Application of Cell Painting for chemical hazard evaluation in support of screening-level chemical assessments.

Jo Nyffeler[1,2], Clinton Willis[1], Felix R. Harris[1,3], MJ Foster[1,3], Bryant Chambers[1], Megan Culbreth[1], Richard E. Brockway[1,3], Sarah Davidson-Fritz[1], Daniel Dawson[1], Imran Shah[1], Katie Paul Friedman[1], Dan Chang[1], Logan J. Everett[1], John F. Wambaugh[1], Grace Patlewicz[1], Joshua A. Harrill[1,*]

ORCIDs:

Nyffeler: https://orcid.org/0000-0002-6155-9743
Willis: https://orcid.org/0000-0002-4996-1780
Harris: https://orcid.org/0000-0002-0256-7452
Foster: NA
Chambers: https://orcid.org/0000-0002-2233-3207
Culbreth: https://orcid.org/0000-0002-3101-1541
Brockway: NA
Davidson-Fritz: https://orcid.org/0000-0002-2891-9380
Dawson: https://orcid.org/0000-0001-9622-4495
Shah: https://orcid.org/0000-0003-0808-0140
Paul Friedman: https://orcid.org/0000-0002-2710-1691
Chang: https://orcid.org/0000-0001-7561-2747
Everett: https://orcid.org/0000-0002-9713-6099
Wambaugh: https://orcid.org/0000-0002-4024-534X
Patlewicz: https://orcid.org/0000-0003-3863-9689
Harrill: https://orcid.org/0000-0003-4317-6391

Affiliations:

[1] Center for Computational Toxicology & Exposure, Office of Research and Development, US Environmental Protection Agency, Durham, NC 27711

[2] Oak Ridge Institute for Science and Education (ORISE) Postdoctoral Fellow, Oak Ridge, TN, 37831

[3] Oak Ridge Associated Universities (ORAU) National Student Services Contractor, Oak Ridge, TN, 37831

* Corresponding Author

Joshua A. Harrill
Center for Computational Toxicology and Exposure (CCTE)
U.S. Environmental Protection Agency
109 TW Alexander Drive, Research Triangle Park, NC 27709
harrill.joshua@epa.gov

Running title: Bioactivity Screening of ToxCast Chemicals in U-2 OS with HTPP.

Key words: High-throughput phenotypic profiling, Cell Painting, concentration-response, computational toxicology

Abstract

'Cell Painting' is an imaging-based high-throughput phenotypic profiling (HTPP) method in which cultured cells are fluorescently labeled to visualize subcellular structures (i.e., nucleus, nucleoli, endoplasmic reticulum, cytoskeleton, Golgi apparatus / plasma membrane and mitochondria) and to quantify morphological changes in response to chemicals or other perturbagens. HTPP is a high-throughput and cost-effective bioactivity screening method that detects effects associated with many different molecular mechanisms in an untargeted manner, enabling rapid *in vitro* hazard assessment for thousands of chemicals. Here, 1201 chemicals from the ToxCast library were screened in concentration-response up to ~100 µM in human U-2 OS cells using HTPP. A phenotype altering concentration (PAC) was estimated for chemicals active in the tested range. PACs tended to be higher than lower bound potency values estimated from a broad collection of targeted high-throughput assays, but lower than the threshold for cytotoxicity. *In vitro* to *in vivo* extrapolation (IVIVE) was used to estimate administered equivalent doses (AEDs) based on PACs for comparison to human exposure predictions. AEDs for 18/412 chemicals overlapped with predicted human exposures. Phenotypic profile information was also leveraged to identify putative mechanisms of action and group chemicals. Of 58 known nuclear receptor modulators, only glucocorticoids and retinoids produced characteristic profiles; and both receptor types are expressed in U-2 OS cells. Thirteen chemicals with profile similarity to glucocorticoids were tested in a secondary screen and one chemical, pyrene, was confirmed by an orthogonal gene expression assay as a novel putative GR modulating chemical. Most active chemicals demonstrated profiles not associated with a known mechanism-of-action. However, many structurally related chemicals produced similar profiles, with exceptions such as diniconazole, whose profile differed from other active conazoles. Overall, the present study demonstrates how HTPP can be applied in screening-level chemical assessments through a series of examples and brief case studies.

Abbreviations

AED	administered equivalent dose
AGP	actin, golgi, plasma membrane
AR	androgen receptor
ATRA	all trans-retinoic acid
BER	bioactivity exposure ratio
BMC	benchmark concentration
BMR	benchmark response
Cl_{int}	intrinsic hepatic clearance
CCTE	Center for Computational Toxicology and Exposure
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
EC50	effective concentration 50%
ER	estrogen receptor
FPR	false positive rate
F_{up}	fraction unbound in plasma
GR	glucocorticoid receptor
GRE	glucocorticoid response element
HTPP	high-throughput phenotypic profiling
HTTr	high-throughput transcriptomics
HTS	high-throughput screening
IVIVE	*in vitro* to *in vivo* extrapolation
LOEC	lowest observed effect concentration
logKoA	octanol air partition coefficient
MAD	median absolute deviation
MOA	mechanism-of-action
Mito	mitochondria
NAMs	new approach methods
nCC	normalized cell count
NR	nuclear receptor
nMAD	normalized median absolute deviation
PAC	phenotype altering concentration
PPAR	peroxisome proliferator-activated receptor
PBS	phosphate-buffered saline
PMI	pointwise mutual information
POD	point-of-departure

104	QSPR	quantitative structure-property relationships
105	RA	retinoic acid
106	RAR	retinoic acid receptor
107	RT-qPCR	reverse transcription quantitative polymerase chain reaction
108	RXR	retinoid X receptor
109	SD	standard deviation
110	SRP	stress response pathways
111	USEPA	United States Environmental Protection Agency

112

Introduction

'Cell Painting' is an imaging-based profiling method in which cultured cells are fluorescently labeled to visualize subcellular structures (i.e., nucleus, nucleoli, endoplasmic reticulum, cytoskeleton, Golgi apparatus / plasma membrane and mitochondria) and changes to the morphology (e.g., their phenotypic appearance) are quantified (Gustafsdottir *et al.*, 2013). Many other bioassays are targeted; that is, they are designed to measure specific, expected effects on a discrete molecular target or biological process. In comparison, Cell Painting is a profiling assay and thus untargeted, meaning that a broad range of biological effects can be observed and evaluated. The method is high throughput and more cost effective than other molecular profiling methods, such as transcriptomics or panels of targeted assays (Chandrasekaran *et al.*, 2021). For these reasons, Cell Painting has been used in many applications, including by the pharmaceutical industry for lead hopping, to establish chemical screening libraries with diverse biological activities, or for functional genomic screens (Caicedo *et al.*, 2016). Different perturbations can result in different phenotypes, which several groups have leveraged to predict outcomes of *in vitro* bioassays (Simm *et al.*, 2018; Hofmarcher *et al.*, 2019; Way *et al.*, 2021), mechanism-of-action (MOA) (Caie *et al.*, 2010; Woehrmann *et al.*, 2013; Ramm *et al.*, 2019; Warchal *et al.*, 2019; Trapotsi *et al.*, 2021) or *in vivo* organ toxicity (Su *et al.*, 2016; Ramm *et al.*, 2019).

The Center for Computational Toxicology and Exposure (CCTE) at the United States Environmental Protection Agency (USEPA) became interested in using the Cell Painting assay for bioactivity screening of chemicals because of its cost-effectiveness (Hughes *et al.*, 2011; Svenningsen and Poulsen, 2019; Chandrasekaran *et al.*, 2021; Marin Zapata *et al.*, 2023) and its broad applicability to different cell types. Scientists at CCTE proposed a tiered testing strategy whereby the first tier comprises *in silico* predictions and *in vitro* bioactivity screening of large chemical collections using two high-throughput profiling methods: image-based high-throughput phenotypic profiling (HTPP) with the Cell Painting assay and high-throughput transcriptomics (HTTr) with targeted RNA-Seq (Thomas *et al.*, 2019). The second tier consists of targeted high-throughput screening (HTS) assays to confirm biological activity observed in Tier 1, followed by higher order tiers including more complex low-throughput systems such as organotypic models and microphysiological systems. The goal of the first tier is to rapidly generate information about the concentration at which a chemical exerts bioactivity, as well as its putative mechanisms. The first-tier screening data from HTPP (and HTTr) could be used in a variety of ways to inform chemical risk assessment, including prioritization of chemicals based on the overlap of biological activity with human exposure predictions, inference of molecular MOA by comparison to reference

chemicals and grouping of chemicals based on similarities in bioactivity profiles to inform chemical read-across.

While other groups notably in the pharmaceutical sector are mostly interested in drug-like chemicals that have one (or a few) specific target(s), the research at CCTE is more focused on characterizing the putative bioactivity of chemicals found in the environment, such as pesticides, herbicides and industrial chemicals in order to inform chemical risk assessment. Such chemicals were typically not designed to have biological activity in humans and in many cases, human health hazard has been incompletely characterized. Moreover, researchers using Cell Painting (or similar profiling assays) in phenotypic drug discovery are often focused on a particular phenotype, or transition between phenotypes, such as reversion of diseased cells to a "healthy" phenotype (Chandrasekaran *et al.*, 2021). In contrast, bioactivity screening of environmental chemicals requires an approach that is capable of measuring the vast array of unknown phenotypes these chemicals may produce in order to detect any type of biological effects. Additionally, when screening environmental chemicals, there is a priority to avoid false negatives, while a certain number of false positives are acceptable in this first testing tier (Nyffeler *et al.*, 2021).

We previously adapted the HTPP approach to be compatible with our available microfluidic instruments and screened 462 chemicals in human U-2 OS osteosarcoma cells (Nyffeler *et al.*, 2020). Upon comparison to mammalian *in vivo* studies, we found that HTPP effect values were lower or comparable to *in vivo* effect values for 68% of chemicals. We further conducted a small study where we profiled U-2 OS cells with both HTPP and HTTr and demonstrated that for one model compound, all-trans retinoic acid (ATRA), a characteristic cellular phenotype co-occurs with transcriptional changes specific to activation of the retinoic acid receptor signaling pathway (Nyffeler *et al.*, 2022).

In the present study, we screened 1201 unique chemicals from the ToxCast chemical library (Richard *et al.*, 2016) in concentration-response mode. We then compared how potency estimates from HTPP related to those from other *in vitro* assays and provided an example of how HTPP can be used for screening-level assessments via the bioactivity exposure ratio (BER) approach. For the first time, we leverage the broad phenotypic information collected with HTPP to identify putative MOAs for environmental chemicals in a human-derived cell model and group chemicals based on similarities in biological activity. These concepts are demonstrated using multiple examples.

Materials and Methods

Materials

Materials and reference chemicals were purchased from various suppliers (Table S1). Lists of screened chemicals are provided as supplementary documents (File S1-S2).

Cell culture

U-2 OS human osteosarcoma cells (HTB-96®, Lot: 64048673, ATCC) were expanded to internal passage number 6 (P6) as described previously (Nyffeler *et al.*, 2020). For the purposes of this study, a subsequent stock of cells was expanded and cryopreserved at P8 using those same methods. Culture media consisted of Dulbecco's Modified Eagle Medium (DMEM) with 10% heat-inactivated fetal bovine serum and penicillin/streptomycin/glutamine. The primary screen was performed with cells expanded from P8 to P10. Follow-up experiments were conducted using P9 cells. Cells were plated at 3000 cells/well in 40 µL media or 9000 cells/well in 120 µL media for 384-well and 96-well plates, respectively, as described previously (Nyffeler *et al.*, 2022).

Chemical treatment

For screening in 384-well format, all test chemicals were received as solutions in dimethyl sulfoxide (DMSO) from the USEPA ToxCast library management contractor (Evotec, Princeton, NJ). Solutions were 20 mM or the highest concentration soluble in DMSO not exceeding 20 mM. An 8-point dilution series of each test chemical was prepared in DMSO using a uniform dilution scheme (half $\log_{10}$-spacing). The dilutions were prepared as 200x the desired nominal test concentrations corresponding to 0.03 to 100 µM for chemicals received as 20 mM stocks. Reference chemicals (e.g., dexamethasone, etoposide, ATRA, staurosporine and trichostatin A) were purchased as neat chemicals and solubilized in DMSO at CCTE laboratories. Dilution series of the first three reference chemicals were prepared as described for test chemicals. Chemical treatments were dispensed 24 h after cell seeding using a LabCyte Echo 550 acoustic dispenser (Beckman-Coulter, Indianapolis, IN) and randomized plate layouts (Fig. S1) as described previously (Nyffeler *et al.*, 2020; Nyffeler *et al.*, 2022). In brief, 200 nL of 200x stock was dispensed into 40 µL of media in the test wells.

For treatments in 96-well plates, 20 µL of media was removed from the cell culture well. Then 750 nL of chemical was dispensed into an empty 384-well plate using the LabCyte Echo 550 acoustic dispenser, manually filled with 25 µL cell culture media and 20 µL of this manually transferred onto the

cells. The final concentration of DMSO in test wells was 0.5% for all experiments, regardless of plate format. The vehicle control treatment for all experiments was 0.5% DMSO.

Experimental design for primary screen

In the primary screen, 1218 chemical samples comprising 1201 unique chemicals were tested at 8 concentrations each, with one technical replicate and four biological replicates (i.e., independent cell cultures). All tested chemicals are listed in File S1 and on the EPA CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard/chemical-lists/HTPP2023_U2OS_SCREEN). The 1218 chemicals were distributed among 29 dose plates. Each dose plate contained up to 42 test chemicals and three reference chemicals (dexamethasone, etoposide, ATRA) at 8 concentrations in half-$\log_{10}$ spacing, as well as three wells of a reference chemical (trichostatin A) and three wells of a cell viability positive control chemical (staurosporine) at a single concentration (1 µM), and eighteen wells of vehicle control (0.5% DMSO). Each dose plate was used to treat four assay plates across four biological replicates (i.e., independent cultures). A set of four assay plates receiving the same group of test chemicals is referred to as a 'plate group' (Fig S1).

