SUPPLEMENTARY MATERIAL FOR:

ENVIRONMENTAL CHEMICAL BURDEN IN DIFFERENTIATED THYROID CANCER

Joshua D. Preston[1,2,3], Yongliang Liang[3], Thomas Szabo Yamashita[4], Sami N. Teeny[3], Jaclyn Weinberg[3,5], William J. Crandall[3,5], Zachery R. Jarrell[3], Xin Hu[3,6], Susan A. Safley[4], Jennifer M. Robertson[7], ViLinh Tran[3], Anee S. Jackson[4], Snehal G. Patel[4], Collin J. Weber[4], Jyotirmay Sharma[4], Neil D. Saunders[4], Young-Mi Go[3], Dean P. Jones[3], and M. Ryan Smith[3,8]

1- Medical Scientist Training Program, Emory University School of Medicine, Atlanta, GA 30322

2- Nutrition and Health Sciences, Laney Graduate School, Emory University, Atlanta, GA 30322

3- Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Department of Medicine, Emory University School of Medicine, Atlanta, GA, USA 30322

4- Division of General and GI Surgery, Department of Surgery, Emory University School of Medicine, Atlanta, GA 30322

5- Molecular and Systems Pharmacology, Laney Graduate School, Emory University, Atlanta, GA 30322

6- Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA 30322

7- Winship Cancer Institute, Emory University, Atlanta, GA 30322

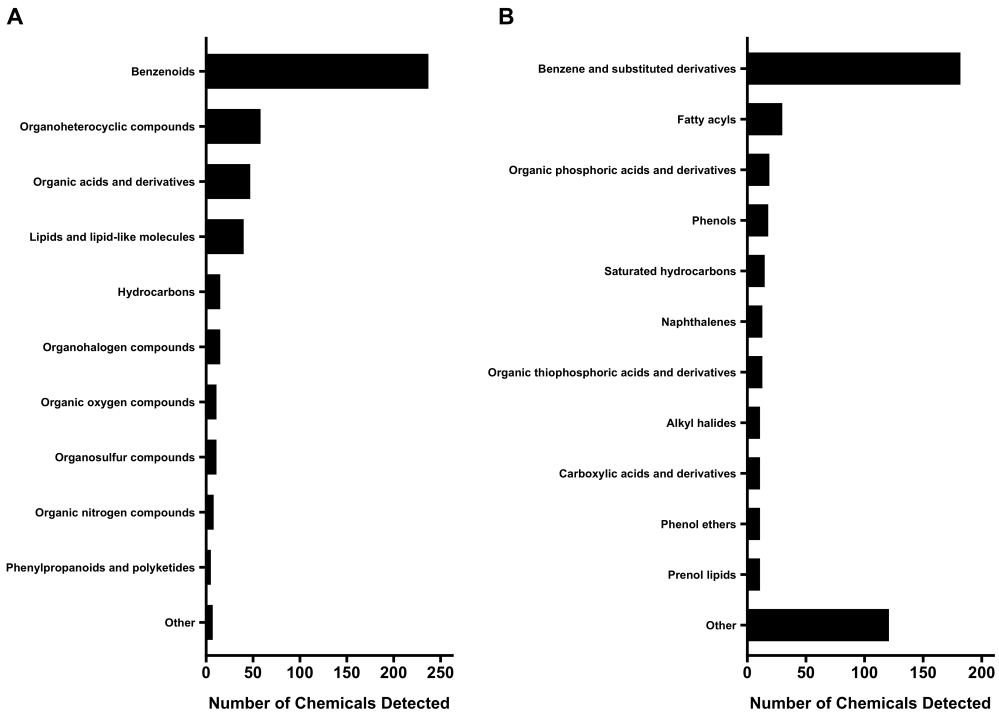
8- VA Healthcare System of Atlanta, Decatur, GA 30033

TABLE OF CONTENTS

SUPPLEMENTARY FIGURE 1 2
SUPPLEMENTARY FIGURE 2 3-4
SUPPLEMENTARY FIGURE 3 5
EXPANDED METHODS 6-11
High-Resolution Exposomics 6
Data Extraction, Feature Annotation, and Signal Identification 6
Preprocessing of Exposomics Data 7
Metadata Annotation of Environmental Chemicals 7
Core Statistical Analysis and Data Visualization 7
Online Code and Data Repository 8
Exposome-Wide Association Study 8
Advanced Carcinogenicity Classification of Select Chemicals 9
Quantitative Estimates of Chemical Concentrations in Tissues 9
Quantitative Comparisons of Tumors and Non-Cancer Cadaver Thyroids 11
Comparing Observed Concentrations to Literature Values for Select Chemicals 11
SUPPLEMENTARY TABLE 1 12-31
SUPPLEMENTARY TABLE 2 32-45
SUPPLEMENTARY TABLE 3 46
SUPPLEMENTARY TABLE 4 47
TABLE ABBREVIATION DICTIONARY 48-51
REFERENCES CITED IN SUPPLEMENTARY MATERIAL 52

