

# Latent Variables Capture Pathway-Level Points of Departure in High-Throughput Toxicogenomic Data

Danilo Basili, Joe Reynolds, Jade Houghton, Sophie Malcomber, Bryant Chambers, Mark Liddell, Iris Muller, Andrew White, Imran Shah, Logan J. Everett, Alistair Middleton,\* and Andreas Bender\*



Cite This: *Chem. Res. Toxicol.* 2022, 35, 670–683



Read Online

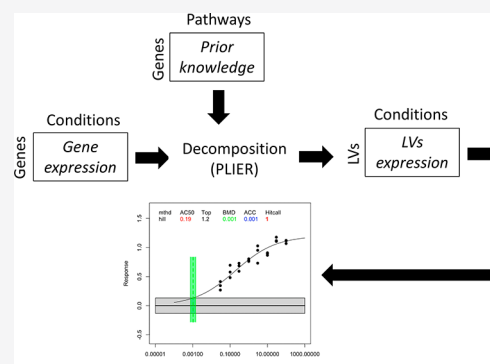
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Estimation of points of departure (PoDs) from high-throughput transcriptomic data (HTTr) represents a key step in the development of next-generation risk assessment (NGRA). Current approaches mainly rely on single key gene targets, which are constrained by the information currently available in the knowledge base and make interpretation challenging as scientists need to interpret PoDs for thousands of genes or hundreds of pathways. In this work, we aimed to address these issues by developing a computational workflow to investigate the pathway concentration–response relationships in a way that is not fully constrained by known biology and also facilitates interpretation. We employed the Pathway-Level Information ExtractoR (PLIER) to identify latent variables (LVs) describing biological activity and then investigated in vitro LVs' concentration–response relationships using the ToxCast pipeline. We applied this methodology to a published transcriptomic concentration–response data set for 44 chemicals in MCF-7 cells and showed that our workflow can capture known biological activity and discriminate between estrogenic and antiestrogenic compounds as well as activity not aligning with the existing knowledge base, which may be relevant in a risk assessment scenario. Moreover, we were able to identify the known estrogen activity in compounds that are not well-established ER agonists/antagonists supporting the use of the workflow in read-across. Next, we transferred its application to chemical compounds tested in HepG2, HepaRG, and MCF-7 cells and showed that PoD estimates are in strong agreement with those estimated using a recently developed Bayesian approach ( $\text{cor} = 0.89$ ) and in weak agreement with those estimated using a well-established approach such as BMDExpress2 ( $\text{cor} = 0.57$ ). These results demonstrate the effectiveness of using PLIER in a concentration–response scenario to investigate pathway activity in a way that is not fully constrained by the knowledge base and to ease the biological interpretation and support the development of an NGRA framework with the ability to improve current risk assessment strategies for chemicals using new approach methodologies.



## INTRODUCTION

Toxicity testing to provide information about chemical hazards and risks has historically relied on apical end points derived from living animals,<sup>1</sup> which is time-consuming, expensive, and the results depend on biological variation. In 2007, the U.S. National Academy of Science envisioned a change in the way toxicity testing is conducted by promoting a transition from an expensive and lengthy in vivo testing to an in vitro high-throughput screening of chemical toxicity by using cell lines.<sup>2</sup> In the last few decades, efforts by academia, industry, and regulatory bodies have driven the development of new approach methodologies (NAMs) that can accelerate the pace of change in chemical risk assessment by providing information on chemical safety without the need for animal testing.<sup>3–7</sup> NAMs encompass any technology or methodology that is able to inform about chemical safety in a high-throughput manner without relying on animal testing and plays a key role in the development of next-generation risk assessment (NGRA).<sup>8</sup> In NGRA, safety decisions are based

largely on the margin of safety, also referred to as the bioactivity exposure ratio,<sup>7,9–11</sup> which is the ratio between an appropriate exposure metric (e.g.,  $C_{\text{max}}$ , AUC) for a given chemical and the point of departure (PoD), defined as the concentration at which the chemical induces bioactivity in relevant in vitro assays.<sup>12–14</sup>

One of the ways to derive data for NGRA is gene expression profiling.<sup>15</sup> While earlier approaches relied on relatively expensive microarrays,<sup>16</sup> recent technological improvements in high-throughput transcriptomics (HTTr) have made it feasible to analyze thousands of chemicals in concentration–response.<sup>11,17</sup> In this context, HTTr is becoming a practical

Received: December 23, 2021

Published: March 25, 2022



Table 1. Panel of Chemicals Selected for the Present Study<sup>ac</sup>

| chemical                    | known biological effects  | CAS number | concentration tested ( $\mu\text{M}$ )  |
|-----------------------------|---|------------|---|
| phenoxyethanol <sup>b</sup> | antimicrobial activity  | 122-99-6   | 0.01–0.1–1–10–100–1000  |
| niacinamide <sup>b</sup>    | antipruritic, antimicrobial, vasoactive, photo-protective, sebostatic | 98-92-0    | 1–4–10–100–1000–10,000  |
| coumarin <sup>b</sup>       | CYP450  | 91-64-5    | 0.001–0.01–0.1–1–10–100   |
| caffeine <sup>b</sup>       | CNS stimulant   | 58-08-2    | 0.01–0.1–1–10–100–1000  |
| andrographolide             | anti-inflammatory   | 5508-58-7  | HepG2: 0.2–0.5–2–5–10–15<br>HepaRG: 0.2–0.5–2–5–15–80<br>MCF-7: 0.2–0.5–2–5–15–80   |
| triclosan                   | fatty acid synthesis inhibitor  | 3380-34-5  | HepG2: 0.1–0.4–1–4–10–20<br>HepaRG: 0.1–0.4–1–4–10–50<br>MCF-7: 0.1–0.4–1–4–10–50   |
| flutamide                   | AR antagonist   | 13311-84-7 | HepG2: 0.2–0.5–2–5–20–50<br>HepaRG: 0.2–0.5–2–5–20–200<br>MCF-7: 0.2–0.5–2–5–20–200 |

<sup>a</sup>For each chemical, the tested concentrations are reported along with the chemical mode of action (MoA) and the CAS number. The differences in concentrations tested across different cell lines were due to different sensitivities assessed by the cell viability assay (see the section Cell Viability Assay for details). <sup>b</sup>The tested concentrations were all the same across the three cell lines.

approach to estimate chemical PoDs used to derive safety decisions in NGRA.<sup>12</sup> Concentration–response modeling of transcriptional data to derive probe-level PoDs is performed by identifying the best fit model for each probe expression profile across the concentration range tested, based on a set of criteria for determining an active call.<sup>18</sup> The estimated PoDs are used to inform safety decisions, and this is usually done by taking the most conservative (i.e., protective) approach, which is to select the gene with the lowest PoD. However, the challenge lies in moving beyond this extremely conservative position where PoDs are instead selected based on both how protective they are and their biological relevance. To facilitate biological interpretation, knowledge bases are usually employed to derive pathway-level PoDs by computing the lowest or median PoDs from the collection of all the responsive probes.<sup>18–20</sup> However, while this approach has proven to be successful in deriving PoDs for known biological pathways,<sup>21</sup> taking the coordinated expression of multiple genes in a pathway is not trivial.<sup>22</sup> Approaches to address this challenge have already been explored and rely on aggregating genes into biologically relevant sets prior to downstream analysis.<sup>23</sup> Aggregating across genes offers many advantages as it improves replication, reduces technical and biological variances across the samples, facilitates biological interpretation, and may also reduce the dimensionality of the data set as well as increase statistical power depending on the size of the gene sets considered.<sup>24–26</sup> Despite these advantages, gene aggregation into biologically relevant gene sets prior to concentration–response modeling has been rarely performed. To our knowledge, the only example of such an approach is the one performed by Harrill and collaborators,<sup>11</sup> who recently developed a novel gene expression signature-based concentration–response modeling approach, which they applied on the HTTr assay screening of a small set of chemicals in the MCF-7 cell model. Harrill and collaborators found out that aggregating signals from individual genes into gene signatures before concentration–response modeling yielded biological pathway altering concentrations that were closely aligned with previous ToxCast assays.<sup>11</sup> While annotating genes with pathways combine both data and prior knowledge, such approaches are restricted by the use of databases created by the curation of scientific literature that are often biased toward genes with prior experimental evidence,<sup>27</sup> and genes not involved in any of the known gene sets are

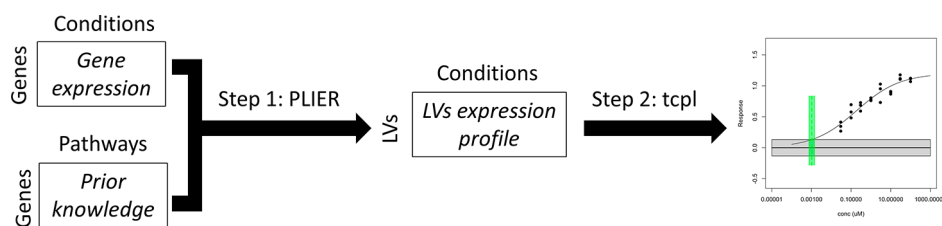
discarded regardless of their level of change. In addition, the scale and redundancy of the existing pathway collections result in the testing of a huge number of gene sets, each with a different estimated PoD, leading to interpretation challenges.

In the present study, we hence set to address these limitations by developing a computational workflow to (1) investigate pathway-level concentration–response relationships in a manner that is not fully constrained by current knowledge bases and (2) ease the biological interpretation. First, we evaluated our workflow on a published transcriptomic data set produced by Harrill et al.<sup>11</sup> and leveraging the TempO-Seq technology<sup>28</sup> for screening 44 chemicals in MCF-7 cells using the concentration–response model.<sup>11</sup> Second, we tested our workflow on a small in-house data set obtained by screening seven chemicals in HepG2, MCF-7, and HepaRG cells using the concentration–response model to compare PoD estimates across different methods.

## MATERIALS AND METHODS

**MCF-7 Data (Public Data Set).** MCF-7 data were obtained from the work performed by Harrill and collaborators.<sup>11</sup> Briefly, this study screened 44 chemicals with different target annotations (Table S1, Supporting Information) in MCF-7 cells across a range of eight concentrations from 0.03 to 100  $\mu\text{M}$  (1/2 log spacing). The MCF-7 count data, which were generated using the TempO-Seq technology,<sup>28</sup> were downloaded from GEO (GSE162855). Flagged samples as identified by Harrill et al.<sup>11</sup> were removed from the analysis. Genes with fewer than 10 counts in more than 90% of the samples were removed. Reproducibility and overall sample quality were assessed by means of PCA and correlation. The data set was adjusted for batch effects due to experimental plate differences using ComBat-Seq (*ComBat\_seq* function from the SVA package, version 3.40.0).<sup>29</sup> More specifically, ComBat-Seq was run by specifying the experimental plate along with the treatment-concentration condition as a biological covariate in order for the signal to be preserved in the adjusted data. Finally, the full data set was normalized by the DESeq2's median of ratios using the *DESeq* function (package DESeq2, version 1.32.0) to account for differences in sequencing the depth and RNA composition.<sup>30</sup>

**HepG2, HepaRG, and MCF-7 Data (In-House Data Set). Chemical Selection.** For the present study, a total of 7 chemicals were selected (Table 1). These chemicals represent a diverse set of substances present in consumer products (including foods and food supplements) and drugs that are relevant for risk assessment. Among these compounds, coumarin, caffeine, phenoxyethanol, and niacinamide have a long history of safe use by humans.<sup>13</sup>



**Figure 1.** Computational workflow overview. First, a concentration–response HTTr data set is used as the input into PLIER along with a collection of predetermined gene sets (step 1). PLIER returns a collection of LVs, which represent patterns of co-regulated genes that may or may not align with any of the gene sets supplied. Next, the concentration–response activity of each LV in each compound is investigated using a new updated version of the ToxCast pipeline (*tcplfit2*).