High-throughput phenotypic profiling assay (HTPP)

Fluorescent labeling

Phenotypic profiling of cells was performed using the 'Cell Painting' method (Bray *et al.*, 2016). Fluorescent labeling was performed exactly as described in Nyffeler *et al.* (2020). In brief, the nucleus, nucleolus, endoplasmic reticulum, action cytoskeleton, Golgi apparatus / plasma membrane and mitochondria were visualized with Hoechst-33342, SYTO14, Alexa Fluor™ 488 Concanavalin A conjugate, Alexa Fluor™ 568 Phalloidin conjugate, Alexa Fluor™ 555 Wheat Germ Agglutinin conjugate and MitoTracker DeepRed, respectively.

Image acquisition & feature extraction

Fluorescent images were acquired using an Opera Phenix High Content Screening System and Harmony® software (v4.9). Five distinct fields-of-view were acquired with a 20× water immersion objective in confocal mode exactly as described in Nyffeler *et al.* (2020). A total of 1300 features per cell were extracted.

Aggregation and normalization

For each feature, data were normalized to vehicle-control treated cells using the median and median absolute deviation (MAD) as described previously (Nyffeler *et al.*, 2020) and aggregated to the

well-level by computing the median of all cells in that well. Well-level data were scaled further by using the vehicle control wells of the corresponding plate group (typically 96 wells), but in slight modification from previous reports (Nyffeler *et al.*, 2022); the mean and standard deviation (SD) were used instead of the median and normalized MAD. These scaled values were used for concentration-response modeling and graphical representations.

Cell viability analysis

Cell counts from the HTPP assay plates were used to evaluate cell viability and cytostasis in the present study. This differs from our previous studies (Nyffeler *et al.*, 2020; Willis *et al.*, 2020) where separate assay plates labeled with Hoechst-33342 and propidium iodide were used to assess cytotoxicity. In the present study, normalized cell counts (nCC) were calculated by dividing the number of cells in each treatment well by the average number of cells in vehicle control wells on the same plate. Concentration-response modeling of nCC was performed using *tcplfit2*. The benchmark response was set at 50% and a cell viability benchmark concentration (CV BMC) was determined. For chemicals with a CV BMC, the lowest test concentration above the CV BMC was defined as the lowest observed effect concentration (CV LOEC) and the highest test concentration below the CV BMC was defined as the no observed effect concentration (CV NOEC). Treatments that produced a greater than 50% decrease in nCC were excluded from analysis of HTPP data.

Concentration-response analysis of HTPP data

Generation of null chemicals

In order to estimate a false positive rate (FPR), a synthetic data set was created consisting of responses of the two lowest concentrations of the tested chemicals. Only responses from chemicals with at least 6 non-cytostatic concentrations were eligible. The responses were randomly assigned to the null chemicals on a per plate basis. In this manner, 10 null chemicals with 8 concentrations each and four biological replicates could be generated for each plate group, resulting in a total of 290 null chemicals. These null chemicals were treated exactly as the test chemicals for concentration-response modeling.

Mahalanobis distance calculation

Two separate Mahalanobis distance modeling approaches were used to reduce the information from the 1300 features to fewer dimensions for concentration-response modeling as described previously (Nyffeler *et al.*, 2021). For the 'Global Mahalanobis' approach, one distance value was

generated for each treatment from the combined 1300 features. For the 'Category-level Mahalanobis' approach, one distance value was generated for each of 49 feature categories for each treatment. A feature category consists of multiple features generated from the same channel, organelle and analysis module. Each of the 1300 features maps to one of 49 categories.

Concentration-response modeling

Mahalanobis distances were used for concentration-response modeling using the R package *tpclfit2* (v0.1.0) (Sheffield *et al.*, 2021) as described previously (Nyffeler *et al.*, 2021; Nyffeler *et al.*, 2022) with minor modification. First, cell count information was used to remove overtly cytostatic concentrations, defined as those that caused a > 50% reduction in nCC compared to vehicle control. Then Mahalanobis distance data were modeled using nine different functions (constant, Hill, poly1, poly2, power, exp2, exp3, exp4, exp5) and setting the cutoff to 1 standard deviation (SD). Chemicals for which the modeled responses did not exceed the cutoff were only modeled with the constant model and thus were deemed inactive. The ~~benchmark concentration~~ BMC was defined as the concentration at which the response exceeded the benchmark response (BMR) of 1 SD. BMCs more than half an order of magnitude below the lowest tested concentration were set to that value. For the 'Category-level Mahalanobis' approach, only curves with a hitcall probability of > 0.9 were considered as active; for the 'Global Mahalanobis' approach no additional restrictions were set.

Hit identification

Chemicals were considered phenotypically active if a BMC could be identified with either the 'Global Mahalanobis' approach or for at least one category using the 'Category-level Mahalanobis' approach. The phenotype altering concentration (PAC) was defined as the lower of the Global BMC or most sensitive Category-level BMC.

Comparison of phenotypic profiles

Feature selection

Of the 1300 acquired features, only a subset was used to compare phenotypic profiles. Features were eliminated in three steps: (1) One feature was excluded that always gave a constant value and thus did not provide any information; (2) 716 features were excluded as their response across biological replicates had low correlation (Pearson correlation <= 0.5), thus indicating low reproducibility, similar as described in Woehrmann *et al.* (2013); and (3) the remaining 583 features underwent stepwise feature elimination (Bray *et al.*, 2016; Warchal *et al.*, 2019). In the process of feature elimination, the feature pair with the highest correlation was identified, one of the two features removed, then this process was

repeated until all remaining features were correlated less than a set threshold value (Kendall correlation < 0.75). This procedure resulted in 289 selected features (listed in File S3). A detailed description of the procedure is provided in the supporting information (Fig. S2).

HTPP profile generation and evaluation of biological similarity

Similarity of biological responses was then calculated by comparing phenotypic profiles using the 289 selected features. First, all bioactive concentrations of test chemicals were identified (i.e., above the PAC and below the CV LOEC). Second, the lowest test concentration that resulted in a reproducible profile (defined in Fig. S3) was identified. As the PAC marks the onset of bioactivity, it may be that a robust reproducible profile is only observable at a slightly higher concentration. Third, after finding the lowest concentration with a reproducible profile, up to two higher concentrations were selected, if they were still below the CV LOEC. These profiles were then modified by setting values that fall within [-1,1] to 0, to reduce noise. Profiles were then compared using Kendall correlation. This procedure resulted in multiple correlation values for each chemical pair; thus, these data were further summarized by retaining the highest correlation value for each chemical pair. A detailed description is provided in the supporting information (Fig. S3).

Combination of HTPP results with other data streams

Generation of chemical structural features and calculation of structural similarity

Chemical structure features aim to represent molecules in machine-readable notation, such as binary or categorial descriptors. Different approaches exist that each have advantages and disadvantages. In the present study, the following descriptors were used to explore whether chemicals with structural similarities produce similar phenotypic profiles in U-2 OS cells:

(1) ToxPrint chemotypes consist of a predefined library of 729 sub-structural features designed to encapsulate a broad range of chemical atoms and scaffolds, which were developed by Altamira and Molecular Networks under contract by the US Food and Drug Administration (Yang *et al.*, 2015). Whilst the binary molecular fingerprint is generated for the set of chemicals utilizing the publicly available ToxPrint chemotype feature set (https://toxprint.org) and the ChemoTyper software (https://chemotyper.org/), given the large number of substances being profiled, the commercial command line version of CORINA Symphony application licensed from Molecular Networks was used. ToxPrint chemotypes were generated for 1173 chemicals.

(2) ClassyFire (Djoumbou Feunang *et al.*, 2016) was used to generate structural features (i.e., substituents) identified through the application's rules-based approach that utilizes its comprehensive chemical structure and structural feature hierarchical chemical taxonomy (ChemOnt) to assign chemical compounds to 4825 categories covering both organic and inorganic chemistries. The chemical taxonomy consists of up to 11 levels (i.e., Kingdom, SuperClass, Class, SubClass, etc.). The terms Kingdom, SuperClass, Class, and SubClass denote the first, second, third and fourth levels of the chemical taxonomy, respectively, where the top level (Kingdom) refers to two distinct classes of chemistries – i.e., organic and inorganic compounds. Each subsequent level in the chemical taxonomy consists of more specific chemical categories and hence, more specific and recognizable structural features (i.e., nodes) as one goes down the taxonomic tree. ClassyFire also returns a list of chemical substituents (i.e., structural features within a corresponding chemical category such as functional groups, substructures or motifs) contained in the chemical structure. Utilizing the molecules' InChIKeys (https://iupac.org/who-we-are/divisions/division-details/inchi/), a list of chemical substituents was generated using the ClassyFire web API (http://classyfire.wishartlab.com/) to programmatically access the stored pre-processed data as a query lookup. The keywords in the column 'substituents' were used to generate binary fingerprints (presence/absence of a given keyword). Fingerprints were available for 1171 chemicals and 1010 features were present in at least one chemical.

(3) The OECD Toolbox (v4.4.1) (Dimitrov *et al.*, 2016) was used to profile the set of chemicals through all available profiling schemes. There are several types of 'profilers' in the Toolbox to facilitate the categorization process. The profilers are grouped into different types from predefined, general mechanistic, endpoint specific, empiric and toxicological. These profilers are effectively structural alert schemes that may assign a chemical into a predefined chemical category. Keywords in the columns 'Organic.functional.groups' (general), 'Organic.functional.groups.nested' (nested), Organic.functional.groups.Norbert.Haider.checkmol' (NH), 'Organic.functional.groups.US.EPA' (USEPA) and 'US.EPA.New.Chemical.Categories' (newCat) were used to generate binary fingerprints (presence/absence of a given keyword). The term in brackets indicates the prefix added to the keyword for graphical representation. Fingerprints were available for 1151 chemicals and 669 features were present in at least one chemical. The Organic.functional.groups (OFG) profiler was developed by LMC in 2012 and comprises 534 categories. The OFG profiler is subdivided into two types, OFG and OFG nested. The difference is that OFG displays all functional groups present in a chemical structure whereas OFG nested does not show functional groups that are nested in larger ones e.g., an aldehyde group would be nested in the carboxylic group and is not shown with OFG nested. The OFG NH was created on the basis

of 204 organic functional groups recognized by the CheckMol program that had been developed by Dr. Norbert Haider of the University of Vienna, Austria. The OFG USEPA comprises 466 categories developed from the 645 structural fragments and correction factors underpinning the EPA KOWWIN fragment library in the EPISuite program. The EPA New Chemical Categories profiler aims to reproduce the original categories cited in the New Chemicals Program Chemical Categories, which was last revised in 2010 (see https://www.epa.gov/reviewing-new-chemicals-under-toxic-substances-control-act-tsca/chemical-categories-used-review-new). Not all categories were implemented since a handful were based on physical considerations and did not include specific structure-based rules. The newCat profiler comprises 66 categories.

The 729 ToxPrints were used to calculate structural similarity of chemicals as this set of fingerprints describes more detailed features as opposed to the other two methods. Structural similarity was calculated using Jaccard/Tanimoto similarity (Jaccard, 1902) with the function *jaccard* from the R package *jaccard*. All structural feature sets were used for the enrichment analysis described below in '*Enrichment of structural and biological properties within phenotypic clusters*'.

Bioactivity-Exposure-Ratio (BER) analysis

For all chemicals, the molecular weight was obtained from the CompTox Chemicals Dashboard, and physicochemical properties were predicted using quantitative structure-property relationships (QSPRs) provided by OPERA v2.6 (Mansouri *et al.*, 2018). Using these properties, chemical-specific values for the fraction unbound (F_{up}) and internal clearance (Cl_{int}) were predicted using the QSPR models from Dawson *et al.* (2021). A total of 1041/1205 chemicals were in the applicability domain of the Dawson et al. 2021 QSPR models. Of those, 412 were active in the HTPP assay. For these 412 chemicals, the administered equivalent dose (AED in units of mg/kg bodyweight/day) was calculated using the *httk* R package (v2.2.2) (Pearce *et al.*, 2017) as described previously (Paul-Friedman *et al.*, 2019; Nyffeler *et al.*, 2020; Breen *et al.*, 2021), based on the predicted F_{up}, Cl_{int}, and octanol-water partitioning values. Briefly, oral equivalents were calculated for human species (using function "calc_mc_oral_equiv") with restrictive clearance set to 'true'. Population variability was simulated using Monte Carlo (Ring *et al.*, 2017) and the 5^{th} percentile (i.e., most sensitive) of the population distribution was considered as the HTPP AED.

General population human exposure predictions (mg/kg bodyweight/day) were obtained from the SEEM3 consensus exposure model (Ring *et al.*, 2019) for all but one chemical. SEEM3 provides distributions of plausible daily intake rates and the upper 95^{th} percentile limit on the estimated median

population intake was used. The BER was calculated as the ratio of the bioactivity estimate (HTPP AED) and the SEEM3 exposure predictions as described (Paul-Friedman *et al.*, 2019; Canada, 2021). A $\log_{10}(BER) < 0$ indicates that the chemical was bioactive at or below the predicted exposures while a $\log_{10}(BER) > 0$ indicates that bioactivity was detected at or above the predicted exposures. More positive values of $\log_{10}(BER)$ indicate a larger margin between those two measures, respectively.

Comparison to targeted *in vitro* assay results

In vitro bioactivity data from the US EPA ToxCast program (Dix *et al.*, 2007; Judson *et al.*, 2010) were obtained from the database *invitrodb* (version 3.4, https://doi.org/10.23645/epacomptox.6062623.v7) and summarized as detailed in Suppl. Fig S4. The database contains potency estimates and activity calls ('hitcall') for chemical x assay endpoint pairs. For chemicals tested more than once in a given assay, results were averaged using a weighted approach that gives less weight to concentration-response curves with more caution flags, as caution flags tend to be indicative of less reliable data/curve fits. Subsequently, an *in vitro* point-of-departure (POD) was calculated from the lowest three potency estimates (i.e., benchmark concentration, BMC) out of all active assays for a given chemical, again by using a weighted average using the number of caution flags as weights.

For comparison with HTPP, only chemicals that were run in at least 400 ToxCast assay endpoints were included (n=1075 chemicals). A chemical was deemed active in the ToxCast battery if it was active in at least 3 assay endpoints, which was the case for 896 chemicals.