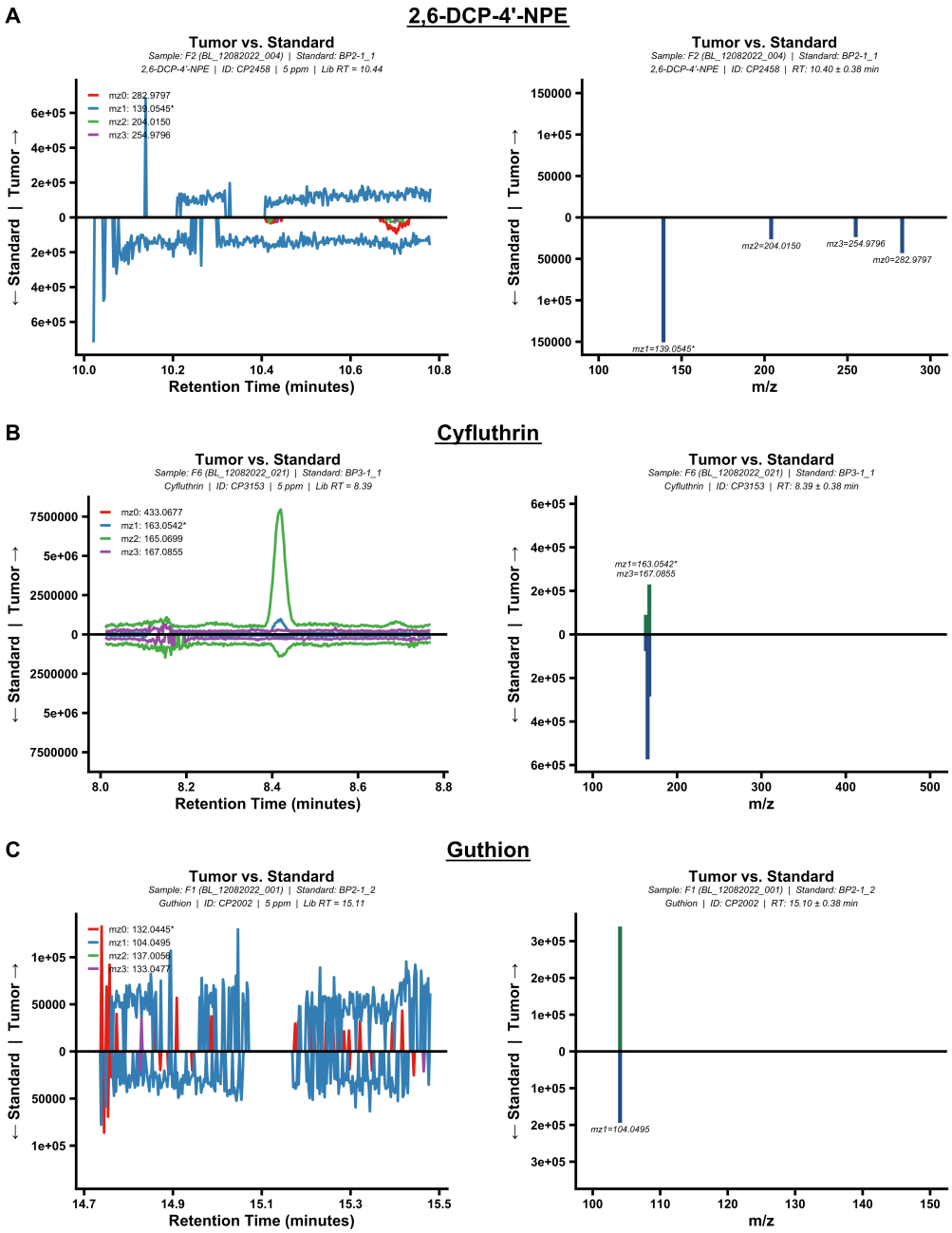
SUPPLEMENTARY FIGURE 1

Classification of detected chemicals. Chemical superclasses (A) and subclasses (B) of all detected chemicals.