**Cell Culture.** HepG2 cells (human hepatoblastoma) were obtained from the Public Health England European Collection of Cell Cultures (ECACC, Salisbury, UK). Cells were cultured in complete EMEM supplemented with 10% fetal bovine serum (FBS), 2 mM GlutaMAX, 1% nonessential amino acids, 53 U/mL penicillin, and 53  $\mu$ g/mL streptomycin. HepG2 cells were maintained in 75 cm<sup>2</sup> cell culture flasks in a humidified atmosphere incubator with 5% CO<sub>2</sub> at 37 °C; the cells were kept at a confluence below 85% and were not maintained in culture for more than 4 weeks (8 passages). Cells were seeded into 384-well, clear-bottom black-walled tissue culture plates at a density of 3000 cells/well and were left overnight to attach.

MCF-7 cells (human Caucasian breast adenocarcinoma) were obtained from ECACC (Salisbury, UK). Cells were cultured in complete RPMI 1640 medium supplemented with 10% FBS, 2 mM GlutaMAX, 53 U/mL penicillin, and 53  $\mu$ g/mL streptomycin. MCF-7 cells were seeded into 384-well, clear-bottom black-walled tissue culture plates at a density of 3000 cells/well and were left overnight for attachment.

HepaRG cells were obtained from Life Technologies and cultured in Williams E medium supplemented with 2 mM L-glutamine and HPRG670 supplement (Lonza, UK), in collagen-coated, 384-well, clear-bottom, black-walled, tissue culture plates, at a density of 20,000 cells/well. HepaRG cells were then transferred to serum-free medium following the initial 24 h seeding procedure (Williams E medium supplemented with 2 mM L-glutamine, 100 units/mL penicillin, 100  $\mu$ g/mL streptomycin, and HPRG640 supplement), for 6 days prior to dosing, with media replenishment every second day.

**Cell Viability Assay.** In order to assess cytotoxicity, HepG2, MCF-7, and HepaRG cells were dosed with each compound at a range of concentrations for 24 h. At the end of the incubation period, the cells were loaded with the relevant dye/antibody for the following cell health markers: cell count, nuclear size, DNA structure, and cellular ATP. The plates were then scanned using an automated fluorescent cellular imager, ArrayScan (Thermo Scientific Cellomics). Cytotoxicity thresholds were defined for each cell line according to the lowest minimum effective concentration (MEC) for all the biomarkers measured and were used to determine the top concentrations tested (Table 1 and Supporting Information: File 1, XLSX).

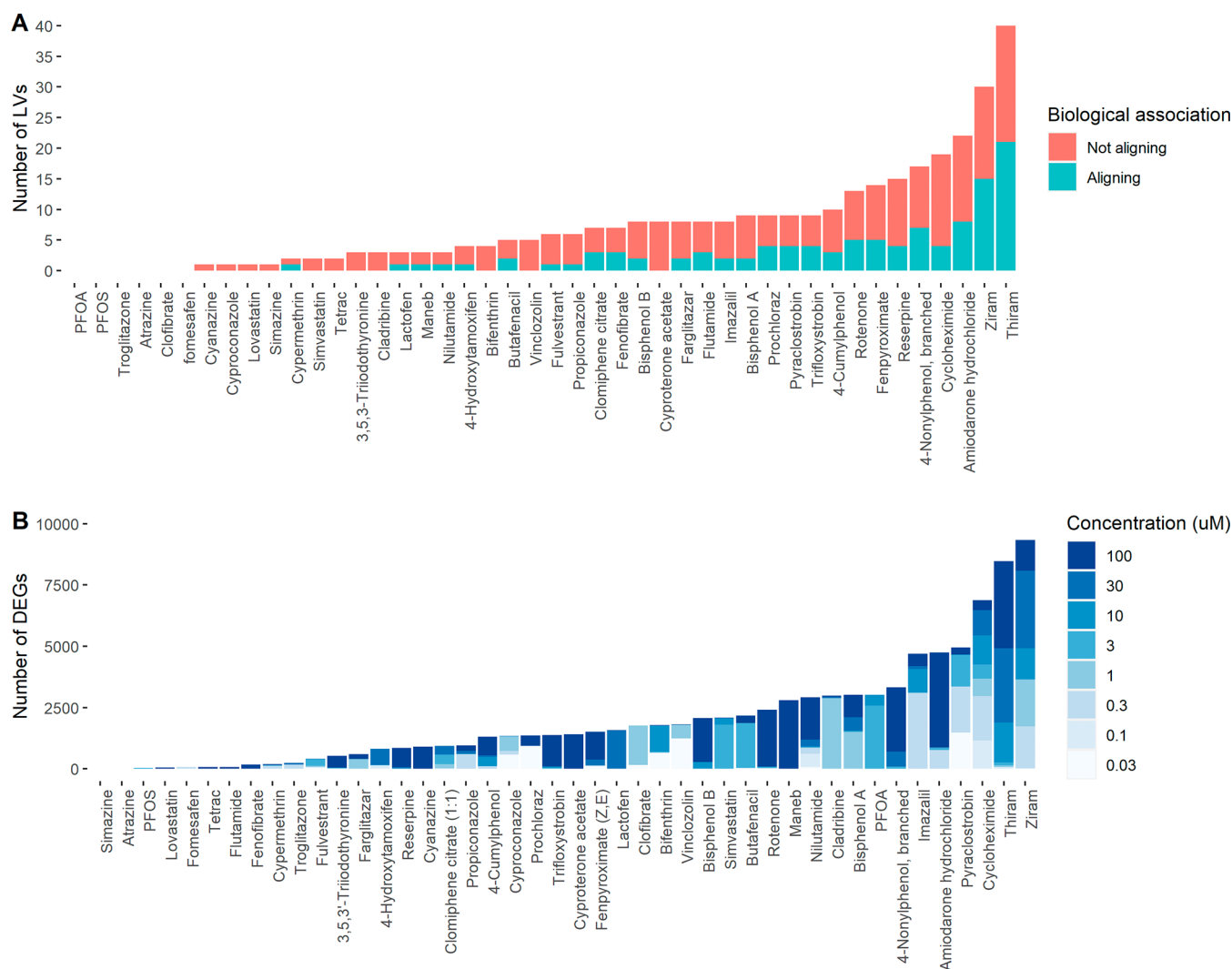
**Chemical Treatments.** Test compounds were prepared as stock solutions in 200 $\times$  higher concentrations than the highest concentration to be tested. [Dimethyl sulfoxide (DMSO) was used as the solvent, and its concentration was maintained at 0.5% v/v.] Serial dilutions were performed using the custom dilution series for each compound. Cells were treated at six concentrations (Table 1) of each test compound, and three biological replicates were generated. Compound treatment was performed for 24 h in a humidified atmosphere with 5% CO<sub>2</sub> at 37 °C. Cells were washed in calcium- and magnesium-free phosphate buffered saline (PBS). With all residual PBS removed, the 2X TempO-Seq lysis buffer (BioSpyder Technologies, proprietary kit) was diluted to 1 $\times$  with PBS and added at a volume of 1  $\mu$ L per 1000 cells with a minimum of 10  $\mu$ L per well and incubated for 10 min at room temperature. Following lysis, the samples were frozen at –80 °C prior to sequencing.

**RNA Sequencing and Data Analysis.** TempO-Seq analysis for the HepG2, MCF-7, and HepaRG studies was performed as described

previously<sup>28</sup> with a targeted sequence depth of 200 mapped read counts per transcript, including the use of the general attenuation panel. Samples with fewer than 500 K reads were removed, leading to the highest concentration of flutamide (200  $\mu$ M) and the third highest concentration of niacinamide (100  $\mu$ M) to be lost in HepaRG. In addition, genes with fewer than 10 counts in more than 90% of the samples were removed. Reproducibility and overall sample quality were assessed by means of PCA and correlation. The second lowest concentration of coumarin (0.01  $\mu$ M) in HepG2 cells was removed due to the low correlation of replicates (Figure S1, Supporting Information). Similar to the MCF-7 study, the data sets were adjusted for batch effects due to plate differences using ComBat-Seq (*ComBat\_seq* function from the SVA package, version 3.40.0).<sup>29</sup> Finally, data were normalized by DESeq2's median of ratios using the *DESeq* function (package DESeq2, version 1.32.0) to account for differences in sequencing depth and RNA composition.<sup>30</sup>

**Gene Aggregation into Latent Variables.** Gene aggregation into latent variables (LVs) was performed using the Pathway-Level Information Extractor (PLIER).<sup>31</sup> An LV is a variable that is inferred using models from observed gene expression data. Briefly, PLIER approximates each gene expression profile as a linear combination of LVs that are similar to eigengenes (e.g., largely independent of one another). Although this deconvolution is driven by prior knowledge where the method optimizes the alignment of LVs to a relevant subset of the available gene sets, some of the LVs do not have any association with biology, allowing for the discovery of activity not captured by the knowledge bases used (Figure 1, step 1). LVs are nonredundant as they capture unique patterns of gene expression; hence, despite the fact that they can align with the same gene set, they will capture the behavior of different genes. It is worth mentioning that the absolute value of an LV has no direct interpretation, while what matters is the value of the LV relative to the other LVs; relatively high values correspond to a subset of correlated genes with relatively high expression within the tissue/cell, while low values highlight patterns of co-regulated genes with a relatively low expression. In the present work, PLIER was run using the *PLIER* function from the PLIER package (version 0.99.0) within the R statistical environment, with default parameters providing as an input the fully normalized data set for each cell line (both the public and in-house data) and using as prior knowledge the MSigDB (sub)collections C2 and H (version 7.2) as these could be mapped to pathways that can be interpreted in meaningful ways. For the MCF-7 public data, we also included the full collection of directional CMAP signature as calculated and used by Harrill et al.<sup>11</sup> in order to perform a more robust comparison, which we excluded from the analysis of the in-house data set as there is uncertainty around its ability to facilitate biological interpretation. The method outputs a matrix of LVs across samples along with a matrix providing the accuracy of the alignment between LVs and predefined gene sets. Gene sets aligning with LVs showing FDR < 0.05 and AUC > 0.7 were considered.

**Concentration–Response Analysis of LVs.** To investigate the activity of LVs for each compound and estimate PoDs of biological pathways, we leveraged the power of the ToxCast pipeline (*tcpl*)<sup>32</sup> using a new updated version, *tcplfit2*<sup>33</sup> (<https://github.com/cran/tcplfit2>) (Figure 1, step 2). Briefly, for each compound, the concentration–response expression profile of each LV is fit to a set

