High Throughput Transcriptomics to Understand Chemical Drivers of Racial Disparities in Breast Cancer

Kimberley E. Sala-Hamrick[1,#], Anagha Tapaswi[1,#], Katelyn M. Polemi[1], Justin A. Colacino[1,2,3,*]

1. Department of Environmental Health Sciences, University of Michigan, Ann Arbor, MI, USA
2. Department of Nutritional Sciences, University of Michigan, Ann Arbor, MI, USA
3. Program in the Environment, University of Michigan, Ann Arbor MI, USA

*. To whom correspondence should be addressed

#. These authors contributed equally to this work

Address for Correspondence:

Justin Colacino, PhD MA MPH
6611D SPH1
1415 Washington Heights
Ann Arbor, MI, 48103
Email: colacino@umich.edu

Abstract

Background
There are profound racial disparities in breast cancer incidence and outcomes in the United States and the impact of chemical exposures remains poorly characterized.

Objectives
Analyzing the US Center for Disease Control's National Health and Nutrition Examination Survey chemical biomarker data, we identified a suite of toxicants, including metals, pesticides, and personal care product compounds, to which non-Hispanic Black women are disproportionately exposed. To characterize the impact of these toxicants on breast cancer pathways, we performed high throughput transcriptomic analysis of toxicant exposed breast cells.

Methods
We treated non-tumorigenic mammary epithelial cells, MCF10A, with 21 chemicals prioritized for exposure disparities, at four doses (25nM, 250nM, 2.5µM, 25µM) for 48 hours. We conducted RNA-sequencing for these 408 samples, adapting the PlexWell plate-based RNA-sequencing method to analyze changes in gene expression resulting from these exposures. For each chemical, we calculated gene and biological pathway specific benchmark doses using BMDExpress2, identifying differentially expressed genes and generating the best fit benchmark dose models for each gene. We employed enrichment testing to test whether each chemical's upregulated or downregulated genes were over-represented in a biological process or pathway.

Results
Overall, significant changes in gene expression varied across doses of each chemical and benchmark dose modeling revealed dose-responsive alterations of thousands of different genes. Comparison of benchmark data to NHANEs chemical biomarker concentrations indicated an overlap between actual exposure levels and levels sufficient to cause a gene expression response. Enrichment and cell deconvolution analyses showed benchmark dose responses correlated with changes in cancer and breast cancer related pathways, including induction of a more basal-like characteristics for some chemicals, including p,p'-DDE, lead, copper, and methyl paraben.

Discussion
These analyses revealed that chemicals with exposure disparities induce significant changes in pathways involved in breast cancer initiation and progression at human exposure relevant doses.

Introduction

There are profound racial disparities in breast cancer subtype incidence and outcomes. [1,2] In the United States, Black women are two to three times more likely to be diagnosed with the most difficult to treat subtype of breast cancer, triple negative breast cancer.[3,4] Triple negative breast cancers are characterized by a lack of expression of estrogen receptor, progesterone receptor, and the human epidermal growth factor receptor 2 (HER2). These tumors often present with a basal or myoepithelial cell-like phenotype.[5] As these aggressive breast cancers form, they acquire the Hallmarks of Cancer, including sustained proliferative signaling, the activation of invasion and metastasis pathways, immortality, and the evasion of growth suppression, the reactivation of stem cell pathways, and the acquisition of cellular plasticity.[6–8] For example, experimental data surprisingly suggest that these basal-like breast cancers develop from dysregulated luminal cells, where these cells acquire phenotypic plasticity and shift from a

luminal-like to a basal-like expression pattern.[9,10] Overall, Black women are also 40% more likely to die following a breast cancer diagnosis compared to White women in the United States.[11] The molecular drivers of these aggressive breast cancers, particularly in the context of cancer disparities, remain poorly understood.[1,12]

The causes of breast cancer disparities are likely complex and multifactorial. Epidemiological studies have quantified the impact of diet, reproductive factors, exercise, social determinants of health and structural racism, unequal access to care, and ancestry associated genetics as etiological factors in breast cancer disparities.[12–18] The role of the chemical environment is less well characterized. We and others have identified profound racial disparities in chemical exposures across a broad range of chemical families including heavy metals, PFAS, phthalates, parabens, and pesticides.[19–25] Many of these chemicals with disproportionate exposures have putative toxicological or epidemiological links to aggressive breast cancers.[26–33] Defining whether and how these chemicals with exposure disparities impact breast cancer risk and outcomes would provide much needed new directions for the prevention and elimination of breast cancer disparities.

The goal of this study was to test the hypothesis that chemicals with characterized exposure disparities in Black women in the United States will dysregulate aggressive breast cancer associated bioactivity and cellular plasticity pathways at human relevant doses. To accomplish this, we performed dose-response high throughput transcriptomic profiling of 21 chemicals with characterized exposure disparities in the commonly used *in vitro* breast carcinogenesis model of non-tumorigenic MCF10A cells. We characterize the dose-response effects of these chemicals on individual genes, breast cancer hallmarks, and on cellular plasticity phenotypes. We also contextualize our dose-response data with respect to detectable biomarker concentrations in a representative sample of the US population. These results highlight potential chemical contributors to aggressive breast cancers and provide a methodological framework for the evaluation of chemical exposures on cancer disparities outcomes.

Methods

Cell Culture*:* The non-tumorigenic mammary epithelial cell line, MCF10A, was obtained from American Type Culture Collection (ATCC, Manassas, VA, USA). Cells were cultured and maintained in Dulbecco's Modified Eagle's Medium/Hams F12 50/50 mix (Corning, Corning, NY, USA) supplemented with 5% Horse Serum (Thermo Fisher, Waltham, MA, USA), 2.5mg/mL HEPES (Thermo Fisher, Waltham, MA, USA), 5ug/mL Insulin (Gibco/Thermo Fisher, Waltham, MA, USA), 96ug/mL Hydrocortisone (StemCell, Vancouver, Canada), 100ug/mL Cholera toxin (Sigma-Aldrich, Saint Louis, MO, USA) and 20ng/mL Epidermal Growth Factor (StemCell, Vancouver, Canada). The cells were maintained at 37°C in a humidified 5% CO_2 condition.