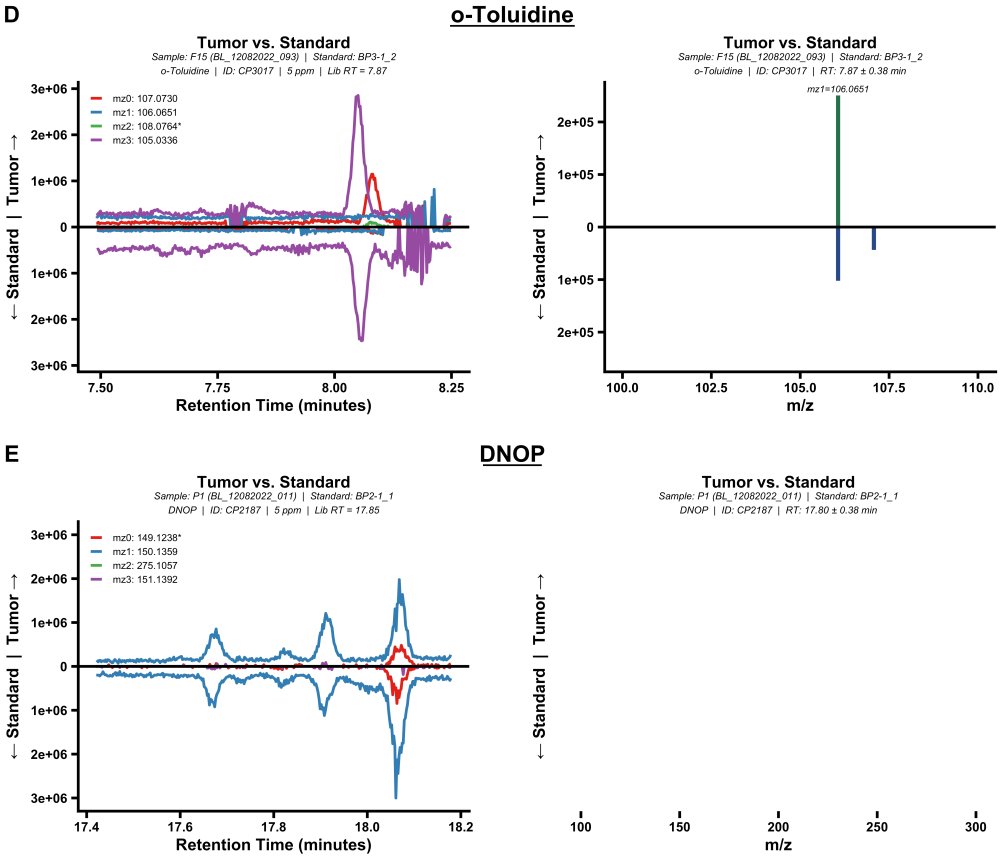


SUPPLEMENTARY FIGURE 2

Spectral validation of select chemical identifications in thyroid tumor samples. Mirrored plots comparing sample spectra (top, positive y-axis) to reference standard spectra (bottom, negative y-axis, flipped) for the top five quantitative mode chemicals from figure 3A: (A) 2,6-DCP-4'-NPE, (B) Cyfluthrin, (C) Guthion, (D) o-Toluidine, and (E) DNOP. For each chemical, the left panel shows retention time (RT) chromatograms with extracted ion chromatograms (EIC) for library m/z fragments, while the right panel shows mass spectra extracted within a RT window around the target peak. The Y axis on all plots represents intensity. Asterisks (*) indicate the fragment used for statistical comparison and displayed in Figure 3A. Figures A-C are displayed below and D-E are displayed on the following page.

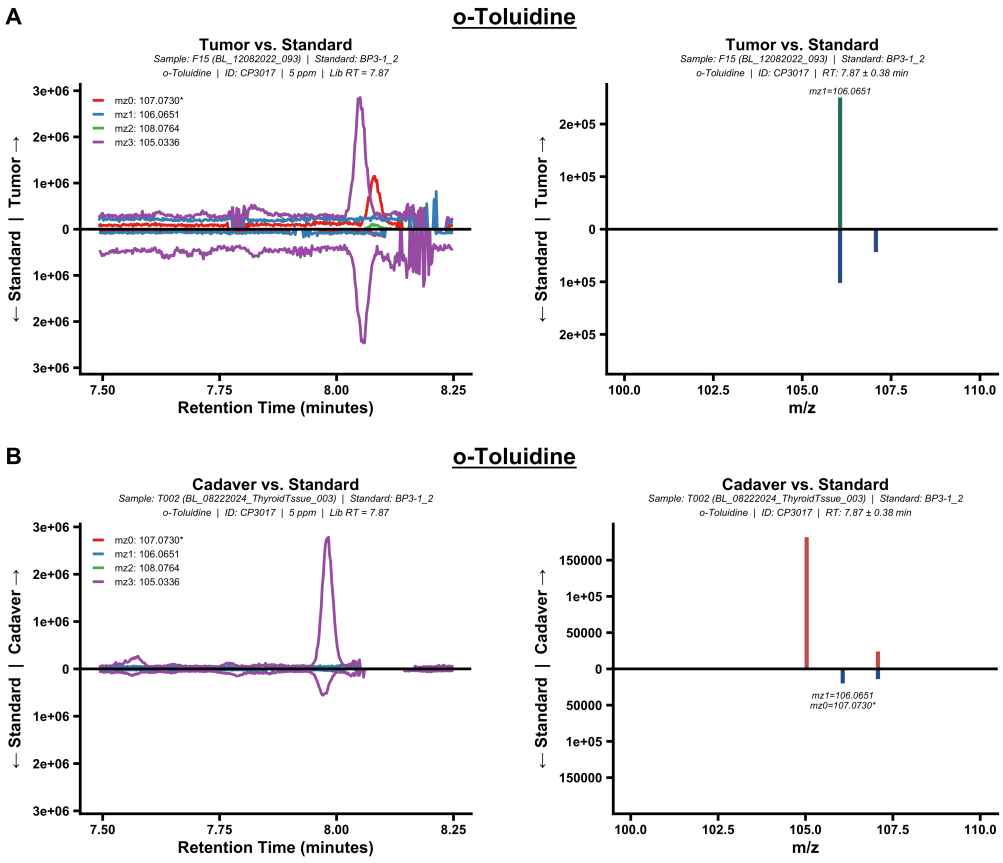


SUPPLEMENTARY FIGURE 2 (CONTINUED)



SUPPLEMENTARY FIGURE 3

Spectral validation of o-Toluidine in tumor and non-cancer cadaver thyroid tissues. Mirrored plots comparing sample spectra (top, positive y-axis) to reference standard spectra (bottom, negative y-axis, flipped) for o-Toluidine detected in (A) thyroid tumor tissue and (B) non-cancer cadaver thyroid tissue. Note that tumor and cadaver samples were analyzed in separate batches. For each tissue type, the left panel shows retention time (RT) chromatograms with extracted ion chromatograms (EIC) for library m/z fragments, while the right panel shows mass spectra extracted within a RT window around the target peak. The Y axis on all plots represents intensity. Asterisks (*) indicate the fragment used for statistical comparison and displayed in Figure 3D.