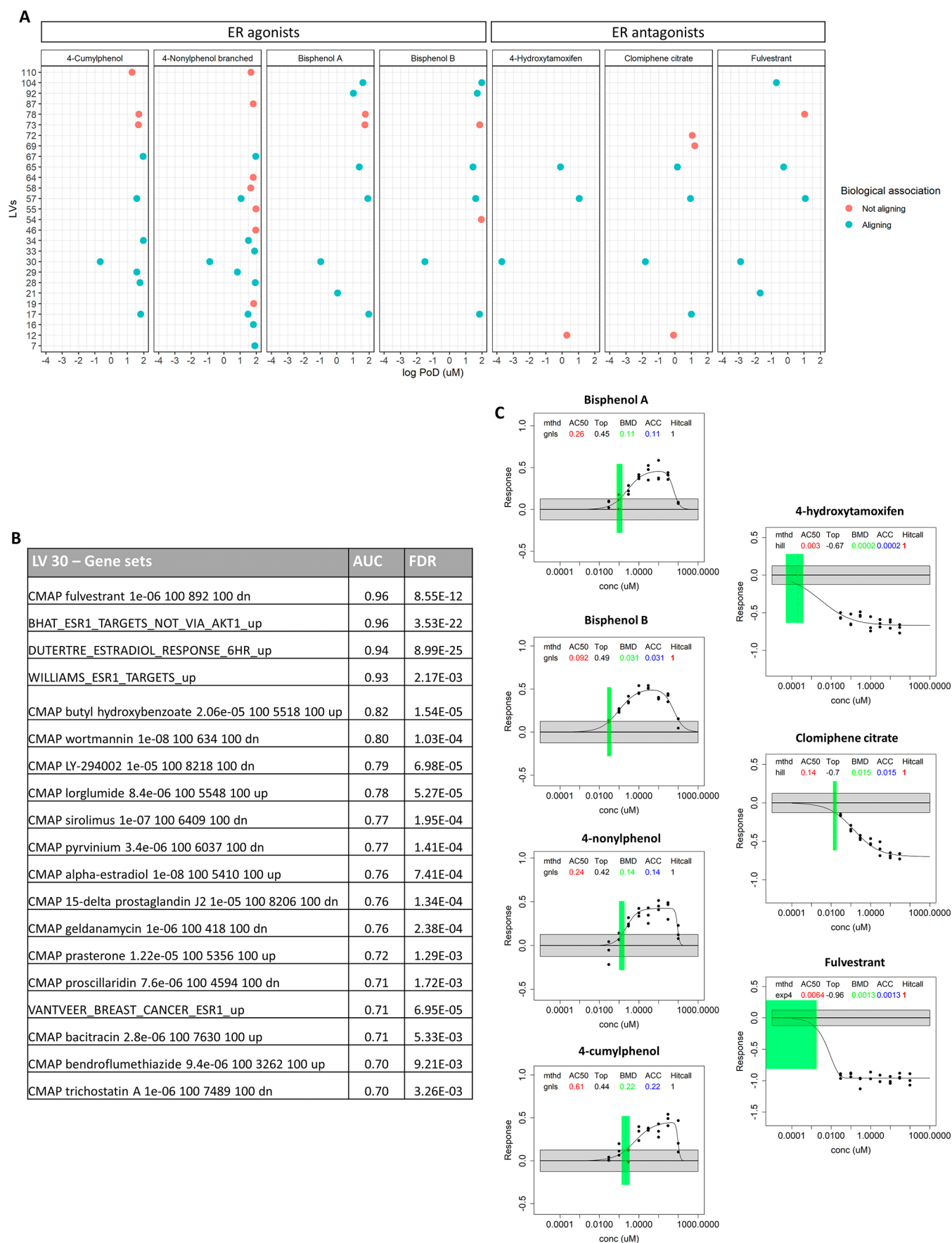


**Figure 2.** Transcriptional activity after compound application. (A) Magnitude of activity in terms of active LVs found by the workflow developed in this work. In cyan, we report those LVs that are found to align with prior knowledge and in red the ones that are found to not align with prior knowledge. (B) Magnitude of perturbed genes across concentrations as identified by Harrill et al.<sup>11</sup>

of models, including the Hill model, a gain–loss model, two polynomial models, a power model, and four exponential models. The LV expression profiles of the control samples are used to determine the baseline expression (noise band). The winning curve fit is selected as the model with the lowest Akaike Information Criterion (AIC).<sup>34</sup> One of the key outputs of *tcp1fit2* over *tcp1* is a “continuous Hitcall” that quantifies the strength of the hits providing confidence for the fitted model. The criteria considered for the calculation of the continuous Hitcall and for the estimation of the benchmark dose (BMD) can be found in the previous study.<sup>11</sup> The only modifications made were that for an active call we expected (1) at least one median response at any test concentration to be greater than 1.5 times the statistically defined baseline expression (noise band) and (2) at least the benchmark dose lower (BMDL) or the benchmark dose upper (BMDU) to be successfully estimated. The BMD value is the PoD, the concentration at which the winning model curve crosses the cutoff, which is set to 2 times the standard deviation (s.d.) of the controls. Models with a continuous Hitcall  $\geq 0.9$  were considered. The compound PoDs from the Harrill et al.<sup>11</sup> work were computed by first ranking all the active signatures in each compound by their PoD and then considering the lowest 5th percentile. PLIER-derived compound’s PoDs were derived by considering the LV with the lowest PoD (provided in Table S2, Supporting Information). For compounds lacking LV activity, no global PoD was assigned.

**Benchmark Dose Modeling with BMDExpress2.** For each data set, normalized counts were used as the input into BMDExpress2 (version 2.2).<sup>35</sup> Hill, power, polynomial 2nd degree (poly 2) and exponential 3rd, 4th, and 5th degree models were all fit to the data. Only probes that passed a William’s Trend test using thresholds of 1.5-fold change and a 0.05 *p* value had models fit to them and the BMDL values calculated. Out of the 6 models fit to the probe data, the best model was determined to be the one with the lowest AIC value, and BMDLs were deemed significant if (1) the BMD was less than or equal to the highest test concentration, (2) the BMDU/BMDL ratio was less than 40, and (3) the model fit *p* values for the best model were more than 0.1. Pathway enrichment was performed using BMDExpress2 by taking all probes analyzed and mapping them to Reactome pathways.<sup>36</sup> A mean of the BMDLs of probes mapped to pathways that were found to have significant dose responses and BMDL values was deduced as the pathway-level PoD. Pathways were deemed as significantly enriched when they had (1) a 2-tailed Fisher’s *p* value less than 0.1, (2) over 2 probes in the input data set found in the pathway, and (3) one or more probes in the pathway that passed the previously listed probe significance criteria. For each chemical cell line data set, the pathway with the lowest BMDL was determined as the compound PoD. The estimation of a global PoD failed for chemicals for which not enough genes had a significant dose responsive model fit to them to significantly enrich any Reactome





**Figure 3.** Concentration-dependent activity of LVs for estrogenic (bisphenol A, bisphenol B, 4-nonylphenol, and 4-cumylphenol) and antiestrogenic (4-hydroxytamoxifen, clomiphene citrate, and fulvestrant) compounds. (A) The plots show LVs found to be active in each

Figure 3. continued

compound along with their PoD across the dose-range tested. The presence or absence of association with biology is reported in cyan and red, respectively. (B) Gene sets aligning with LV 30, which was found to be active in all the estrogenic and antiestrogenic compounds along with the AUC and the FDR set at 0.7 and 0.05, respectively. (C) Concentration–response models for LV30 in response to ER agonists and ER antagonists. For each panel, the noise band is represented by the gray band spanning zero, while the estimated PoD (green line) is reported along with its confidence intervals (green box). mthd = best-fit concentration–response model; Hitcall = confidence for the fitted model; top = top parameter for the fitted model.

pathway. Global PoDs obtained using this method are provided in Table S2, [Supporting Information](#).

**Bayesian Concentration–Response Analysis.** A modification of the approach developed by Reynolds et al.<sup>37</sup> was used to estimate global PoDs for the chemicals andrographolide, caffeine, coumarin, flutamide, niacinamide, phenoxyethanol, and triclosan in HepaRG, HepG2, and MCF-7 cell lines. Briefly, the approach fits a Bayesian statistical model to the counts for a single gene. The posterior distribution was used to derive a distribution encompassing the uncertainty in the estimate of a gene-level PoD. This was repeated for all genes with a mean or median raw count greater than 5. A global PoD was derived using the method outlined earlier.<sup>37</sup> The statistical model was modified by switching the sampling distribution of counts from a negative binomial to a log-Student's *t*-Poisson compound distribution to increase kurtosis, thereby reducing sensitivity to outliers. PoD samples for the Bayesian concentration–response method were obtained using the Hamiltonian Monte Carlo algorithm implemented in PyStan 2.19.<sup>38</sup> Data processing was performed with Python 3.8, utilizing the packages matplotlib 3.3, NumPy 1.19, pandas 1.2, and SciPy 1.6. Global PoDs obtained using this method are provided in Table S2, [Supporting Information](#).

**Concordance Correlation Analysis.** To assess the concordance between PoD estimates across methods, we used the concordance correlation coefficient (CCC), an index that measures the correlation between variables with the ability to capture bias and under-/overdispersion in the estimates.<sup>39</sup> The CCC was computed by using the CCC function from the DescTools R package. PoDs of estrogen activity from the Harrill et al.<sup>11</sup> work were computed by first ranking all the active estrogen-related signatures in each ER agonist and ER antagonist by their PoD and then considering the lowest 5th percentile.

## RESULTS

### LV Activity Reflects the Known Mechanism of Action.

We first assessed the ability of our computational workflow to identify known mechanisms of toxicity by leveraging the information covered in the data set developed by Harrill and collaborators.<sup>11</sup> Results obtained with the PLIER workflow (Figure 2A) displayed thiram, ziram, amiodarone hydrochloride, cycloheximide, and 4-nonylphenol to be among the compounds with a wider spectrum of bioactivity, being able to modulate 40, 30, 22, 19, and 17 LVs, respectively. It is interesting to note that compounds with the highest number of active LVs also had the highest number of perturbed genes as identified by Harrill and collaborators<sup>11</sup> (Figure 2B). The complete list of active LVs found for each compound and their biological association are provided in [Supporting Information: File 1 \(XLSX\)](#).

Herbicides were found to have low activity, with just a few LVs found to be modulated in some of the compounds (Figure S2, [Supporting Information](#)), and this was expected given the absence of the chemical class targets in the MCF-7 cells. On the other hand, chemicals with a specific mode of action (MoA), as estrogenic and antiestrogenic compounds, were able to modulate a higher number of LVs (Figure 3A and [Supporting Information: File 1, XLSX](#)). While fulvestrant, 4-hydroxytamoxifen, and clomiphene citrate had most of the

PoDs occurring in the range 0–20  $\mu\text{M}$ , 4-cumylphenol, 4-nonylphenol, bisphenol A, and bisphenol B were found to have PoDs up to 100  $\mu\text{M}$ . Interestingly, we found both estrogenic and antiestrogenic compounds to share activity for LV 30, which had the lowest PoDs across all the compounds (Figure 3A). This LV was found to be associated with estrogen receptor gene sets and some CMAP signatures underlying estrogen-related activity (Figure 3B). We hypothesized this LV was able to capture specific MoA for compounds targeting the estrogen receptor, which was supported by the concentration–response data showing effects in the opposite direction for ER agonists and ER antagonists (Figure 3C). Of note, estrogenic activity was found to have PoDs at or below 0.1  $\mu\text{M}$  for all the agonists and antagonists tested, highlighting a high potency. In addition, the LV30 PoDs were found to be in concordance with the PoDs of estrogenic activity found by Harrill et al.<sup>11</sup> (CCC = 0.98 with CI [0.96, 0.99]), highlighting the effectiveness of the present workflow to estimate relevant PoDs while easing biological interpretation as redundant gene sets are captured by a single LV associated with a single PoD.

To assess the specificity of the method, we explored whether the LV30 capturing estrogen-related activity was also found to be active in compounds which are not well-established ER agonists/antagonists. Indeed, we could identify activity for LV30 in another 17 chemicals ([Supporting Information File 1](#)) of which five had a PoD below 0.5  $\mu\text{M}$  (Table 2). It is