Compound preparation: The 21 chemicals were purchased from Sigma-Aldrich (Saint Louis, MO, USA) and Cayman Chemicals (Ann Arbor, MI, USA) (Supplementary Table 1). The chemicals were weighed and dissolved in Dimethyl sulfoxide (DMSO) (for most) and in Sterile Water (for some) at a concentration of 5mg/mL and stored at -20°C for long term storage. For dosing, we prepared an intermediate stock concentration of 5mM for each of the chemicals, which was further diluted into final concentrations for dosing the MCF10A cells: 25uM, 2.5uM, 0.25uM and 0.025uM. These final concentrations were prepared fresh in MCF10A media for each experiment. For chemicals dissolved in DMSO, the final DMSO concentration was 0.5%,

as were the vehicle controls. For chemicals dissolved in water, an equivalent amount of water was included in the vehicle control.

Cell plating: MCF10A cells were cultured and expanded at passage 105 for all rounds of cell culture. Cells were plated in black, clear flat bottom 384 well plates (Corning, Corning, NY, USA) using a 2.5-125uL multi-channel pipette. First, 10uL of plain media was added to pre-warmed plates. 500 cells in 30 uL of media were added to each well at a slow speed to make up a final volume of 40 uL. The cells were left undisturbed to allow the cells to attach and distribute uniformly over the well. Each microplate was incubated at 37°C humidified incubator with 5% CO_2 for 24 hours prior to dosing. Cells were then exposed to test chemicals for 48 hours with 4 biological replicates per dose. Experiments were run in 2 batches to include experimental replicates for all chemical-dose combinations.

plexWell RNA-Sequencing Overview: To conduct high throughput RNA-sequencing, we developed an adapted version of the plexWell (SeqWell, Beverly, MA, USA) plate-based sequencing method to make the method applicable for high throughput *in vitro* toxicological analysis. In the plexWell protocol, sequencing reactions are prepared in custom 96 well plates, where each well contains a unique oligo barcode. Plates are also assigned a unique oligo barcode. Cells are lysed and cDNA is directly prepared from the cellular lysate. A tagmentation reaction adds the well specific barcode to the cDNA from each well. The tagmented cDNAs are then pooled prior to library preparation, where plate specific barcodes can also be incorporated for ultrahigh sample multiplexing prior to sequencing. For example, 384 wells of treated samples would be processed across 4 custom 96 well plates, where each of the samples receives a well specific barcode and each of the 4 plates receives a plate specific barcode. This leads to each of the 384 samples receiving a unique combination of well and plate barcodes, which can be bioinformatically deconvoluted following sequencing. This protocol decreases the cost of library preparation by approximately 90% relative to traditional RNA-sequencing, enabling unbiased high throughput toxicological analyses.

Our key modification to the manufacturer protocol to streamline the process was to avoid the step of RNA purification prior to cDNA preparation. The plexWell protocol calls for the input of 1uL of RNA for cDNA preparation. To eliminate the step of RNA extraction, we devised a method of lysing the cells directly in the 384-well microplate. After 48 hours of incubation with the test chemicals, the cells were washed twice with 40uL of 1X HBSS. After the 2nd wash, 5uL of 1X HBSS was left behind, to avoid cell loss. We added 10uL of plexWell lysis buffer (with RNase inhibitor) to each well using a 2.5-125uL multi-channel micropipette. The buffer was pipetted up and down to mix 10 times and the cell lysates were collected in 1mL eppendorf tubes on ice to avoid RNA degradation. Cell lysis was confirmed via examination by microscopy.

plexWell cDNA preparation: The plexWell manufacturer's protocol was followed to prepare cDNA for subsequent library preparation. In brief, 1uL of cell lysate was used as an input for cDNA preparation and oligo dT annealing in a PCR 96 well plate (Dot Scientific, Burton, MI, USA) The cDNA synthesis reaction was run for 12 cycles on a Bio-Rad CFX96 Touch Real-Time PCR Detection System. Equivalent amount of MAGwise Paramagnetic Beads was added to the amplified cDNA in each well of the 96 well plate. The cDNA was allowed to bind for 5 minutes. The plate was placed on a magnet to allow the beads to pellet on the inner walls of each well. Supernantant was discarded and the beads were washed once using 80% ethanol. The

cDNA was eluted using 20uL of 10mM Tris solution. Purified cDNA was stored at -20°C for short term storage.

Picogreen assay: As per plexWell protocol, we quantified the cDNA concentration of a subset of samples using th eQuant-iT Picogreen dsDNA Assay kit (Thermo Fisher, Waltham, MA, USA). Standards ranging from 25pg/mL to 25ng/mL were prepared using the Lambda Standard DNA (100ug/mL) provided in the kit. 1uL of prepared dsDNA was diluted in 99uL of 1X TE buffer to get a 1:100 dilution. A 200-fold aqueous Quant-iT Picogreen working solution was prepared in 1X TE buffer at a volume of 100uL/well. The standards and samples were plated in a flat bottom Corning 96 well plate using a 125-1250uL multi-channel pipette. Quant-iT Picogreen working solution added at 100uL/well using a 125-1250uL multi-channel pipette. The samples were incubated at room temperature for 5 minutes, protected from light. The fluorescence from Quant-iT picogreen working solution was read on SpectraMax M5e microplate reader (Molecular Devices, San Jose, CA). The preset protocol on SoftMaxPro software version 5.4 for Picogreen Assay for Nucleic acid was used for analysis.

As per the plexWell protocol, 6 samples per 96 well plate were analyzed on Agilent High Sensitivity DNA Bioanalyzer (up to 15000 base pairs) at the University of Michigan Advanced Genomics Core. The analyzed data showed electropherograms with a summary of fragment sizes. The electropherogram for submitted samples were compared to the example electropherograms provided in the plexWell protocol to check for fragment size discrepancies. A pool of 22 random samples was submitted for QuBit analysis at the Advanced Genomics Core per round of experiment to obtain an average concentration. According to the plexWell protocol, the reagents in the kit were formulated to tolerate up to 10-fold difference in sample input across 96 samples. After fragment distribution and concentration checks, we calculated a global dilution factor using the QuBit concentrations (one per library preparation) as described in the plexWell protocol. This global dilution factor gives the volume for dilution of cDNA prior to tagmentation. 4uL of prepared cDNA was diluted in the calculated Tris-HCl volume.