EXPANDED METHODS

High-Resolution Exposomics

Prior to processing samples, several individual reference standards were prepared by spiking pure chemical standards into pooled reference plasma to achieve a common concentration of 0·47 ng/mL. A majority of all chemical standards were purchased from AccuStandard (New Haven, CT, USA), Restek (Bellefonte, PA, USA), MilliporeSigma (Burlington, MA, USA), Santa Cruz Biotechnology (Dallas, TX, USA), Wellington Laboratories (Guelph, ON, Canada), Cambridge Isotope Laboratories (Tewksbury, MA, USA). Plasma reference standards and tissue samples were processed using a modified version of the express liquid extraction (XLE) method.[1] The XLE is optimized for biofluids; however, it was necessary to modify it to ensure analytical consistency across plasma standards and tumor tissue samples, as detailed below.

First, ∼∼50 mg of tumor or non-tumor cadaver thyroid tissue was cut and placed into a glass vial (13$×$100 mm), and exact tissue weights were recorded (mean ± SD = 52·7 ± 11·3 mg for tumor tissue, 43·1 ± 4·5 mg for non-cancer cadaver thyroid tissue). Next, 200 μL of each pooled reference plasma standard (an in-house standard referred to as QStd; see Go et al.[2] for a full description of a previous iteration of this standard) was transferred to glass vials (13$×$100 mm). For tissues, 500 μL of an extraction buffer comprised of acetone and petroleum ether (1:1 v/v) (MilliporeSigma, Catalog #184519) with ∼∼2% internal standard mix, was added to the glass vials containing tissues. For plasma, 200 μL of extraction buffer containing 2% internal standard mix was added to the glass vials containing the pooled reference plasma. 50 μL of formic acid (Emprove® Essential DAC, 98-100% pure, MilliporeSigma, Catalog #1002631000) was then added to tissue and plasma standard vials. To the tissue vials, 20 mg of NaCl was added, and 45 ± 5 mg of $MgSO_4$ ($≥$99·99% trace metals basis, Sigma-Aldrich, Catalog #203726) was added to tissue and plasma standard vials.

Tissue samples were then homogenized in this solution using a tissue-tearor, followed by water bath sonication for 30 minutes at room temperature. Plasma samples were shaken vigorously on ice using a multitube vortexer (VWR VX-2500) for 60 min. Following this, both tissue and plasma homogenates were centrifuged at 3000 rpm × 4 °C × 10 min. Supernatants were then transferred to new glass vials (13 × 100 mm) and dried using a vacuum centrifuge for 60 min at 35 °C. Finally, dried samples were reconstituted in 50 μL of isooctane. Reconstituted samples were transferred to autosampler vials with 150 μL vial inserts for analysis. Samples were prepared in batches containing 20 samples along with standard reference material (SRM) samples (NIST 1957, 1958, and several in-house SRMs discussed above). In addition, pooled reference plasma samples, a retention-time batch comparison standard, a method solvent blank that underwent solvent extraction, and a solvent blank were prepared in tandem as part of quality control measures. It should be noted that tumor tissues and non-cancer cadaver thyroid tissues were analyzed at different timepoints and in separate batches, although the same method was used for each tissue. As a result, direct comparison of raw spectral intensities between non-cancer thyroids and tumors was limited; however, quantitative estimates of chemical concentrations in tissues (see the section 'Quantitative Estimates of Chemical Concentrations in Tissues' below) did enable some degree of direct comparison, though these comparisons should be interpreted conservatively and cautiously given the high potential for batch effects.

Following preparation, samples were analyzed in duplicate with a Q Exactive GC hybrid quadrupole Orbitrap mass spectrometer (ThermoScientific), allowing for 25 minutes of data collection on retention times and spectral intensities from m/z 85-850 collected at 60k resolution. Samples were injected at 2 μL and subjected to a capillary DB-5MS column (15 m × 0·25 mm × 0·25 μm film thickness) with a gradient as follows: 75 °C for 1 min then 25 °C/min to 180 °C, 6 °C/min to 250 °C, 20 °C/min to 300 °C, with a final 5 min hold. Helium gas flow rate was 1 mL/min. Positive electron ionization was set at 70 eV, the ion source was set at 250 °C, and the transfer line was set at 280 °C.

Data Extraction, Feature Annotation, and Signal Identification

Raw data were extracted on a sample-by-sample basis using MZmine2.[3] Data were interpreted and combined using an in-house algorithm for feature annotation and the identification and quantification of chemicals in GC-MS exposomics data. Briefly, the algorithm utilizes a large library of environmental chemicals (EC) with known concentrations in standards, each of which has been manually quantified to ascertain retention times within the context of the laboratory procedure. In total, 738 unique standards are in the library; however, 28 of these represent chemicals that can be considered largely endogenous and not ECs. The remaining standards can be considered entirely exogenous (700) or both endogenous and exogenous (10) chemicals, resulting in a total of 710 ECs in the library, which can be screened by parallel analysis of standards with samples. This library is compared against sample data to find the best matching signal for ECs if any appropriate candidates exist. Shifts in elution time due to column degradation and the presence of co-occurring isomers pose significant challenges for standard annotation procedures; however, the algorithm addresses this issue by employing time-warping techniques, along with corrective algorithms