**Table 2. Additional Compounds Targeting Estrogen Signaling<sup>a</sup>**

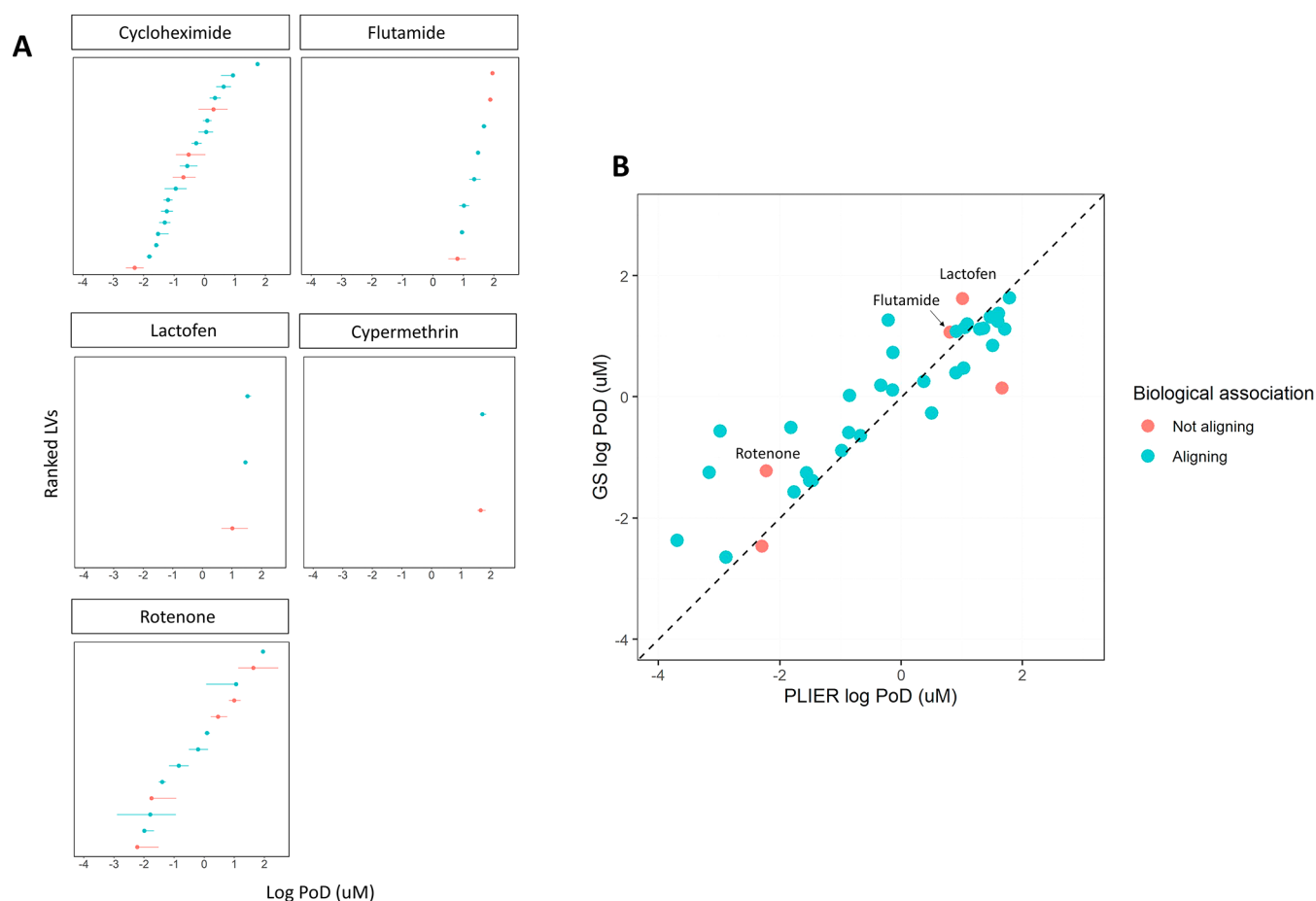
| compound               | PoD ( $\mu\text{M}$ ) | target annotation   | ER reference |
|------------------------|-----------------------|---|--------------|
| fenpyroximate          | 0.007                 | mitochondrial electron transport inhibitor                  | 40           |
| rotenone               | 0.02                  | mitochondria (complex I inhibitor)                          | 40, 41       |
| tetrac                 | 0.03                  | T4 synthesis inhibitor                                      | 42           |
| ziram                  | 0.04                  | inhibition of metal-dependent and sulfhydryl enzyme systems | 43           |
| 3-5-3-triiodothyronine | 0.41                  | thyroid hormone receptor agonist                            | 42           |

<sup>a</sup>Compounds that are not well-established ER agonists/antagonists eliciting LV30 with underlying estrogenic/antiestrogenic activity and whose PoDs fall within the three lowest concentrations.

interesting to note that while these compounds are not well-established ER agonists/antagonists, their ability to modulate estrogen signaling, either directly or indirectly, has been already demonstrated (see Table 2 for citations).

### Activity of LVs Not Aligned to Biological Knowledge.

One of the advantages of PLIER is its ability to identify patterns of co-regulated genes in a data-driven fashion that do not necessarily need to align with predefined gene sets, allowing the discovery of biological activity not captured by current knowledge bases, but whose activity may be of concern



**Figure 4.** PoD estimates of LVs not aligning with prior knowledge. (A) Distribution of PoDs for compounds having the lowest LV not associated with prior knowledge. The plots display the distribution of the different PoDs (each dot, ordered from lowest to highest) along with their confidence intervals, estimated using the approach developed in this study. PoDs are colored in cyan or red depending on whether they have an association with existing gene sets (aligning) or not (not aligning), respectively. (B) The scatterplot shows the agreement between PoD estimates across the two methods: the gene signature (GS) and PLIER. Chemicals are colored depending on whether the LV with the lowest PoD aligns (cyan) or not (red) with prior knowledge.

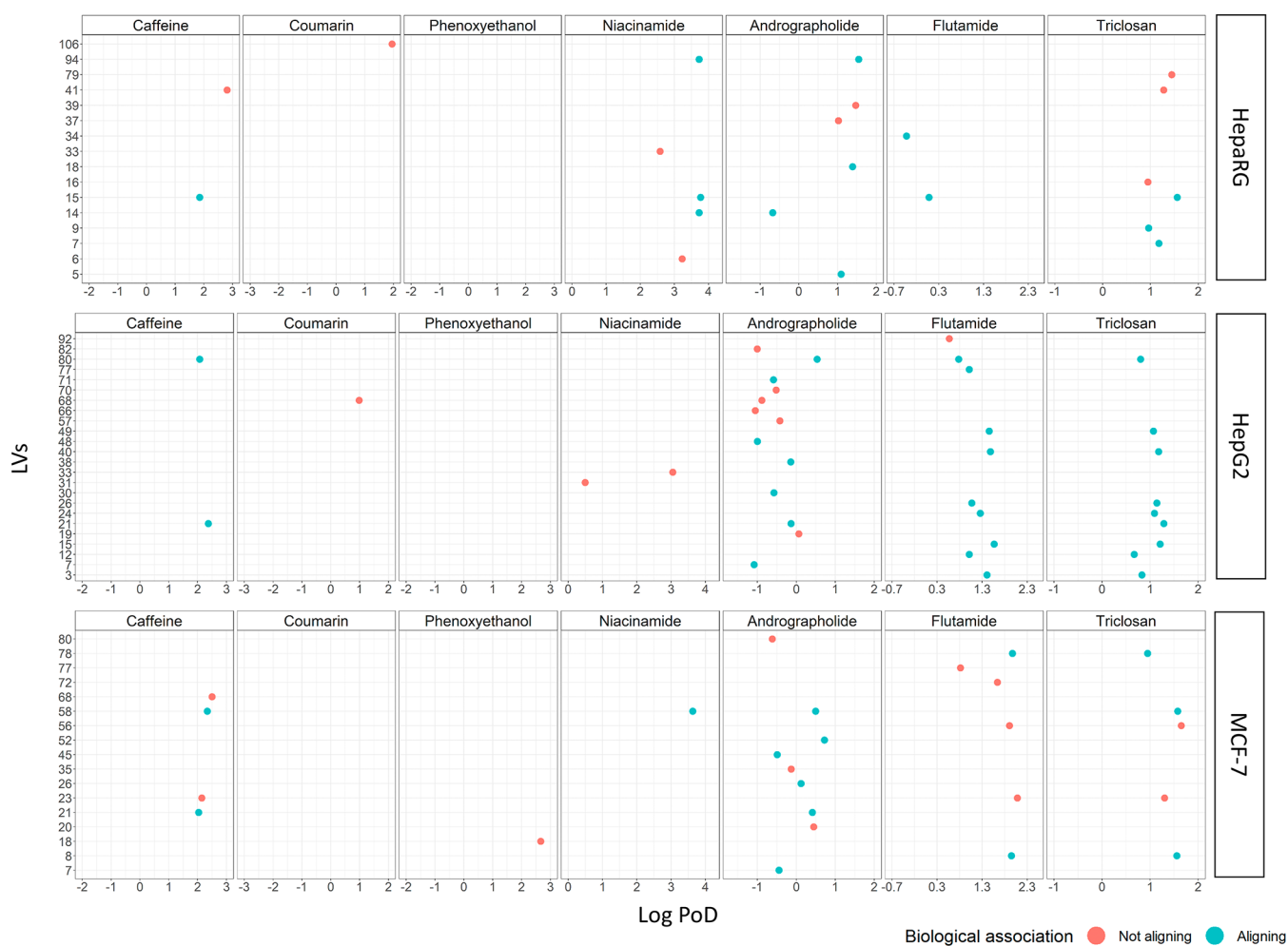
in a risk assessment scenario. Our analysis revealed the presence of concentration–responsive LVs not aligned with any of the gene sets present in the prior knowledge supplied, as shown in Figure 2. For cycloheximide, flutamide, lactofen, cypermethrin, and rotenone, we could identify active LVs lacking an association with prior knowledge to have the lowest PoDs among all the active LVs identified (Figure 4A). This suggests that those LVs capture biological activity not present in the knowledge base used as prior knowledge. An inspection of the concentration–response profiles of these LVs revealed that cycloheximide (LV 12), flutamide (LV 69), lactofen (LV 69), cypermethrin (LV 69), and rotenone (LV 12) displayed potential biological activity that may be important for the risk assessment (Figure S3, Supporting Information).

Interestingly, rotenone, lactofen, and flutamide LVs had a PoD lower than the PoD of any of the concentration–responsive gene signatures, highlighting the ability of PLIER to identify activity that could be potentially missed by the gene-signature approach developed by Harrill et al.<sup>11</sup> (Figure 4B). More precisely, the activity of LVs not aligning with any predetermined gene set may be driven by probes that, despite modulated as a result of chemical exposure, are discarded by the gene-signature approach because they do not belong to any of the gene sets considered. Nevertheless, a good overall

agreement of PoD estimates was found between the two methods (CCC = 0.85 with CI [0.75, 0.91]).

**Characterizing the Bioactivity of Compounds Profiled across Cell Lines.** We next applied the developed concentration–response modeling approach to a set of seven compounds present in consumer products (including foods and food supplements) and drugs that are relevant for risk assessment and applied it to HepG2, HepaRG, and MCF-7 cells, which is a novel data set generated in the context of the current study.

Coumarin, niacinamide, caffeine, and phenoxyethanol were all found to modulate none or a small number of LVs across all cell lines whose PoDs occurred at high concentrations (Figure 5). The small spectrum of activities detected along with the highest PoDs of the LVs found to be active highlights the low potency of these chemicals in this part of bioactivity space for this class of consumer good compounds, a finding already described by Judson et al.<sup>44</sup> On the other hand, triclosan, flutamide, and andrographolide were all found to elicit a higher number of concentration–responsive LVs, suggesting the modulation of a wider spectrum of biological processes (Figure 5). While PoDs for both triclosan and flutamide were mostly close to the highest tested concentrations (at 20/50 and 50/200  $\mu\text{M}$ , respectively), activity for andrographolide was found



**Figure 5.** LV activity. The plot shows LVs found to be active in each compound across the different cell lines. The y-axis reports the different LVs found to be active, while the x-axis shows the concentration range tested in the log scale. The presence or absence of association with biology is reported in cyan and red, respectively.

even at the lowest concentration tested ( $0.2 \mu\text{M}$ ). The list of LVs found to be active for each compound in each cell line, along with their biological association, is provided in [Supporting Information: File 1 \(XLSX\)](#).

Next, we set to further explore the transcriptional activity of andrographolide, flutamide, and triclosan in order to assess whether active LVs captured some known MoA for these compounds.