Library Preparation: Post dilution, 4uL of the diluted cDNA per sample was used for library preparation. Each sample was individually barcoded in a hard skirted Sample Barcode plate provided in the plexWell 384 Rapid Single Cell RNA Library Prep Kit (SeqWell, Beverly, MA, USA). Each sample was labeled with an i7 index, referred to as a Sample Barcode, using a tagmentation reaction. Post i7 tagging, the 96 samples were pooled to a final volume of 800-860uL. Equivalent amount of MAGwise Paramagnetic Beads was added to the pooled SB (Sample Barcode) reactions. The cDNA was allowed to bind to the beads for 5 minutes. The beads then formed a pellet on the inner wall of the tube, as it was placed on a magnetic stand. The pellet was washed two times with 80% Ethanol. 40uL of 10mM Tris was used to elute out purified SB reaction pool. Next, the purified SB reaction pool was labelled with an i5 index, referred to as a Pool Barcode, using a tagmentation reaction. Post i5 tagging, an equivalent amount of MAGwise Paramengnetic beads was added to the PB (Pool Barcode) reaction. The DNA was allowed to bind to the beads. The beads then formed a pellet on the magnetic stand and was washed two times using 80% ethanol. Purified PB product was amplified using Bio-Rad CFX96 Touch Real-Time PCR Detection System for 12 cycles. Post amplification, the total volume for the reaction was measured and the DNA was diluted to a total volume of 205uL with 10mM Tris. 200uL of the diluted product was transferred to a new 1.5mL LoBind tubes for purification. 5uL of the unpurified product was stored as control. 0.8 equivalents of MAGwise

paramagnetic beads was added to the PB product and DNA was allowed to bind to the beads. The beads were allowed to form a pellet on the inner wall of the tube. The pellet was washed two times using 80% ethanol. Purified library was eluted using 32uL of 10mM of Tris. 28uL of the purified product was transferred to a new 1.5mL of LoBind tube. 4 libraries were prepared, each having a unique Pool Barcode. Purified libraries were stored at -20°C for short term storage. Library QC was done on the Agilent Bioanalyzer (High Sensitivity DNA 5000 kit) at the Advanced Genomics Core.

RNA-Sequencing and Data Processing: Libraries were sequenced at Illumina NovaSeq 6000 l at the Advanced Genomics Core. FASTQ reads were demultiplexed back into individual samples based on the i7 and i5 indices. Sequencing data were transferred to the University of Michigan Great Lakes high performance computing cluster for analysis. Sequencing read quality was assessed via FastQC and MultiQC. Reads were aligned to a splice junction aware build of the human genome (GRCh38) using STAR. Aligned reads were assigned to genes using featureCounts, where multimapping and multi-overlapping reads were not counted.

Differential Gene Expression Testing: Read count matrices from featureCounts were loaded into edgeR, and samples with fewer than 1,000,000 mapped reads were excluded from downstream analysis. Genes with low expression were excluded from analysis using the edgeR filterByExpr() function with default settings. Normalization factors and dispersion were calculated prior to generating a log2-transformed normalized counts per million (cpm) matrix for downstream analysis. Differential gene expression between each treatment and the relevant controls was calculated using quasi-likelihood negative binomial generalized log-linear modeling in edgeR. Genes were considered differentially expressed between a treatment and control at a false discovery rate (FDR) adjusted p-value less than 0.05.

Benchmark Dose Analyses: Gene specific benchmark doses and best fit benchmark dose model were identified using BMDExpress v.2.3, a free software package developed by the EPA and available for download (https://github.com/auerbachs/BMDExpress-2/releases). Best practices for dose-response modeling for each chemical were conducted according BMDExpress online documentation (https://bmdexpress-2.readthedocs.io/en/feature-readthedocs/basic_workflow/). Normalized counts per million reads were imported into BMDExpress and prefiltered using One-Way ANOVA for significance at a p-value of < 0.05 to identify genes showing significant increasing or decreasing concentration responses. To determine concentration-response relationships, the filtered data were modeled in BMDExpress with hill, power, linear, polynomial (2°, 3°, 4°), and exponential (2, 3, 4, 5) models and the best fit model was chosen. The benchmark response (BMR) was set to 1 standard deviation relative to control response, maximum iterations of the model was set to 250, a confidence level of 0.95 was used, and a power was restricted to greater than or equal to 1 for the power model. Best fit models were chosen for each dose response relationship per gene using nested chi square to select for the best model followed by lowest Akaike information criterion (AIC). Hill models were flagged if its 'k' parameter was smaller than the lowest positive dose and flagged hill models were included in the data. Benchmark concentrations with values above the highest noncytotoxic dose (25 µM) were excluded from further analysis. We conducted "Defined Category" analyses to additionally assess the dose-response relationship to the Molecular Signature Database "Hallmark" gene sets, embryonic stem cell genes, breast cancer molecular signatures, and breast stem cell genes.[34–37] All original expression data, ANOVA filtered results, BMD results, and subsequent

BMDExpress analysis for all chemicals are included in a .bm2 file available in supplementary information.

For hypergeometric enrichment testing of BMD data, R package HypeR developed by the Monti lab at Boston University was used.[38] Gene sets were derived from Molecular Signatures Database's "Hallmark" and "CGP" (Chemical and Genetic Perturbations) gene sets and user-defined gene sets from Pal et al. (2021) and Grashow et al. (2018).[39,40] Background gene expression included all human genes (23467). Results were filtered for significance at an FDR of < 0.05 and visualized by heatmap using heatmap.2 function from the R package gplots.

Comparison of NHANEs exposure levels to benchmark doses: We compared bioactive doses of disparities associated chemicals to human biomarker concentrations measured as part of the National Health and Nutrition Examination Survey following our previously established methodology.[30] Briefly, chemical biomarker concentrations in 9,116 non-Hispanic Black women in NHANES 1999-2014 were converted to molarity units to compare to the benchmark dose concentrations found through benchmark dose analysis of RNA Sequencing data to understand the biologically relevant exposure doses. Overlapping NHANES concentrations and BMD concentrations indicate that relevant biomarkers measured in non-Hispanic Black women in NHANES occur within the benchmark doses and were visualized using boxplots.

Secondary analysis of normal breast single cell data for cell type composition prediction: To quantify whether treatments altered the cellular composition of our samples, we performed bioinformatic deconvolution of our bulk RNA-seq data based on a single cell atlas of the normal human mammary gland (Pal et al., 2021), where 10 normal mammary gland samples were processed for single cell analysis using the 10x Genomics Chromium platform.[39] Sample specific count matrices of the single cell RNA-seq data were downloaded from the Gene Expression Omnibus and processed via Seurat. Briefly, we excluded cells which express fewer than 200 or more than 5000 genes, along with cells with greater than 15% mitochondrial reads. Data were log-normalized and scaled, and then clustered with 0.02 resolution, following Pal et al's methods, to identify four cell clusters. Clusters were then assigned to "luminal progenitor", "mature luminal", and "myoepithelial" identities based on marker gene expression. Gene expression signatures for each of the cell types were calculated using the FindAllMarkers() function in Seurat.

