic acid), PFNA (perfluoro-n-nonanoic acid), PFOS (perfluorooctane sulfonic acid), 4-DCB (4-dichlorobenzne), BDCM (bromodichloromethane)

Supplementary Table 2. All enrichment test statistics in human data

Chemical	Chi-squared p-value	Fisher's p-value	PRR (CI)
Arsenic	$2x10^{-80}$	$5x10^{-53}$	4.4 (3.8-5.2)
Copper	$4x10^{-99}$	$1x10^{-55}$	5.7 (4.8-6.8)
Lead	$2x10^{-31}$	$3x10^{-22}$	3.2 (2.6-3.9)
Mercury	$4x10^{-50}$	$2x10^{-22}$	7.6 (5.6-10.3)
DEHP	$3x10^{-54}$	$3x10^{-33}$	4.4 (3.6-5.4)
PCB126	$2x10^{-142*}$	$1x10^{-45}$	14.1 (11.0-18.0)
PCB153	$4x10^{-50*}$	$4x10^{-20}$	9.4 (6.7-13.1)
BPS	$1x10^{-28}$	$7x10^{-19}$	3.6 (2.8-4.6)
Methylparaben	$2x10^{-32*}$	$7x10^{-12}$	11.4 (7.2-18.3)
Propylparaben	$2x10^{-38*}$	$2x10^{-12}$	14.1 (8.8-22.6)
DDE	$8x10^{-87*}$	$2x10^{-25}$	16.0 (11.5-22.2)
DEET	$2x10^{-6}$	$8x10^{-5}$	3.0 (1.9-4.9)
PFOS	$2x10^{-40}$	$6x10^{-22}$	5.3 (4.1-6.9)

*Chi-squared "N-1" test
PRR: Proportional reporting ratio
CI: Confidence interval

Supplementary Table 3. All enrichment test statistics in mouse data

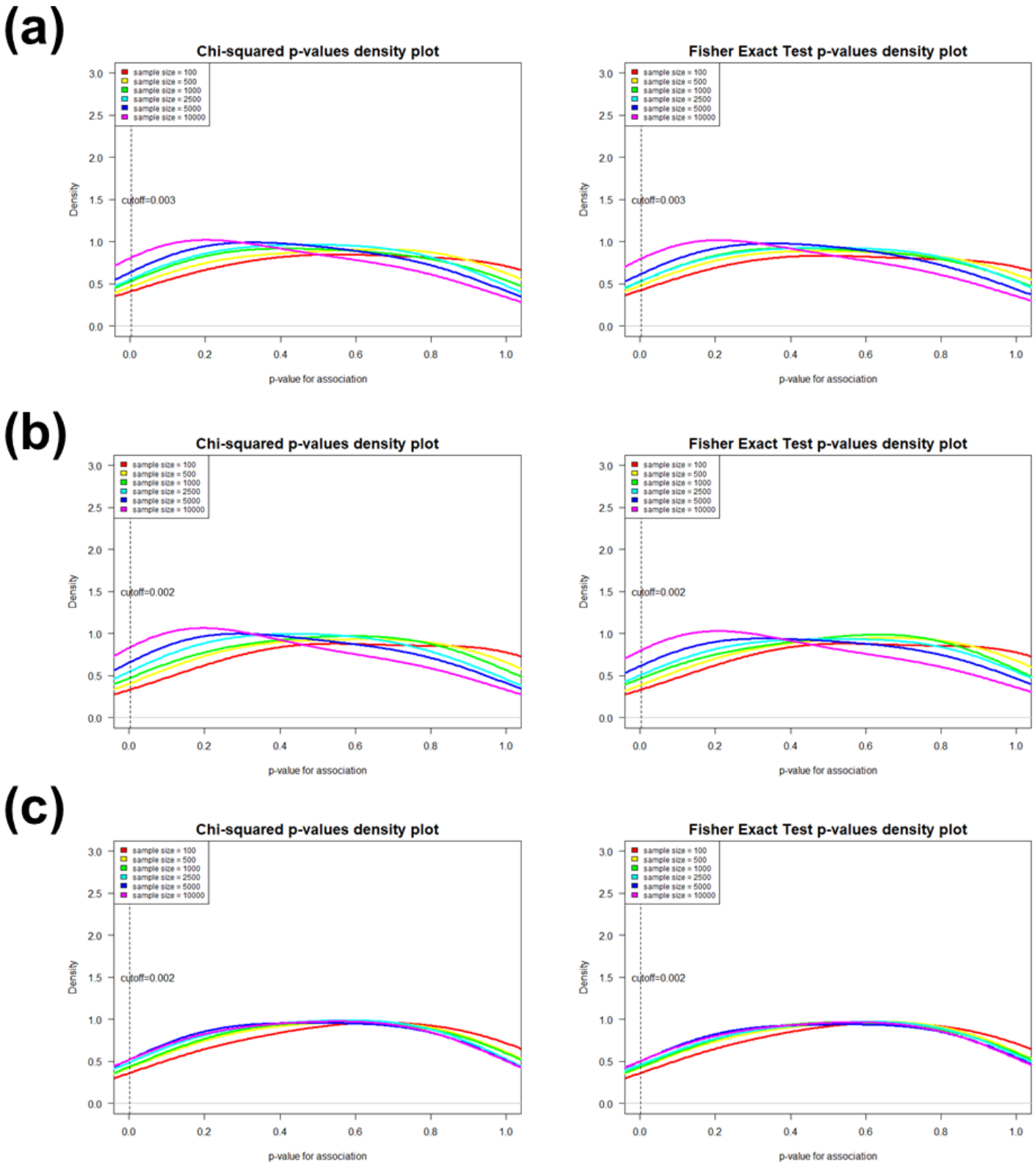
Chemical	Chi-squared p-value	Fisher's p-value	PRR (CI)
Arsenic	$1x10^{-52}$	$1x10^{-24}$	7.3 (5.5-9.6)
Copper	$2x10^{-37}$	$1x10^{-20}$	5.3 (4.0-6.9)
Lead	$3x10^{-65}$	$2x10^{-19}$	16.0 (11.0-23.0)
Mercury	$3x10^{-10}$	$6x10^{-6}$	5.1 (2.9-8.8)
DEHP	$1x10^{-52}$	$5x10^{-33}$	4.4 (3.6-5.3)
PCB118	0.01	0.02	1.6 (1.1-2.2)
PCB126	$6x10^{-65}$	$8x10^{-26}$	9.7 (7.2-13.1)
PCB138	$6x10^{-13}$	$6x10^{-10}$	2.5 (1.9-3.2)
PCB153	$2x10^{-34}$	$4x10^{-23}$	3.5 (2.8-4.3)
BPS	$1x10^{-49}$	$1x10^{-33}$	3.8 (3.2-4.6)
Beta-hex.	$1x10^{-5}$	$8x10^{-4}$	4.1 (2.1-8.1)
PFOS	$6x10^{-49}$	$5x10^{-26}$	5.7 (4.4-7.4)