A subset of ToxCast assay endpoints (n=75) were annotated as relating to cytotoxicity and cell health and were used to calculate a 'burst' estimate, i.e., a lower bound potency measure for impact on general cell health using the function *tcplCytoPt* of the *tcpl* package (version 2.0.2). Only chemicals that were run in at least 40 'burst' assays were included (n=1071 chemicals). A chemical was deemed active in the 'burst' assays if the lower bound potency estimate was < 1000 µM.

Annotation of chemicals to known MOA

For all chemicals that were active in the screen, annotations were retrieved from RefChemDB (Judson *et al.*, 2019). Only entries for nuclear receptors (NRs) as targets were considered. Fifty-six chemicals had a support of five or more for one of 21 NRs. As some chemicals were annotated to more than one receptor family, this list was manually curated to retain only the most significant target for

each chemical. Despite these efforts, two chemicals could not be resolved and were assigned to two receptor families.

Enrichment of structural and biological properties within phenotypic clusters

To cluster the phenotypic profiles of all active chemicals (n=547), only the profile from the highest of the three lowest active concentrations was selected. For these profiles (consisting of the 289 selected features), similarity was calculated using the function *get_dist* from the package *factoextra* (v1.0.7) with the option method="kendall" followed by clustering with the function *hclust* from the *stats* package (v.3.6.2). The resulting tree was then cut into 15 clusters.

Enrichment of chemical fingerprints: Chemical fingerprints were generated with ToxPrints, ClassyFire, and the OECD Toolbox as described in '*Generation of chemical structural features and calculation of structural similarity*'. A Chi-squared test was performed comparing the number of chemicals with/without a feature and their presence/absence in a given cluster. The p-value was calculated with the function *chisq.test* from the *stats* package with options correct=TRUE and simulate.p.value=F.

Enrichment of activity in ToxCast/Tox21 bioassays: Activity data was summarized to binary hitcalls as explained in '*Comparison to targeted in vitro assay results*'. A total of 1191 chemicals had activity data for at least one bioassay. A Chi-squared test was performed comparing the number of chemicals that were active/inactive in the bioassay and their presence/absence in a given cluster. The p-value was calculated with the function *chisq.test* from the *stats* package with options correct=TRUE and simulate.p.value=F.

Enrichment of stress response pathways (SRPs): For each chemical, a literature search was conducted using PubMed (NCBI, 2018) to count how many references co-occurred between a chemical and a keyword for a specific SRP (e.g., "DNA damage response", "oxidative stress", …), relative to the overall occurrence of the chemical name and the key word. A list with all key words is included in the File S4. An SRP could have multiple keywords. For each keyword x chemical combination, the pointwise mutual information (PMI) score (Church and Hanks, 1990) was calculated as follows: PMI = $-\log_2\{$(abstracts mentioning chemical name AND keyword)/((abstracts mentioning chemical name)*(abstracts mentioning keyword)) (Chambers et al. 2022, *submitted*). PMIs below zero were set to zero. Subsequently, a one-sided Wilcoxon rank sum test using these PMIs was performed to inform whether chemicals within a given cluster had higher PMIs than chemicals outside of the cluster.

Experimental designs and analysis for follow-up experiments

Conazoles case study

The primary screen included 18 conazoles. For the follow-up study, the ToxCast chemical inventory was searched for chemicals with a high structural similarity to at least one of the 18 tested conazoles. Structural similarity was measured as the Jaccard similarity (also known as Tanimoto similarity) (Jaccard, 1902) of 729 binary ToxPrints using the R function *jaccard* from the *jaccard* package (v0.1.0). Chemicals were selected if (1) they had a Jaccard similarity of $\geq$ 0.5 with one of the tested conazoles; (2) they were among the three closest neighbors for any tested conazole, and (3) if their chemical name contained the string "conazole". This led to 69 chemicals being selected (File S2, https://comptox.epa.gov/dashboard/chemical-lists/HTPP2023_U2OS_CONAZOLES). For the 18 chemicals tested during the primary screen, the same sample (i.e., from the same source bottle) was requested, where possible, from the ToxCast chemical collection. For diniconazole, two additional samples were requested from the ToxCast chemical collection. Additionally, diniconazole, ketoconazole and difenoconazole were purchased from Sigma-Aldrich, solubilized in DMSO at the testing laboratory and tested as additional samples. All conazoles were randomly distributed among two 'plate groups', with the three purchased chemicals and the three diniconazole samples from the chemical inventory tested on both plate groups. Additionally, each plate contained three reference chemicals (dexamethasone, etoposide, ATRA) and two negative control chemicals (saccharin, sorbitol). All chemicals were tested at 8 concentrations with half-log spacing. Each assay plate contained 24 vehicle control wells (0.5% DMSO). Four biological replicates were conducted.

Samples were processed as described above, with slight modifications to the analysis: For computing global Mahalanobis distances, the eigenfeatures generated in the primary screen were used. For concentration-response modeling, the lowest two concentrations of test chemicals and negative control chemicals were used to calculate the mean and SD to define the BMR. Chemicals were considered active if they had a BMC with the 'Global Mahalanobis' approach.

Secondary screen for suspected glucocorticoid-like chemicals

Seven chemicals known to be agonists of the glucocorticoid receptor (GR) and eleven chemicals with high profile similarity were re-tested. For this purpose, new aliquots were requested from the ToxCast chemical collection, if possible, from the same sample/bottle. Chemicals were tested at six concentrations (highest concentration was 100 µM) at half-log spacing. The assay plate contained twenty vehicle control wells (0.5% DMSO). Four biological replicates were conducted.

Samples were processed as described above, with slight modifications to the analysis: For computing global Mahalanobis distances, the eigenfeatures generated in the primary screen were used. For concentration-response modeling, the solvent control wells were used to calculate the mean and SD. As opposed to the other experiments, the BMR was set at 2 SD. Chemicals were considered active if they had a BMC with the 'Global Mahalanobis' approach.

Orthogonal assay: Reverse Transcription Quantitative Polymerase Chain Reaction (RT-qPCR)

The seven chemicals known to be agonists of the GR receptor and the eleven chemicals with high profile similarity in the primary screen were tested in RT-qPCR. For each chemical, a high, phenotypically active concentration that was not cytostatic was selected (concentrations are listed in Table S2).

Experiments were conducted in 96-well format as previously described (Nyffeler *et al.*, 2022), with three vehicle control wells. Cells were harvested 24 h after applying the test chemicals, RNA reverse transcribed, and qPCR conducted using the SYBR Green Fast Advanced Cells-to-CT Kit (Invitrogen) following the manufacturer's recommendations. The qPCR was carried out in 384-well MicroAmp Optical 384-well reaction plates in duplicates. Reactions contained 1 µL cDNA, 5 µL PowerUp SYBR Green Master Mix and 250 nM of each forward and reverse primer in a total of 10 µL. Primers were designed to amplify a segment spanning two exons using the freeware All in One software (Karreman, 2002). Primers were designed for five genes (*ALOX5AP, ARL4C, CRACDL, FKBP5, PER1*) whose expression was strongly affected by dexamethasone in a previous study in U-2 OS cells (Nyffeler *et al.*, 2022). *GAPDH* served as the housekeeping gene and *CTNNB1* and *PTK2* were used as negative control genes whose expression was not affected by dexamethasone treatment. For the later three genes, the primers have been used in previous studies (Nyffeler *et al.*, 2017; Nyffeler *et al.*, 2018). All primer sequences are listed in Table S3.

Data were analyzed using the $\Delta\Delta$Ct method (Livak and Schmittgen, 2001), with GAPDH as the reference gene and using one solvent control cell culture well as reference condition. Three biological replicates were conducted; the $\Delta\Delta$Ct from these three replicates were averaged.

Supplemental data files for all phases of this study can found at https://doi.org/10.23645/epacomptox.21183481.

Results

Monitoring assay performance

To screen all 1218 chemical samples in concentration-response, a total of 116 assay plates were assessed with up to 42 chemicals per plate group, 29 plate groups and four biological replicates (Fig S1). On each plate, three reference chemicals were run in concentration-response to monitor assay performance. The chemicals were chosen from a previous study (Nyffeler *et al.*, 2022) to induce either a weak (dexamethasone), medium (ATRA), or strong (etoposide) phenotypic response in U-2 OS cells. Concentration-response modeling of phenotypic responses demonstrated repeatable concentration-response curves for all three reference chemicals (Fig S5A). For all reference chemicals and both analysis approaches (i.e., Global and Category-level) the SD of PACs was $\leq$ 0.5 an order of magnitude (Fig S5B). The PACs obtained by the Category-level Mahalanobis approach were slightly lower (i.e., more sensitive) than the ones obtained with the Global Mahalanobis approach (two-sided Wilcoxon test, $p < 0.05$). Additionally, 17 test chemicals were screened in duplicate. For both analysis approaches, the concordance of active / inactive hit calls was 94% (16/17 chemicals) for chemicals screened in duplicate (Fig S5C).

Moreover, phenotypic responses were also monitored for the reference chemicals. Each reference chemical produced a characteristic phenotypic profile. These profiles were qualitatively similar across all 29 plate groups for a given chemical (Fig S6). Biological similarity was assessed by measuring Kendall correlation of the 289 selected features for a single concentration of each reference chemical (Fig S7A). For ATRA, etoposide, and trichostatin A, the median similarity was > 0.75, while it was lower for dexamethasone (0.55) (Fig S7B). Overall, these data illustrate reproducible assay performance and response of U-2 OS cells across all plate groups in the primary screen.

Screening results

For the remainder of the analysis, results from the chemicals tested in duplicate were combined to yield results for 1199 unique chemicals. For 1151 chemicals, the number of viable cells relative to vehicle control was $\geq$ 50% at all tested concentrations. The remainder of the chemicals had between three and seven tested concentrations that were non-cytostatic (Fig 1A). Most active chemicals were active only at 10 µM or higher (Fig 1B). To estimate a false positive rate, a synthetic dataset was generated. Only 3% (9/290) of "null" chemicals were considered active with either analysis approach, and their PAC was at the upper range of the concentration series, as expected (Nyffeler *et al.*, 2021). The

two analysis approaches gave consistent results, with 41.1% and 45.8% of test chemicals being active according to the Global and Category-level Mahalanobis approach, respectively (Fig 1C). Overall, 49.3% (590/1196) of chemicals were active with at least one approach. For the remainder of the analysis, a chemical was considered active if it was active with either of the approaches. A subset of chemicals (n=437) was tested in a previous study in U-2 OS cells, but at a lower cell density (Nyffeler *et al.*, 2020). When data from the previous screen were re-analyzed using parameters consistent with the current study, PACs were within half an order of magnitude from each other for the majority of chemicals, with only few chemicals having a difference of more than one order of magnitude (Fig S8). Chemicals that were active in one screen but not the other tended to have PACs at the upper end of the tested concentration range (10 – 100 µM).

Many test chemicals in this screen have been categorized with regards to use class and described in lists on the USEPA CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard/). We hypothesized that certain chemical use groups would produce larger biological effects than other use groups in the HTPP screen. Chemicals annotated as pharmaceuticals and as active ingredients in pesticides were overall more active than chemicals not annotated as such (Fig 1D, Fig S9). On the contrary, chemicals annotated as inert, food additives and food contact chemicals tended to be less active than chemicals not annotated for these uses.

We also hypothesized that physicochemical properties would affect whether a chemical will be identified as active or not in the HTPP assay. To test this hypothesis, we developed a machine learning model using random forest modeling to predict the HTPP activity calls from 13 predicted physicochemical and environmental fate properties listed in Fig 1E (see also Supplementary Method 1). Overall, using several cross-validation methods, the model had a balanced accuracy of 76-77%. This indicates that predicted physicochemical and fate properties may relate to which chemicals are active in the HTPP assay. As these 13 input features were all predicted from SMILES and are interrelated, the value of the most important features in the model may be limited. Boiling point was the most important feature in the model, but this feature is well-correlated with average mass (0.68) and octanol air partition coefficient (logKoA) (0.84), suggesting that chemicals with higher boiling point, greater mass, and thereby greater partitioning to octanol (higher logKoA) are chemicals that are more amenable to cell-based systems and more likely to be identified as active in HTPP.

Examining the data from a more localized perspective, chemicals that were structurally similar tended to have a concordant bioactivity call more often than expected by chance (Fig S10). For example,

the long chain alkane hydrocarbons (dodecane, hexadecane, tetradecane, tridecane, pentadecane, heptadecane) were overall inactive. Though these structure-informed approaches may provide useful information *a priori* in future screens regarding chemical amenability to aqueous cell-based screening and the chances of observing bioactivity, these approaches may not predict what morphological phenotypes will be observed.

Application of potency estimates for BER analysis

One application of potency estimates derived from HTPP is their use in derivation of a lower-bound estimate for bioactivity, which in turn can be compared to human exposure predictions to identify chemicals of relatively higher concern. For all active chemicals, physicochemical properties were predicted and subsequently used to predict F_{up} and Cl_{int} (Fig 2A). These two properties are required in reverse dosimetry to convert the PAC (in µM) into an AED (in mg/kg-bw/day). The AED can then be compared to exposure estimates by computing the ratio of the two. This ratio is termed BER (Paul Friedman *et al.*, 2020; Canada, 2021).

Overall, the median $\log_{10}$(BER) was 2.9 (i.e., bioactivity occurred at least at a 1000-fold higher dose than estimated human exposures) (Fig 2B). However, for 18/412 chemicals, the $\log_{10}$(BER) was negative, indicating the potential for humans to be exposed to bioactive concentrations of these chemicals. Fig 2C plots the AED (*x*-axis) versus the predicted exposure (*y*-axis) and identifies the 18 chemicals with negative $\log_{10}$(BER) values. These included etoposide, the GR agonist dexamethasone, a pair of retinoic acid receptor agonists (ATRA, arotinoid acid), two polycyclic aromatic hydrocarbons (benzo(a)pyrene, benzo(k)fluoranthene), a food-use fungicide (cyazofamid), a triarylmethane dye (gentian violet), a disinfectant and metal chelator (sodium dimethyldithiocarbamate) and several chemicals commonly found in consumer goods (octinoxate, 4-chloro-3,5-dimethyl phenol (i.e., PCMX), 2-Ethylhexyl 4-(dimethylamino)benzoate (i.e., Padimate O) and 1,2-Dibromo-2,4-dicyanobutane (i.e., MDBGN)).