We then used these single cell gene expression data to estimate cell type proportions in our data using the Multi-subject Single Cell (MuSiC) deconvolution method. MuSiC predicts cell type proportions in bulk RNA-seq data based on annotated single cell RNA expression data from multiple donors. Using MuSiC, we estimated the proportions of myoepithelial, mature luminal, and luminal progenitor cells in each sample to test the hypothesis that chemical treatments may induce cellular plasticity or differences in cell type marker expression. Differences in cell type proportions by chemical dose were calculated with generalized linear modeling in R, comparing back to the relevant control samples (DMSO or water) as the reference group. Differences in estimated cell type proportions by treatment were compared to control using generalized linear modeling.

Results

To understand the impacts of exposure disparities chemicals on dysregulated breast cancer biology, we treated non-tumorigenic MCF10A cells with chemicals found through previous analysis of population disparity data that represent exposures from metals, pesticides, personal care products, and other industrial and environmental sources, including per- and polyfluoroalkyl substances.[21,30] Additionally, two of the chemicals for this study (PFOA and BPF) were based on chemical similarity to those identified with the biomarker data. After filtering expression data for genes with significant alterations, fold change and significance of gene expression varied by chemical (Figure 1, Table 1, Supplemental Figure 1, Supplemental Table 2). For example, p,p'-DDE showed over 700 differentially expressed genes (DEGs) at all 4 doses, while mercury only had over 10 DEGs at the highest dose of 25 µM. Arsenic and cadmium both had DEGs at all four doses, but had over 8000 and 4000 DEGs at 25 µM, respectively.

Shown in Figure 1 are plots for 2.5 µM of each chemical, with other doses included in Supplemental Figure 1. The log(fold change) ranged from -10 to 10, with most genes falling between -5 and 5 and significant genes often falling in a narrower range. For the 2.5 µM dose, cadmium and p,p'-DDE showed log(fold change) ranging from -10 to 10, arsenic, mercury, copper, lead, methyl paraben, propyl paraben, BPA, BPS, BPF, PFOA, thiram, 2,5-dichlorophenol, 1,4-dichlorobenzene, PCB153, and PCB187 showed log(fold change) ranging from -8 to 8, and phenanthrene, PFDA, PFNA, PFOS, and p,p'-DDE showed log(fold change) ranging from -6 to 6.

Particular genes related to breast cancer were found differentially expressed by our disparity chemicals. Of note, G protein alpha subunit (GNAS) , a gene related to breast cancer cell proliferation and migration, expression changes were seen across 28 exposures.[41] Additionally, aldo-keto reductases AKR1C1, AKR1C2, and AKR1C3 expression changed across 30, 38, and 6 exposures, respectively, and these genes are associated with hormone signaling and breast cancer proliferation.[42] Keratins KRT14, KRT15, and KRT17 expression levels also changed across 17, 9, and 18 exposures, and their dysregulation is associated with breast and other cancers.[43-45] Serine protease inhibitor, clade E member 1 (SERPINE1), a gene associated with TNBC chemotherapy resistance, also showed changes in expression across 12 chemical treatments.[46]

To characterize the dose-response relationship of each exposure on gene expression, we performed benchmark dose modeling, to define the range of benchmark doses for different chemicals (Figure 2). While each gene corresponded to its own best fit model, median benchmark doses were calculated for each gene regulated by each chemical (Figure 2A). The total number of genes filtered by ANOVA and fit to benchmark models was different for each chemical, with some chemicals (arsenic, cadmium, and thiram) inducing dose-responsive alterations of over 1000 different genes. Dose-response fit was assessed via examination of the best fit dose-response curves for individual gene/chemical combinations. Two example gene BMD curves are displayed in Figure 2 B and C, showing the gene expression of *ARC* following exposure to a range of arsenic doses in a linear best fit model (Figure 2B) and thiram's dose-response regulation of gene *HSPA6* in a hill best fit model (Figure 2C) across the range of treated dose. All benchmark dose response data for each dysregulated gene for every chemical tested is included in a supplementary bm2 file.

We next wanted to contextualize the doses required *in vitro* to dysregulate gene expression with biomarker levels detectable in non-Hispanic Black women in the United States. We compared benchmark dose results from our *in vitro* experiments to National Health and Nutrition Examination Survey (NHANES) biomarker concentration data for each chemical, plotting the range of blood or urine chemical biomarker concentrations in non-Hispanic Black women in the United States compared to the range of BMDs for each chemical from our *in vitro* data (Figure 3). For many of the chemicals, the levels detectable in non-Hispanic Black women are sufficient to alter gene expression in MCF10A cells, reflected by an overlap in the distributions of NHANES biomarker concentrations and benchmark doses for a given chemical. For many chemicals (phenanthrene, PFNA, PFDA, PCB187, PCB153, mercury, cadmium, bisphenol S, bisphenol F, bisphenol A, arsenic), chemical biomarker concentration in the population were below the vast majority of benchmark doses for altered gene expression. Conversely, for a number of chemicals, population biomarker concentrations were comparable to, or above, the bulk of the benchmark doses. These chemicals included propyl paraben, PFOA, p,p'-DDE, methyl paraben, lead, copper, and 2,5-dichlorophenol.

We next wanted to quantify the biological processes altered by disparities associated chemical exposure and characterize the doses at which these processes are dysregulated. BMD modeled genes were assessed using MSigDB's Hallmark gene sets consisting of 50 biological processes. The supplemental bm2 file contains benchmark dose gene data for every hallmark category for each chemical organized under "functional classifications." Figure 4A displays an accumulation plot depicting the median BMD at which Hallmark gene sets were significantly impacted. The accumulation plot demonstrates low dose effects present for many of the hallmarks. Figure 4B shows a heatmap of significantly changed hallmark processes - 9 of the chemicals showed significant enrichment or downregulation for 1 or more of the hallmark gene sets. Notably, many of these processes are changes that are related to carcinogenesis, including the epithelial mesenchymal transition, inflammatory response, and reactive oxygen species pathway. Epithelial to mesenchymal transition was upregulated by arsenic, cadmium, copper, lead, and thiram at median BMD dose of 7.58 µM, 9.83 µM, 0.60 µM, 0.55 µM, and 9.01 µM, respectively. Inflammatory response was upregulated by arsenic, cadmium, copper, and lead at median BMD dose of 8.26 µM, 12.28 µM, 0.27 µM, and 0.19 µM, respectively. Reactive oxygen species was upregulated by arsenic, copper, lead, and thiram at median BMD dose of 8.31 µM, 0.18 µM, 0.12 µM, and 10.59 µM, respectively. p,p'-DDE downregulated epithelial mesenchymal transition genes (median BMD of 0.0079 µM) and upregulated cell cycle related processes, including E2f targets and G2m checkpoint (median BMD of 0.011 µM and 0.015 µM, respectively), both of which were downregulated by arsenic, cadmium, and thiram (median BMDs of 14.09 µM, 10.77 µM, and 8.27 µM, respectively, for E2f targets and of 8.91 µM, 10.68 µM, and 8.96 µM, respectively, for G2m checkpoint), highlighting the potential for distinct mechanisms of promoting carcinogenesis depending on the chemical involved.