*Chi-squared "N-1" test
PRR: Proportional reporting ratio
CI: Confidence interval

Supplementary Table 4. All enrichment test statistics in rat data

Chemical	Chi-squared p-value	Fisher's p-value	PRR (CI)
Arsenic	$1x10^{-46}$	$1x10^{-17}$	11.0 (7.6-15.7)
Copper	$4x10^{-24}$	$8x10^{-19}$	2.6 (2.2-3.2)
Lead	$9x10^{-50}$	$3x10^{-20}$	9.4 (6.8-13.1)
Mercury	$9x10^{-21}$	$2x10^{-9}$	8.2 (5.0-13.1)
Methylmercury	$1x10^{-10}$	$5x10^{-6}$	5.6 (3.2-9.7)
DEHP	$5x10^{-32}$	$4x10^{-21}$	3.7 (3.0-4.7)
DEP	$5x10^{-7}$	$7x10^{-5}$	3.7 (2.2-6.2)
DiBP	0.001	0.005	2.8 (1.5-5.3)
PCB126	$3x10^{-25}$	$3x10^{-17}$	3.5 (2.7-4.4)
PCB138	0.008	0.01	1.8 (1.2-2.7)
PCB153	$3x10^{-29}$	$7x10^{-14}$	7.2 (5.0-10.5)
BPS	$2x10^{-15}$	$7x10^{-7}$	8.7 (4.9-15.8)
DDE	$1x10^{-21}$	$1x10^{-11}$	5.7 (3.9-8.4)
PFOS	$2x10^{-34}$	$2x10^{-20}$	4.6 (3.6-6.0)
Chloroform	$2x10^{-11}$	$5x10^{-6}$	6.3 (3.5-11.2)

*Chi-squared "N-1" test
PRR: Proportional reporting ratio
CI: Confidence interval



Supplementary Figure 2. Validation of preterm birth enrichment findings. In order to assess the preterm birth gene list for susceptibility to false-positive results in enrichment testing, we performed validation using simulated gene lists. Genes lists of varying sizes (100, 500, 1000, 5000, and 10,000 genes) were randomly selected. We permuted the random gene selection 1000 times per gene list size and tested for enrichment with preterm birth genes using Fisher's exact test and Chi-squared tests. Tests were conducted separately using **(a)** human, **(b)** mouse and **(c)** rat data.

Supplementary Table 5 (Excel file). This table shows all 640 preterm birth genes from the Preterm Birth Database and chemical gene interactions for all 19 chemicals evaluated in human data from the Comparative Toxicogenomics Database. 0=no curated chemical gene interaction, 1=curated chemical gene interaction

Supplementary Table 6 (Excel file). Results for pathway enrichment analysis of the 640 preterm birth genes. Enrichment analysis was conducted using DAVID online enrichment software.

Supplementary Table 7 (Excel file). Results for DAVID pathway enrichment analysis for all 12 chemicals evaluated in human data from the Comparative Toxicogenomics Database. Chemical gene interactions for all 12 chemicals were obtained from the Comparative Toxicogenomics Database.