Identifying the Link Between Chemical Exposures and Breast Cancer in African American Women via ToxCast High Throughput Screening Data

Katelyn Polemi[1], Vy Nguyen[1,2], Julien Heidt[1], Adam Kahana[1], Olivier Jolliet[1,3], Justin A. Colacino[1,3,4]

[1] Department of Environmental Health Sciences; [2] Department of Computational Medicine and Bioinformatics; [3] Center for Computational Medicine and Bioinformatics; [4] Department of Nutritional Sciences; University of Michigan, Ann Arbor, MI, USA

Abstract:

Among women, breast cancer is the most prevalent form of cancer worldwide and has the second highest mortality rate of any cancer in the United States. The breast cancer related death rate is 40% higher in African American women compared to European American women in the US. The incidence of triple negative breast cancer (TNBC), an aggressive subtype of breast cancer for which there is no targeted therapy, is approximately three times higher in non-Hispanic Black women (NHBW) compared to non-Hispanic White women (NHWW). The drivers of these differences in breast cancer incidence and mortality are still poorly understood, and likely lie in an interaction between genetic and environmental factors. Here, we aimed to identify chemical exposures which may play a role in breast cancer disparities. Using chemical biomonitoring data from the National Health and Nutrition Examination Survey (NHANES) and biological activity data from the EPA's ToxCast program, we assessed the toxicological profiles of chemicals with higher biomarker concentrations in US NHBW. We conducted a literature search to identify a gene set of breast cancer targets included in ToxCast to analyze the response of prioritized chemicals in these assays. Forty-four chemical biomarkers are significantly higher in NHBW. Investigation of these chemicals in ToxCast resulted in a total of 33,645 assays for analysis, 5,301 of which contained nonzero values for ACC (modl_acc: the concentration at which the dose-response fitted model reaches the cutoff considered "active") and modl_tp (scaled top value of dose response curve) data. Of these chemicals BPA, PFOS, and thiram are most comprehensively assayed. 2,5-dichlorophenol, 1,4-dichlorobenzene, and methyl and propyl parabens had higher biomarker concentrations in NHBW and moderate testing and activity in ToxCast. The distribution of active concentrations for these chemicals in ToxCast assays are comparable to biomarker concentrations in NHBW. Through this integrated analysis, we have identified that multiple chemicals, including thiram, propylparaben, and p,p' DDE, with disproportionate exposures in NHBW, have breast cancer associated biological activity at human exposure relevant doses.

1. Introduction:

Breast cancer accounts for more than 1 in 10 new cancer diagnoses each year and is the second leading cause of cancer-related mortality in women (Simon and Robb, 2014; Power *et al.*, 2018; Siegel *et al.*, 2019). Breast cancer outcomes, however, are significantly different across women of different races/ethnicities. African American women are 40% more likely to die from breast cancer than women of any other race (Williams *et al.*, 2016). Furthermore, the incidence of triple negative breast cancer (TNBC), an aggressive subtype of breast cancer for which there is no targeted therapy, is approximately three times higher in non-Hispanic Black women compared to non-Hispanic White women (Carey *et al.*, 2006; Stark *et al.*, 2010). Relative to non-Hispanic White women, non-Hispanic Black women are also 2.4 times more likely to die of

breast cancer after being diagnosed with the pre-invasive lesion, ductal carcinoma *in situ* (Narod *et al.*, 2015).

The mechanisms driving these differences in breast cancer outcomes are likely due to interactions between genetic and environmental factors, although these interactions are complex and remain poorly understood. Genetic epidemiology cohort studies have identified a small number of genetic polymorphisms linked to breast cancer in African American women (Murphy *et al.*, 2017; Feng *et al.*, 2017). However, the contribution of genetic variations appear minor in explaining these cancer disparities (Braun, 2002; Diez Roux, 2012; Cooper *et al.*, 2003; Jing *et al.*, 2014; Rodgers *et al.*, 2018). A number of additional non-genetic contributing factors such as disparities in income, barriers to screening, differences in treatment, and higher stage of disease at diagnosis have also been identified (Dietze *et al.*, 2015). The role of chemical exposures has been less explored in the context of racial disparities in breast cancer outcomes; however, differences in chemical exposures have been hypothesized to be important etiologic factors in racial disparities for multiple diseases (Hoover *et al.*, 2012; Juarez and Matthews-Juarez, 2018; Ruiz *et al.*, 2018; Wang *et al.*, 2016; Zota and Shamasunder, 2017). Mounting evidence from toxicology and epidemiology studies, and an increased understanding of the mechanisms linking toxicant exposures with breast cancer, suggests that many common chemical exposures may alter breast cancer risk (Gray *et al.*, 2017).

We recently have examined racial disparities in chemical biomarker concentrations in US women. Using data from the National Health and Nutrition Examination Survey (NHANES), an ongoing population-based health study conducted by the US Centers for Disease Control and Prevention, we identified stark differences in chemical exposure biomarker concentrations across women of various racial/ethnic groups (Nguyen *et al.*, 2020). We identified a set of chemicals with concentrations significantly higher, on average, in African American women including pesticide metabolites (2,5-dichlorophenol, 1,4-dichlorobenzene), chemicals of personal care products (methyl paraben, propyl paraben, monoethyl phthalate), and heavy metals (mercury and lead). Here, our goal is to assess the biological activity of these chemicals identified as highly disparate in non-Hispanic Black women using data from the U.S Environmental Protection Agency's (EPA) ToxCast program (now CompTox). Numerous studies have highlighted the utility of ToxCast data to identify toxicant effects on cell stress and cytotoxicity (Judson *et al.*, 2016) and cancer (Iyer *et al.*, 2019), as well as to inform adverse outcome pathway development (Fay *et al.*, 2018a). In this study, we integrate human population chemical biomarker concentrations from NHANES with biological activity data from ToxCast. We identify a suite of chemicals with high biologic activity and significant exposure in African American women for further toxicologic and epidemiologic assessment for their relationship to breast cancer disparities.

2. Methods

In the current study, we integrate chemical biomonitoring data from NHANES and biological activity data from the EPA's ToxCast program to assess the toxicological relevance of chemicals with higher exposures in non-Hispanic Black women to breast cancer.

The publicly available ToxCast database offers high throughput in vitro toxicity information generated for over 8,000 different chemicals and over 1,000 biological endpoints (Brunner *et al.*, 2019). Through screening a large chemical library, ToxCast aims to identify biological pathways relevant for response to toxic stressors, develop high throughput screening (HTS) assays for these pathways, generate comprehensive chemical dose-response data for these HTS assays, identify points of departure from the HTS data, and link HTS results to adverse effects. These HTS assays expose living cells, isolated proteins, and other biological molecules to chemicals to screen for biological activity and thus assess toxic effects of chemicals. Resulting endpoints from these assays indicate quantifiable effects linked to known biological processes which are then used to determine the prioritization level of the chemical.