To assess specific pathways related to breast cancer, additional enrichment analyses assessed additional breast cancer associated biological pathways in MSigDB's Chemical and Genetic Perturbations (CGP) gene sets (Figure 4C and Supplemental Figure 1): Gene sets from three studies related to breast cancer and stemness (Wong et al., 2008; Lim et al., 2010; & Smid et al., 2008). Eight chemicals differentially regulated these features at various doses. Genes upregulated in mammary stem cells were found upregulated by copper, lead, and thiram (median

BMDs of 0.59 µM, 0.40 µM and 11.78 µM, respectively) and genes upregulated in embryonic stem cells were upregulated by p,p'-DDE (median BMD of 0.012 µM). Additionally, genes upregulated in basal subtypes of breast cancer were found upregulated by lead and PFOA (median BMDs of 0.49 µM and 0.0026 µM), while genes upregulated in luminal A breast cancers were found upregulated by arsenic and thiram (median BMD of 10.36 µM and 8.98 µM). PFOA also upregulated genes that are upregulated in normal like breast cancers (median BMD of 0.00093 µM) and arsenic upregulated genes upregulated in breast cancer relapse in the lung (median BMD of 10.16 µM).

To assess potential breast cancer specific alterations regulated by our chemicals of interest, we tested for enrichment with a publicly available breast carcinogenesis gene panel, BCScreen.[40] The overlapping genes within different categories of breast carcinogenesis are shown in supplemental table 3. The gene panel includes 500 genes total divided into 14 categories, with 33 genes per category other than the "mammary" category which includes 71 genes. Of these, arsenic, cadmium, and thiram showed BMD gene alterations in all 14 categories, and copper, lead, and p,p'-DDE showed at least one gene alteration in most categories.

We next wanted to test the hypothesis that disparities associated chemicals induce changes consistent with a luminal-to-basal transition. We performed cell type deconvolution based on a normal breast single cell RNA-seq reference to predict the proportions of myoepithelial, mature luminal, and luminal progenitor cells in each of our exposures (Figure 5 and Supplemental Figures 2-4). [39] Interestingly, PFOA showed a significant decrease in myoepithelial cell proportion from around 65% to 52% across 3 doses (Supplemental Figure 2). Arsenic, copper, and lead showed a significant increase in the proportion of myoepithelial cells across all doses compared to control, while p,p'-DDE showed a significant increase in this proportion at the 3 highest doses (Figure 5). These data suggest that these disparities associated chemicals induce an expression signature consistent with a luminal-to-basal transition, a characteristic of the basal-like breast cancers by which Black women are disproportionately affected.

Discussion

Across all 21 chemicals representing exposures disproportionately affecting non-Hispanic Black women in the United States, nearly 30,000 differentially expressed genes were found dysregulated in exposed breast cells compared to controls. While many of these alterations were found at the highest doses, lower dose effects were also seen, with median BMDs falling near the range of NHANES biomarker levels for all chemicals. This suggests that the gene expression effects seen in this study are representative of effects seen in humans, and so our analysis links disproportional chemical exposures to disproportional incidence and characteristics of breast cancer. Interestingly, of the chosen set of chemicals, only nine have been classified by IARC as possible, probable, or definite carcinogens to humans, making this study both novel and relevant in assessing underlying chemical influence on the biology of cancers.

Specific gene alterations included some commonalities across chemical treatments showing expression changes in genes related to breast cancer progression. This included GNAS, aldo-keto reductases, keratins, and SERPINE1. GNAS has been implicated in breast cancer cell proliferation and epithelial to mesenchymal transition via modulation of the

PI3K/AKT/Snail1/E-cadherin signaling axis.[41] Aldo-keto reductases are involved in estrogen and progesterone production and are thus involved in pro-proliferative signaling of hormone responsive tumors.[42] Dysregulation of keratin expression levels has also been associated with breast cancer, with increased KRT14 expression associated with invasiveness, decreased KRT15 expression associated with poor prognosis, and increased KRT17 associated with the development of TNBC.[43-45] SERPINE1 is associated with resistance of TNBC to paclitaxel and knockdown of SERPINE1 reversed resistance of breast cancer to paclitaxel through downregulation of vascular endothelial growth factor A (VEGFA).[46]

In addition to common gene expression changes across chemical treatments, our enrichment analyses suggest that chemicals alter breast biology by several distinct mechanisms. Metals including arsenic, cadmium, copper, and lead were found to induce epithelial to mesenchymal transition, inflammatory response, and reactive oxygen species pathways, consistent with previous literature on metal-induced carcinogenesis that places particular emphasis on oxidative stress related carcinogenesis.[47–49] In breast cancer, arsenic and cadmium have been associated with oxidative stress related pathways of malignant transformation.[50] While these metals as well as many of the other chemicals in our analysis are known to play an endocrine disrupting role that could lead to breast tumorigenesis, these findings propose alternative routes of carcinogenesis.[51–53] p,p'-DDE downregulated epithelial to mesenchymal transition in our study but altered cell cycle processes. Previous literature has found that a mixture of organochlorine pesticides including p,p'-DDE had potential to modulate proliferation of breast cancer cell lines partially through induction of cell cycle entry.[54]

Alterations in stemness features were also seen across eight chemicals analyzed, important because the cancer stem cell hypothesis states that there is a small subset of cancer stem cells that result in cancer initiation and invasion, and so alterations in stemness features can lead to increased tumorigenicity and metastasis.[55] We found stemness signatures enhanced by copper, lead, thiram, and p,p'-DDE, findings that could indicate novel mechanisms of chemically induced carcinogenesis.