that mimic manual chemical identification procedures, to overcome these challenges. Chemical identification and accuracy are achieved by leveraging the co-elution and correlation of chemical fragments generated from the same chemical. Standards are run in tandem with samples, which allows not only for the identification and relative quantification of all chemicals in the library of standards but also for the absolute quantification of identified metabolites using the comparative intensity of peak areas. The algorithm specifically assigns "quality fractions" based on the simultaneous detection of multiple fragments of the same chemical. Following the application of the algorithm for annotation, the quality fraction was used to label features as "annotations" (quality fraction of 0) while the remaining signals were labeled "identifications." It is important to note that identification versus annotation status was determined solely in reference to tumor samples and not non-cancer cadaver thyroid samples. Thus, when annotations or identifications are denoted in any figures or tables, these refer to the determination made in the data from thyroid tumors rather than non-cancer cadaver samples. Information on annotation versus identification status in non-cancer cadaver samples can be found in the "primary_data.xlsx" spreadsheet located in the GitHub repository (see below for further details). The algorithm will be published as a Python or R package in the coming 1-2 years. For further details on the algorithm, readers are welcome to contact the corresponding author.

Preprocessing of Exposomics Data

A targeted exposomics feature table (see the associated GitHub repository) was derived from the feature annotation and signal identification process described above. The proportion of missing values (PMV) was then calculated for each feature (total number of missing values/60 total samples). Next, features with PMV $>$ 30% missing values were separated for analysis in "qualitative" mode. The individual spectral intensity values for qualitative features were converted to "1", representing detection, whereas missing values were converted to "0", representing non-detection. Any qualitative features with 100% "0" values (indicating detection in reference standards but complete non-detection in samples) were eliminated and not considered for further analysis. Alternatively, the remaining features with PMV $\leq$ 30% were separated for analysis in "quantitative" mode. The missing values for quantitative features were imputed using the half minimum method (study-wide).[4] Following imputation, all spectral intensity values for quantitative features were $\log_2$-transformed before further analysis. Finally, to determine the total number of unique chemicals detected per sample, features were consolidated and grouped by their respective Chemical Abstracts Service registry numbers (CAS number), regardless of analytical mode (qualitative versus quantitative), such that the detection of at least one fragment for a given CAS number was considered as detection of that unique chemical. This list of chemicals was then compiled, and chemicals that were mainly endogenous (i.e., carnitines, cholesterol, arachidonic acid metabolites, bile acids, hormones, and hormone precursors) were removed and not considered for exposomics analysis (23 total). Chemicals that had both exogenous and endogenous dispositions (7 total) were retained for exposomics analysis. The number of unique chemicals detected was then tallied for each sample. The median number of chemicals detected per variant (prior to imputation) was compared using the Kruskal-Wallis test.

Metadata Annotation of Environmental Chemicals

The name, IUPAC name, CAS number, PubChem Compound Identification (CID), SMILES, InChIKey, InChI, Toxin and Toxin-Target Database[5] ID (when applicable), monoisotopic mass, and formula were obtained for each chemical detected in at least one sample. For the purposes of data visualization and simplified naming in the manuscript text, a short name or abbreviation was assigned to a large portion of chemicals. The Toxin-Toxin-Targeted Database[5] (T3DB) ID and the International Agency for Research on Cancer (IARC) carcinogen group[6] were assigned to chemicals when applicable and possible (when this information could not be ascertained, the T3DB ID was listed as "NA" and the IARC Group was listed as "Not Classified"). Classification as either a potential endocrine-disrupting chemical (EDC) or non-EDC was assigned to each chemical, per the PARCEDC list.[7] Inclusion on the list resulted in chemicals being marked as potential EDCs; non-inclusion resulted in classification as a non-EDC. The single most common use or best-fit chemical class was identified for each chemical. To facilitate data visualization, broader categories were also assigned to each chemical. For example, if a chemical's common-use class was "Insecticide/Pesticide (Pyrethroid)", this was simplified to "Insecticide/Pesticide" for data display. In the broader categories, metabolites or degradation products of certain chemicals were simply considered as members of the class to which their precursor belonged. For example, aldicarb sulfone, a breakdown product of the carbamate pesticide aldicarb, was classified as an insecticide/pesticide. However, the inclusion of breakdown products is indicated in figures and tables where applicable. ClassyFire[8] was used to designate superclasses, classes, subclasses, direct parents, and molecular frameworks. All chemical metadata are available below in Supplementary Table 1.

Core Statistical Analysis and Data Visualization

All data were compiled and structured using Microsoft Excel (Mac v16·96, Microsoft Corporation, 2025). All code

was written and executed using R (v4·3·1)[9] in Visual Studio Code (v1·82·2, Microsoft Corporation, 2025). Source code can be found on the GitHub repository (source_code.R; see below for details). Data visualization was performed using GraphPad Prism (Mac v10·4·0, GraphPad Software, 2024) and the R package, ggplot2.[10] All figures were compiled and edited using BioRender.com (see 'BioRender Publication Licenses' below).

Online Code and Data Repository

All source code is available in the accompanying GitHub repository, accessible at the following URL: https://github.com/jdpreston30/thyroid-exposomics-2025.