2.1 Listing of Top 44 Chemicals:

Figure 1 shows a graphical workflow of the overall approach. We previously identified chemical biomarkers disproportionately found in non-Hispanic Black women compared to non-Hispanic White women in NHANES (Nguyen *et al.*, 2020). NHANES is a cross-sectional study designed to collect data on demographic, socioeconomic, dietary, and health-related characteristics in the non-institutionalized, civilian US population. The analysis used the continuous data on chemical biomarkers and demographics, collected from 1999-2014. The analysis focused on measuring the chemical disparities in US women. Briefly, using a generalized linear regression-based approach, we quantified the relative difference in biomarker concentration between non-Hispanic Black and non-Hispanic White women across 143 chemicals, 44 of which are higher in non-Hispanic Black women. Through the NHANES biomarker conversion tool (https://wwwn.cdc.gov/nchs/nhanes/search/default.aspx), the exposure biomarkers were manually converted to identify the parent compound and CASRN. Some metabolites detected in NHANES had insufficient ToxCast data, therefore parent compounds are used in place. For the chemicals found to be disproportionately higher in non-Hispanic Black women, we extracted high-throughput chemical assay data from the ToxCast database (version invitroDBv3.1) for further analysis in R. Bisphenol A (BPA), though not significantly higher in non-Hispanic Black women, was also included in the analysis as a model chemical due to the large number of assays in which it has been tested and the association between BPA exposure and breast cancer (Wang *et al.*, 2017; Zhang *et al.*, 2015; Clément *et al.*, 2017). Metals elevated in non-Hispanic Black women, such as lead or mercury, are considered in their biologically relevant forms. For example, assay data for Triphenyllead acetate and lead (II) acetate trihydrate are extracted from the ToxCast database rather than elemental lead.

2.2 Collection of Assays for Analysis:

We extract publicly available ToxCast summary assay information for as many chemicals as possible in 44 of which are higher in Non-Hispanic White Women. Assay information includes intended target family, defined as the target family of the objective target for the assay such as ‘cytokine,’ and the intended target gene name/symbol. From the composite ToxCast assay results dataset, modl_tp (scaled top of winning dose response curve) and ACC (modl_acc: the concentration at which the dose-response fitted model reaches the cutoff considered “active”; referred to as “ACC”) are used as measures of efficacy and potency, respectively. Any assay-chemical pairs with missing modl_tp or ACC values are included for the purpose of generating summary statistics but excluded from all subsequent analyses. The hitcall variable indicates whether the assay was active or inactive. According to the ToxCast owner manual, a dose-

response series must meet three criteria to have an active hitcall: the Hill or Gain-Loss model must win, the modeled curve fit top must exceed the efficacy cutoff, and for at least one concentration the median response value must also exceed the efficacy cutoff (EPA, 2018). Chemicals are first analyzed based on a ratio of active (defined as hitcall=1) to total assays. Because of limited data and varying degrees of flag severity, consistent with Judson et al. 2016, no assays were omitted due to flags.

2.3 Identifying breast cancer relevant assays in ToxCast:

Assay data was further filtered by selecting assays with Intended Target Family Genes considered to play a role in breast cancer. A review of breast cancer literature using Google Scholar was conducted to identify these target genes of interest. Keywords and phrases used for the literature review included, “molecular genetics of breast cancer,” “molecular characteristics of breast cancer,” “genes involved in TNBC pathogenesis,” “genes involved in breast cancer metastasis,” “genes involved in TNBC metastasis,” “TNBC and epithelial-mesenchymal transition,” “molecular mechanism of breast cancer,” “molecular pathways of breast cancer,” “molecular pathways of TNBC,” and “Molecular characteristics of TNBC.” A list of 45 genes were produced from the literature search with the requirement that they also had to be tested in ToxCast. **Supplemental table 1** defines the function of each gene and provides references to support the relationship of this gene with breast cancer. The genes identified are involved in critical processes such as immune response, cell cycle regulation, epithelial/mesenchymal transition (EMT), metastasis, and DNA repair mechanisms. This list was used to understand the activity of ToxCast assays with TNBC gene targets to assess the prioritization of our 44 chemicals and further analyze the mechanism of action of these chemicals in TNBC. ToxCast gene symbols are all converted to human gene symbols. Unsupervised hierarchical clustering was performed to visualize and group chemicals with similar activation patterns for gene specific assays.

2.4 Evaluating Relevant Chemical Concentrations:

Chemical biomarker concentrations in non-Hispanic Black women in NHANES are converted to molarity units to compare to ToxCast assay ACC concentrations in order to understand the biological activity of human relevant exposure doses.

3. Results:

Our previous analyses of chemical biomarker concentrations measured in NHANES identified a suite of chemical biomarkers significantly elevated in in non-Hispanic Black women (Nguyen *et al.*, 2020). To characterize the biological activity of these chemical exposures, we linked these exposures to ToxCast data, comparing differences in biomarker concentrations to the ratio of active assays to total ToxCast assays performed per chemical, a surrogate readout of biological activity. **Figure 2A** represents the biological activity of chemicals found disproportionately in African American women by indicating the ratio of active to total assays in ToxCast compared with the relative percent difference of chemical biomarker concentrations detected in NHBW compared to NHWW in NHANES. On average, 2,5-Dichlorophenol is found more than 3-fold higher in non-Hispanic Black women compared to non-Hispanic White women. 1,4-Dichlorophenol is found approximately 2-fold higher in non-Hispanic Black women, and methylparaben, propylparaben, and 2,4-Dichlorophenol are more than twice as high in non-Hispanic Black women. In addition, BPS, Diethyl phthalate, p,p’-DDE, o,p’-DDT, and o,p’-DDE

are all found to be nearly twice as high. Of the chemicals that are approximately twice as high in non-Hispanic Black women, p,p'-DDE, o,p'-DDT, o,p'-DDE, propylparaben, and BPS have higher ratios of active assays to total assays in ToxCast. **Supplemental table 2** shows the numerical value for the biological activity of each chemical found disproportionately in African American women.

Next, data exploration and visualization are performed across all assays obtained from ToxCast; no assays are excluded based on missing data or flags. The waterfall plot shown in **Figure 2B** displays the number of total assays run per chemical in ToxCast by activity. The y-axis represents the number of assays in ToxCast for each chemical (x-axis) (**Figure 2B**). The bars are colored by ratio of active to total assays. BPA and PFOS are the most frequently tested chemicals with 3389 assays and 2915 assays, respectively. The next most tested chemicals included thiram (1250), 1,4-dichlorobenzene (1097), methylparaben (1088), 2-naphthalenol (1074), and cotinine (1011). The least tested chemicals are shown to be arsenic oxide (79) and PCB 153 (104). The chemicals that showed the highest ratio activity are metals: Triphenyllead acetate (0.738), mercury (II) acetate (0.682), mercury (II) iodide (0.650). The chemicals which showed the lowest ratio of activity include tris(2-chloroethyl) phosphate (0.002), cotinine (0.003), cis-chlordane (0.004), 1-bromopropane (0.005), and cadmium oxide (0.007). BPA has the highest numbers of assays run but a low ratio of activity. Toxcast's method of prioritization is evident here. The metals with a high ratio of activity are not frequently tested because their toxicity is more established. Pesticides such as DDT/DDE and chlordane, with a slightly lower ratio of activity, are being tested in more assays because their toxicity is still being understood.

After completing the comprehensive literature search of TNBC genes represented with ToxCast assays, we then assessed the biological activity of assays in ToxCast with TNBC targets for these chemicals with exposure disparities. **Supplemental Figure 1** shows the total number of TNBC gene targets assays run, along with the ratio of active to total assays, for each chemical. BPA had the highest numbers of assays run and a high ratio of activity. Many other of the highly tested chemicals, such as PFOS, thiram, chlordane, PFDA, DDE and DDT had a high ratio of activity as well across this set of TNBC gene assays. It was noted that although metals had a higher ratio of activity in the full dataset, they rank lower amongst the subset of TNBC gene assays.