Breast cancer is heterogenous in nature, with diverse subtypes characterized by different molecular features, which can be simplified into luminal versus basal subtypes.[56] Breast cancers exhibiting triple negative phenotype (negative for HER2, estrogen, and progesterone receptors) often overlap with basal-like breast cancer features.[56] Non-Hispanic black women are three times more likely to be diagnosed with these basal-like cancers.[3,4] Interestingly, enrichment analysis found upregulation of basal subtype genes by lead and PFOA and cell type proportion analysis found upregulated basal proportions by arsenic, copper, lead, and p,p'-DDE and a decrease in basal cell proportions by PFOA. Previous literature has shown that arsenic is capable of inducting transition from luminal to basal features, but this association has not been studied for the other chemicals.[57]

One major limitation of the present study is that all exposures were done on an acute timescale (48 hours) while actual human exposures to the chemicals analyzed here are likely chronic, over the course of years or a lifetime. Another limitation is that only one cell type was used for this study, normal breast cells taken from a white woman. Future work should consider treatment on a more chronic timescale as well as different breast cell lines representing diverse donors. Additionally, future work could consider treatment of diverse breast cancer cell lines, ranging

from more basal to more luminal cell types to investigate changes in luminal or basal features related to chemical exposure. Another limitation in our analyses integrating benchmark dose levels with NHANES biomarker concentrations is that the NHANES chemical biomarkers are measured in urine and blood. These levels may or may not reflect mammary tissue levels. As additional research is conducted on biomonitoring of mammary tissues, or physiologically based toxicokinetic models reflecting mammary gland concentrations are developed, these data and methods will substantially inform our understanding of how chemical exposures impact breast cancer risk.

Overall, our transcriptomic analysis uncovers molecular mechanisms by which chemicals disproportionately affecting Non-Hispanic Black women may induce breast cancer disparities in severity and outcomes. This includes both features related to known carcinogenic properties of these chemicals as well as potentially novel mechanisms, opening the door to future avenues of investigation to elucidate how different exposures relate to breast carcinogenesis and breast cancer disparities.

Acknowledgements

This work was supported by grants from the National Institutes of Health (R01 ES028802, R01 AG072396, T32 ES 007062, P30 ES017885, P30 CA046592).

References

1. Dietze EC, Sistrunk C, Miranda-Carboni G, O'Regan R, Seewaldt VL. Triple-negative breast cancer in African-American women: disparities versus biology. *Nat Rev Cancer*. 2015;15(4):248-254. Doi:10.1038/nrc3896

2. Yedjou CG, Sims JN, Miele L, et al. Health and Racial Disparity in Breast Cancer. *Adv Exp Med Biol*. 2019;1152:31-49. Doi:10.1007/978-3-030-20301-6_3

3. McCarthy AM, Friebel-Klingner T, Ehsan S, et al. Relationship of established risk factors with breast cancer subtypes. *Cancer Med*. 2021;10(18):6456-6467. Doi:10.1002/cam4.4158

4. Carey LA, Perou CM, Livasy CA, et al. Race, Breast Cancer Subtypes, and Survival in the Carolina Breast Cancer Study. *JAMA*. 2006;295(21):2492-2502. Doi:10.1001/jama.295.21.2492

5. Yin L, Duan JJ, Bian XW, Yu SC. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Res BCR*. 2020;22(1):61. Doi:10.1186/s13058-020-01296-5

6. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100(1):57-70. Doi:10.1016/s0092-8674(00)81683-9

7. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-674. Doi:10.1016/j.cell.2011.02.013

8. Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discov*. 2022;12(1):31-46. Doi:10.1158/2159-8290.CD-21-1059

9. Lim E, Vaillant F, Wu D, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med*. 2009;15(8):907-913. Doi:10.1038/nm.2000

10. Rädler PD, Wehde BL, Triplett AA, et al. Highly metastatic claudin-low mammary cancers can originate from luminal epithelial cells. *Nat Commun*. 2021;12(1):3742. Doi:10.1038/s41467-021-23957-5

11. Yedjou CG, Tchounwou PB, Payton M, et al. Assessing the Racial and Ethnic Disparities in Breast Cancer Mortality in the United States. *Int J Environ Res Public Health*. 2017;14(5):E486. Doi:10.3390/ijerph14050486

12. Stringer-Reasor EM, Elkhanany A, Khoury K, Simon MA, Newman LA. Disparities in Breast Cancer Associated With African American Identity. *Am Soc Clin Oncol Educ Book Am Soc Clin Oncol Annu Meet*. 2021;41:e29-e46. Doi:10.1200/EDBK_319929

13. Bandera EV, Alfano CM, Qin B, Kang DW, Friel CP, Dieli-Conwright CM. Harnessing Nutrition and Physical Activity for Breast Cancer Prevention and Control to Reduce

Racial/Ethnic Cancer Health Disparities. *Am Soc Clin Oncol Educ Book Am Soc Clin Oncol Annu Meet*. 2021;41:1-17. Doi:10.1200/EDBK_321315

14. Martini R, Chen Y, Jenkins BD, et al. Investigation of triple-negative breast cancer risk alleles in an International African-enriched cohort. *Sci Rep*. 2021;11(1):9247. Doi:10.1038/s41598-021-88613-w

15. Newman LA, Jenkins B, Chen Y, et al. Hereditary Susceptibility for Triple Negative Breast Cancer Associated With Western Sub-Saharan African Ancestry: Results From an International Surgical Breast Cancer Collaborative. *Ann Surg*. 2019;270(3):484-492. Doi:10.1097/SLA.0000000000003459

16. Gehlert S, Hudson D, Sacks T. A Critical Theoretical Approach to Cancer Disparities: Breast Cancer and the Social Determinants of Health. *Front Public Health*. 2021;9:674736. Doi:10.3389/fpubh.2021.674736

17. Freeman HP. Poverty, culture, and social injustice: determinants of cancer disparities. *CA Cancer J Clin*. 2004;54(2):72-77. Doi:10.3322/canjclin.54.2.72

18. Hall IJ, Moorman PG, Millikan RC, Newman B. Comparative analysis of breast cancer risk factors among African-American women and White women. *Am J Epidemiol*. 2005;161(1):40-51. Doi:10.1093/aje/kwh331

19. 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):161-180. Doi:10.1007/s40471-016-0073-9