Andrographolide was found to trigger the widest spectrum of activities. Interestingly, the LVs with the lowest PoDs across all the cell lines were all found to modulate cell stress adaptive responses ([Figure 6](#)). LVs 7 and 48 in HepG2 were found to be associated with the activity at the level of the mitochondria, a well-known target organelle for this compound.<sup>45</sup> The activity of these LVs was driven by genes taking part in the electron transport chain as subunits of cytochrome *c* oxidase (*cox6b1*, *cox7a2*, *cox7a2l*, *cox7b*, and *cox7c*) and subunits of NADH dehydrogenase ubiquinone (*ndufa2*, *ndufa4*, *ndufa13*, *ndufab1*, *ndufb7*, and *ndufc1*) as well as inorganic pyrophosphatase 2 (*ppa2*), which is essential for the correct regulation of mitochondrial membrane potential and mitochondrial organization and function<sup>46,47</sup> ([Figure S4, Supporting Information](#)). In MCF-7, LVs 7 and 80 had the lowest PoDs, and while LV 80 was found not to align with any biological knowledge, LV 7 was found to be associated with oxidative

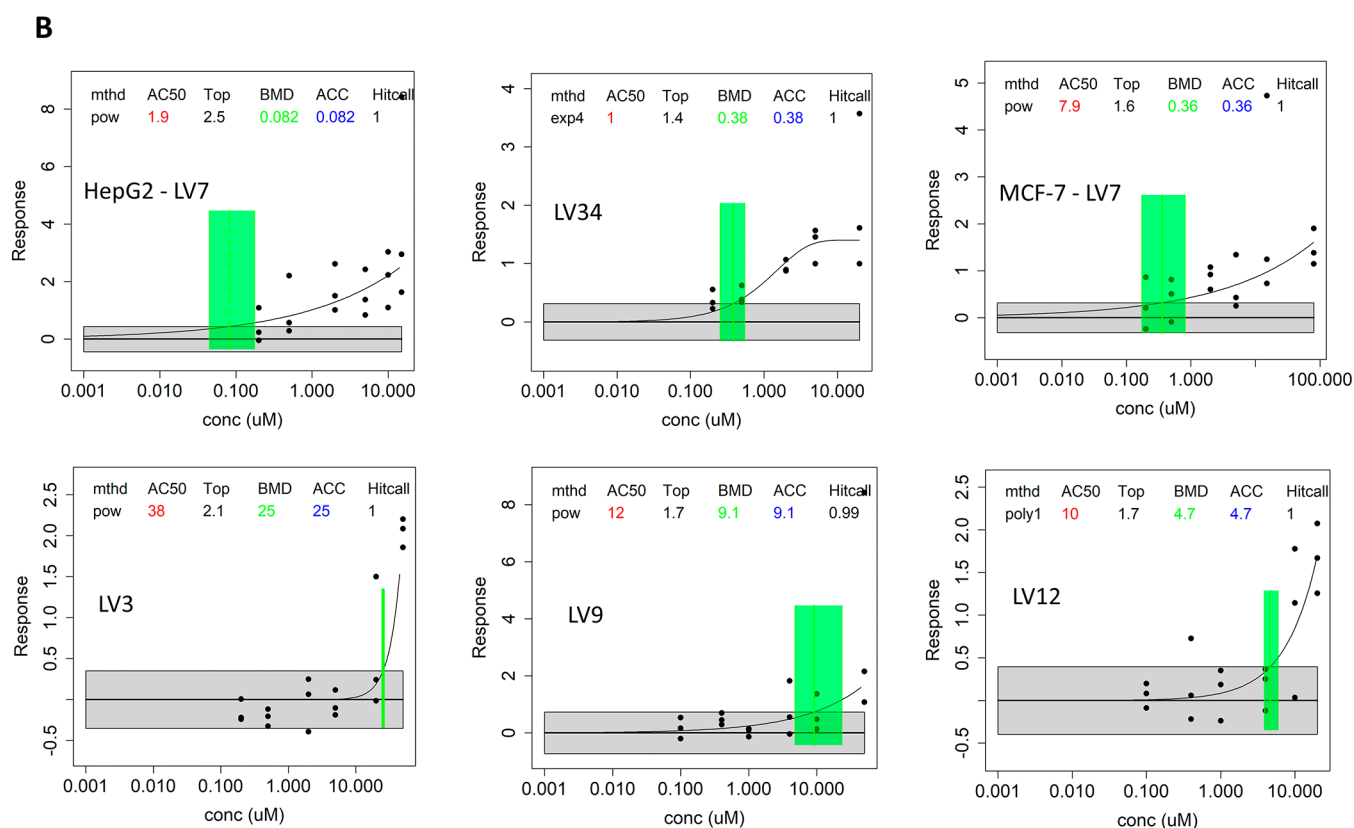
stress response and, specifically, with the ability of this compound to augment antioxidant defense<sup>48</sup> ([Figure 6](#)). LV 14 in HepaRG was instead found to align with hypoxia, an association supported by the ability of this compound to inhibit the hypoxia-inducible factor 1  $\alpha$  (*HIF-1 $\alpha$* ).<sup>49,50</sup>

*Flutamide* was found to trigger LVs with high PoDs in HepG2 and MCF-7 cells, while in HepaRG, we detected a smaller spectrum of activity with only two active LVs identified (LV 15 and 34) and whose PoDs occurred in the range of the two lowest concentrations tested ( $0.2$  and  $0.5 \mu\text{M}$ ). LV 34 had the lowest PoD and was found to align with functions including steroid biosynthesis, oxidative phosphorylation, and xenobiotic response ([Figure 6](#) and [Supporting Information: File 1, XLSX](#)), well-known activities found to be modulated by this compound.<sup>51–53</sup> The LV with the lowest PoDs in HepG2 was LV 92, which did not align with any predefined gene set ([Figure S5, Supporting Information](#)). LVs 3, 15, and 80 were found to represent biological processes, including hypoxia, endoplasmic reticulum (ER) stress, and immune response ([Figure 6](#) and [Supporting Information: File 1, XLSX](#)), the well-documented adverse effects of this compound,<sup>54–56</sup> which all occurred at the onset of cytotoxicity (MEC =  $4.62 \mu\text{M}$ , [Supporting Information: File 1, XLSX](#)). No sub-cytotoxic flutamide effects were identified in MCF-7 as all the active LVs were found to occur at the onset of cytotoxicity (MEC =  $3.03$



**A**

| Cell Line | Treatment               | LV | Gene Set   | AUC  | FDR    |
|-----------|-------------------------|----|--|------|--------|
| HepG2     | Andrographolide         | 7  | mootha_voxphos   | 0.74 | 1.1e-2 |
|           |                         |    | reactome_respiratory_electron_transport                              | 0.72 | 1.0e-2 |
|           |                         |    | wp_electron_transport_chain_oxphos_system_in_mitochondria            | 0.71 | 9.3e-3 |
|           |                         | 48 | Kegg_oxidative_phosphorylation                                       | 0.74 | 4.6e-3 |
|           | Flutamide               | 3  | pid_hif1_tfpathway   | 0.74 | 1.9e-2 |
| HepaRG    | Flutamide/<br>Triclosan | 15 | reactome_srp_dependent_cotranslational_protein_targeting_to_membrane | 0.71 | 1.7e-3 |
|           |                         |    | mori_immature_b_lymphocyte_dn  | 0.93 | 7.9e-9 |
|           | Triclosan               | 12 | gavin_foxp3_targets_cluster_p7                                       | 0.71 | 4.2e-2 |
|           |                         |    | elvidge_hif1a_and_hif2a_targets_dn                                   | 0.82 | 2.8e-5 |
| MCF-7     | Andrographolide         | 7  | kegg_steroid_hormone_biosynthesis                                    | 0.75 | 2.9e-2 |
|           |                         |    | reactome_xenobiotics   | 0.99 | 4.4e-3 |
|           |                         |    | moserle_ifna_response  | 0.99 | 9.6e-4 |
|           | Triclosan               | 9  | wp_nrf2_pathway  | 0.71 | 1.5e-2 |
|           |                         |    | fardin_hypoxia_11  | 0.99 | 8.7e-5 |

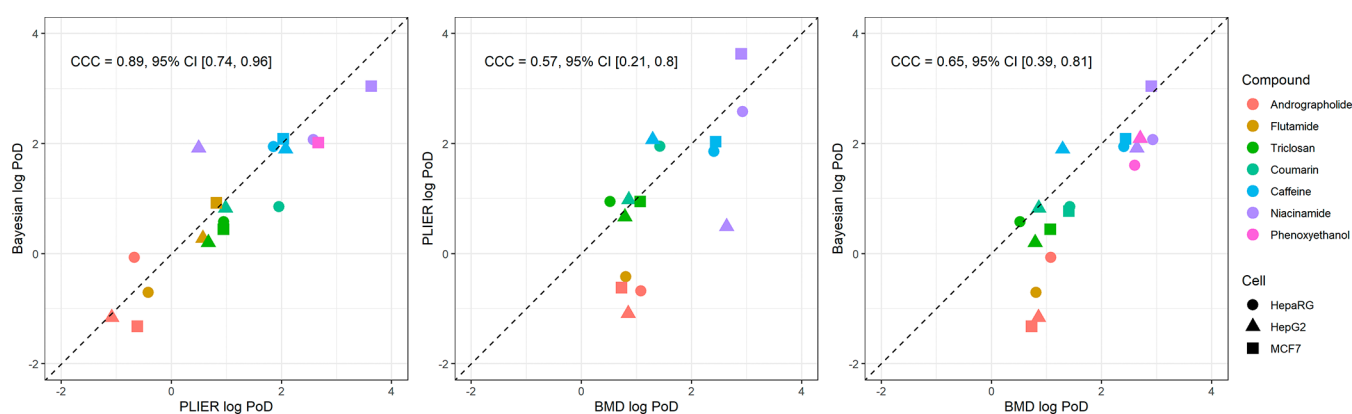


**Figure 6.** Summary of the biological activity for andrographolide, flutamide, and triclosan. (A) The table displays LVs found to be active in each treatment/cell line along with the AUC and the FDR. (B) The concentration–response plots show example activities for some of the LVs listed in (A), modeled using tcplfit2. For each plot, the noise band is represented by the gray band spanning zero, while the estimated PoD (green line) is reported along with its confidence intervals (green box). mthd = best-fit concentration–response model; Hitcall = confidence for the fitted model; top = top parameter for the fitted model.

$\mu\text{M}$ , Supporting Information: File 1, XLSX). This difference in potency between cell lines can be explained by the fact that HepaRG is the most metabolically competent cell line and the toxicity of flutamide is mediated by metabolites produced by members of the cytochrome P450 family;<sup>57,58</sup> indeed, the LV

with the lowest PoD in HepaRG for flutamide (LV 34) was found to be associated with the “REACTOME\_XENOBOTICS”, which represents metabolic activity (Figure 6).

Triclosan was found to trigger a similar number of LVs across all cell lines (Figure 5), and its PoDs mostly occurred at



**Figure 7.** Agreement between PoD estimates across methods. The CCC is reported along with the CI.

the highest concentrations tested ( $>10 \mu\text{M}$ ). Interestingly, the LVs with the lowest PoDs aligning with a predetermined gene set, across all cell lines, were all found to be linked to inflammatory responses with HepaRG (LV 9) and MCF-7 (LV 78), which align with gene sets characterized by high levels of chemokines and interleukins,<sup>59</sup> and HepG2 (LV 12), which align with a gene set of FOXP3 targets underlying Tr cell homeostasis<sup>60</sup> (Figure 6). The ability of triclosan to modulate the inflammatory response in HepG2 is in agreement with a previous work where we observed pro-inflammatory responses at similar concentrations.<sup>13</sup> In addition, LV 15 in HepG2 was found to describe ER stress, a known target of triclosan<sup>61</sup> whose PoD ( $16.1 \mu\text{M}$ ) coincided with the onset of cytotoxicity (MEC =  $13.9 \mu\text{M}$ , Supporting Information: File 1, XLSX).

Overall, we can conclude that these chemicals show similar potency across cell lines with major differences driven by metabolic competence as in the case of HepaRG.

**Assessing PoD Estimate Concordance across Methods.** We next explored the agreement between PoDs estimated with our workflow and those estimated by using a well-established method as BMDEExpress2<sup>35</sup> and a recently developed Bayesian approach,<sup>37</sup> the results of which are shown in Figure 7. An overall good agreement was found between the PLIER-derived global PoDs and the Bayesian ones (CCC = 0.89 with CI [0.74, 0.96]), while the agreement of both methods with the BMD estimates was found to be lower (Bayesian CCC = 0.65 with CI [0.39, 0.81]; PLIER CCC = 0.57 with CI [0.21, 0.8]).