Comparison to targeted *in vitro* assay results

The majority of chemicals tested in the primary screen (89%) have already been assessed with a large battery of targeted *in vitro* assays (both cell-based and cell-free) from the ToxCast and Tox21 screening efforts (Judson *et al.*, 2010). It was of interest to us to compare the activity calls and potency estimates from HTPP to these legacy data. Of interest, chemicals that were active in the HTPP assay

595 tended to be active in a greater percentage of ToxCast assays than chemicals that were inactive in HTPP
596 (Fig 3A).

597 Overall, 83.4% of chemicals were active in the ToxCast assay battery, but only 38.2% were active
598 in HTPP (Fig 3B). Only 7/1072 (1.8%) chemicals were active in HTPP but not ToxCast. For chemicals active
599 in both platforms, the POD calculated from ToxCast assays was nearly always lower than the HTPP PAC
600 (Fig 3C). A subset of ToxCast assays are indicative of overall cytotoxicity and cell stress, commonly
601 referred to as 'burst assays'. Approximately 56.4% of chemicals were active in the burst assays (Fig 3D).
602 Of the chemicals active in HTPP, 9.2% (38/409) were not active in the burst assays. For chemicals having
603 both a burst estimate and PAC, the HTPP PAC tended to be lower by orders of magnitude for some
604 chemicals (Fig 3E). Chemicals inactive in HTPP but active in the burst assays tended to be active in only a
605 few burst assays and/or had a potency estimate close to the upper bound of the tested concentration
606 range.

607 To summarize, chemicals active in HTPP were also active in many ToxCast assays. The ToxCast
608 POD tended to be lower than the HTPP PAC, but the PAC was typically lower than the 'burst' estimate.

609 ## Feature Selection for Profile Correlation

610 Measurements from the HTPP assay can also be leveraged to compare profiles produced by
611 different chemicals. The hypothesis is that two chemicals that share the same MOA will produce the
612 same phenotypic changes, thus comparing phenotypic profiles can be used to group chemicals and infer
613 mechanism when comparisons involve mechanistic reference chemicals. Mechanistic hypotheses can
614 then be confirmed using follow-up assays as described (Thomas *et al.*, 2019). The remainder of this
615 manuscript will illustrate how this can be achieved and provide some examples and applications.

616 Not all extracted features in the HTPP assay are informative. For this study, we chose to do a
617 simple feature selection by retaining features that were reproducible among biological replicates and
618 did not have a high correlation to another feature (Fig S2). The second step in this process was
619 accomplished using stepwise feature elimination. This resulted in 289 features.

620 Next, results associated with the retained features (named a profile) were modified further by
621 censoring values that were above or below a certain threshold. The generated profiles can then be
622 compared with different correlation methods, such as Pearson or Kendall correlation or cosine
623 similarity. For this study, we compared performance of different censoring thresholds and correlation
624 methods using the reference chemicals, test chemicals tested in duplicate and "mismatches", i.e.,

correlation of profiles of non-identical chemicals (Fig S11). Based on these results, Kendall correlation was selected with a censoring threshold of 1.

Profile comparison of nuclear receptor (NR) modulators

The approach of leveraging Cell Painting profiles of reference chemicals to generate hypothesis regarding MOA of test chemicals is a well-accepted practice with multiple examples in the literature (Hughes *et al.*, 2011; Melillo *et al.*, 2018; Schneidewind *et al.*, 2020; Rohban *et al.*, 2022). To determine whether the HTPP assay in U-2 OS cells can be used to identify chemicals that share the same MOA, a chemical set with annotated mechanisms was needed. Among the tested chemicals, 58 chemicals were annotated in the literature as modulators of a NR. A modulator can be either an agonist, antagonist, or an unspecified direction of the interaction. Of those 58 chemicals, 50 were active in HTPP and could be phenotypically compared. Modulators of GR and modulators of the retinoic acid and retinoid X receptors (RAR/RXR) each produced a characteristic profile that was different from the profiles observed for other NRs (Fig. 4A). For some NRs, such as androgen receptor (AR), estrogen receptor (ER) or peroxisome proliferator-activated receptor (PPAR), no characteristic pattern was observed.

We hypothesized that some of these NRs might not be expressed in U-2 OS cells. Upon inspection of transcriptome data from the Human Protein Atlas (https://www.proteinatlas.org/) (Uhlen *et al.*, 2017), we confirmed that GR, several RAR/RXR isoforms, and several PPAR isoforms were expressed, but ER isoforms were not expressed (Fig. 4B). If a receptor is not expressed, the test chemicals cannot interact with the receptor; thus, the observed phenotypes are likely the result of 'off-target' effects. Despite ER not being expressed in U-2 OS cells, many tested chemicals with known estrogenic activity were active in HTPP. We then investigated how the HTPP PAC related to potency estimates from the ToxCast assay battery in such cases. As expected, HTPP PACs were higher than the ToxCast POD for chemicals modulating ER, but were similar for chemicals modulating RAR/RXR, and partly similar for GR modulators (Fig. 4C).

To summarize, we saw characteristic profiles for GR and RAR/RXR modulators, but not for modulators of other NRs. For those two receptor types, HTPP gave potency estimates comparable to the ToxCast assay battery.

Application: Identification of potential GR modulating chemicals

It is well established that image-based phenotypic profiling can identify MOA for drug-like chemicals, and in the previous section, we showed that this is also true for HTPP in U-2 OS cells.

However, to apply HTPP in chemical hazard evaluation for environmental chemicals, it is of greater interest to identify possible specific mechanisms of non-drug-like chemicals. As an example, we chose to focus on chemicals targeting the GR pathway. Thirteen chemicals induced profiles similar to the seven known glucocorticoids (Fig 5, left), albeit with a lower correlation than the known glucocorticoids to each other. To confirm these initial findings, the HTPP assay was repeated with a fresh chemical library. Overall, the same pattern was observed (Fig 5, middle). The highest correlation values were obtained with pyrene, followed by N-hydroxybenzamide and benfluralin. Profile similarity is not definitive proof of interaction with the GR pathway; hence, we performed qPCR on selected genes. The seven known glucocorticoids produced an upregulation of *ALOX5AP*, *CRACDL*, *FKBP5* and *PER1*, and a downregulation of *ARL4C* (Fig. 5, right, Fig S12). Only one test chemical, pyrene, induced a similar change in gene expression, indicating that it likely acts in a similar manner to glucocorticoids in U-2 OS cells.

Structural similarity translates to biological similarity

The results presented thus far showcase how HTPP can be used to identify chemical bioactivity that is exerted through a specific MOA, such as activation of a NR. However, our interest is more in environmental chemicals. The chemical library contained many structurally related chemicals, therefore, we set out to investigate whether structurally related chemicals would produce biologically similar responses in U-2 OS cells. For this purpose, we clustered all active chemicals using structural information. We found that chemicals that co-clustered based on structural information had higher biological similarity than expected by chance (Fig 6A-C). Some examples are highlighted in the following (Fig 6D-H, Fig S13).

For example, cluster 22 comprises five organochlorine pesticides, all of which have pair-wise biological similarity scores of greater than 0.6 (Fig 6D). Their phenotypic profile indicated that the HTPP DNA channel was particularly affected (Fig. S13A). Inspection of the microscopy images revealed that cells treated with dieldrin have brighter and less bright areas within the same nucleus (i.e., greater texture), while control cells tended to have a more even pixel intensity within nuclei (Fig. 6E). Cluster 23 contains four chloroacetamide herbicides and the benzamide fungicide zoxamide. The minimum pair-wise biological similarity score (0.4-0.6) between chemicals in this cluster was lower than for organochlorines, but it was still greater than random (Fig 6F, Fig S13B).

However, not all biological activity was explained by structural similarity. For example, cluster 13 contained phthalates, a group of chemicals typically used as plasticizers. Only two of the five structures (dibutyl phthalate and benzyl butyl phthalate) resulted in high biological similarity (0.68), with all other

profiles having low similarity with one another (-0.07 – 0.26) (Fig 6G, Fig S13C). A peculiar example is carvone in cluster 15 (Fig. 6H, Fig S13D). The two stereoisomers produce completely different profiles in U-2 OS cells, with S-(+)-carvone being phenotypically similar to the other structurally related chemicals, and R-(-)-carvone inducing a different profile. Taken together, these results demonstrate that structural similarity can be associated with similarity in biological activity in the HTPP assay, but not all biologically similar effects can be explained by chemical structure.

Grouping of chemicals based on phenotypic profiles

Next, we wanted to examine the phenotypic results from a broader perspective. We clustered all active chemicals based on their profiles (Fig S14) and displayed them in a correlation matrix (Fig 7A). Qualitatively, two to three large clusters containing dozens of chemicals were present along with a few small clusters with high correlation within them. We hypothesized that chemicals in specific clusters would share certain features such as chemical fingerprints, activity in certain targeted ToxCast assays or activation of stress response pathways (SRPs). For this purpose, the dendrogram was cut into 15 clusters. Then, each cluster was evaluated with regards to whether chemicals within the cluster were enriched in chemical fingerprints, ToxCast assay activity or association with SRPs compared to all other tested chemicals that were not in that particular cluster. Results are summarized qualitatively in a table (Fig S15) and quantitatively in heatmaps (Fig S16).

For example, cluster 10 contained glucocorticoids, and the cluster was enriched in chemicals with a 21-hydroxysteroid structure. Consistently, chemicals in that cluster were more likely to be active in GR agonism assays than chemicals not in that cluster. Similarly, cluster 11 contained retinoids and ToxCast assays targeting RARs were enriched. Cluster 6 was the largest cluster and contained multiple conazoles, parabens and phenols. The chemotype 1,2,4-triazole was enriched as well as bioassays measuring CYP activation from multiple platforms (i.e., assay providers). Cluster 3 had significant similarity to cluster 6 in the correlation matrix. It also contained conazoles, but different chemotypes and bioassays were enriched. Cluster 2 was enriched in quinones and epoxides and included chemicals like etoposide, hydroxyquinone, hydroxyurea and carbaryl. Enriched bioassays came mostly from the Bioseek platform, and multiple assays measured downregulation of VCAM1. We hypothesized that the large clusters (i.e., 2, 3, 4 and 6) might be characteristic for different stress responses. However, of those only clusters 2 and 3 were enriched in chemicals known to exert these responses. Additionally, clusters 10 and 13 were enriched in stress responses. Interestingly, these two clusters were located next to cluster 2 in the heatmap.

To summarize, we observed several different clusters of profiles. But not all chemicals resulted in unique profiles – many resulted in more general, broad profiles. However, these phenotypic groupings contained information, such as common underlying chemotypes or activity in the same targeted bioassay. Interestingly, some clusters were characterized by more shared chemotypes, while others shared activity in bioassays or stress responses (Fig 7B).

Application: Grouping of conazoles

The chemical test set contained 18 conazoles, of which 16 were bioactive in U-2 OS cells. Upon closer inspection of Figure 7, we noted that diniconazole was located in cluster 7, while most other conazoles were located in clusters 3 and 6. This finding was confirmed when looking at the phenotypic profiles and biological similarity: diniconazole produced a different profile in U-2 OS cells compared to all other (bioactive) conazoles that had a similar profile amongst them (Fig 8A+B). We are not aware of any structural features that would explain this drastic difference. Investigation of results from targeted *in vitro* assays and *in vivo* data did not indicate a different bioactivity pattern for diniconazole (data not shown). We were curious whether other conazole-like chemicals would be phenotypically similar to diniconazole. Thus, we selected an additional 41 chemicals with structural similarity to one of the previously tested 18 conazoles and conducted the HTPP assay in U-2 OS cells as a secondary screen. Of the 69 tested chemicals, 42 were bioactive. However, none of the other tested conazole-like chemicals produced a similar profile to diniconazole using a threshold value of 0.6 (Fig 8C). In this screen, diniconazole had the highest biological similarity with triadimefon (0.60), followed by hexaconazole (0.55) and difenoconazole (0.52). Of note, to exclude the possibility that the diniconazole sample from the primary screen was contaminated or wrongly annotated, we obtained multiple samples of diniconazole from two different providers. All diniconazole samples consistently reproduced the distinct profile (Fig S17). As diniconazole did not produce a phenotypic profile similar to other conazoles, we sought to probe its MOA in U-2 OS cells. Leveraging data from the primary screen, we identified the thirty chemicals with the highest phenotypic similarity to diniconazole (Fig 8D). Phenolphthalein was the most similar (0.55), with multiple estrogens (17alpha-ethinylestradiol, 17beta-estradiol, diethylstilbestrol) and inhibitors of mitochondrial respiration (fenpyroximate, rotenone) being in the top 20. Inspection of the microscopy images confirmed, that the phenotype produced by diniconazole is similar to the phenotype produced by phenolphthalein (AGP stain aggregated around the nucleus, cells are large and polygonal) and different from the one produced by ketoconazole (Fig 8E). Many of the listed chemicals are known to be modulators of microtubules (Metzler and Pfeiffer, 1995; Kipp and

748 Ramirez, 2003; Srivastava and Panda, 2007; George *et al.*, 2008; Heard *et al.*, 2013; Adamakis *et al.*,
749 2019). Hence, we hypothesize that diniconazole might affect microtubule dynamics in U-2 OS cells.
750 Overall, this case study shows how structural and phenotypic information can be combined to group
751 chemicals with similar activity in cells, as well as to highlight chemicals that produce a different activity.