Exposome-Wide Association Study

Electron ionization generates multiple fragments for each chemical prior to detection; thus, in many cases, multiple fragments were annotated for the same chemical. In keeping with our exploratory and descriptive approach and to maximize coverage, we treated each fragment as an individual observation during statistical analysis. For qualitative features, Fisher's exact test for count data was employed on the binary exposomics data to determine if any chemical was overrepresented in its detection within any specific variant. Corresponding detection fractions were calculated for each variant (sum of samples with detection of chemical/20 samples per variant). For quantitative features, one-way ANOVA was used to assess differences in mean spectral intensities across variants, with corresponding p-values calculated for each chemical. Post-hoc testing was performed using Tukey's HSD test, which included correction for multiple comparisons. Given the pilot design and exploratory nature of this study, no false discovery rate corrections were applied beyond those used in post-hoc testing. Mean spectral intensities were scaled by converting them to z-scores within each chemical. Chemicals with a $p < 0{\cdot}05$ from either test were considered significant and were compiled, tabulated, and further classified as described below. Four chemicals had multiple fragments showing significant differences between variants. In these cases, the quantitative fragment was selected if available; otherwise, the fragment with the highest frequency of detection across all samples, irrespective of variant, was selected. The top 5 quantitative features were visualized via violin plots (of z-scored spectral intensities), and the top 10 qualitative features via a heatmap (of detection fraction/percent). A balloon plot was constructed based on these results, wherein significant ECs were grouped by usage classes and variants, being `counted' once each time it was`highest,' whether by qualitative or quantitative measures in a variant. In cases where the detection percentages were equally high in two variants, a count was assigned to each variant.

Advanced Carcinogenicity Classification of Select Chemicals

To identify any potential carcinogens that have not been evaluated by the IARC, additional research was performed on the chemicals that showed significant differences in presence or concentration between variants (63 in total). To accomplish this, the Global Harmonized System of Classification and Labelling of Chemicals (GHS) statements were screened for any listed H350 ("may cause cancer"), H350i ("may cause cancer by inhalation"), or H351 ("suspected of causing cancer") status (and the corresponding consensus/confidence percentages of data sources as listed on PubChem in section 12·1.1, "GHS Classification"). Systematic logic that considered both the IARC grouping and the GHS carcinogen statement was then applied to categorize chemicals into the groups "Known Carcinogen", "Likely Carcinogen", "Possible Carcinogen", or "Uncertain Risk" (for the function accomplishing this, see the blocks of code under "#+ Carcinogen classification based on GHS and IARC (3E)" in the source_code.R file available in the GitHub repository). Chemicals with no IARC group and GHS statement were not classified. Briefly, the following rules were applied for classification:

- IARC group 1 chemicals were automatically assigned "Known Carcinogen", and group 2A "Likely Carcinogen". These conditions overrode any other considerations.
- If IARC grouping was 2B, 3, or not evaluated, but H350 or H350i 50%, then the chemical was assigned "Likely Carcinogen".
- If IARC grouping was 2B, 3, or not evaluated, but H351 $>$ 0% and/or 0% $<$ H350/H350i $<$ 50%, then the chemical was assigned "Possible Carcinogen".
- IARC group 3 chemicals with no supporting GHS H350, H350i, or H351 statement listed were assigned "Uncertain Risk".
- Chemicals with no IARC evaluation and no supporting GHS H350, H350i, or H351 statement listed were considered "Unclassified" as to their carcinogenicity.

These classifications, along with other relevant chemical metadata, can be found on the GitHub repository (chemical_metadata.xlsx).

Quantitative Estimates of Chemical Concentrations in Tissues

The SRM used for annotation and identification had a concentration of 0·47 ng/mL for each individual EC. Thus, we used the spectral intensities associated with these known concentrations to establish estimates of chemical concentrations within tissues in parts per million (PPM) and parts per billion (PPB). It should be noted that the original data processing pipeline used a nominal concentration of 0·5 ng/mL for calibration calculations; therefore, a correction factor of 0·94 (0·47/0·5) was applied to all concentration estimates prior to PPM and PPB calculations to account for the true SRM concentration. Importantly, we relied on the following experimental features and assumptions:

- Tissues are processed via solvent extraction and subsequent solvent evaporation, followed by reconstitution in 50 μL of isooctane. For the purposes of calculating concentrations in the original samples, we assume 100% extraction efficiency from tissues and plasma, such that the entire amount of the EC present in the sample is recovered in the isooctane.
- A chemical concentration of 0 ng/mL results in an intensity of 0 (i.e., the origin of (0,0)). This is used in combination with the intensity from the known concentration in the SRM (0·47 ng/mL) to establish a two-point standard curve, which assumes linearity.
- While each chemical has a concentration in the SRM of 0·47 ng/mL, the concentration in 50 μL of isooctane is 1·88 ng/mL. This is because 200 μL of SRM plasma is used, and therefore, 0·094 ng of the total EC of interest is extracted. This is dried after extraction and reconstituted in 50 μL of isooctane, resulting in a concentration of 1·88 ng/mL in the injected solution.

Our calculations to estimate tissue concentrations were performed as follows for each individual chemical:

1. A linear model is fit based on the assumptions discussed above, such that:

$$y = \beta x$$

   where y is the known concentration of the EC in the SRM (0·47 ng/mL for all ECs analyzed), x is the observed spectral intensity for the given EC, and β is the slope of the calibration curve, defined by the line through the origin (0,0) and the point (x, y).