After examining the breadth of assay data available from ToxCast, we aimed to compare biologically active chemical concentrations in ToxCast to the biomarker concentrations detected in NHANES (**Figure 3**). Rows represent chemicals in Figure 2 (excluding metals) found to have active assays in ToxCast. The observed concentrations of chemical biomarkers in the NHANES population (blue) and the distribution of ACCs for active ToxCast assays (red) are displayed. The green boxes indicate the concentrations across active ToxCast assays assessing TNBC gene targets. Overlap indicates that assay ACCs for the chemical occurs within the concentration of the relevant chemical biomarkers measured in non-Hispanic Black women in NHANES. Overlapping chemicals include: Methylparaben, p,p'-DDE, Propylparaben, 2,5-Dichlorophenol, 2-Naphthalenol, Dieldrin, beta-Hexachlorobiphenyl, Diethyl phthalate, 2,4-Dichlorophenol, MEHP, thiram, phenanthrene, 1-bromopropane, 1,4-dichlorobenzene, cotinine, and arsenobetaine. **Supplementary Table 3** indicates the amount of overlap.

We next created a heatmap to visualize the unbiased hierarchal clustering of ToxCast assays for TNBC genes based on ratio of active to total assays run per chemical (**Figure 4**) to group chemicals by biological activity. Grey boxes indicate no assays are run for that chemical with that specific gene target. White indicates assays are run, but none are active. The ratio of

activity ranges from 0.0-1.0. Genes most frequently tested with highest biological activity across chemicals of interest include *TP53*, *PPARG*, *ESR1*, *AR*, and *HIF1A*. Chemicals grouped in the same node as our model chemical, BPA, include PFOS, chlordane, thiram, p,p'-DDE and 2-naphthalenol. The metals all cluster near each other, demonstrating similar activation profiles. Interestingly, o,p'-DDE clusters far away from p,p'-DDE and o,p'-DDT.

Finally, we focused on those chemicals with the most significant overlap between the active assay concentrations and measured biological concentrations in non-Hispanic Black women and the most active TNBC assays. Graphs in **Figure 5** show the modl_tp and ACC of active assays testing TNBC genes for propylparaben, p,p'-DDE, and thiram. For propylparaben, assays with high modl_tp are those testing *ESR1* and *AR* (**Figure 5A**). In **Figure 3,** the median biomarker concentrations of propylparaben for NHBW is about 0.1 μM. **Figure 5A** indicates that most of the assays are showing a gain or loss in function at this or slightly higher concentrations. Most of the *ESR1* assays are agonists (gain of function). Assays for thiram with high modl_tp are those testing *ESR1*, *AR*, *HIF1A*, *PPARG*, and *TP53*. This indicated that Thiram is active in many immune assays including *CXCL10*, *CXCL9*, and *IL1A*. These assays indicate a loss in function for these immune genes (**Figure 5B**). According to our concentration box plots, these assays seem to be active at concentrations right above the observed concentrations in NHBW. Assays associated with p,p'-DDE are mostly gain of function for *ESR1*, and those assays with a high modl_tp and modl_acc are testing for *AR* (**Figure 5C**). In **Figure 3**, the median biomarker concentration of p,p'-DDE for NHBW is about 0.32 μM, a concentration at which we observe an increase of function in *AR* in **Figure 5C**. Additionally, *ESR1* is within the range of the observed biomarker concentrations. **Supplemental table 4, 5,** and **6** indicate the plot values for the chemicals represented in Figure 5. Graphs for the modl_tp and ACC for the other chemicals are shown in **Supplemental Figure 2**.

Discussion:

Advancing methods to screen candidate chemicals for associations with specific disease outcomes is critical for prioritizing chemicals for further experimental and epidemiological investigation, as well as to design interventions in highly exposed populations. In this study, we aimed to identify chemical exposures which may play an etiologic role in breast cancer racial disparities and prioritize chemicals for further experimental investigation. By performing an analysis which integrates human biomonitoring data from NHANES and biological activity data from the EPA's ToxCast program, we identified chemicals with significant exposure disparities by race with biological activity relevant to breast cancer. We found that methylparaben, propylparaben, 2,5-dichorophenol, DDT and DDT metabolites (p,p'-DDE), and thiram are chemicals of particular interest for further experimental investigation based on our observations of breast cancer associated biological activity at concentrations relevant to human exposures.

The prioritization method of determining which assays are run for specific chemicals in ToxCast is based on a hazard prediction, where chemicals with low assay activity are categorized as "low priority" and not further tested (Dix *et al.*, 2007). Chemicals with intermediate assay activity undergo further testing to re-evaluate prioritization. Finally, chemicals with high assay activity are prioritized as hazardous and recommended for further screening and testing outside of ToxCast (Dix et al. 2007). The high activity of assays that are related to breast cancer genes at relevant concentrations to non-Hispanic Black women provides evidence that these chemicals may be associated with breast cancer disparities.

Several other studies have used ToxCast for chemical prioritization including environmental chemicals for obesity outcomes, evaluation of food-relevant chemicals, assessment of human indoor exposome of chemicals in dust, and the evaluation of novel flame retardants (Auerbach S. *et al.*, 2016; Karmaus *et al.*, 2016; Dong *et al.*, 2019; Bajard *et al.*, 2019). Most of these studies maintained a similar methodology to our current study, where assay information from ToxCast is analyzed for a predefined set of chemicals. However, the study conducted by (Auerbach S. *et al.*, 2016) used a different approach where experts were used to determine assays testing for biological processes that play important roles in the mechanism of diabetes and obesity. After selecting assay targets relevant in the biological process for diabetes and obesity, they were able to identify chemicals which perturbed these targets and pathways. Differing approaches to prioritize chemicals and assess biological activity and relevant doses demonstrates the broad utility of ToxCast.

TNBCs and aggressive breast cancers with poor prognosis have been shown to have a stem cell-like biological phenotype (Thong *et al.*, 2020; Malta *et al.*, 2018), and developmental pathways are commonly dysregulated by toxicant exposures (Thong *et al.*, 2019). African American women are disproportionately likely to be diagnosed with TNBC and are more likely to die of breast cancer across all subtypes. Here, we also found that stemness related genes, including *AHR*, *SOX1*, *GLI1*, and *HIF-1A*, are targets of chemicals with characterized exposure disparities in African American women. *AHR* has been shown to regulate immunity and maintain cell differentiation (Hao and Whitelaw, 2013; Kawajiri and Fujii-Kuriyama, 2017), deregulation of this protein may contribute to tumor initiation (Gasiewicz *et al.*, 2008). Over expression of HIF1A has been shown to promote tumor growth in breast cancer (Schwab *et al.*, 2012). *HIF-1A* activation in cancer cells regulates expression of stemness transcription factors, including OCT4, NANOG, SOX2, and KLF4 (Mathieu *et al.*, 2011). A recent study found that chemotherapy induced *HIF-1A* controlled pathway led to epigenetic activation of pluripotency factor transcription (Lu *et al.*, 2020), suggesting that this pathway can be activated by stressor exposures. We previously showed that *HIF-1A* is required for stem cell proliferation from normal human mammary gland *in vitro*, including both spheroid formation and organoid growth (Rocco *et al.*, 2018). Determining which chemicals can push cells into a stemness phenotype can help in testing prioritization as well as identify novel mechanisms by which chemical exposures can promote the development of aggressive breast cancers.