752

Discussion

One of CCTE's objectives is to generate publicly accessible chemical bioactivity data to inform chemical hazard evaluation and risk assessment. Previously, the ToxCast program had used a large assortment of targeted HTS assays to generate *in vitro* bioactivity data across many molecular targets (Kavlock *et al.*, 2012), primarily using contract mechanisms. This was an expensive and time-consuming endeavor that took many years. More recently, a revised strategy for *in vitro* hazard evaluation was proposed that envisions the use of more comprehensive and efficient molecular profiling assays such as Cell Painting (and HTTr), as opposed to large collections of targeted HTS assays, as a first-tier approach for generating bioactivity data for environmental chemicals (Thomas *et al.*, 2019). As these assays are compatible with many different human-derived cell lines, they can be used to broadly survey different aspects of human biology using cell line collections and potentially detect chemical bioactivity at molecular targets not represented in ToxCast. For laboratories with appropriate infrastructure (i.e., a high content imaging platform, automation-assisted sample preparation and networked data storage), using the Cell Painting assay to screen large chemical collections (100s to 1000s of chemicals at multiple concentrations) across multiple cell lines may be more a time and cost-effective (Hughes *et al.*, 2011; Svenningsen and Poulsen, 2019; Chandrasekaran *et al.*, 2021; Marin Zapata *et al.*, 2023) means of establishing *in vitro* PODs compared to testing those same sets of chemicals in a collection of targeted HTS assays. We previously evaluated the Cell Painting assay towards that purpose (Nyffeler *et al.*, 2020). Here, we expand upon that work by further demonstrating the reproducibility and robustness of the Cell Painting assay and demonstrate its utility for use in risk assessment in three different contexts: 1) BER analysis, 2) inference of molecular MOA and grouping of chemicals to inform chemical read-across.

Screening considerations

Inclusion of reference chemicals on each assay plate allowed for monitoring of assay performance and indicated that the HTPP assay yielded highly reproducible results within the screen. Repeated estimation of the PAC in each plate group resulted in values typically within half an order of magnitude. This is consistent with our previous study (Nyffeler *et al.*, 2020). Moreover, PACs were also consistent among two studies conducted in U-2 OS cells at two different cell seeding densities (Fig S8).

Overall, half (49.3%) of the tested chemicals were active in the HTPP assay. While this hit rate may seem high, this was expected because HTPP is intended to be a broad Tier 1 screening assay that can detect many types of biological activity. Moreover, the HTPP analysis approach was geared toward sensitivity (Nyffeler *et al.*, 2021) and thus a few false active calls are to be expected in the list of active

784 chemicals. However, using information from the 'null chemicals', each inactive chemical had only about
785 a 3% chance of being falsely called active. This becomes apparent when comparing the hit calls for the
786 global and category-level analysis approaches (Fig 1C). There were some discordant hit calls among the
787 two methods, but they tended to be chemicals with PACs close to the upper bound of the tested
788 concentration range. These chemicals were likely 'borderline actives', i.e., chemicals with low
789 magnitudes of effect that barely exceed the threshold for biological activity applied during
790 concentration-response modeling. Depending on the analysis approach these chemicals were either
791 determined to be active or inactive. As expected, pharmaceuticals and pesticides had overall higher
792 effect size compared to chemicals found in foods (Fig 1D). It should be noted that a chemical being
793 active in HTPP is not proof of biological adversity or hazard. Moreover, many chemicals were active at >
794 10 µM, a concentration that might not be reached in plasma at steady-state in real-life human exposure
795 scenarios. Hence, a comparison of bioactive concentrations with exposure estimates was important (Fig
796 2).

797 Using just 13 physicochemical properties enabled the prediction of binary HTPP activity (active
798 vs. inactive) with 76% balanced accuracy. This could indicate that some chemicals have physicochemical
799 properties outside the domain of applicability for screening in this assay. For example, semi-volatile
800 chemicals or chemicals with poor solubility in DMSO would be expected to appear inactive in the assay
801 despite their potential to be bioactive. Alternatively, the U-2 OS cells may lack expression of important
802 molecular target(s) for some chemicals or classes of chemicals. Therefore, the machine learning model,
803 which relies on physicochemical properties of the test chemicals but not biological properties of the *in*
804 *vitro* model, may be insufficient for predicting activity versus inactivity with absolute accuracy. Future
805 inclusion of additional information in a machine learning model, such as presence or absence of
806 moieties in the test chemicals, presence or absence of molecular target expression in the U-2 OS cells,
807 and how the moieties and molecular targets interact may improve the predictions made by the machine
808 learning approach.

809 ## Application of HTPP for potency estimation and BER analysis

810 One potential application of HTPP screening results is their use as input for a BER analysis. Such
811 an analysis has already been undertaken using HTS data (Paul-Friedman *et al.*, 2019; Canada, 2021). The
812 previous study using HTS data found 11/448 chemicals with a negative $\log_{10}$(BER), while in the present
813 study using HTPP data 18/412 active chemicals had a negative $\log_{10}$(BER). It should be noted that for a
814 subset of chemicals, exposure at bioactive concentrations is expected, namely for drugs (e.g., etoposide,

dexamethasone). Most of the other chemicals with negative log_{10}(BER) identified in the present study are found in consumer products which would also contribute to comparatively higher exposure estimates in human populations. Whether a negative log_{10}(BER) is of concern depends on the specific chemical, chemical use scenario, and on the exposed population. For example, for a pesticide, overlap of bioactivity with human exposure is likely undesirable. For a food additive, overlap of bioactivity with human exposure may be expected and may be benign or beneficial. In contrast to Paul-Friedman *et al.* (2019), who used measured toxicokinetics information, we used predicted physicochemical properties to calculate F_{up} and Cl_{int}, even for chemicals where empirical F_{up} and Cl_{int} were available. Thus, all chemicals were considered uniformly in this regard. This approach was chosen as HTPP is of higher throughput than generation of toxicokinetic data. Hence, in order to put those high-throughput bioactivity data into perspective, high-throughput toxicokinetics approaches are needed, which at the moment are only *in silico* based. The advantage is that such predicted values are available for most chemicals. One disadvantage is that the predicted F_{up} and Cl_{int} values have more uncertainty compared to those that are derived empirically, which in turn results in additional uncertainty in the IVIVE prediction. Additional uncertainty comes from the IVIVE modeling procedure. Other sources of uncertainty are the choice of the lower 5^{th} population percentile for exposure (instead of, e.g., a median) and, as outlined above, the possibility of false actives. For example, 3-hydroxy-2-naphthanilide is used in textile coloring (Roed-Petersen *et al.*, 1990) and is also used as a fluorescent dye in science (Vaughan *et al.*, 1971), hence it might be a false active due to autofluorescence. Despite known uncertainties and limitations in the IVIVE approach, calculating the BER can help prioritize chemicals for additional information gathering or higher tier testing.

Many of the chemicals tested in the present study were bioactive in Toxcast assays, a collection of hundreds of HTS assays of various types conducted over many years. Despite HTPP (in a single cell type) being less sensitive (Fig 3), the results indicate that HTPP could be used as a less expensive and less time-consuming alternative to the ToxCast assay battery approach for identification of bioactive chemicals of various types (Fig 1D). HTPP screening in additional cell types beyond U-2 OS is anticipated to increase the proportion of chemicals identified as bioactive and improve concordance with ToxCast. HTPP was more sensitive than the lower bound of the ToxCast burst estimate which indicated that HTPP measures cellular changes in the absence of cytotoxicity. Most chemicals that had a burst estimate but were inactive in HTPP were active in only a few burst assays, which is indicative of a higher degree of uncertainty as to whether these are authentic cytotoxicants or may require activation by metabolic enzymes that U-2 OS cells lack in order to exert cytotoxicity. Overall, ToxCast assays were more sensitive

than HTPP, which was expected. First, ToxCast assays measure biological activity at many different molecular targets and are thus capable of detecting MOAs that are absent in U-2 OS cells. Second, several ToxCast assays are cell-free assays, which may be more sensitive with respect to bioactive potency than assays where chemicals can bind to various cell components (e.g., proteins, lipids), or assays using intact cells with some resilience / adaptive response capacity. Third, it is consistent with our previous study, where we found that ToxCast assays produced more conservative (i.e., lower) AED estimates than the HTPP assay relative to *in vivo* effect data. The HTPP assay yielded a conservative or comparable surrogate for *in vivo* effects for 68% of chemicals (Nyffeler *et al.*, 2020). Interestingly chemicals inactive in HTPP were often active in only a few ToxCast assays (Fig 3A). This suggests that these chemicals are either biologically inactive and are false actives in the respective ToxCast assays, or they could have a very specific MOA that is not capable of being detected in the U-2 OS HTPP assay. Conversely, chemicals active in HTPP, particularly those with large effect sizes in HTPP, were often active in many ToxCast assays. This could indicate that the chemicals are 'pan-active', i.e., affecting multiple cellular pathways in a relatively unspecific manner.

Application of HTPP for identification of specific MOA

Multiple studies have leveraged phenotypic profiling for identification of MOA (Ramm *et al.*, 2019; Warchal *et al.*, 2019; Trapotsi *et al.*, 2021), to predict outcomes of *in vitro* bioassays (Simm *et al.*, 2018; Hofmarcher *et al.*, 2019; Way *et al.*, 2021) or organ toxicity (Su *et al.*, 2016; Ramm *et al.*, 2019). Way et al. concluded that a wealth of information is captured in Cell Painting, even without staining for a particular cell health event (Way *et al.*, 2021). This is consistent with findings by Simm et al. who repurposed data from a targeted high content screen to predict biological activity for other, not related targets (Simm *et al.*, 2018). Similarly, Hofmarcher et al. predicted outcomes of 66 assays from cell morphology changes and Trapotsi et al. predicted 90 targets (MOA) using Cell Painting data (Hofmarcher *et al.*, 2019; Trapotsi *et al.*, 2021). Our present study is in line with these findings: we identified a novel glucocorticoid-acting chemical using HTPP, and have supporting evidence that diniconazole may affect microtubules, without staining for these two specific mechanisms.

Many of these studies used drug-like chemicals with a specific, known MOA (Woehrmann *et al.*, 2013; Hofmarcher *et al.*, 2019; Warchal *et al.*, 2019; Trapotsi *et al.*, 2021). To our knowledge, the present study is the first one to apply HTPP to environmental chemicals in order to derive mechanistic information. While our approach measures 1300 features, not all features were reproducible or informative. A multitude of strategies for feature selection/reduction and profile correlation are

878 possible. Here, we performed a feature selection where features with low reproducibility were
879 removed, similar to the procedure described in Woehrmann *et al.* (2013), followed by a stepwise feature
880 elimination, similar to the procedure as described (Bougen-Zhukov *et al.*, 2017; Caicedo *et al.*, 2017;
881 Rohban *et al.*, 2017; Warchal *et al.*, 2019; Chen *et al.*, 2020). We chose this feature selection approach
882 because the interpretability of the remaining features is retained, as opposed to approaches such as
883 feature reduction that produces new latent variables. We further selected Kendall correlation followed
884 by unsupervised hierarchical clustering to facilitate grouping of chemicals without prior knowledge of
885 the expected profile (see 'Introduction' for more explanation). Unlike many other studies, we had to
886 integrate data from multiple chemical concentrations. We chose to use information from the lowest
887 three active test concentrations that gave reproducible phenotypic profiles (across the independent
888 cultures). The goal was to (1) select a high enough concentration to obtain a profile that is characteristic
889 for that chemical, and (2) select low enough concentrations to capture the most potent mechanism, as
890 opposed to general cell stress effects that might be induced at higher chemical concentrations. We
891 chose three concentrations as this covers one order of magnitude and we observed that for the
892 reference chemicals the profiles were fairly consistent in that range but became more robust with
893 increasing concentration.

894 To gain confidence in our approach, we first applied it to "drug-like" chemicals, i.e., chemicals
895 with a specific, known MOA, such as NR modulators. We observed that multiple glucocorticoids and
896 retinoids each shared a characteristic profile, but this was not the case for modulators of some other
897 receptors (e.g., ER, PPAR) (Fig 4). We further determined that receptors for retinoids and glucocorticoids
898 are expressed in U-2 OS cells, whereas ERs are not. Of note, many estrogen modulators produced a
899 phenotypic profile in U-2 OS, i.e., were bioactive, albeit at concentrations much higher (3 – 100 µM)
900 than anticipated for potent ER modulators (e.g., 17alpha-ethinylestradiol, 17beta-estradiol). With the
901 receptor being absent, this indicates that the observed profiles are likely a consequence of "off-target"
902 activity. PPAR modulators did not produce a characteristic profile although several PPAR subtypes are
903 expressed in U-2 OS cells. It should be noted that 4/11 PPAR modulators were inactive in HTPP
904 altogether. The tested modulators target different receptor subtypes, either as agonists or antagonists.
905 Hence there might not have been many chemicals with the same molecular target and directionality of
906 effect. Moreover, it is possible that other components of the PPAR signaling pathway are not expressed
907 in U-2 OS cells or that the modulators affect the cells (e.g., changing gene expression) but without a
908 change in profile. From all these findings, we conclude that in order to have confidence in predicting
909 receptor-mediated MOAs and associated potencies, two conditions need to be fulfilled: (1) the

corresponding target has to be expressed in the tested cell type; and (2) multiple reference chemicals targeting the same receptor need to result in the same profile to give confidence that the observed profile is indeed a result of an on-target receptor-mediated effect. Secondary (Tier 2) targeted assays could then be used to confirm the MOAs predicted from the HTPP profile comparisons.

As a proof-of-concept, we identified 13 chemicals that produced a similar profile to glucocorticoids in the primary screen and confirmed the phenotypic response in an independent HTPP experiment as a secondary screen. For most tested chemicals, the results between the two experiments were consistent (Fig 5). In both experiments, pyrene had the most consistent similarity with the seven glucocorticoids. We conducted qPCR as a targeted, orthogonal (Tier2) assay and confirmed that pyrene upregulated the same genes as the known GR modulators.

Pyrene is a planar polycyclic aromatic hydrocarbon consisting of four fused benzene rings. It has some structural similarity with dexamethasone and other GR modulators in Fig 4 and Fig 5, which all contain a steroid backbone (i.e., four fused non-aromatic rings in a planar structure). We did not find literature evidence of pyrene's activity on the glucocorticoid pathway. Pyrene was tested in 988 ToxCast assay endpoints and was active in 57 of them. It was inactive in the GR assay from the Tox21 and Attagene platform and showed some activity in the H295R steroidogenesis assay. Of interest, Wang *et al.* (2009) reported that dexamethasone and polycyclic aromatic hydrocarbons crosstalk between the aryl hydrocarbon receptor and the GR. While pyrene did not induce the glucocorticoid response element (GRE) alone, in the presence of dexamethasone it induced the GRE more than dexamethasone alone. Thus, one could speculate that in the U-2 OS cell system a minimal level of glucocorticoids is present (Honn *et al.*, 1975; Cao *et al.*, 2009) that allows observation of GR pathway activation upon exposure to pyrene. This hypothesis is consistent with pyrene's inactivity in ToxCast GR assays: these assays measure direct interaction with GR or activation of the GRE in an engineered cell system. However, according to Wang *et al.* (2009), polycyclic aromatic hydrocarbons do not interact with the GR pathway in a direct manner.