2. Once the slope (β) for the specific chemical is established, it is then used to estimate concentration (ng/mL) in the sample (C_e) using the following equation:

$$C_e = \beta I$$

   where I is the observed spectral intensity for the given EC in the sample. However, to determine the true concentration in the isooctane solvent (C_{es}), we must scale C_e by a factor of 4 to account for the concentration that occurs during extraction. Specifically, the 200 μL of plasma standard extracted contains 0·094 ng total of the EC (0·47 ng/mL). Assuming 100% extraction, this 0·094 ng was then resuspended in 50 μL isooctane (1·88 ng/mL), Therefore, we use a 4× scaling factor:

$$C_{es} = 4C_e$$

3. Next, the C_{es} is converted from ng/mL to mg/L:

$$C_{es}\ \text{(mg/L)} = C_{es}\ \text{(ng/mL)} \times \frac{10^3\ \text{mL}}{\text{L}} \times \frac{10^{-6}\ \text{mg}}{\text{ng}}$$

$$C_{es}\ \text{(mg/L)} = C_{es}\ \text{(ng/mL)} \times 10^{-3}$$

4. Next, the total mass of the chemical of interest present in 50 μL of isooctane (M_{EC}) is determined:

$$M_{EC}\text{ (mg)} = C_{es}\text{ (mg/L)} \times 50\ \mu\text{L solvent} \times \frac{10^{-6}\text{ L}}{\mu\text{L}}$$

$$M_{EC}\text{ (mg)} = 5 \times 10^{-5}\text{ (L)} \times C_{es}\text{ (mg/L)}$$

5. To derive the mass fraction of the EC in tissue, we divide M_{EC} (mg) by M_T (mg) and scale to PPM or PPB:

$$\text{PPM} = \frac{M_{EC}}{M_T} \times 10^6$$

$$\text{PPB} = \frac{M_{EC}}{M_T} \times 10^9$$

Alternatively, when algorithm outputs are used directly without unit conversions, the following formula can be used, provided C_e is in ng/mL and M_T is in mg:

$$\text{PPM} = \frac{C_e \times 10^2}{M_T}$$

$$\text{PPB} = \frac{C_e \times 10^5}{M_T}$$

Quantitative Comparisons of Tumors and Non-Cancer Cadaver Thyroids

Targeted quantitative feature tables, which were additionally filtered for detection in relevant standards, were generated for tumors and non-cancer cadaver thyroids, and conversion to PPM was performed as described above. Once spectral intensities were converted to PPM, direct comparisons of chemical concentrations were made only for fragments that were detected in both the tumor and non-cancer thyroid tissues. In cases where multiple fragments were annotated for the same chemical in both tumors and non-cancer thyroids, an ideal fragment was first selected based on the highest percentage of detection in tumors, then the highest percentage of detection in non-cancer thyroids, and finally the greatest mean intensity in tumors. For all matching fragments, the mean, maximum, and theoretical minimum (i.e., half the minimum detectable concentration in either the tumors or non-cancer thyroids) were calculated for tumors (all combined) and non-cancer thyroids. These data, along with percentage detection in tumors and non-cancer thyroids, are reported as PPB below in Supplementary Table 3. It is important to note that means were calculated using half-minimum imputed data, even if features were detected at 'qualitative' analysis thresholds. Finally, these values were converted to PPM for a direct comparison of tumors to non-cancer thyroids of all detected IARC Group 1 carcinogens (Figure 3D); however, this comparison was only reported for features that met quantitative criteria in both the tumor and non-cancer thyroid datasets. T-tests were run on $\log_2$-transformed data to compute p-values for comparing the means between tumors and non-cancer thyroids for these chemicals.

Comparing Observed Concentrations to Literature Values for Select Chemicals

To compare the quantification estimates generated as described above with estimated values in other tissues and matrices, we identified values for select chemicals in the published literature and tabulated our data alongside these reported concentrations (Supplementary Table 2). We prioritized chemicals that were IARC Group 1 carcinogens and polycyclic aromatic hydrocarbon combustion byproducts. For quantitative estimates of 245 individual chemicals, see Supplementary Table 3.

SUPPLEMENTARY TABLE 1

The full library of xenobiotic chemicals employed for chemical identification. The library of 710 confirmed xenobiotic chemicals employed for chemical identification. All chemicals were present in pooled reference plasma at a concentration of 0·47 ng/mL. There are 710 total unique chemicals (i.e., unique CAS numbers), but for some chemicals, there are multiple fragments from different standards used for identification, thus resulting in 892 total rows in the table. The individual mz columns indicate typical fragments observed for the given chemical.

[INSERT ST1 HERE - TO BE GENERATED PROGRAMMATICALLY]

SUPPLEMENTARY TABLE 2

Metadata for all chemicals detected in samples. Both the long-form chemical name and alias or abbreviation are listed if there is sufficient space. However, for chemical names that are too long or redundant, only the alias or abbreviation has been listed. The column 'variant diff.' specifies if the chemical had differential abundance or detection between the three variants.

[INSERT ST2 HERE - TO BE GENERATED PROGRAMMATICALLY]

SUPPLEMENTARY TABLE 3

Observed concentrations versus reported literature values. All values originally published as ng/g, μg/L, ng/mL are listed as PPB. Values originally published as pg/mL were converted to ng/mL and then listed as PPB. All values are rounded to the nearest integer or are listed as $<$ 1 PPB when applicable. The corresponding reference from which the comparison value is derived is cited next to the listed concentration.

[INSERT ST3 HERE - TO BE GENERATED PROGRAMMATICALLY]

SUPPLEMENTARY TABLE 4

Quantitative estimates of chemicals in non-cancer thyroids and tumors. Data table containing quantitative estimates of chemical concentrations in thyroid tissues.