When considering the correlation between Bayesian and PLIER-derived PoDs, it is interesting to note how compound PoDs tend to cluster close to each other regardless of the cell type, with the exception of flutamide, where its PoD in HepaRG is  $\sim 1$ -fold change lower due to the fact that the toxicity of flutamide is often mediated by metabolites produced by members of the cytochrome P450 family,<sup>57,58</sup> and niacinamide, where the PLIER PoD in HepG2 is influenced by the activity of LV 31, which does not align with any of the predetermined gene sets (Figure S6, Supporting Information). On the other hand, the lower concordance of both methods with BMDEExpress2 estimates can be partially explained by two different reasons. First is the fact that pathway-level PoDs from BMDEExpress2 are an average over gene-level PoDs in a gene set, and hence for high potency chemicals, many gene-level PoDs correspond with the onset of cytotoxicity, dragging the average further away from the minimum effect concentration. Second, the different compendium of biological knowledge

used to derive PoDs between the two methods, which for BMDEExpress2 took into account only the Reactome database.

These results highlight a good overall agreement between the PLIER and Bayesian estimates, supporting the use of the present workflow for PoD estimation in addition to an easier biological interpretation as the method scales down to just few LVs to be interpreted.

## DISCUSSION

High-throughput profiling technologies allow the measurement of thousands of genes across multiple conditions in a single experiment. While the high dimensionality of these experiments allows scientists to fully explore the landscape of transcripts in a given condition, it also poses challenges of extracting meaningful biological insights. Approaches aggregating genes into biologically relevant sets prior to downstream analysis partially addresses this issue,<sup>23</sup> however, they still leave scientists to focus on thousands of gene sets, each associated with a different PoD estimate, leading to challenges in interpretation. Moreover, they are completely constrained by the curated gene signatures and do not enable the identification of biological activity that is not yet well-characterized.

Here, we applied a matrix factorization approach (PLIER) to aggregate genes into LVs in a data-driven way that also incorporates predefined gene sets. This methodology offers some important advantages over the gene-aggregation methods that have been so far employed in concentration–response studies: (1) it allows the discovery of nonspecific biology of potential concern as not all the latent variables inferred are associated with existing gene sets and (2) it eases biological interpretation as multiple gene sets may align with the same LV whose activity is described by a single PoD.

Our results demonstrate that our approach is able to capture the activity of known biological functions despite the fact that we have reduced the dimensionality of the problem substantially, as in the case of chemicals targeting the estrogen receptor (ER) (Figure 3). In agreement with the results obtained in Harrill et al.,<sup>11</sup> we were able to identify an LV whose gene loadings aligned with gene sets related to ER activity and CMAP signatures underlying estradiol-related functions. As expected, this LV was found to respond in opposite directions for ER agonists and antagonists. Estimated PoDs for this LV were found to occur at low test concentrations, as found in Harrill et al., highlighting a high potency, and showed a high degree of concordance across

methods. In addition, the method also identified estrogenic activity in compounds that are not well-established ER agonists/antagonists but for which evidence in literature was available, showing the potential of the method to be applied in read-across (Table 1).

Interestingly, our approach also identified some activity that did not align with any of the gene sets supplied as prior knowledge, highlighting the presence of potential biological activity that may have been missed by the gene-signature scoring approach that focused only on known gene sets (Figure 4). The activity not aligning with predefined gene sets can be relevant to investigating a chemical's MoA and the associated adverse events, especially when the associated PoD is the lowest among all the estimated activities. Indeed, our results revealed that cycloheximide, flutamide, lactofen, rotenone, and cypermethrin had LV activity not aligned with prior knowledge whose PoDs were the smallest among all the LVs identified in each compound (Figure 4). More importantly, flutamide, rotenone, and lactofen had a global PoD estimate lower than the one estimated by Harrill et al.<sup>11</sup> derived using only existing gene sets,<sup>11</sup> highlighting the ability of the present method to identify the activity that may be missed by approaches where PoDs are estimated from a set of predefined gene signatures (covering the known biology). Investigating this transcriptional activity to fully understand whether there is any activity that is relevant for use in risk assessment is challenging, and integrating this approach with more specific assays to investigate cellular stress may represent a suitable strategy.<sup>13</sup>

We found that the workflow developed in this work was also successful in characterizing and differentiating the toxicity profile of low-risk compounds from those with higher bioactivity, when applied to a small newly generated data set obtained by screening HepG2, MCF-7, and HepaRG cells in concentration–response. Indeed, coumarin, caffeine, niacinamide, and phenoxyethanol were found to modulate none or a small number of active LVs, confirming the low bioactivity of these compounds in the concentration–response range tested. On the other hand, andrographolide, flutamide, and triclosan were found to trigger a higher number of LVs, highlighting their ability to affect a wider spectrum of biological activities, and some of the LVs were able to capture known adverse events. As an example, the lowest andrographolide PoDs across all the cell lines were found to be represented by LVs describing the known cell stress activity for this compound, including mitochondrial dysfunction,<sup>45,62</sup> oxidative stress,<sup>48</sup> and hypoxia.<sup>49,50</sup>

We demonstrate that this approach is successful in disentangling the biological activities underlying chemical toxicity (including the known MoA). Moreover, one of the primary advantages of the method is its ability to ease biological interpretation as a single LV can capture multiple redundant gene sets describing similar biological activities, biological pathways with significant cross talk, and pathways converging on similar transcriptional patterns. Each LV is associated with a single PoD estimate, reducing the effort needed to interpret thousands of gene sets with individual PoDs. This was the case of the ER agonists/antagonists in the MCF-7 public data set, where a single LV (LV 30) was found to represent estrogenic activity, and many similar gene sets describing estrogen-related activity were found to align with this LV. Moreover, by exploring the gene contribution to each LV, it is also possible to identify drivers of the associated biological activity, making PLIER a useful approach for

biomarker discovery, as highlighted for mitochondrial activity in andrographolide.

The fact that our workflow did not capture some of the known biology may depend on multiple factors. First, because PLIER is a dimensionality reduction technique, there is a chance that weaker signals are lost during the decomposition. Second, the ability to capture the activity of genes or pathways strongly depends on the cutoff used in concentration–response modeling. In the present study, we used  $2\times$  s.d. of the controls to define the cutoff as this value is stringent enough to provide high confidence that an LV is truly active. Other recently developed pipelines for HTTr concentration–response modeling drove PoD estimation by considering 1 s.d. as the cutoff<sup>17</sup> or by simulating signature scores in the absence of correlation between fold changes to derive a noise band and the corresponding cutoff.<sup>11</sup> Defining a sensitive and reliable cutoff for concentration–response analysis is still under debate, but consensus needs to be achieved to make results comparable and useful for risk assessment. Third, the ability of the method to identify meaningful biological processes for the different conditions also depends on the degree of coverage of the biological knowledge used as prior knowledge. This means that some of the LV activity we identified that did not align with any of the gene sets used in the prior knowledge may still align with different gene sets curated in knowledge bases not taken into consideration in the analysis. In the present study, we used a collection of gene sets or pathways that provide a good coverage of the biological activity at a cellular level. However, gene sets describing nonspecific stress response still lack in the currently available knowledge base, and manually curating this information represents a key step in further advancing our understanding on how cells respond to toxicity insults. Fourth and last, the ability of PLIER to identify LVs describing the full landscape of biological activity represented in the data set also depends on the heterogeneity of the data set itself. Indeed, the more diversified the experimental condition in the data set, the better the method is able to disentangle the different underlying biological processes. In this context, the limited number of conditions tested in our data set may have reduced the ability of PLIER to capture the biological activity underlying the toxicity profile of the compounds tested. In this regard, applying our method to future studies with bigger and more heterogeneous data (greater chemical space, more time points, and different cell culture media) may help address this issue. As an example, Harrill and collaborators were able to identify some weak activity for troglitazone, while the PLIER workflow developed here failed (Figure 2). We figured out that this was due to the different noise band calculation as we were able to capture LV activity for troglitazone when using  $1\times$  s.d. for the noise band. Another example, most strictly related to the biological aspects, is the partial inability to capture proper stress response activity known to be elicited by most of the compounds. The main reason for this is a current lack of curated information about stress response pathways in the available knowledge bases due to the complex nature of these biological processes. Indeed, one of the greatest challenges in order to develop accurate signatures is to ensure both their sensitivity and specificity,<sup>63</sup> especially due to the cross talk between stress response pathways that produce similar patterns of effector genes, hence limiting their specificity.<sup>64</sup> The inclusion of curated stress response signatures<sup>65</sup> would provide a better and more reliable biological coverage, allowing the present workflow to

successfully identify the underlying chemically adaptive stress response.

When comparing whether PoD estimates calculated with the workflow developed in this work were in agreement with recently developed or more established methods, we found that the PLIER-derived PoDs were generally in agreement with the Bayesian PoDs, while the concordance of both the aforementioned methods with BMD estimates was lower.

Taken together, these results strongly demonstrate the effectiveness of aggregating transcriptional changes into LVs prior to the concentration–response analysis and show the potential of the method to be employed toward the development of a framework with the ability to improve current risk assessment strategies for chemicals using NAMs by allowing the identification of the most biologically relevant PoD.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.chemrestox.1c00444>.

Panel of chemicals screened in Harrill's work; correlation plot of coumarin exposure in HepG2 cells; global PoDs obtained using the Bayesian approach developed by Reynolds et al.; BMDEExpress2 and the workflow developed in this study; LV activity of herbicides in the MCF7 study; and concentration–response plots of activity for the LVs not aligning with predetermined gene sets, for some of the leading genes involved in oxidative phosphorylation and mitochondrial dysfunction, for the flutamide LV 92 in HepG2, and for the niacinamide LV 31 in HepG2 (PDF)

Cell viability assay data, cell viability plots for all chemicals/cell lines, and active LVs along with their biological associations for each data set, chemical, and cell line (XLSX)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Alistair Middleton** – Unilever, Safety and Environmental Assurance Centre (SEAC), Colworth Science Park, Bedfordshire MK44 1LQ, U.K.; Email: [alistair.middleton@unilever.com](mailto:alistair.middleton@unilever.com)

**Andreas Bender** – Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, U.K.; [orcid.org/0000-0002-6683-7546](https://orcid.org/0000-0002-6683-7546); Email: [ab454@cam.ac.uk](mailto:ab454@cam.ac.uk)

### Authors

**Danilo Basili** – Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, U.K.; Unilever, Safety and Environmental Assurance Centre (SEAC), Colworth Science Park, Bedfordshire MK44 1LQ, U.K.; [orcid.org/0000-0003-4830-7848](https://orcid.org/0000-0003-4830-7848)

**Joe Reynolds** – Unilever, Safety and Environmental Assurance Centre (SEAC), Colworth Science Park, Bedfordshire MK44 1LQ, U.K.

**Jade Houghton** – Unilever, Safety and Environmental Assurance Centre (SEAC), Colworth Science Park, Bedfordshire MK44 1LQ, U.K.

**Sophie Malcomber** – Unilever, Safety and Environmental Assurance Centre (SEAC), Colworth Science Park, Bedfordshire MK44 1LQ, U.K.

**Bryant Chambers** – Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711, United States

**Mark Liddell** – Unilever, Safety and Environmental Assurance Centre (SEAC), Colworth Science Park, Bedfordshire MK44 1LQ, U.K.

**Iris Muller** – Unilever, Safety and Environmental Assurance Centre (SEAC), Colworth Science Park, Bedfordshire MK44 1LQ, U.K.

**Andrew White** – Unilever, Safety and Environmental Assurance Centre (SEAC), Colworth Science Park, Bedfordshire MK44 1LQ, U.K.

**Imran Shah** – Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711, United States; [orcid.org/0000-0003-0808-0140](https://orcid.org/0000-0003-0808-0140)

**Logan J. Everett** – Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711, United States

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.chemrestox.1c00444>

### Author Contributions

A.M. and A.B. conceived the project. S.M. supervised the experimental work. D.B. developed the computational workflow. J.H. and J.R. performed the concentration–response analysis using BMDEExpress2 and the Bayesian approach, respectively. D.B., B.C., M.L., J.R., I.M., A.W., L.E., I.S., A.M., and A.B. contributed to the data interpretation and the writing of the manuscript.

### Funding

This work was funded by Unilever PLC.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

Experiments were performed at Cyprotex PLC, and the authors thank Paul Walker and team.