20. Zota AR, Shamasunder B. The environmental injustice of beauty: framing chemical exposures from beauty products as a health disparities concern. *Am J Obstet Gynecol*. 2017;217(4):418.e1-418.e6. doi:10.1016/j.ajog.2017.07.020

21. Nguyen VK, Kahana A, Heidt J, et al. A comprehensive analysis of racial disparities in chemical biomarker concentrations in United States women, 1999-2014. *Environ Int*. 2020;137:105496. Doi:10.1016/j.envint.2020.105496

22. Bloom MS, Wenzel AG, Brock JW, et al. Racial disparity in maternal phthalates exposure; Association with racial disparity in fetal growth and birth outcomes. *Environ Int*. 2019;127:473-486. Doi:10.1016/j.envint.2019.04.005

23. Varshavsky JR, Zota AR, Woodruff TJ. A Novel Method for Calculating Potency-Weighted Cumulative Phthalates Exposure with Implications for Identifying Racial/Ethnic Disparities among U.S. Reproductive-Aged Women in NHANES 2001-2012. *Environ Sci Technol*. 2016;50(19):10616-10624. Doi:10.1021/acs.est.6b00522

24. Wesselink AK, Fruh V, Hauser R, et al. Correlates of urinary concentrations of phthalate and phthalate alternative metabolites among reproductive-aged Black women from Detroit,

Michigan. *J Expo Sci Environ Epidemiol*. 2021;31(3):461-475. Doi:10.1038/s41370-020-00270-9

25. Silva MJ, Barr DB, Reidy JA, et al. Urinary levels of seven phthalate metabolites in the U.S. population from the National Health and Nutrition Examination Survey (NHANES) 1999-2000. *Environ Health Perspect*. 2004;112(3):331-338. Doi:10.1289/ehp.6723

26. Stiel L, Adkins-Jackson PB, Clark P, Mitchell E, Montgomery S. A review of hair product use on breast cancer risk in African American women. *Cancer Med*. 2016;5(3):597-604. Doi:10.1002/cam4.613

27. Llanos AAM, Rabkin A, Bandera EV, et al. Hair product use and breast cancer risk among African American and White women. *Carcinogenesis*. 2017;38(9):883-892. Doi:10.1093/carcin/bgx060

28. Brinton LA, Figueroa JD, Ansong D, et al. Skin lighteners and hair relaxers as risk factors for breast cancer: results from the Ghana breast health study. *Carcinogenesis*. 2018;39(4):571-579. Doi:10.1093/carcin/bgy002

29. Eberle CE, Sandler DP, Taylor KW, White AJ. Hair dye and chemical straightener use and breast cancer risk in a large US population of black and white women. *Int J Cancer*. 2020;147(2):383-391. Doi:10.1002/ijc.32738

30. Polemi KM, Nguyen VK, Heidt J, Kahana A, Jolliet O, Colacino JA. Identifying the link between chemical exposures and breast cancer in African American women via integrated in vitro and exposure biomarker data. *Toxicology*. 2021;463:152964. Doi:10.1016/j.tox.2021.152964

31. Rodgers KM, Udesky JO, Rudel RA, Brody JG. Environmental chemicals and breast cancer: An updated review of epidemiological literature informed by biological mechanisms. *Environ Res*. 2018;160:152-182. Doi:10.1016/j.envres.2017.08.045

32. Gonzalez TL, Rae JM, Colacino JA. Implication of environmental estrogens on breast cancer treatment and progression. *Toxicology*. 2019;421:41-48. Doi:10.1016/j.tox.2019.03.014

33. Koual M, Tomkiewicz C, Cano-Sancho G, Antignac JP, Bats AS, Coumoul X. Environmental chemicals, breast cancer progression and drug resistance. *Environ Health Glob Access Sci Source*. 2020;19(1):117. Doi:10.1186/s12940-020-00670-2

34. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MsigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417-425. Doi:10.1016/j.cels.2015.12.004

35. Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, Chang HY. Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell*. 2008;2(4):333-344. Doi:10.1016/j.stem.2008.02.009

36. Smid M, Wang Y, Zhang Y, et al. Subtypes of breast cancer show preferential site of relapse. *Cancer Res*. 2008;68(9):3108-3114. Doi:10.1158/0008-5472.CAN-07-5644

37. Lim E, Wu D, Pal B, et al. Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res BCR*. 2010;12(2):R21. Doi:10.1186/bcr2560

38. Federico A, Monti S. hypeR: an R package for geneset enrichment workflows. *Bioinformatics*. 2020;36(4):1307-1308. Doi:10.1093/bioinformatics/btz700

39. Pal B, Chen Y, Vaillant F, et al. A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. *EMBO J*. 2021;40(11):e107333. Doi:10.15252/embj.2020107333

40. Grashow RG, De La Rosa VY, Watford SM, Ackerman JM, Rudel RA. BCScreen: A gene panel to test for breast carcinogenesis in chemical safety screening. *Comput Toxicol Amst Neth*. 2018;5:16-24. Doi:10.1016/j.comtox.2017.11.003

41. Jin X, Zhu L, Cui Z, Tang J, Xie M, Ren G. Elevated expression of GNAS promotes breast cancer cell proliferation and migration via the PI3K/AKT/Snail1/E-cadherin axis. *Clin Transl Oncol*. 2019;21(9):1207-1219. Doi:10.1007/s12094-019-02042-w

42. Penning TM, Byrns MC. Steroid hormone transforming aldo-keto reductases and cancer. *Ann N Y Acad Sci*. 2009;1155:33-42. Doi:10.1111/j.1749-6632.2009.03700.x

43. Zhong P, Shu R, Wu H, Liu Z, Shen X, Hu Y. Low KRT15 expression is associated with poor prognosis in patients with breast invasive carcinoma. *Experimental and Therapeutic Medicine*. 2021;21(4):1-1. Doi:10.3892/etm.2021.9736

44. Zhang H, Zhang Y, Xia T, et al. The Role of Keratin17 in Human Tumours. *Frontiers in Cell and Developmental Biology*. 2022;10. Accessed October 18, 2022. https://www.frontiersin.org/articles/10.3389/fcell.2022.818416

45. Hanley CJ, Henriet E, Sirka OK, Thomas GJ, Ewald AJ. Tumor resident stromal cells promote breast cancer invasion through regulation of the basal phenotype. *Mol Cancer Res*. 2020;18(11):1615-1622. Doi:10.1158/1541-7786.MCR-20-0334