All other candidate chemicals did not induce characteristic gene expression changes associated with GR activation (Fig 5). These results demonstrate that a shared profile does not necessarily mean a shared underlying mechanism. In fact, there could be different molecular effects that lead to a change in some of the same phenotypic features, and thus could lead to a seemingly high correlation. This is consistent with previous findings by Willis *et al.* (2020): amperozide, fluphenazine and tetrandrine produced similar profiles but have different known molecular targets and MOA. To summarize, from

1200 chemicals, we identified 13 candidates of which one chemical was confirmed for novel GR activity. This illustrates the power of HTPP for screening environmental chemicals: thousands of chemicals can be screened for multiple mechanisms at the same time, and candidates can be followed up with targeted assays in a hypothesis-driven manner, which is cost- and resource efficient.

Application of HTPP for grouping of chemicals

As shown in Fig 1 and Suppl. Fig S10, physicochemical properties and chemical structure can predict whether a chemical is active in the HTPP assay. But moreover, chemical structure also affects to some extent the profile that is produced (Fig 6): Chemicals that are structurally similar tend to produce similar profiles. However, this is not always the case. For example, R-(-)-carvone and S-(+)-carvone are stereoisomers. With the chosen structural feature descriptions, which do not take into account stereochemistry, they are 100% identical, but their phenotypic profiles are different from each other. This is interesting, because the two stereoisomers have different biological activities: R-(-)-carvone has a spearmint-like odor, while S-(+)-carvone has a caraway-like odor (Leitereg *et al.*, 1971) due to the differential interaction of these molecules with olfactory receptors (Krautwurst *et al.*, 1998). A similar phenomenon may be occuring in U-2 OS cells where R-(-)-carvone and S-(+)-carvone interact with distinct, and currently unknown, molecular targets to produce distinct profiles. This is a powerful example demonstrating how HTPP can enrich structural information to predict biological activity.

When looking from a broader perspective, we found that while there are profiles specific for certain MOAs, many of the active chemicals produced one of a few profiles (Fig. 7). This is consistent with findings from Young *et al.* (2008) who profiled chemicals for effects on nuclear morphology. We hypothesized that the larger clusters would be representative of more generalized forms of cellular stress. This seemed to be the case for clusters 2,10 and 13, which are adjacent clusters in Fig 7. This is interesting, because the clustering utilized only phenotypic information, and the literature search for positive associations with cellular stress was an independent exercise. In contrast to our expectation, the largest cluster (cluster 6) was not enriched in stress response pathways. Chemicals in cluster 6 were overrepresented among the actives in CYP assays (Fig. S15, S16). Hence, we hypothesize that cluster 6 is representative of chemicals activating xenobiotic metabolism in the absence of a generalized stress response. Overall, the automated literature search supported the hypothesis and, hence increased confidence that the HTPP assay can distinguish stress-driven phenotypic responses from responses occurring due to other mechanisms. Future work will aim to further characterize these stress phenotypes using other data streams, such as transcriptomics.

Using conazoles as a case study, we illustrated how HTPP could be used to group environmental chemicals (Fig 8). In the primary screen, diniconazole's phenotypic profile was different from that of all other active conazoles. We confirmed this finding in an independent follow-up experiment including multiple samples of diniconazole, as well as additional structurally related conazoles. The observations did not provide information on the MOA of the underlying profile for diniconazole, but leveraging the breadth of data obtained from HTPP allowed us to form a hypothesis: many of the chemicals with high profile similarity to diniconazole have literature evidence for modulating tubulin polymerization. Phenolphthalein, the chemical with the highest profile similarity, induces tubulin polymerization abnormalities (Heard *et al.*, 2013). Thiophanate methyl (ranked 6th most similar) is a fungicide whose MOA is inhibition of microtubule assembly (Casida, 2009). Rotenone (ranked 12th) was found to inhibit microtubule assembly in cell lines through tubulin binding at sub-micromolar concentrations (Srivastava and Panda, 2007). Several estrogenic compounds (17alpha-Ethinylestradiol, 17beta-Estradiol, diethylstilbestrol, bisphenol B) had high profile similarity with diniconazole. However, we do not think that diniconazole acts in an estrogenic manner in U-2 OS cells. First, estrogen receptors are not expressed in U-2 OS cells (Fig 4B), second, not all estrogenic compounds produced the same profile (Fig. 4A) and third, estrogens were active at relatively high concentrations (3-100 µM) compared to the concentration typically needed to activate its receptors (pM to nM). Instead, there is evidence for estrogens, diethylstilbestrol and bisphenol A to directly affect microtubule polymerization (Metzler and Pfeiffer, 1995; Kipp and Ramirez, 2003; George *et al.*, 2008; Adamakis *et al.*, 2019). We did not find literature evidence for an effect of diniconazole on microtubules. Thus, future experiments using targeted assays that measure microtubule dynamics should explore this hypothesis. With regards to other conazoles, two hypotheses are possible: (A) diniconazole has a different MOA in U-2 OS cells than the other active conazoles; and (B) diniconazole has the same MOA compared to other active conazoles and in addition has a secondary biological activity, such as affecting microtubule polymerization. According to Fig 8D, there is some limited similarity between diniconazole and other conazoles, which would lend support to the second hypothesis.

Considerations for future application of HTPP

Using multiple examples, we have outlined in the present study how HTPP could be used in next generation risk assessment, specifically BER analysis, generating MOA predictions that could be used in weight-of-evidence assessments, and grouping chemicals based on profile similarity – a procedure with potential applications in chemical read-across. The described concepts can all be applied to various cell

types; indeed, Thomas *et al.* (2019) envisioned the application of HTPP to multiple cell types. This will provide coverage of more biological space compared to just one cell line, as different receptors and pathways are expressed in different cell types. As shown in Fig 4, expression of the pathway of interest is necessary but not sufficient to see characteristic profiles that can point to putative MOA. The ability to infer biological activity at multiple receptors in using HTPP is advantageous from an efficiency perspective, as multiple targeted HTS assays would be needed to achieve a similar amount of information on the biological activity of a chemical set.

In the near future, HTPP could be used in a tiered hazard evaluation strategy, where structural information is used as the starting point in tier 0. Chemicals that fall within the applicability domain of HTPP (Fig 1) could then be tested in multiple cell types in HTPP. One could also consider leveraging the structural information to test only a subset of structurally related chemicals, as structurally related chemicals tend to have related biological activity (Fig S10, Fig 6). HTTr is another high-throughput profiling platform envisioned as a part of Tier 1 hazard evaluation, but it is less cost-effective compared to HTPP. Previous studies comparing Cell Painting and transcriptomics datasets have observed that they contain both redundant and unique information that be leveraged for MOA prediction (Nassiri and McCall, 2018; Haghighi *et al.*, 2022; Seal *et al.*, 2022; Way *et al.*, 2022). A previous study with a small collection of 11 chemicals demonstrated that potency thresholds from HTPP and HTTr in U-2 OS cells tended to be close to one another, within order of magnitude (Nyffeler *et al.*, 2022). However, further work is needed to establish whether HTTr can be used to test chemicals or treatments with no activity in HTPP or whether HTPP can be used to reduce the testing burden of HTTr within Tier 1 of a tiered hazard evaluation strategy. One possibility is conducting HTPP screening first as a "Tier 1a" assay to help determine appropriate chemical-by-chemical concentration ranges for more expensive HTTr screening as a "Tier 1b", using the HTPP result to benchmark the upper end of the HTTr dose range based on biological activity. HTPP chemical activities would fall into one of three groups: (1) no bioactivity detected at the tested concentrations; (2) specific bioactivity detected at concentrations below 'general cytotoxicity', e.g., activities at certain receptors, such as GR in the present study (Fig 4, 5); (3) nonspecific bioactivity detected. Information from the two first tier platforms could then be combined to identify which chemicals should be selected for targeted follow-up assays. Chemicals in group (3) would be of lesser priority for investigation of putative receptor-based mechanisms, because according to Fig 3A, they are most likely pan-active chemicals and despite being active in many targeted assays might not have a defined target. Chemicals in group (2) would be of highest priority, as there is evidence for specific bioactivity that should be confirmed with targeted assays. Chemicals in group (1) may also be

1035 further investigated. Absence of bioactivity in Tier 1 screening in a particular cell type is not necessarily
1036 predictive of absence of bioactivity in a different cell type. If a chemical continues to test as negative for
1037 Tier 1 bioactivity across a variety of biologically diverse cell types, the weight of evidence grows that the
1038 chemical is not bioactive within the concentration ranges tested and does not pose a substantial hazard
1039 to humans. Such an accumulation of negative bioactivity data is not available in the scientific literature
1040 at present. For all three groups, the efforts in higher tier studies should be integrated with exposure
1041 information (Fig 2) (Baltazar *et al.*, 2020) to provide context for these screening data in next-generation
1042 risk assessments.

1043

Funding Information

The USEPA through its Office of Research and Development provided funding for this research. J.N. was supported by an appointment to the Research Participation Program of the USEPA, Office of Research and Development, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the USEPA.

Conflict of Interest

The authors declare no conflict of interest. This manuscript has been reviewed by the Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Acknowledgements

The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. This research was funded by the US EPA Office of Research and Development under the Chemical Safety for Sustainability Strategic Research Action Plan FY2019-FY2022.

The authors would also like to thank Terri Fairley, Daniel Hallinger, and Sandra Roberts for operations support activities during conduct of this research. Finally, the authors would like to thank Drs. Joseph Bundy, Richard Judson, Kimberly Slentz-Kelser and E. Sidney Hunter for their insightful comments during review of this manuscript.

Figure Legends

Figure 1: Overview of screening results.

(A) Graphical representation of the number of non-cytostatic concentrations for each test chemical. Each chemical was tested at 8 concentrations (typically from 0.03 – 100 µM). A reduction of cell counts by more than 50% was considered cytostatic. Most chemicals (1151/1199) were not cytostatic in the tested concentration range. (B) PACs for all active test chemicals (blue data points) with the two analysis approaches. A synthetic data set consisting of 290 "null" chemicals was generated (gray circles). PACs extrapolated below the tested concentration range are depicted as triangles. (C) Concordance of the two analysis approaches (Global and Category-level Mahalanobis). The scatter plot displays the PACs for both approaches for all chemicals active according to at least one approach. The inset Venn diagram indicates the number of active chemicals for each approach. (D) Distribution of effect size values for chemicals depending on their presence ("Y") or absence ("-") in chemical lists. The value below the box indicates the number of tested chemicals that are a member of the indicated list. Presence in lists information was obtained from the CompTox Chemicals Dashboard for 1207 chemicals. The effect size value corresponds to the 'top over cutoff' from the Global Mahalanobis curve fit. For each list, a two-sided Wilcoxon rank sum test was performed by comparing the distribution of effect size values of chemicals in the list with the ones from chemicals not in that list. *P*-values in red indicate a higher group median of chemicals in the list vs. chemicals not in the list; green *p*-values indicate the opposite. Note that values >20 were adjusted for graphical display, but the statistical analysis was performed on the original values. (E) OPERA predictions for 13 physicochemical properties were used to build Random Forest (RF) models to predict activity in the HTPP assay. Multiple RF models were built and tuned (see Supporting Information). The confusion matrix for the winning model is shown, which lead to a balanced accuracy of 76%.

Figure 2: Application of potency estimates for BER analysis.

(A) Schematic overview outlining the prediction of metabolic parameters (F_{up}, Cl_{int}) to use in reverse dosimetry to convert the HTPP PAC (*in vitro* concentration, in µM) to an administered equivalent dose (AED, in mg/kg-bw/day). The BER is then calculated from the ratio of the lower bound (5^{th} percentile) of the HTPP AED and the upper bound (95^{th} percentile) of the exposure estimate. (B) Distribution of $\log_{10}$(BER) for 412 chemicals that were active in HTPP, had the physicochemical parameters available to estimate the AED, and had exposure estimates available. The gray dotted line indicates the median of the distribution. The black dashed line indicates the unity line. Chemicals to the

left of the unity line have an AED below the upper bound of the exposure estimate. (C) Scatter plot of the bioactivity estimate (*x*-axis) versus exposure estimate (*y*-axis) for the chemicals in B. Points are colored by their $\log_{10}$(BER). Chemicals with a negative $\log_{10}$(BER) are labeled. The solid, dashed, and dotted lines indicate $\log_{10}$(BER) of 0, 1, and 2, respectively. Points with asterisks correspond to PACs that were extrapolated below the test concentration range. Abbreviations: F_{up}: fraction unbound; Cl_{int}: internal clearance; BER: bioactivity-exposure ratio.

Figure 3: Comparison to targeted *in vitro* assay results from the ToxCast assay battery.
(A) Boxplot comparing the HTPP activity calls (*x*-axis) with the number of ToxCast assays where a chemical was found active (*y*-axis). Each dot represents a chemical. The color code corresponds to the effect size ('top over cutoff' of the 'Global Mahalanobis' approach) in the HTPP assay. The box indicates the median with the first and third quartile and the whiskers indicate 1.5 * inter-quartile range. (B) Venn diagram of the number of chemicals active in HTPP (purple) and ToxCast (blue). (C) Scatter plot displaying the potency estimates for all chemicals active in either HTPP or ToxCast assays. A chemical was deemed active in ToxCast if it was active in at least three assays. The solid line indicates unity; the dashed lines marks ½ an order of magnitude. (D) Venn diagram of the number of chemicals active in HTPP (purple) and ToxCast burst assays (gray). (E) Scatter plot displaying the potency estimates for all chemicals active in either HTPP or ToxCast burst assays. The solid line indicates unity; the dashed lines marks ½ an order of magnitude.