Through an unbiased approach, we observed activation patterns in multiple breast cancer relevant biological pathways for different chemicals with characterized exposure disparities. A significant advantage of ToxCast data is the wide range of biological targets assessed for each chemical. For example, immune system-associated genes, including *IL1A, CXCL10, VCAM1*, and *STAT3* are commonly targeted by many of these chemicals, including chlordane, PFOS, DDT, PFNA, and PFDA. The immune system plays an important role in cancer development and prognosis through numerous mechanisms, such as cancer associated inflammation and immune tolerance. Studies have correlated tumors with high immune cell infiltration with more aggressive subtypes of breast cancer such as basal-like or TNBC (O'Meara *et al.*, 2019; Acerbi *et al.*, 2015). Estrogen receptor (*ESR1*) and androgen receptor (*AR*) are also common targets of chemicals with exposure disparities. For example, p,p'-DDE, propylparaben, and thiram exposure all increased AR and ESR1 function, although at very different doses, with thiram activating these receptors at a much lower dose (**Figure 5**). AR expression has been shown to stimulate cellular proliferation through MYC regulation and induce invasiveness through PI3-K signaling in TNBC development. Additionally, studies indicate that AR may be a key factor in

chemoresistance (Giovannelli *et al.*, 2018, 2019). Nearly 70% of breast tumors are shown to express ER, where high expression results in increased proliferation and increased expression of phosphorylated AKT (Gonzalez *et al.*, 2019).

We encountered methodological challenges in performing this integrated NHANES and ToxCast analysis which provides our study with several limitations. First, we are limited by the assays run per chemical, which varied widely. Since there are several chemicals that did not have assays run for specific gene targets, deriving pathway analysis results from these data using Gene Ontology or Gene Set Enrichment Analysis was not successful. Additionally, we recognize that multiple chemicals may lead to the same metabolite. In our data, this would result in an over-representation of that metabolite. Second, due to the possibility of false positive and negative hit calls, Toxcast added a processing step to assign warnings known as "flags." Flags are recommended to be considered with discretion, but overall fewer flags indicates more confidence in results (Judson *et al.*, 2016). Here, we did not exclude data based on flags. Third, scientists using ToxCast data have identified a phenomenon known as cytotoxic burst (CTB) where many of the assays begin to respond at cytotoxic concentrations of the chemicals (Fay *et al.*, 2018b). The bounds of the CTB were determined to identify those assays that elicit cytotoxic responses at concentrations lower than the CTB, therefore prioritizing those assays with high specificity but low responsivity. In the current study, we did not take CTB into account, although we did determine those chemicals with responsive assays within the biological concentrations observed in NHANES (**Figure 3**).

Another limitation of our study is that we are using concentrations of chemicals in NHANES participants using urine and blood measures, rather than concentrations of chemicals measured directly in breast tissue. Persistent lipophilic chemicals are preferentially stored in the mother's breast adipose tissue (Mead, 2008). The female body can mobilize lifetime fat stores to produce milk for her infant, therefore transmitting the persistent environmental contaminants to her newborn during breastfeeding (Lehmann *et al.*, 2014). Chemicals such as BPA, methylparaben (MP), and propylparaben (PP), dichloro-diphenyl-trichloroethane (DDT) and metabolite dichlorodiphenyldichloroethylene (DDE) have all been detected in breast milk with concentrations ranging from 0.0005μM – 0.0033μM (PP), 0.0033μM – 0.015 μM (MP), 0.0013μM – 0.0455μM (BPA), 0.069μM – 50.31μM (DDE) to 0.0031μM – 19.46μM (DDT). (Hines *et al.*, 2015; Mendonca *et al.*, 2014; Agency for Toxic Substances and Disease Registry, 2008). The development of physiologically based toxicokinetic models designed to predict breast tissue concentrations of chemicals based on biomarkers commonly assessed in blood and urine would significantly improve our ability to estimate the impacts of these exposures on breast biology in diverse human populations.

In conclusion, our results demonstrate that African American women are disproportionately exposed to chemicals with breast cancer associated biological activity at doses relevant to human exposure. Future studies should aim to analyze pathways and genes identified as active at biologically relevant concentrations as more ToxCast assay data becomes available. Additionally, ongoing work in our laboratory is using *in vitro* assessment of chemicals in primary human mammary cells collected from non-Hispanic Black and non-Hispanic White women to validate the biological activity which we identified here. These experiments will help to inform whether integration of exposure data from NHANES with biological activity data from Toxcast is a relevant methodology to identify hazardous chemicals that may be involved in the development and prognosis of breast cancer.

Figures:

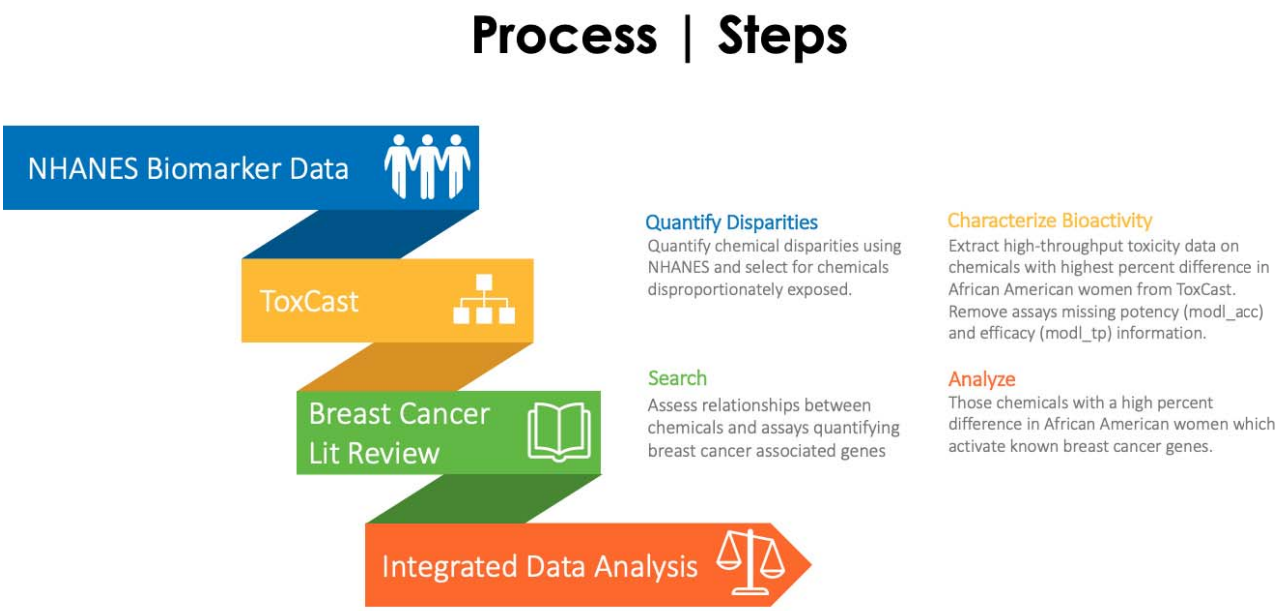
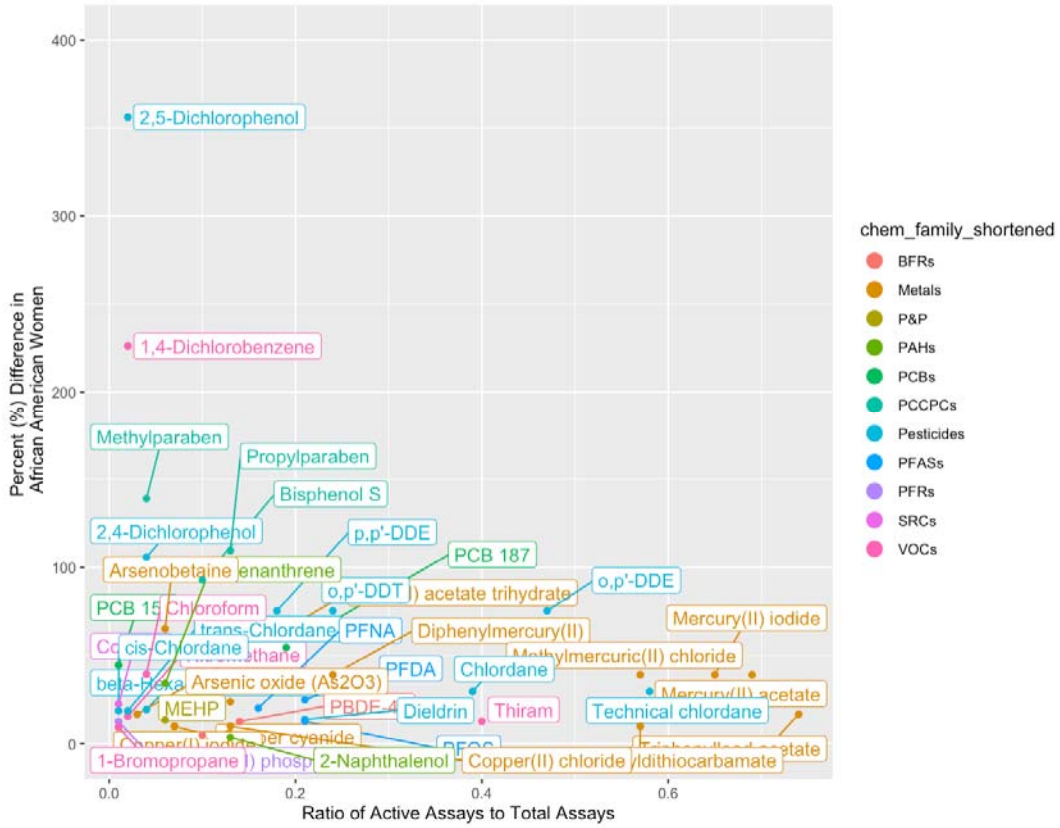