## ■ REFERENCES

- (1) Fischer, I.; Milton, C.; Wallace, H. Toxicity Testing Is Evolving! *Toxicol. Res.* **2020**, *9*, 67–80.
- (2) National Research Council. *Toxicity Testing in the 21st Century*; National Academies Press: Washington, D.C., 2007.
- (3) Zuang, V.; Dura, A.; Asturiol Bofill, D.; Batista Leite, S.; Berggren, E.; Bopp, S.; Carpi, D.; Casati, S.; Coecke, S.; Corvi, R.; Deceuninck, P.; Franco, A.; Gribaldo, L.; Halder, M. E.; Holloway, M.; Kienzler, A.; Langezaal, I.; Madia, F.; Milcamps, A.; Munn, S.; Paimi, A.; Piergiorganni, M.; Pistollato, F.; Price, A.; Prieto Peraita, M. D. P.; Viegas Barroso, J. F.; Wittwehr, C.; Worth, A.; Whelan, M. *Non-Animal Methods in Science and Regulation*, 2020.
- (4) Westmoreland, C.; Carmichael, P.; Dent, M.; Fentem, J.; MacKay, C.; Maxwell, G.; Pease, C.; Reynolds, F. Assuring Safety without Animal Testing: Unilever's Ongoing Research Programme to Deliver Novel Ways to Assure Consumer Safety. *ALTEX* **2010**, *27*, 207–211.
- (5) Carmichael, P.; Davies, M.; Dent, M.; Fentem, J.; Fletcher, S.; Gilmour, N.; MacKay, C.; Maxwell, G.; Merolla, L.; Pease, C.; Reynolds, F.; Westmoreland, C. Non-Animal Approaches for Consumer Safety Risk Assessments: Unilever's Scientific Research Programme. *ATLA, Altern. Lab. Anim.* **2009**, *37*, 595–610.



- (6) Desprez, B.; Dent, M.; Keller, D.; Klaric, M.; Ouédraogo, G.; Cubberley, R.; Duplan, H.; Eilstein, J.; Ellison, C.; Grégoire, S.; Hewitt, N. J.; Jacques-Jamin, C.; Lange, D.; Roe, A.; Rothe, H.; Blaauboer, B. J.; Schepky, A.; Mahony, C. A Strategy for Systemic Toxicity Assessment Based on Non-Animal Approaches: The Cosmetics Europe Long Range Science Strategy Programme. *Toxicol. Vitro* **2018**, *50*, 137–146.
- (7) Thomas, R. S.; Bahadori, T.; Buckley, T. J.; Cowden, J.; Deisenroth, C.; Dionisio, K. L.; Frithsen, J. B.; Grulke, C. M.; Gwinn, M. R.; Harrill, J. A.; Higuchi, M.; Houck, K. A.; Hughes, M. F.; Hunter, E. S.; Isaacs, K. K.; Judson, R. S.; Knudsen, T. B.; Lambert, J. C.; Linnenbrink, M.; Martin, T. M.; Newton, S. R.; Padilla, S.; Patlewicz, G.; Paul-Friedman, K.; Phillips, K. A.; Richard, A. M.; Sams, R.; Shafer, T. J.; Setzer, R. W.; Shah, I.; Simmons, J. E.; Simmons, S. O.; Singh, A.; Sobus, J. R.; Strynar, M.; Swank, A.; Tornero-Valez, R.; Ulrich, E. M.; Villeneuve, D. L.; Wambaugh, J. F.; Wetmore, B. A.; Williams, A. J. The Next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency. *Toxicol. Sci.* **2019**, *169*, 317–332.
- (8) U.S. EPA. *Next Generation Risk Assessment: Incorporation of Recent Advances in Molecular, Computational, and Systems Biology (Final Report)*: Washington, DC, 2014.
- (9) Gannon, A. M.; Moreau, M.; Farmahin, R.; Thomas, R. S.; Barton-Maclaren, T. S.; Nong, A.; Curran, L.; Yauk, C. L. Hexabromocyclododecane (HBCD): A Case Study Applying Tiered Testing for Human Health Risk Assessment. *Food Chem. Toxicol.* **2019**, *131*, 110581.
- (10) Paul Friedman, K.; Gagne, M.; Loo, L.-H.; Karamertzanis, P.; Netzeva, T.; Sobanski, T.; Franzosa, J. A.; Richard, A. M.; Lougee, R. R.; Gissi, A.; Lee, J.-Y. J.; Angrish, M.; Dorne, J. L.; Foster, S.; Raffaele, K.; Bahadori, T.; Gwinn, M. R.; Lambert, J.; Whelan, M.; Rasenberg, M.; Barton-Maclaren, T.; Thomas, R. S. Utility of In Vitro Bioactivity as a Lower Bound Estimate of In Vivo Adverse Effect Levels and in Risk-Based Prioritization. *Toxicol. Sci.* **2020**, *173*, 202–225.
- (11) Harrill, J. A.; Everett, L. J.; Haggard, D. E.; Sheffield, T.; Bundy, J. L.; Willis, C. M.; Thomas, R. S.; Shah, I.; Judson, R. S. High-Throughput Transcriptomics Platform for Screening Environmental Chemicals. *Toxicol. Sci.* **2021**, *181*, 68–89.
- (12) Baltazar, M. T.; Cable, S.; Carmichael, P. L.; Cubberley, R.; Cull, T.; Delagrangue, M.; Dent, M. P.; Hatherell, S.; Houghton, J.; Kukic, P.; Li, H.; Lee, M.-Y.; Malcomber, S.; Middleton, A. M.; Moxon, T. E.; Nathanail, A. V.; Nicol, B.; Pendlington, R.; Reynolds, G.; Reynolds, J.; White, A.; Westmoreland, C. A Next-Generation Risk Assessment Case Study for Coumarin in Cosmetic Products. *Toxicol. Sci.* **2020**, *176*, 236–252.
- (13) Hatherell, S.; Baltazar, M. T.; Reynolds, J.; Carmichael, P. L.; Dent, M.; Li, H.; Ryder, S.; White, A.; Walker, P.; Middleton, A. M. Identifying and Characterizing Stress Pathways of Concern for Consumer Safety in Next-Generation Risk Assessment. *Toxicol. Sci.* **2020**, *176*, 11–33.
- (14) Li, H.; Yuan, H.; Middleton, A.; Li, J.; Nicol, B.; Carmichael, P.; Guo, J.; Peng, S.; Zhang, Q. Next Generation Risk Assessment (NGRA): Bridging In Vitro Points-of-Departure to Human Safety Assessment Using Physiologically-Based Kinetic (PBK) Modelling – A Case Study of Doxorubicin with Dose Metrics Considerations. *Toxicol. Vitro* **2021**, *74*, 105171.
- (15) Alexander-Dann, B.; Pruteanu, L. L.; Oerton, E.; Sharma, N.; Berindan-Neagoe, I.; Módos, D.; Bender, A. Developments in Toxicogenomics: Understanding and Predicting Compound-Induced Toxicity from Gene Expression Data. *Mol. Omics* **2018**, *14*, 218–236.
- (16) Joseph, P. Transcriptomics in Toxicology. *Food Chem. Toxicol.* **2017**, *109*, 650–662.
- (17) House, J. S.; Grimm, F. A.; Jima, D. D.; Zhou, Y.-H.; Rusyn, I.; Wright, F. A. A Pipeline for High-Throughput Concentration Response Modeling of Gene Expression for Toxicogenomics. *Front. Genet.* **2017**, *8*, 168.
- (18) Farmahin, R.; Williams, A.; Kuo, B.; Chepelev, N. L.; Thomas, R. S.; Barton-Maclaren, T. S.; Curran, I. H.; Nong, A.; Wade, M. G.; Yauk, C. L. Recommended Approaches in the Application of Toxicogenomics to Derive Points of Departure for Chemical Risk Assessment. *Arch. Toxicol.* **2017**, *91*, 2045–2065.
- (19) Webster, A. F.; Chepelev, N.; Gagné, R.; Kuo, B.; Recio, L.; Williams, A.; Yauk, C. L. Impact of Genomics Platform and Statistical Filtering on Transcriptional Benchmark Doses (BMD) and Multiple Approaches for Selection of Chemical Point of Departure (PoD). *PLoS One* **2015**, *10*, No. e0136764.
- (20) Thomas, R. S.; Wesselkamper, S. C.; Wang, N. C. Y.; Zhao, Q. J.; Petersen, D. D.; Lambert, J. C.; Cote, I.; Yang, L.; Healy, E.; Black, M. B.; Clewell, H. J.; Allen, B. C.; Andersen, M. E. Temporal Concordance Between Apical and Transcriptional Points of Departure for Chemical Risk Assessment. *Toxicol. Sci.* **2013**, *134*, 180–194.
- (21) Farmahin, R.; Gannon, A. M.; Gagné, R.; Rowan-Carroll, A.; Kuo, B.; Williams, A.; Curran, L.; Yauk, C. L. Hepatic Transcriptional Dose-Response Analysis of Male and Female Fischer Rats Exposed to Hexabromocyclododecane. *Food Chem. Toxicol.* **2019**, *133*, 110262.
- (22) Mootha, V. K.; Lindgren, C. M.; Eriksson, K.-F.; Subramanian, A.; Sihag, S.; Lehar, J.; Puigserver, P.; Carlsson, E.; Ridderstråle, M.; Laurila, E.; Houstis, N.; Daly, M. J.; Patterson, N.; Mesirov, J. P.; Golub, T. R.; Tamayo, P.; Spiegelman, B.; Lander, E. S.; Hirschhorn, J. N.; Altshuler, D.; Groop, L. C. PGC-1 $\alpha$ -Responsive Genes Involved in Oxidative Phosphorylation Are Coordinately Downregulated in Human Diabetes. *Nat. Genet.* **2003**, *34*, 267–273.
- (23) Khatri, P.; Sirota, M.; Butte, A. J. Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges. *PLoS Comput. Biol.* **2012**, *8*, No. e1002375.
- (24) Sinnott, J. A.; Cai, T. Pathway Aggregation for Survival Prediction via Multiple Kernel Learning. *Stat. Med.* **2018**, *37*, 2501–2515.
- (25) Zheng, X.; Amos, C. I.; Frost, H. R. Comparison of Pathway and Gene-Level Models for Cancer Prognosis Prediction. *BMC Bioinf.* **2020**, *21*, 76.
- (26) Guo, Z.; Zhang, T.; Li, X.; Wang, Q.; Xu, J.; Yu, H.; Zhu, J.; Wang, H.; Wang, C.; Topol, E. J.; Wang, Q.; Rao, S. Towards Precise Classification of Cancers Based on Robust Gene Functional Expression Profiles. *BMC Bioinf.* **2005**, *6*, 58.
- (27) Haynes, W. A.; Tomczak, A.; Khatri, P. Gene Annotation Bias Impedes Biomedical Research. *Sci. Rep.* **2018**, *8*, 1362.
- (28) Yeakley, J. M.; Shepard, P. J.; Goyena, D. E.; VanSteenhouse, H. C.; McComb, J. D.; Seligmann, B. E. A Trichostatin A Expression Signature Identified by TempO-Seq Targeted Whole Transcriptome Profiling. *PLoS One* **2017**, *12*, No. e0178302.
- (29) Zhang, Y.; Parmigiani, G.; Johnson, W. E. ComBat-Seq: Batch Effect Adjustment for RNA-Seq Count Data. *NAR: Genomics Bioinf.* **2020**, *2*, lqaa078.
- (30) Love, M. I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15*, 550.
- (31) Mao, W.; Zaslavsky, E.; Hartmann, B. M.; Sealfon, S. C.; Chikina, M. Pathway-Level Information Extractor (PLIER) for Gene Expression Data. *Nat. Methods* **2019**, *16*, 607–610.
- (32) Filer, D. L.; Kothiyi, P.; Setzer, R. W.; Judson, R. S.; Martin, M. T. Tcpl: The ToxCast Pipeline for High-Throughput Screening Data. *Bioinformatics* **2017**, *33*, 618.
- (33) Sheffield, T.; Brown, J.; Davidson, S.; Friedman, K. P.; Judson, R. Tcplfit2: An R-Language General Purpose Concentration-Response Modeling Package. *Bioinformatics* **2021**, *38*, 1157.
- (34) Akaike, H. A New Look at the Statistical Model Identification. *IEEE Trans. Automat. Control* **1974**, *19*, 716–723.
- (35) Phillips, J. R.; Svoboda, D. L.; Tandon, A.; Patel, S.; Sedykh, A.; Mav, D.; Kuo, B.; Yauk, C. L.; Yang, L.; Thomas, R. S.; Gift, J. S.; Davis, J. A.; Olszyk, L.; Merrick, B. A.; Paules, R. S.; Parham, F.; Saddler, T.; Shah, R. R.; Auerbach, S. S. BMDExpress 2: Enhanced Transcriptomic Dose-Response Analysis Workflow. *Bioinformatics* **2019**, *35*, 1780–1782.
- (36) Jassal, B.; Matthews, L.; Viteri, G.; Gong, C.; Lorente, P.; Fabregat, A.; Sidiropoulos, K.; Cook, J.; Gillespie, M.; Haw, R.;