Figure 4: Profile comparison of nuclear receptor modulators
(A) Correlation matrix of phenotypic profiling data for 50 nuclear receptor (NR) modulating chemicals. Chemicals are ordered by their main target, as annotated in the RefChemDB collection. Two chemicals (marked with **) were assigned to two targets (AR, ESR). Abbreviations: GW0742: 4-[(2-[3-Fluoro-4-(trifluoromethyl)phenyl]-4-methyl-1,3-thiazol-5-yl}methyl)sulfanyl]-2-methylphenoxy}acetic acid; L-165041: 4-[3-(4-Acetyl-3-hydroxy-2-propylphenoxy)propoxy]phenoxy-acetic acid. (B) Expression data for U-2 OS cells obtained from the Human Protein Atlas (http://www.proteinatlas.org) (Uhlen *et al.*, 2017). Genes with an NX value < 1 are considered not expressed (indicated by the dotted line). (C) Scatter plot displaying the potency estimates for NR modulators active in HTPP (see Fig 3C for more details). The color code corresponds to the different receptor families and is the same for all subfigures.

Figure 5: Identification of potential GR modulating chemicals
(A) Correlation matrix of phenotypic profiles of seven known glucocorticoids (GC) and a subset of test chemicals with results from the primary screen. (B) Correlation matrix of the same chemicals tested in a

repeated experiment. Of note, the chemicals were selected based on preliminary results. Upon re-analysis, the numbers slightly changed, leading to two chemicals not being tested in the secondary screen and one chemical being tested that didn't pass the threshold in the final analysis. (C) Results from the orthogonal assay, qPCR, testing for changes in gene expression for GC target genes. The indicated values correspond to $\Delta\Delta$Ct values. A value of +1 indicates a twofold upregulation.

Figure 6: Similarity of phenotypic profiles for chemicals with structural similarity.

(A) Schematic overview of the procedure. ToxPrints were obtained for active chemicals (n=536) and used to create a dendrogram. The dendrogram was cut at a tree height of 0.6, which resulted in 234 clusters. Of these, 137 clusters contained at least two chemicals. Next, for clusters with more than two chemicals, two chemicals were selected randomly (n=274). (B) For each of these 137 chemical pairs, the biological similarity (Kendall correlation) was calculated. (C) To discern whether chemicals within a cluster are more similar than expected by chance, a randomized data set was created. For this purpose, the 274 chemicals were randomly paired 10 times, generating 1370 biological similarity values. The graph displays the biological similarity values of the randomized data set (gray) and the real data set (red). The box indicates the first and third quantile and the median, while the whiskers extend to 1.5 times the inter-quartile range. The indicated p-value above the graph was obtained from a one-sided Wilcoxon rank sum test. (D,F,G,H) Exemplary correlation matrices of selected clusters. The bottom left half displays structural similarity calculated as Jaccard similarity; the top right half displays profile similarity calculated as Kendall correlation of the 289 selected features. (E) Representative images of Hoechst-33342 labeling in U-2 OS cells treated with DMSO (0.5 %) (left panel) or 100 µM dieldrin (right panel). Differences in nucleus texture compared to control were observable in dieldrin treated cells (arrows). Scale bar is 50 µm.

Figure 7: Grouping of phenotypic profiles of all active chemicals.

(A) The profile of the third lowest active concentration for each chemical (n=547) was used for hierarchical clustering using Kendall correlation to obtain the dendrogram, which was then cut into 15 clusters. The colors in the matrix are pairwise Kendall correlations between two chemicals. (B) For each cluster, it was determined whether chemicals in the cluster were enriched in certain chemical fingerprints, activity in targeted bioassays or literature evidence for activation of stress response pathways. The bar graph indicates the number of entities that were enriched in any given cluster.

Figure 8: Case study of conazole-like chemicals.

(A) Phenotypic profiles of all active conazoles in the primary screen. For each chemical the highest active, non-cytotoxic concentration is shown. The number in parentheses indicates the test concentration in µM. Features are clustered within the respective fluorescent channels (indicated by the column colors). Hierarchical clustering was performed using Kendall correlation of profiles. (B) Correlation matrix of all active conazoles in the primary screen. Profiles of the three lowest concentrations of each chemical were compared using Kendall correlation; the displayed color indicates the value for the condition pair with the largest value. (C) Correlation matrix of all active conazoles of the follow-up study. The bottom left part displays structural similarity, as calculated with Jaccard/Tanimoto similarity. The top right part displays profile similarity, as calculated with Kendall correlation. (D) Ranked list of chemicals tested in the primary screen with the largest profile similarity with diniconazole. DTXSID are unique substance identifiers associated with ToxCast chemicals, as referenced on the CompTox Chemicals Dashboard (Williams *et al.*, 2017). Abbreviations: conc: test concentration in µM; biol sim: biological similarity. (E) Representative images of Hoechst-33342 (blue) and multiplexed WGA-Alexa 555 and phalloidin-Alexa 568 conjugates (gold) in U-2 OS cells treated with DMSO (0.5 %), 100 µM diniconazole, 100 µM ketoconazole or 100 µM phenolphthalein (from left to right). Note that the phenotype of diniconazole is qualitatively different from that of DMSO and ketoconazole and more similar to phenolphthalein.

Bibliography

Adamakis, I.-D.S., Panteris, E., Eleftheriou, E.P., 2019. Tubulin Acetylation Mediates Bisphenol A Effects on the Microtubule Arrays of Allium cepa and Triticum turgidum. Biomolecules **9**, 185.

Baltazar, M.T., Cable, S., Carmichael, P.L., Cubberley, R., Cull, T., Delagrange, 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., 2020. A Next-Generation Risk Assessment Case Study for Coumarin in Cosmetic Products. Toxicological sciences : an official journal of the Society of Toxicology **176**, 236-252.

Bougen-Zhukov, N., Loh, S.Y., Lee, H.K., Loo, L.H., 2017. Large-scale image-based screening and profiling of cellular phenotypes. Cytometry A **91**, 115-125.

Bray, M.A., Singh, S., Han, H., Davis, C.T., Borgeson, B., Hartland, C., Kost-Alimova, M., Gustafsdottir, S.M., Gibson, C.C., Carpenter, A.E., 2016. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc **11**, 1757-1774.

Breen, M., Ring, C.L., Kreutz, A., Goldsmith, M.R., Wambaugh, J.F., 2021. High-throughput PBTK models for in vitro to in vivo extrapolation. Expert Opin Drug Metab Toxicol **17**, 903-921.

Caicedo, J.C., Cooper, S., Heigwer, F., Warchal, S., Qiu, P., Molnar, C., Vasilevich, A.S., Barry, J.D., Bansal, H.S., Kraus, O., Wawer, M., Paavolainen, L., Herrmann, M.D., Rohban, M., Hung, J., Hennig, H., Concannon, J., Smith, I., Clemons, P.A., Singh, S., Rees, P., Horvath, P., Linington, R.G., Carpenter, A.E., 2017. Data-analysis strategies for image-based cell profiling. Nat Methods **14**, 849-863.

Caicedo, J.C., Singh, S., Carpenter, A.E., 2016. Applications in image-based profiling of perturbations. Curr Opin Biotechnol **39**, 134-142.

Caie, P.D., Walls, R.E., Ingleston-Orme, A., Daya, S., Houslay, T., Eagle, R., Roberts, M.E., Carragher, N.O., 2010. High-content phenotypic profiling of drug response signatures across distinct cancer cells. Mol Cancer Ther **9**, 1913-1926.

Canada, H., 2021. Science approach document - Bioactivity exposure ratio: Application in priority setting and risk assessment. In Canada, H., (Ed.), pp.

Cao, Z., West, C., Norton-Wenzel, C.S., Rej, R., Davis, F.B., Davis, P.J., Rej, R., 2009. Effects of Resin or Charcoal Treatment on Fetal Bovine Serum and Bovine Calf Serum. Endocrine Research **34**, 101-108.

Casida, J.E., 2009. Pest toxicology: the primary mechanisms of pesticide action. Chem Res Toxicol **22**, 609-619.

Chandrasekaran, S.N., Ceulemans, H., Boyd, J.D., Carpenter, A.E., 2021. Image-based profiling for drug discovery: due for a machine-learning upgrade? Nat Rev Drug Discov **20**, 145-159.

Chen, Q., Meng, Z., Su, R., 2020. WERFE: A Gene Selection Algorithm Based on Recursive Feature Elimination and Ensemble Strategy. Frontiers in Bioengineering and Biotechnology **8**.

Church, K.W., Hanks, P., 1990. Word-Association Norms, Mutual Information, and Lexicography. 27th Annual Meeting of the Association for Computational Linguistics, 76-83.

Dawson, D.E., Ingle, B.L., Phillips, K.A., Nichols, J.W., Wambaugh, J.F., Tornero-Velez, R., 2021. Designing QSARs for Parameters of High-Throughput Toxicokinetic Models Using Open-Source Descriptors. Environmental Science & Technology **55**, 6505-6517.

Dimitrov, S.D., Diderich, R., Sobanski, T., Pavlov, T.S., Chankov, G.V., Chapkanov, A.S., Karakolev, Y.H., Temelkov, S.G., Vasilev, R.A., Gerova, K.D., Kuseva, C.D., Todorova, N.D., Mehmed, A.M., Rasenberg, M., Mekenyan, O.G., 2016. QSAR Toolbox – workflow and major functionalities. SAR and QSAR in Environmental Research **27**, 203-219.

Dix, D.J., Houck, K.A., Martin, M.T., Richard, A.M., Setzer, R.W., Kavlock, R.J., 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci **95**, 5-12.

Djoumbou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., Fahy, E., Steinbeck, C., Subramanian, S., Bolton, E., Greiner, R., Wishart, D.S., 2016. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. Journal of Cheminformatics **8**, 61.

George, O., Bryant, B.K., Chinnasamy, R., Corona, C., Arterburn, J.B., Shuster, C.B., 2008. Bisphenol A directly targets tubulin to disrupt spindle organization in embryonic and somatic cells. ACS Chem Biol **3**, 167-179.

Gustafsdottir, S.M., Ljosa, V., Sokolnicki, K.L., Anthony Wilson, J., Walpita, D., Kemp, M.M., Petri Seiler, K., Carrel, H.A., Golub, T.R., Schreiber, S.L., Clemons, P.A., Carpenter, A.E., Shamji, A.F., 2013. Multiplex cytological profiling assay to measure diverse cellular states. PLoS One **8**, e80999.

Haghighi, M., Caicedo, J.C., Cimini, B.A., Carpenter, A.E., Singh, S., 2022. High-dimensional gene expression and morphology profiles of cells across 28,000 genetic and chemical perturbations. Nat Methods **19**, 1550-1557.

Heard, P.L., Rubitski, E.E., Spellman, R.A., Schuler, M.J., 2013. Phenolphthalein induces centrosome amplification and tubulin depolymerization in vitro. Environ Mol Mutagen **54**, 308-316.

Hofmarcher, M., Rumetshofer, E., Clevert, D.A., Hochreiter, S., Klambauer, G., 2019. Accurate Prediction of Biological Assays with High-Throughput Microscopy Images and Convolutional Networks. J Chem Inf Model **59**, 1163-1171.

Honn, K.V., Singley, J.A., Chavin, W., 1975. Fetal Bovine Serum: A Multivariate Standard. Proceedings of the Society for Experimental Biology and Medicine **149**, 344-347.

Hughes, J.P., Rees, S., Kalindjian, S.B., Philpott, K.L., 2011. Principles of early drug discovery. Br J Pharmacol **162**, 1239-1249.

Jaccard, P., 1902. Lois de distribution florale dans la zone alpine. Bull Soc Vaudoise Sci Nat **38**, 69-130.

Judson, R.S., Houck, K.A., Kavlock, R.J., Knudsen, T.B., Martin, M.T., Mortensen, H.M., Reif, D.M., Rotroff, D.M., Shah, I., Richard, A.M., Dix, D.J., 2010. In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect **118**, 485-492.

Judson, R.S., Thomas, R.S., Baker, N., Simha, A., Howey, X.M., Marable, C., Kleinstreuer, N.C., Houck, K.A., 2019. Workflow for defining reference chemicals for assessing performance of in vitro assays. ALTEX - Alternatives to animal experimentation **36**, 261-276.

Karreman, C., 2002. AiO, combining DNA/protein programs and oligo-management. Bioinformatics **18**, 884-885.

Kavlock, R., Chandler, K., Houck, K., Hunter, S., Judson, R., Kleinstreuer, N., Knudsen, T., Martin, M., Padilla, S., Reif, D., Richard, A., Rotroff, D., Sipes, N., Dix, D., 2012. Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. Chem Res Toxicol **25**, 1287-1302.

Kipp, J.L., Ramirez, V.D., 2003. Estradiol and Testosterone Have Opposite Effects on Microtubule Polymerization. Neuroendocrinology **77**, 258-272.

Krautwurst, D., Yau, K.-W., Reed, R.R., 1998. Identification of Ligands for Olfactory Receptors by Functional Expression of a Receptor Library. Cell **95**, 917-926.

Leitereg, T.J., Guadagni, D.G., Harris, J., Mon, T.R., Teranishi, R., 1971. Chemical and sensory data supporting the difference between the odors of the enantiomeric carvones. Journal of Agricultural and Food Chemistry **19**, 785-787.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. Methods **25**, 402-408.

Mansouri, K., Grulke, C.M., Judson, R.S., Williams, A.J., 2018. OPERA models for predicting physicochemical properties and environmental fate endpoints. Journal of Cheminformatics **10**, 10.

Marin Zapata, P.A., Méndez-Lucio, O., Le, T., Beese, C.J., Wichard, J., Rouquié, D., Clevert, D.-A., 2023. Cell morphology-guided de novo hit design by conditioning GANs on phenotypic image features. Digital Discovery **2**, 91-102.

Melillo, B., Zoller, J., Hua, B.K., Verho, O., Borghs, J.C., Nelson, S.D., Jr., Maetani, M., Wawer, M.J., Clemons, P.A., Schreiber, S.L., 2018. Synergistic Effects of Stereochemistry and Appendages on the Performance Diversity of a Collection of Synthetic Compounds. J Am Chem Soc **140**, 11784-11790.

Metzler, M., Pfeiffer, E., 1995. Effects of estrogens on microtubule polymerization in vitro: correlation with estrogenicity. Environmental Health Perspectives **103**, 21-22.