Figure 1: A Comprehensive Workflow for the Methodology of this Study.

A



B

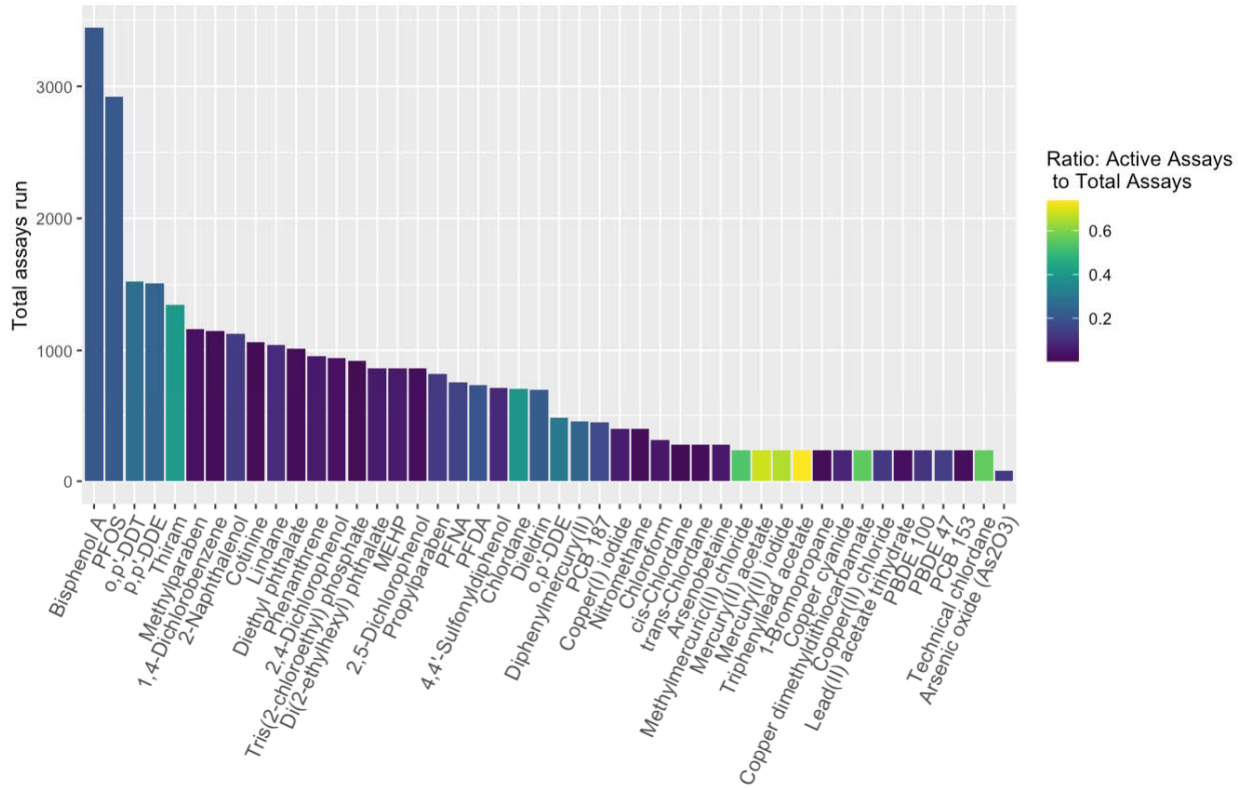


Figure 2: A) Biological Activity of Chemicals Disproportionately Found in African American Women. Ratio of active (hitcall=1) to total assays compared with percent (%) difference in non-Hispanic Black women. **B: Total Assays Run per Chemical by Activity.** Visualization of total number of assays run for each chemical colored by ratio of active to total assays.