[INSERT ST4 HERE - TO BE GENERATED PROGRAMMATICALLY]

TABLE ABBREVIATION DICTIONARY

The supplementary tables have a substantial number of abbreviations, largely for chemical names. A full dictionary of abbreviations relevant to all supplementary tables can be found below:

- 9Cl-PF3ONS = 9-Chlorohexadecafluoro-3-oxanone-1-sulfonic acid
- BDCPP = bis(1,3-Dichloro-2-propyl) phosphate
- BDE = brominated diphenyl ether
- BDPP = Bis(2,3-dibromopropyl) hydrogen phosphate
- Bromo-TMP-Phenol = 2-Bromo-4-(2,4,4-trimethylpentan-2-yl)phenol
- CAS = Chemical Abstracts Service (Number)
- CDC = Centers for Disease Control and Prevention
- CID = Compound ID (PubChem)
- Compds. = compounds
- DBahA = Dibenz(a,h)anthracene
- DCP = Dichlorophenyl
- DCPMNB = 4-(2,4-dichlorophenoxy)-2-methyl-1-nitrobenzene
- DDD = Dichlorodiphenyldichloroethane
- DDE = Dichlorodiphenyldichloroethylene
- DDT = Dichlorodiphenyltrichloroethane
- DFTPP = Decafluorotriphenylphosphine
- DTPAs = Dithiophosphoric Acids
- EPN = Ethyl p-nitrophenyl phenylphosphorothioate
- EtFOSAA = N-Ethylperfluoro-1-octanesulfonamidoacetic acid (linear)
- Furaneol = 4-Hydroxy-2,5-dimethyl-3(2H)-furanone
- HpCDD = Heptachlorodibenzo-p-dioxin
- HpCDF = Heptachlorodibenzofuran
- HxCDD = Hexachlorodibenzo-P-dioxin
- HxCDF = Hexachlorodibenzofuran
- IARC = International Agency for Research on Cancer
- IMHP = 2-Isopropyl-6-methyl-4-pyrimidinol
- Lin. = linear
- LLMs = lipid-like molecules
- LOD = limit of detection
- MBOT = 4,4’-Methylenebis(o-toluidine)
- MCPA = 2-Methyl-4-chlorophenoxyacetic acid
- MEcPP = Mono(5-carboxy-2-ethylpentyl) phthalate

- MEHHP = Mono(2-ethyl-5-hydroxyhexyl) phthalate
- MEOHP = Mono(2-ethyl-5-oxohexyl) phthalate
- MGK-264 = McLaughlin Gormley King-264 (also known as N-2-Ethylhexylbicycloheptenedicarboximide)
- min = minutes
- MOCA = 4,4'-Methylenebis(2-chloroaniline)
- mz = mass-to-charge ratio
- N-MeFOSAA = N-Methylperfluoro-1-octanesulfonamidoacetic acid (linear)
- NHANES = National Health and Nutrition Examination Survey
- NPE = nitrophenyl ether
- o-Dianisidine = 3,3'-Dimethoxybenzidine
- OD-PABA = Octyl-dimethyl-p-aminobenzoic acid
- Org = organic
- Org. Heterocycl. = organoheterocyclic
- p-Chlorocresol = 4-Chloro-3-methylphenol
- PAH = polycyclic aromatic hydrocarbon
- PBB = polybrominated biphenyl
- PCB = polychlorinated biphenyl
- PCDF = Pentachlorodibenzofuran
- PeCDD = pentachlorodibenzo-p-dioxin
- PKs = polyketides
- PPB = parts per billion
- RT = retention time
- SDs = Substituted Derivatives
- TBBPA-BAE = Tetrabromobisphenol A bis(allyl ether)
- TCDD = tetrachlorodibenzo-p-dioxin
- TCDF = tetrachlorodibenzofuran
- TCP = Trichlorophenyl
- TCP-4'-NPE = TCP-4'-NPE
- TCPP = Tris(1-chloro-2-propyl) phosphate
- TDCPP = Tris(1,3-dichloro-2-propyl)phosphate
- TEEP = Tetraethyl ethylenediphosphonate
- TPAs = Thiophosphoric acids
- TTBNPP = Tris(tribromoneopentyl) phosphate

REFERENCES CITED IN SUPPLEMENTARY MATERIAL

1 Hu X, Walker DI, Liang Y, *et al.* A scalable workflow to characterize the human exposome. *Nature Communications* 2021; **12**: 5575.

2 Go Y-M, Walker DI, Liang Y, *et al.* Reference Standardization for Mass Spectrometry and High-resolution Metabolomics Applications to Exposome Research. *Toxicological Sciences* 2015; **148**: 531–43.

3 Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 2010; **11**: 395.

4 Wei R, Wang J, Su M, *et al.* Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. *Scientific Reports* 2018; **8**: 663.

5 Wishart D, Arndt D, Pon A, *et al.* T3DB: The toxic exposome database. *Nucleic Acids Research* 2015; **43**: D928–34.

6 Agents Classified by the IARC Monographs, Volumes 1–136.

7 Andres S, Dulio V. S109 | PARCEDC | List of 7074 potential endocrine disrupting compounds (EDCs) by PARC T4.2. 2024; published online April. DOI:10.5281/ZENODO.10944198.

8 Djoumbou Feunang Y, Eisner R, Knox C, *et al.* ClassyFire: Automated chemical classification with a comprehensive, computable taxonomy. *Journal of Cheminformatics* 2016; **8**: 61.

9 (2023) RCT. R: A Language and Environment for Statistical Computing.

10 Wickham H. Ggplot2: Elegant graphics for data analysis, Second edition. Switzerland: Springer, 2016.