46. Zhang Q, Lei L, Jing D. Knockdown of SERPINE1 reverses resistance of triple negative breast cancer to paclitaxel via suppression of VEGFA. *Oncol Rep*. 2020;44(5):1875-1884. Doi:10.3892/or.2020.7770

47. Harris GK, Shi X. Signaling by carcinogenic metals and metal-induced reactive oxygen species. *Mutat Res*. 2003;533(1-2):183-200. Doi:10.1016/j.mrfmmm.2003.08.025

48. Eckstein M, Rea M, Fondufe-Mittendorf YN. Transient and permanent changes in DNA methylation patterns in inorganic arsenic-mediated epithelial-to-mesenchymal transition. *Toxicol Appl Pharmacol*. 2017;331:6-17. Doi:10.1016/j.taap.2017.03.017

49. Chen QY, DesMarais T, Costa M. Metals and Mechanisms of Carcinogenesis. *Annu Rev Pharmacol Toxicol*. 2019;59:537-554. Doi:10.1146/annurev-pharmtox-010818-021031

50. Zimta AA, Schitcu V, Gurzau E, et al. Biological and molecular modifications induced by cadmium and arsenic during breast and prostate cancer development. *Environ Res*. 2019;178:108700. Doi:10.1016/j.envres.2019.108700

51. Davey JC, Bodwell JE, Gosse JA, Hamilton JW. Arsenic as an endocrine disruptor: effects of arsenic on estrogen receptor-mediated gene expression in vivo and in cell culture. *Toxicol Sci Off J Soc Toxicol*. 2007;98(1):75-86. Doi:10.1093/toxsci/kfm013

52. Bimonte VM, Besharat ZM, Antonioni A, et al. The endocrine disruptor cadmium: a new player in the pathophysiology of metabolic diseases. *J Endocrinol Invest*. 2021;44(7):1363-1377. Doi:10.1007/s40618-021-01502-x

53. Wan MLY, Co VA, El-Nezami H. Endocrine disrupting chemicals and breast cancer: a systematic review of epidemiological studies. *Crit Rev Food Sci Nutr*. Published online April 5, 2021:1-27. Doi:10.1080/10408398.2021.1903382

54. Aubé M, Larochelle C, Ayotte P. Differential effects of a complex organochlorine mixture on the proliferation of breast cancer cell lines. *Environ Res*. 2011;111(3):337-347. Doi:10.1016/j.envres.2011.01.010

55. Kakarala M, Wicha MS. Implications of the cancer stem-cell hypothesis for breast cancer prevention and therapy. *J Clin Oncol Off J Am Soc Clin Oncol*. 2008;26(17):2813-2820. Doi:10.1200/JCO.2008.16.3931

56. Yersal O, Barutca S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World J Clin Oncol*. 2014;5(3):412-424. Doi:10.5306/wjco.v5.i3.412

57. Danes JM, de Abreu ALP, Kerketta R, et al. Inorganic arsenic promotes luminal to basal transition and metastasis of breast cancer. *FASEB J Off Publ Fed Am Soc Exp Biol*. 2020;34(12):16034-16048. Doi:10.1096/fj.202001192R

	25nM vs. Ctrl		250nM vs. Ctrl		2.5uM vs. ctrl		25uM vs. Ctrl		
Chemical	Up	Down	Up	Down	Up	Down	Up	Down	Total DEGs
Arsenic	8	13	41	34	137	211	3604	4701	8749
Mercury	0	0	0	1	0	4	6	10	21
Copper	9	17	72	53	246	182	288	467	1334
Cadmium	1	4	0	7	0	67	2539	2087	4705
Lead	0	15	29	48	78	85	235	343	833
MPB	8	12	30	39	32	24	111	60	316
PPB	5	21	5	23	3	10	1	0	68
BPA	52	13	20	20	98	16	12	1	232
BPS	20	16	3	10	2	3	5	4	63
BPF	6	15	12	9	0	22	5	2	71
Phenanthrene	59	5	72	6	12	4	7	3	168
PFOA	96	78	92	53	56	66	52	72	565
PFDA	65	5	32	22	38	2	57	6	227
PFNA	20	2	24	4	76	3	68	3	200
PFOS	21	6	17	3	21	7	41	5	121
Thiram	0	1	3	5	0	32	2521	2549	5111
DCP_25	0	0	0	0	0	1	0	2	3
DCB_14	0	0	0	0	1	0	1	0	2
PCB153	1	0	1	50	0	0	1	0	53
PCB187	1	1	0	0	0	0	4	1	7
p,p'-DDE	467	312	852	572	645	562	960	749	5119
									27968

Table 1: Differentially expressed genes (DEGs) (FDR < 0.05 relative to control) by dose for each chemical.

Figure Legends

Figure 1: Differential gene expression for 21 disparities associated chemicals at 2.5 µM. Differential gene expression between each chemical dose and controls was calculated using quasi-likelihood negative binomial generalized log-linear modeling in edgeR. Red lines mark a false discovery rate (FDR) adjusted p-value of 0.05.

Figure 2: Benchmark dose accumulation plots and individual benchmark dose modeling graphs. In the accumulation plot of median best benchmark doses (A), each plotted point represents a differentially expressed gene and each differently colored line represents individual chemicals. On the right, example individual best fit benchmark dose graphs are shown for the gene *ARC* in arsenic treated cells (B) and for the gene *HSPA6* in thiram treated cells (C). For B and C, experimental data is shown in red, best fit model in blue, and lower bound of the 95% confidence interval of the benchmark dose (BMDL), benchmark dose (BMD), and upper bound of the 95% confidence interval of the benchmark dose (BMDU) labeled on each graph.

Figure 3: Comparison of benchmark dose modeling results with NHANEs exposure levels. Median BMD for each differentially expressed gene (red) and NHANES chemical biomarker concentration data, converted to molarity units, for non-Hispanic Black women (blue) plotted for each chemical.

Figure 4: Enrichment analyses. All of MSigDB's "Hallmark" biological processes as regulated by each chemical across different median BMDs (A). Nine of the chemicals showed significant enrichment or downregulation for 1 or more of 50 biological processes defined by hallmark gene sets (B). Eight chemicals showed significant alterations of breast cancer or stemness specific gene pathways as defined by some of MSigDB's chemical and genetic perturbations sets (C).

Figure 5: MuSiC deconvolution results. Comparing the proportion of estimated myoepithelial cells in MCF10A cultures treated with varying doses of arsenic (A), copper (B), lead (C), and p,p'-DDE (D) all showed significant increases in myoepithelial cell proportions relative to control.