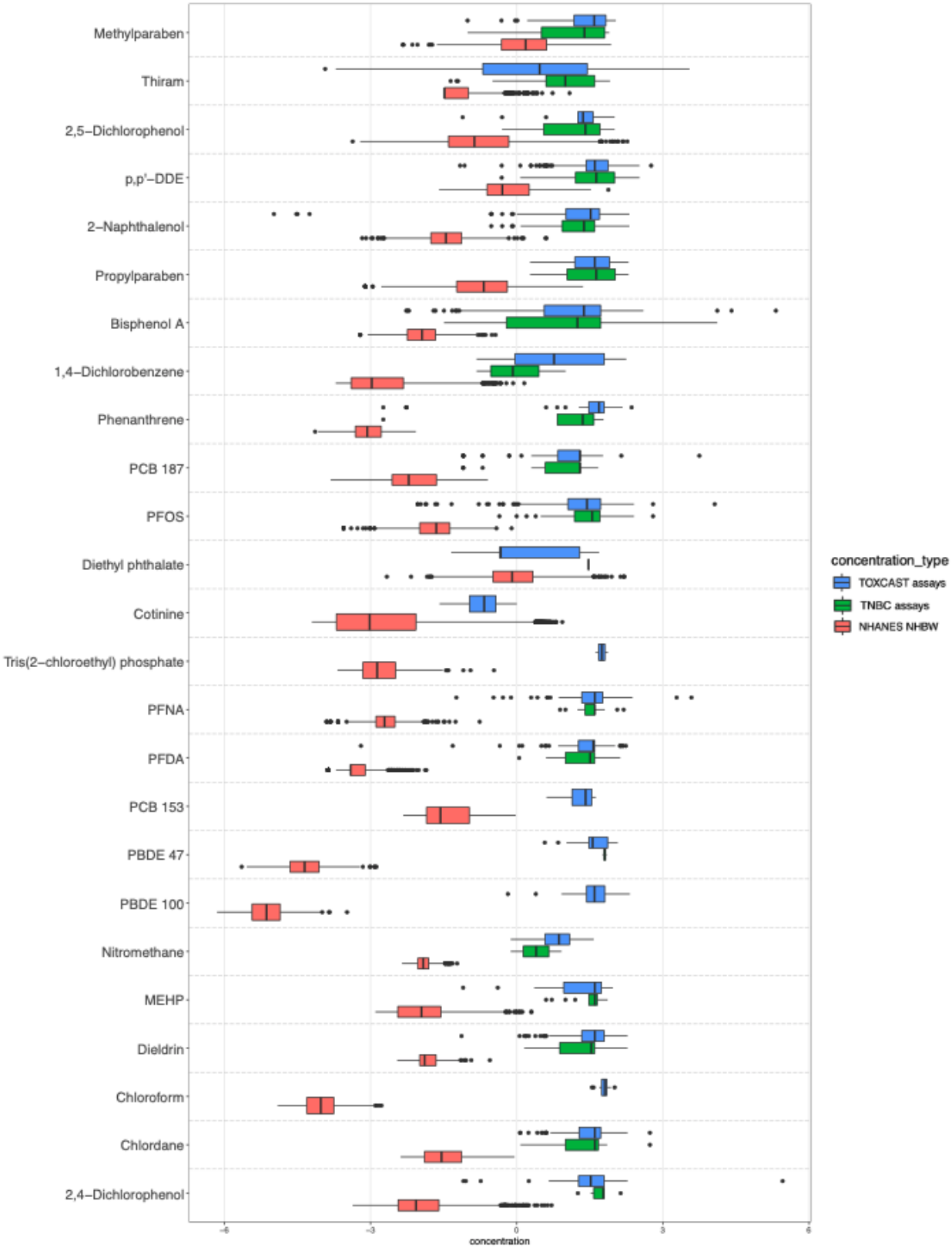


Figure 3: Comparing Chemical Concentrations in ToxCast and NHANES. Log10 (Modl_acc) concentrations across all active assays from ToxCast (blue) and active TNBC gene ToxCast assays (green) are compared to the concentrations of the chemical biomarkers measured in NHB women in NAHNES (red). Modl_acc is the concentration at which the dose-response fitted model reaches the cutoff considered “active”. Overlapping distributions indicate that chemical concentrations detectible in NHB women produce a response in ToxCast assays. The chemicals are ordered from most concentration types with the most overlap to the concentration types with the least overlap.

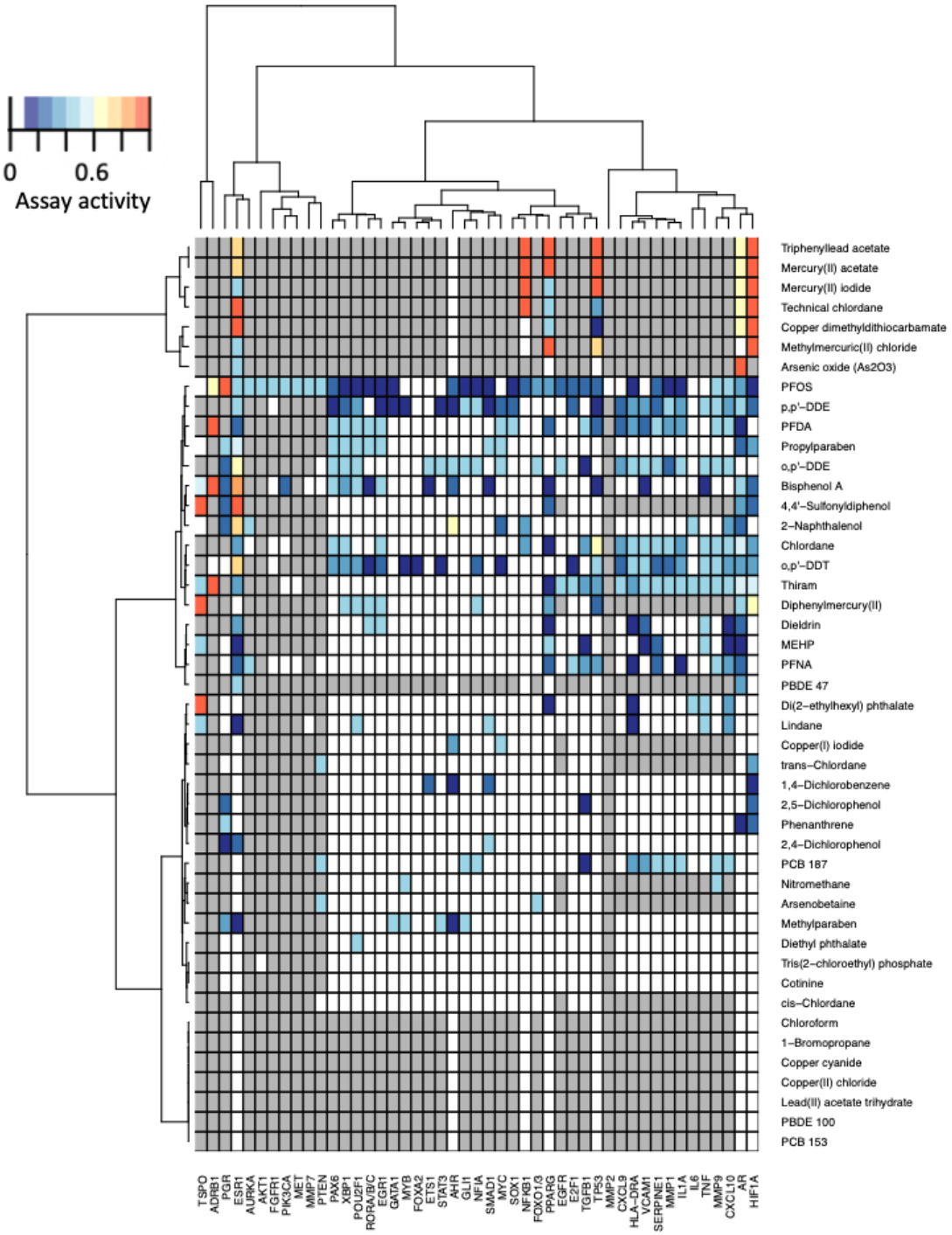


Figure 4: Biological Activity of Assays with Breast Cancer Gene Targets. Unbiased hierarchical clustering of selected TNBC genes and chemicals based on ratio of active (hitcall=1) to total assays run. Breast Cancer genes were identified through a comprehensive literature review. White indicates no assays are run for that chemical with that specific gene target. Gray indicates assays are run but none are active.