- Loney, F.; May, B.; Milacic, M.; Rothfels, K.; Sevilla, C.; Shamovsky, V.; Shorser, S.; Varusai, T.; Weiser, J.; Wu, G.; Stein, L.; Hermjakob, H.; D'Eustachio, P. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **2019**, *48*, D498.
- (37) Reynolds, J.; Malcomber, S.; White, A. A Bayesian Approach for Inferring Global Points of Departure from Transcriptomics Data. *Comput. Toxicol.* **2020**, *16*, 100138.
- (38) Carpenter, B.; Gelman, A.; Hoffman, M. D.; Lee, D.; Goodrich, B.; Betancourt, M.; Brubaker, M.; Guo, J.; Li, P.; Riddell, A. Stan: A Probabilistic Programming Language. *J. Stat. Software* **2017**, *76*, 1.
- (39) Lin, L. I.-K. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* **1989**, *45*, 255.
- (40) Lynch, C.; Zhao, J.; Sakamuru, S.; Zhang, L.; Huang, R.; Witt, K.; Merrick, B.; Teng, C.; Xia, M. Identification of Compounds That Inhibit Estrogen-Related Receptor Alpha Signaling Using High-Throughput Screening Assays. *Molecules* **2019**, *24*, 841.
- (41) Datta, S.; He, G.; Tomilov, A.; Sahdeo, S.; Denison, M. S.; Cortopassi, G. In Vitro Evaluation of Mitochondrial Function and Estrogen Signaling in Cell Lines Exposed to the Antiseptic Cetylpyridinium Chloride. *Environ. Health Perspect.* **2017**, *125*, 087015.
- (42) Duarte-Guterman, P.; Navarro-Martín, L.; Trudeau, V. L. Mechanisms of Crosstalk between Endocrine Systems: Regulation of Sex Steroid Hormone Synthesis and Action by Thyroid Hormones. *Gen. Comp. Endocrinol.* **2014**, *203*, 69–85.
- (43) Chen, L.; Chen, X.; Chen, X.; Hu, Z.; Li, X.; Su, Y.; Li, X.; Ge, R.-S. Ziram Inhibits Aromatase Activity in Human Placenta and JEG-3 Cell Line. *Steroids* **2017**, *128*, 114–119.
- (44) Judson, R.; Houck, K.; Martin, M.; Richard, A. M.; Knudsen, T. B.; Shah, I.; Little, S.; Wambaugh, J.; Woodrow Setzer, R.; Kothya, P.; Phuong, J.; Filer, D.; Smith, D.; Reif, D.; Rotroff, D.; Kleinstreuer, N.; Sipes, N.; Xia, M.; Huang, R.; Crofton, K.; Thomas, R. S. Editor's Highlight: Analysis of the Effects of Cell Stress and Cytotoxicity on In Vitro Assay Activity Across a Diverse Chemical and Assay Space. *Toxicol. Sci.* **2016**, *152*, 323–339.
- (45) Mussard, E.; Cesaro, A.; Lespessailles, E.; Legrain, B.; Berteina-Raboin, S.; Toumi, H. Andrographolide, A Natural Antioxidant: An Update. *Antioxidants* **2019**, *8*, 571.
- (46) Lundin, M.; Baltscheffsky, H.; Ronne, H. Yeast PPA2 Gene Encodes a Mitochondrial Inorganic Pyrophosphatase That Is Essential for Mitochondrial Function. *J. Biol. Chem.* **1991**, *266*, 12168–12172.
- (47) Guimier, A.; Gordon, C. T.; Godard, F.; Ravenscroft, G.; Oufadem, M.; Vasnier, C.; Rambaud, C.; Nitschke, P.; Bole-Feyso, C.; Masson, C.; Dautant, S.; Longman, C.; Laing, N. G.; Kugener, B.; Bonnet, D.; Bouvagnet, P.; Di Filippo, S.; Probst, V.; Redon, R.; Charron, P.; Rötig, A.; Lyonnet, S.; Dautant, A.; de Pontual, L.; di Rago, J.-P.; Delahodde, A.; Amiel, J. Biallelic PPA2 Mutations Cause Sudden Unexpected Cardiac Arrest in Infancy. *Am. J. Hum. Genet.* **2016**, *99*, 666–673.
- (48) Tan, W. S. D.; Liao, W.; Peh, H. Y.; Vila, M.; Dong, J.; Shen, H.-M.; Wong, W. S. F. Andrographolide Simultaneously Augments Nrf2 Antioxidant Defense and Facilitates Autophagic Flux Blockade in Cigarette Smoke-Exposed Human Bronchial Epithelial Cells. *Toxicol. Appl. Pharmacol.* **2018**, *360*, 120–130.
- (49) Li, G.-f.; Qin, Y.-h.; Du, P.-q. Andrographolide Inhibits the Migration, Invasion and Matrix Metalloproteinase Expression of Rheumatoid Arthritis Fibroblast-like Synoviocytes via Inhibition of HIF-1 $\alpha$  Signaling. *Life Sci.* **2015**, *136*, 67–72.
- (50) Lin, H.-C.; Su, S.-L.; Lin, W.-C.; Lin, A.-H.; Yang, Y.-C.; Lii, C.-K.; Chen, H.-W. Andrographolide Inhibits Hypoxia-Induced Hypoxia-Inducible Factor 1 $\alpha$  and Endothelin 1 Expression through the Heme Oxygenase 1/CO/CGMP/MKP-5 Pathways in EA.Hy926 Cells. *Environ. Toxicol.* **2018**, *33*, 269–279.
- (51) Brzoskwini, M.; Pardyak, L.; Kaminska, A.; Tworzyl, W.; Hejmej, A.; Marek, S.; Bilinski, S. M.; Bilinska, B. Flutamide Treatment Reveals a Relationship between Steroidogenic Activity of Leydig Cells and Ultrastructure of Their Mitochondria. *Sci. Rep.* **2021**, *11*, 13772.
- (52) Kashimshetty, R.; Desai, V. G.; Kale, V. M.; Lee, T.; Moland, C. L.; Branham, W. S.; New, L. S.; Chan, E. C. Y.; Younis, H.; Boelsterli, U. A. Underlying Mitochondrial Dysfunction Triggers Flutamide-Induced Oxidative Liver Injury in a Mouse Model of Idiosyncratic Drug Toxicity. *Toxicol. Appl. Pharmacol.* **2009**, *238*, 150–159.
- (53) Gao, X.; Xie, C.; Wang, Y.; Luo, Y.; Yagai, T.; Sun, D.; Qin, X.; Krausz, K. W.; Gonzalez, F. J. The Antiandrogen Flutamide Is a Novel Aryl Hydrocarbon Receptor Ligand That Disrupts Bile Acid Homeostasis in Mice through Induction of Abcc4. *Biochem. Pharmacol.* **2016**, *119*, 93–104.
- (54) Mercader, M.; Bodner, B. K.; Moser, M. T.; Kwon, P. S.; Park, E. S. Y.; Manecke, R. G.; Ellis, T. M.; Wojcik, E. M.; Yang, D.; Flanigan, R. C.; Waters, W. B.; Kast, W. M.; Kwon, E. D. T cell infiltration of the prostate induced by androgen withdrawal in patients with prostate cancer. *Proc. Natl. Acad. Sci.* **2001**, *98*, 14565–14570.
- (55) Czaja, E.; Wiecech-Rojkiewicz, I.; Knapczyk-Stwora, K.; Grzesiak, M.; Slomczynska, M. The Impact of Antiandrogen Flutamide on the Hypoxia Inducible Factor 1 $\alpha$  and Vascular Endothelial Growth Factor A Gene and Protein Expression in the Pig Placenta during Late Pregnancy. *Folia Histochem. Cytobiol.* **2017**, *55*, 159–167.
- (56) Azhary, J. M. K.; Harada, M.; Takahashi, N.; Nose, E.; Kunitomi, C.; Koike, H.; Hirata, T.; Hirota, Y.; Koga, K.; Wada-Hiraie, O.; Fujii, T.; Osuga, Y. Endoplasmic Reticulum Stress Activated by Androgen Enhances Apoptosis of Granulosa Cells via Induction of Death Receptor 5 in PCOS. *Endocrinology* **2019**, *160*, 119–132.
- (57) Shet, M. S.; McPhaul, M.; Fisher, C. W.; Stallings, N. R.; Estabrook, R. W. Metabolism of the Antiandrogenic Drug (Flutamide) by Human CYP1A2. *Drug Metab. Dispos.* **1997**, *25*, 1298–1303.
- (58) Matsuzaki, Y.; Nagai, D.; Ichimura, E.; Goda, R.; Tomura, A.; Doi, M.; Nishikawa, K. Metabolism and Hepatic Toxicity of Flutamide in Cytochrome P450 1A2 Knockout SV129 Mice. *J. Gastroenterol.* **2006**, *41*, 231–239.
- (59) Blanco-Melo, D.; Nilsson-Payant, B. E.; Liu, W.-C.; Uhl, S.; Hoagland, D.; Möller, R.; Jordan, T. X.; Oishi, K.; Panis, M.; Sachs, D.; Wang, T. T.; Schwartz, R. E.; Lim, J. K.; Albrecht, R. A.; TenOever, B. R. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **2020**, *181*, 1036–1045.
- (60) Gavin, M. A.; Rasmussen, J. P.; Fontenot, J. D.; Vasta, V.; Manganiello, V. C.; Beavo, J. A.; Rudensky, A. Y. Foxp3-Dependent Programme of Regulatory T-Cell Differentiation. *Nature* **2007**, *445*, 771–775.
- (61) Weatherly, L. M.; Gosse, J. A. Triclosan Exposure, Transformation, and Human Health Effects. *J. Toxicol. Environ. Health, Part B* **2017**, *20*, 447–469.
- (62) Dai, Y.; Chen, S.-R.; Chai, L.; Zhao, J.; Wang, Y.; Wang, Y. Overview of Pharmacological Activities of Andrographis paniculata and Its Major Compound Andrographolide. *Crit. Rev. Food Sci. Nutr.* **2019**, *59*, S17–S29.
- (63) Liu, B.; Lindner, P.; Jirmo, A. C.; Maus, U.; Illig, T.; DeLuca, D. S. A Comparison of Curated Gene Sets versus Transcriptomics-Derived Gene Signatures for Detecting Pathway Activation in Immune Cells. *BMC Bioinf.* **2020**, *21*, 28.
- (64) Schlage, W. K.; Westra, J. W.; Gebel, S.; Catlett, N. L.; Mathis, C.; Frushour, B. P.; Hengstermann, A.; Van Hooser, A.; Poussin, C.; Wong, B.; Lietz, M.; Park, J.; Drubin, D.; Veljkovic, E.; Peitsch, M. C.; Hoeng, J.; Deehan, R. A. Computable Cellular Stress Network Model for Non-Diseased Pulmonary and Cardiovascular Tissue. *BMC Syst. Biol.* **2011**, *5*, 168.
- (65) Chambers, B.; Shah, I. Evaluating Adaptive Stress Response Gene Signatures Using Transcriptomics. *Comput. Toxicol.* **2021**, *20*, 100179.