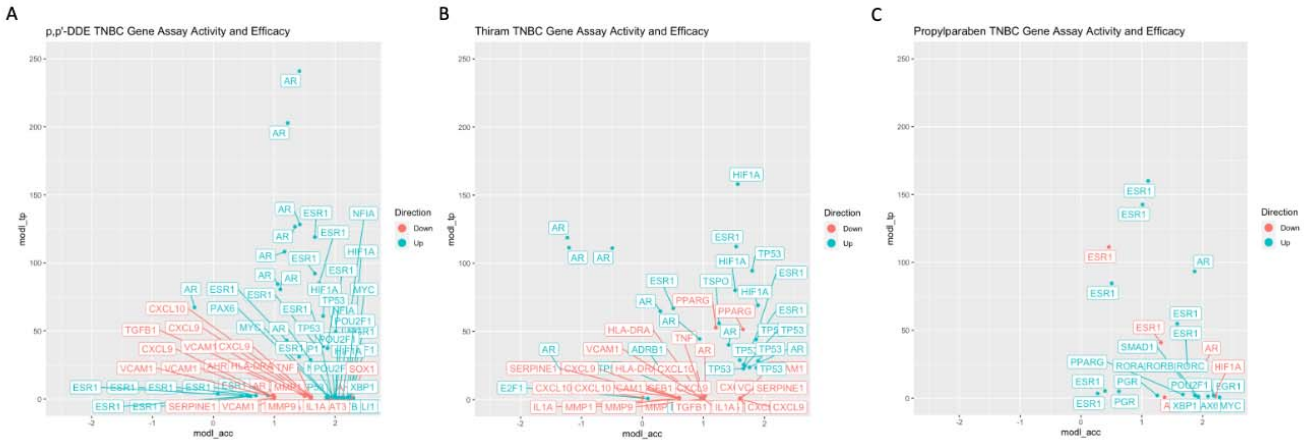


Figure 5: Breast Cancer Gene Assay Activity and Efficacy for Selected Chemicals. These graphs show the modl_tp and modl_acc of active assays testing TNBC genes for certain chemicals. P,p'-DDE (A), thiram (B), and propylparaben (C) had significant overlap between the active assay concentrations and measured biological concentrations in non-Hispanic Black women and the most active Breast Cancer assays. A higher modl_tp indicates a greater response/efficacy. A higher ACC means a higher concentration at which the dose-response fitted model reaches the cutoff considered "active". Direction of the assay is reported in red (loss of function) or blue (gain of function).

Supplemental Tables:

Table 1: Breast Cancer Genes and References from Literature Search

Table 2: Biological Activity of Chemicals Disproportionately Found in African American Women

Table 3: Concentration Type Overlap for Each Chemical

Table 4: Breast Cancer Gene Assay Activity and Efficacy for p,p'- DDE

Table 5: Breast Cancer Gene Assay Activity and Efficacy for thiram

Table 6: Breast Cancer Gene Assay Activity and Efficacy for propylparaben

References:

Acerbi, I. *et al.* (2015) Human breast cancer invasion and aggression correlates with ECM stiffening and immune cell infiltration. *Integr. Biol.*, **7**, 1120–1134.

Agency for Toxic Substances and Disease Registry (2008) Addendum To the Toxicological Profile for Dinitrophenols.

Alluri, P. and Newman, L.A. (2014) Basal-like and triple-negative breast cancers. Searching for positives among many negatives. *Surg. Oncol. Clin. N. Am.*, **23**, 567–577.

Auerbach, S. *et al.* (2016) Prioritizing environmental chemicals for obesity and diabetes outcomes research: A screening approach using ToxCast^{TM} high-throughput data. *Environ. Health Perspect.*, **124**, 1141–1154.

Bailes, A.A. *et al.* (2013) Impact of race and ethnicity on features and outcome of ductal carcinoma in situ of the breast. *Cancer*, **119**, 150–157.

Bajard, L. *et al.* (2019) Prioritization of hazards of novel flame retardants using the mechanistic toxicology information from ToxCast and Adverse Outcome Pathways. *Environ. Sci. Eur.*, **31**, 14.

Bertucci, F. *et al.* (2011) Basal Breast Cancer: A Complex and Deadly Molecular Subtype. *Curr. Mol. Med.*, **12**, 96–110.

Braun, L. (2002) Race, ethnicity, and health: Can genetics explain disparities? *Perspect. Biol. Med.*, **45**, 159–174.

Brunner, A.M. *et al.* (2019) Prioritizing anthropogenic chemicals in drinking water and sources through combined use of mass spectrometry and ToxCast toxicity data. *J. Hazard. Mater.*, **364**, 332–338.

Carey, L.A. *et al.* (2010) Intrinsic breast tumor subtypes, race, and long-term survival in the carolina breast cancer study. *Clin. Cancer Res.*, **16**, 6100–6110.

Carey, L.A. *et al.* (2006) Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *J. Am. Med. Assoc.*, **295**, 2492–2502.

Cooper, R.S. *et al.* (2003) Race and Genomics. *N. Engl. J. Med.*, **348**, 1166–1170.

Dietze, E.C. *et al.* (2015) Triple-negative breast cancer in African-American women: Disparities versus biology. *Nat. Rev. Cancer*, **15**, 248–254.

Diez Roux, A. V. (2012) Conceptual approaches to the study of health disparities. In, *Annual Review of Public Health*. Annu Rev Public Health, pp. 41–58.

Dix, D.J. *et al.* (2007) The ToxCast Program for Prioritizing Toxicity Testing of Environmental Chemicals. *Toxicol. Sci.*, **95**, 5–12.

Dong, T. *et al.* (2019) Human Indoor Exposome of Chemicals in Dust and Risk Prioritization Using EPA's ToxCast Database. *Environ. Sci. Technol.*, **53**, 7045–7054.

EPA, U.S. (2018) ToxCast Owner ' s Manual - Guidance for Exploring Data.

Fallahpour, S. *et al.* (2017) Breast cancer survival by molecular subtype: a population-based analysis of cancer registry data. *C. Open*, **5**, E734–E739.

Fay, K.A. *et al.* (2018) Differentiating pathway-specific from nonspecific effects in high-throughput toxicity data: A foundation for prioritizing adverse outcome pathway development. *Toxicol. Sci.*, **163**, 500–515.

Feng, Y. *et al.* (2017) Characterizing genetic susceptibility to breast cancer in women of african ancestry. *Cancer Epidemiol. Biomarkers Prev.*, **26**, 1016–1026.

Gasiewicz, T.A. *et al.* (2008) Expression and activity of aryl hydrocarbon receptors in development and cancer. *Crit. Rev. Eukaryot. Gene Expr.*, **18**, 279–321.

Giovannelli, P. *et al.* (2018) The androgen receptor in breast cancer. *Front. Endocrinol. (Lausanne).*, **9**.

Giovannelli, P. *et al.* (2019) Androgens Induce Invasiveness of Triple Negative Breast Cancer Cells Through AR/Src/PI3-K Complex Assembly. *Sci. Rep.*, **9**, 1–14.

Gong, Y. *et al.* (2017) Impact of molecular subtypes on metastatic breast cancer patients: A SEER population-based study. *Sci. Rep.*, **7**.

Gonzalez, T.L. *et al.* (2019) Implication of environmental estrogens on breast cancer treatment and progression. *Toxicology*, **421**, 41–48.

Gray, J.M. *et al.* (2017) State of the evidence 2017: An update on the connection between breast cancer and the environment. *Environ. Heal. A Glob. Access Sci. Source*, **16**.

Hao, N. and Whitelaw, M.L. (2013) The emerging roles of AhR in physiology and immunity. *Biochem. Pharmacol.*, **86**, 561–570.

Hines, E.P. *et al.* (2015) Concentrations of environmental phenols and parabens in milk, urine and serum of lactating North Carolina women. *Reprod. Toxicol.*, **54**, 120–128.

Hoover, E. *et al.* (2012) Indigenous peoples of North America: Environmental exposures and reproductive justice. *Environ. Health Perspect.*, **120**, 1645–1649.

Howlader, N. *et al.* (2018) Differences in breast cancer survival by molecular subtypes in the United States. *Cancer Epidemiol. Biomarkers Prev.*, **27**, 619–626.

Iyer, S. *et al.* (2019) An integrated approach using publicly available resources for identifying and characterizing chemicals of potential toxicity concern: Proof-of-concept with chemicals that affect cancer pathways. *Toxicol. Sci.*, **169**, 14–24.

Jiagge, E. *et al.* (2018) Triple-Negative Breast Cancer, Stem Cells, and African Ancestry. *Am. J. Pathol.*, **188**, 271–279.

Jing, L. *et al.* (2014) Ethnic Background and Genetic Variation in the Evaluation of Cancer Risk: A Systematic Review. *PLoS One*, **9**, e97522.

Juarez, P.D. and Matthews-Juarez, P. (2018) Applying an Exposome-Wide (ExWAS) Approach to Cancer Research. *Front. Oncol.*, **8**.

Karmaus, A.L. *et al.* (2016) Evaluation of food-relevant chemicals in the ToxCast high-throughput screening program. *Food Chem. Toxicol.*, **92**, 188–196.

Kawajiri, K. and Fujii-Kuriyama, Y. (2017) The aryl hydrocarbon receptor: A multifunctional chemical sensor for host defense and homeostatic maintenance. *Exp. Anim.*, **66**, 75–89.

Krewski, D. *et al.* (2010) Toxicity testing in the 21st century: A vision and a strategy. *J. Toxicol. Environ. Heal. - Part B Crit. Rev.*, **13**, 51–138.

Lehmann, G.M. *et al.* (2014) Improving the risk assessment of lipophilic persistent environmental chemicals in breast milk. *Crit. Rev. Toxicol.*, **44**, 600–617.

Li, C.I. (2005) Racial and ethnic disparities in breast cancer stage, treatment, and survival in the United States. *Ethn. Dis.*, **15**.

Livasy, C.A. *et al.* (2007) Identification of a basal-like subtype of breast ductal carcinoma in situ. *Hum. Pathol.*, **38**, 197–204.

Lu, H. *et al.* (2020) Chemotherapy-induced S100A10 recruits KDM6A to facilitate OCT4-mediated breast cancer stemness. *J. Clin. Invest.*

Mead, N. (2008) Contaminants in human milk: Weighing the risks against the benefits of breastfeeding. *Environ. Health Perspect.*, **116**, 426–434.

Mendonca, K. *et al.* (2014) Bisphenol A concentrations in maternal breast milk and infant urine. *Int. Arch. Occup. Environ. Health*, **87**, 13–20.

Millis, S.Z. *et al.* (2015) Predictive biomarker profiling of > 6000 breast cancer patients shows heterogeneity in TNBC, with treatment implications. *Clin. Breast Cancer*, **15**, 473-481.e3.

Murphy, M.E. *et al.* (2017) A functionally significant SNP in TP53 and breast cancer risk in African-American women. *npj Breast Cancer*, **3**, 5.

N.C., K. *et al.* (2011) Environmental impact on vascular development predicted by high-throughput screening. *Environ. Health Perspect.*, **119**, 1596–1603.

Narod, S.A. *et al.* (2015) Breast cancer mortality after a diagnosis of ductal carcinoma in situ. *JAMA Oncol.*, **1**, 888–896.

Nassar, H. *et al.* (2009) Ductal carcinoma in situ in African American versus Caucasian American women. *Cancer*, **115**, 3181–3188.

Nguyen, V.K. *et al.* (2020) A comprehensive analysis of racial disparities in chemical biomarker concentrations in United States women, 1999–2014. *Environ. Int.*, **137**, 105496.

Oh, D.S. *et al.* (2006) Estrogen-Regulated Genes Predict Survival in Hormone Receptor-Positive Breast Cancers. *J Clin Oncol*, **24**, 1656–1664.

O'Meara, T. *et al.* (2019) Immune microenvironment of triple-negative breast cancer in African-American and Caucasian women. *Breast Cancer Res. Treat.*, **175**, 247–259.

Power, E.J. *et al.* (2018) Breast Cancer Incidence and Risk Reduction in the Hispanic Population. *Cureus.*

Prat, A. and Parker, J.S. (2019) Standardized versus research-based PAM50 intrinsic subtyping of breast cancer. *Clin. Transl. Oncol.*

Rocco, S.A. *et al.* (2018) Cadmium Exposure Inhibits Branching Morphogenesis and Causes Alterations Consistent With HIF-1a Inhibition in Human Primary Breast Organoids. *Toxicol. Sci.*, **164**, 592–602.

Rodgers, K.M. *et al.* (2018) Environmental chemicals and breast cancer: An updated review of epidemiological literature informed by biological mechanisms. *Environ. Res.*, **160**, 152–182.

Ruiz, D. *et al.* (2018) Disparities in environmental exposures to endocrine-disrupting chemicals and diabetes risk in vulnerable populations. *Diabetes Care*, **41**, 193–205.

Siegel, R.L. *et al.* (2019) Cancer statistics, 2019. *CA. Cancer J. Clin.*, **69**, 7–34.

Simon, A. and Robb, K. (2014) Cancer: Breast. In, *Cambridge Handbook of Psychology, Health and Medicine, Second Edition*. Cambridge University Press, pp. 577–580.

Sisti, J.S. *et al.* (2016) Reproductive risk factors in relation to molecular subtypes of breast cancer: Results from the nurses' health studies. *Int. J. Cancer*, **138**, 2346–2356.

Song, D. *et al.* (2014) Pathway-based analysis of breast cancer. *Am. J. Transl. Res.*, **6**, 302–311.

Stark, A. *et al.* (2010) African ancestry and higher prevalence of triple-negative breast cancer. *Cancer*, **116**, 4926–4932.

Thike, A.A. *et al.* (2010) Triple-negative breast cancer: Clinicopathological characteristics and relationship with basal-like breast cancer. *Mod. Pathol.*, **23**, 123–133.

Waks, A.G. and Winer, E.P. (2019) Breast Cancer Treatment: A Review. *JAMA - J. Am. Med. Assoc.*, **321**, 288–300.

Wang, A. *et al.* (2016) Environmental influences on reproductive health: the importance of chemical exposures. *Fertil. Steril.*, **106**, 905–929.

Williams, D.R. *et al.* (2016) Understanding and effectively addressing breast cancer in African American women: Unpacking the social context. *Cancer*, **122**, 2138–2149.

Yang, X.R. *et al.* (2007) Differences in risk factors for breast cancer molecular subtypes in a population-based study. *Cancer Epidemiol. Biomarkers Prev.*, **16**, 439–443.

Zota, A.R. and Shamasunder, B. (2017) The environmental injustice of beauty: framing chemical exposures from beauty products as a health disparities concern. *Am. J. Obstet. Gynecol.*, **217**, 418.e1-418.e6.