Nassiri, I., McCall, M.N., 2018. Systematic exploration of cell morphological phenotypes associated with a transcriptomic query. Nucleic Acids Res **46**, e116.

NCBI, 2018. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res **46**, D8-D13.

Nyffeler, J., Chovancova, P., Dolde, X., Holzer, A.K., Purvanov, V., Kindinger, I., Kerins, A., Higton, D., Silvester, S., van Vugt-Lussenburg, B.M.A., Glaab, E., van der Burg, B., Maclennan, R., Legler, D.F., Leist, M., 2018. A structure-activity relationship linking non-planar PCBs to functional deficits of neural crest cells: new roles for connexins. Arch Toxicol **92**, 1225-1247.

Nyffeler, J., Haggard, D.E., Willis, C., Setzer, R.W., Judson, R., Paul-Friedman, K., Everett, L.J., Harrill, J.A., 2021. Comparison of Approaches for Determining Bioactivity Hits from High-Dimensional Profiling Data. SLAS Discov **26**, 292-308.

Nyffeler, J., Karreman, C., Leisner, H., Kim, Y.J., Lee, G., Waldmann, T., Leist, M., 2017. Design of a high-throughput human neural crest cell migration assay to indicate potential developmental toxicants. ALTEX **34**, 75-94.

Nyffeler, J., Willis, C., Harris, F.R., Taylor, L.W., Judson, R., Everett, L.J., Harrill, J.A., 2022. Combining phenotypic profiling and targeted RNA-Seq reveals linkages between transcriptional perturbations and chemical effects on cell morphology: Retinoic acid as an example. Toxicol Appl Pharmacol **444**, 116032.

Nyffeler, J., Willis, C., Lougee, R., Richard, A., Paul-Friedman, K., Harrill, J.A., 2020. Bioactivity screening of environmental chemicals using imaging-based high-throughput phenotypic profiling. Toxicol Appl Pharmacol **389**, 114876.

Paul-Friedman, K., Gagne, M., Loo, L.H., Karamertzanis, P., Netzeva, T., Sobanski, T., Franzosa, J., Richard, A., Lougee, R., Gissi, A., Lee, J.J., Angrish, M., Dorne, J.L., Foster, S., Raffaele, K., Bahadori, T., Gwinn, M.R., Lambert, J.C., Whelan, M., Rasenberg, M., Barton-Maclaren, T.S., Thomas, R.S., 2019. Examining the utility of in vitro bioactivity as a conservative point of departure: a case study. Toxicol Sci.

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.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., 2020. Utility of In Vitro Bioactivity as a Lower Bound Estimate of In Vivo Adverse Effect Levels and in Risk-Based Prioritization. Toxicol Sci **173**, 202-225.

Pearce, R.G., Setzer, R.W., Strope, C.L., Wambaugh, J.F., Sipes, N.S., 2017. httk: R Package for High-Throughput Toxicokinetics. J Stat Softw **79**, 1-26.

Ramm, S., Todorov, P., Chandrasekaran, V., Dohlman, A., Monteiro, M.B., Pavkovic, M., Muhlich, J., Shankaran, H., Chen, W.W., Mettetal, J.T., Vaidya, V.S., 2019. A Systems Toxicology Approach for the Prediction of Kidney Toxicity and Its Mechanisms In Vitro. Toxicol Sci **169**, 54-69.

Richard, A.M., Judson, R.S., Houck, K.A., Grulke, C.M., Volarath, P., Thillainadarajah, I., Yang, C., Rathman, J., Martin, M.T., Wambaugh, J.F., Knudsen, T.B., Kancherla, J., Mansouri, K., Patlewicz,

G., Williams, A.J., Little, S.B., Crofton, K.M., Thomas, R.S., 2016. ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. Chem Res Toxicol **29**, 1225-1251.

Ring, C.L., Arnot, J.A., Bennett, D.H., Egeghy, P.P., Fantke, P., Huang, L., Isaacs, K.K., Jolliet, O., Phillips, K.A., Price, P.S., Shin, H.-M., Westgate, J.N., Setzer, R.W., Wambaugh, J.F., 2019. Consensus Modeling of Median Chemical Intake for the U.S. Population Based on Predictions of Exposure Pathways. Environmental Science & Technology **53**, 719-732.

Ring, C.L., Pearce, R.G., Setzer, R.W., Wetmore, B.A., Wambaugh, J.F., 2017. Identifying populations sensitive to environmental chemicals by simulating toxicokinetic variability. Environ Int **106**, 105-118.

Roed-Petersen, J., Batsberg, W., Larsen, E., 1990. Contact dermatitis from Naphthol AS. Contact Dermatitis **22**, 161-163.

Rohban, M.H., Fuller, A.M., Tan, C., Goldstein, J.T., Syangtan, D., Gutnick, A., DeVine, A., Nijsure, M.P., Rigby, M., Sacher, J.R., Corsello, S.M., Peppler, G.B., Bogaczynska, M., Boghossian, A., Ciotti, G.E., Hands, A.T., Mekareeya, A., Doan, M., Gale, J.P., Derynck, R., Turbyville, T., Boerckel, J.D., Singh, S., Kiessling, L.L., Schwarz, T.L., Varelas, X., Wagner, F.F., Kafri, R., Eisinger-Mathason, T.S.K., Carpenter, A.E., 2022. Virtual screening for small-molecule pathway regulators by image-profile matching. Cell Syst **13**, 724-736 e729.

Rohban, M.H., Singh, S., Wu, X., Berthet, J.B., Bray, M.A., Shrestha, Y., Varelas, X., Boehm, J.S., Carpenter, A.E., 2017. Systematic morphological profiling of human gene and allele function via Cell Painting. Elife **6**.

Schneidewind, T., Brause, A., Pahl, A., Burhop, A., Mejuch, T., Sievers, S., Waldmann, H., Ziegler, S., 2020. Morphological Profiling Identifies a Common Mode of Action for Small Molecules with Different Targets. Chembiochem **21**, 3197-3207.

Seal, S., Carreras-Puigvert, J., Trapotsi, M.A., Yang, H., Spjuth, O., Bender, A., 2022. Integrating cell morphology with gene expression and chemical structure to aid mitochondrial toxicity detection. Commun Biol **5**, 858.

Sheffield, T., Brown, J., Davidson, S., Friedman, K.P., Judson, R., 2021. tcplfit2: an R-language general purpose concentration-response modeling package. Bioinformatics.

Simm, J., Klambauer, G., Arany, A., Steijaert, M., Wegner, J.K., Gustin, E., Chupakhin, V., Chong, Y.T., Vialard, J., Buijnsters, P., Velter, I., Vapirev, A., Singh, S., Carpenter, A.E., Wuyts, R., Hochreiter, S., Moreau, Y., Ceulemans, H., 2018. Repurposing High-Throughput Image Assays Enables Biological Activity Prediction for Drug Discovery. Cell Chem Biol **25**, 611-618 e613.

Srivastava, P., Panda, D., 2007. Rotenone inhibits mammalian cell proliferation by inhibiting microtubule assembly through tubulin binding. The FEBS Journal **274**, 4788-4801.

Su, R., Xiong, S., Zink, D., Loo, L.H., 2016. High-throughput imaging-based nephrotoxicity prediction for xenobiotics with diverse chemical structures. Arch Toxicol **90**, 2793-2808.

Svenningsen, E.B., Poulsen, T.B., 2019. Establishing cell painting in a smaller chemical biology lab - A report from the frontier. Bioorg Med Chem **27**, 2609-2615.

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., 2019. The Next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency. Toxicol Sci **169**, 317-332.

Trapotsi, M.A., Mervin, L.H., Afzal, A.M., Sturm, N., Engkvist, O., Barrett, I.P., Bender, A., 2021. Comparison of Chemical Structure and Cell Morphology Information for Multitask Bioactivity Predictions. J Chem Inf Model **61**, 1444-1456.

Uhlen, M., Zhang, C., Lee, S., Sjostedt, E., Fagerberg, L., Bidkhori, G., Benfeitas, R., Arif, M., Liu, Z., Edfors, F., Sanli, K., von Feilitzen, K., Oksvold, P., Lundberg, E., Hober, S., Nilsson, P., Mattsson, J., Schwenk, J.M., Brunnstrom, H., Glimelius, B., Sjoblom, T., Edqvist, P.H., Djureinovic, D., Micke, P., Lindskog, C., Mardinoglu, A., Ponten, F., 2017. A pathology atlas of the human cancer transcriptome. Science **357**.

Vaughan, A., Guilbault, G.G., Hackney, D., 1971. Fluorometric methods for analysis of acid and alkaline phosphatase. Anal Chem **43**, 721-724.

Wang, S.-H., Liang, C.-T., Liu, Y.-W., Huang, M.-C., Huang, S.-C., Hong, W.-F., Su, J.-G.J., 2009. Crosstalk between activated forms of the aryl hydrocarbon receptor and glucocorticoid receptor. Toxicology **262**, 87-97.

Warchal, S.J., Dawson, J.C., Carragher, N.O., 2019. Evaluation of Machine Learning Classifiers to Predict Compound Mechanism of Action When Transferred across Distinct Cell Lines. Slas Discovery **24**, 224-233.

Way, G.P., Kost-Alimova, M., Shibue, T., Harrington, W.F., Gill, S., Piccioni, F., Becker, T., Shafqat-Abbasi, H., Hahn, W.C., Carpenter, A.E., Vazquez, F., Singh, S., 2021. Predicting cell health phenotypes using image-based morphology profiling. Mol Biol Cell **32**, 995-1005.

Way, G.P., Natoli, T., Adeboye, A., Litichevskiy, L., Yang, A., Lu, X., Caicedo, J.C., Cimini, B.A., Karhohs, K., Logan, D.J., Rohban, M.H., Kost-Alimova, M., Hartland, K., Bornholdt, M., Chandrasekaran, S.N., Haghighi, M., Weisbart, E., Singh, S., Subramanian, A., Carpenter, A.E., 2022. Morphology and gene expression profiling provide complementary information for mapping cell state. Cell Syst **13**, 911-923 e919.

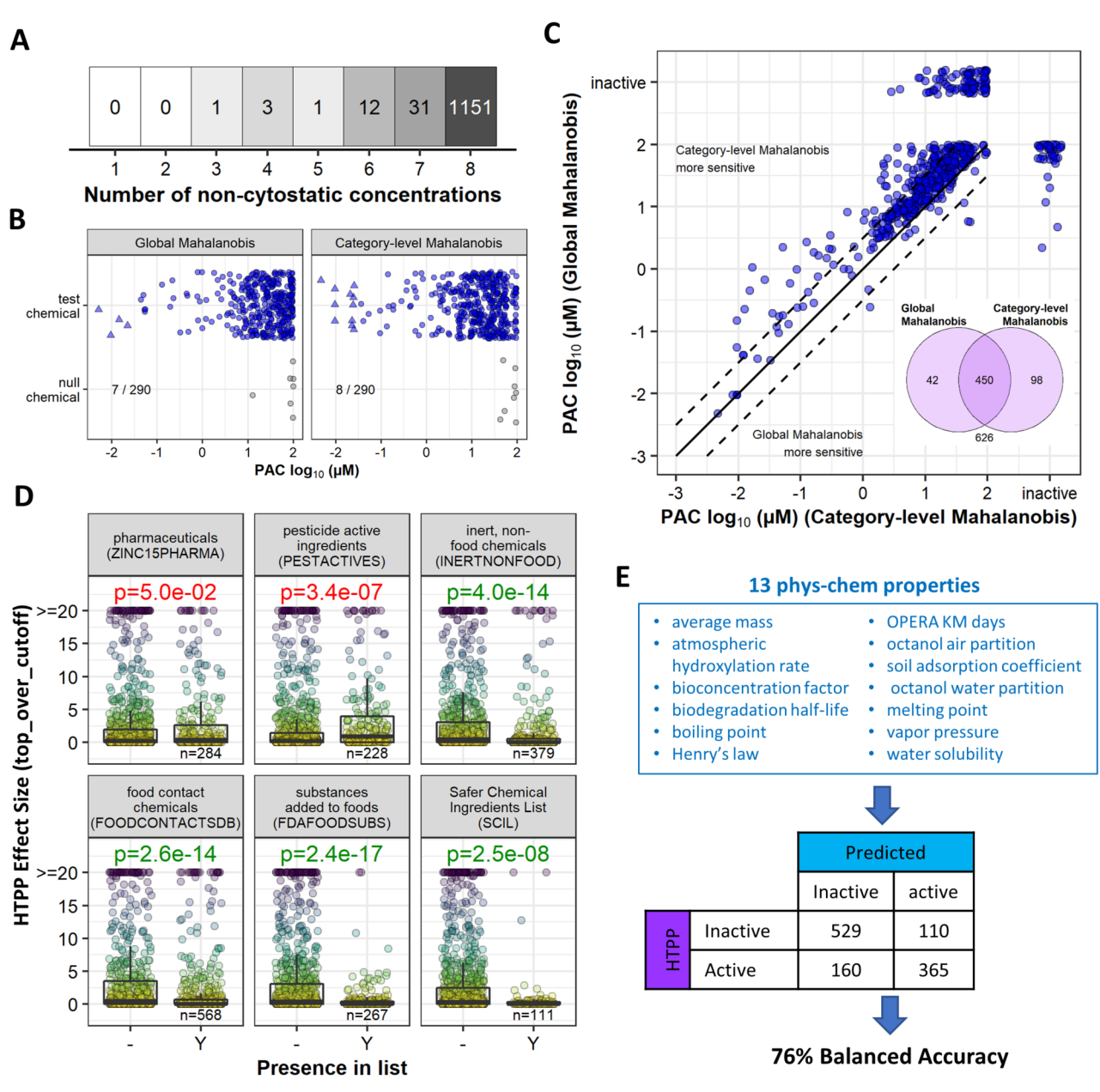
Williams, A.J., Grulke, C.M., Edwards, J., McEachran, A.D., Mansouri, K., Baker, N.C., Patlewicz, G., Shah, I., Wambaugh, J.F., Judson, R.S., Richard, A.M., 2017. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. J Cheminform **9**, 61.

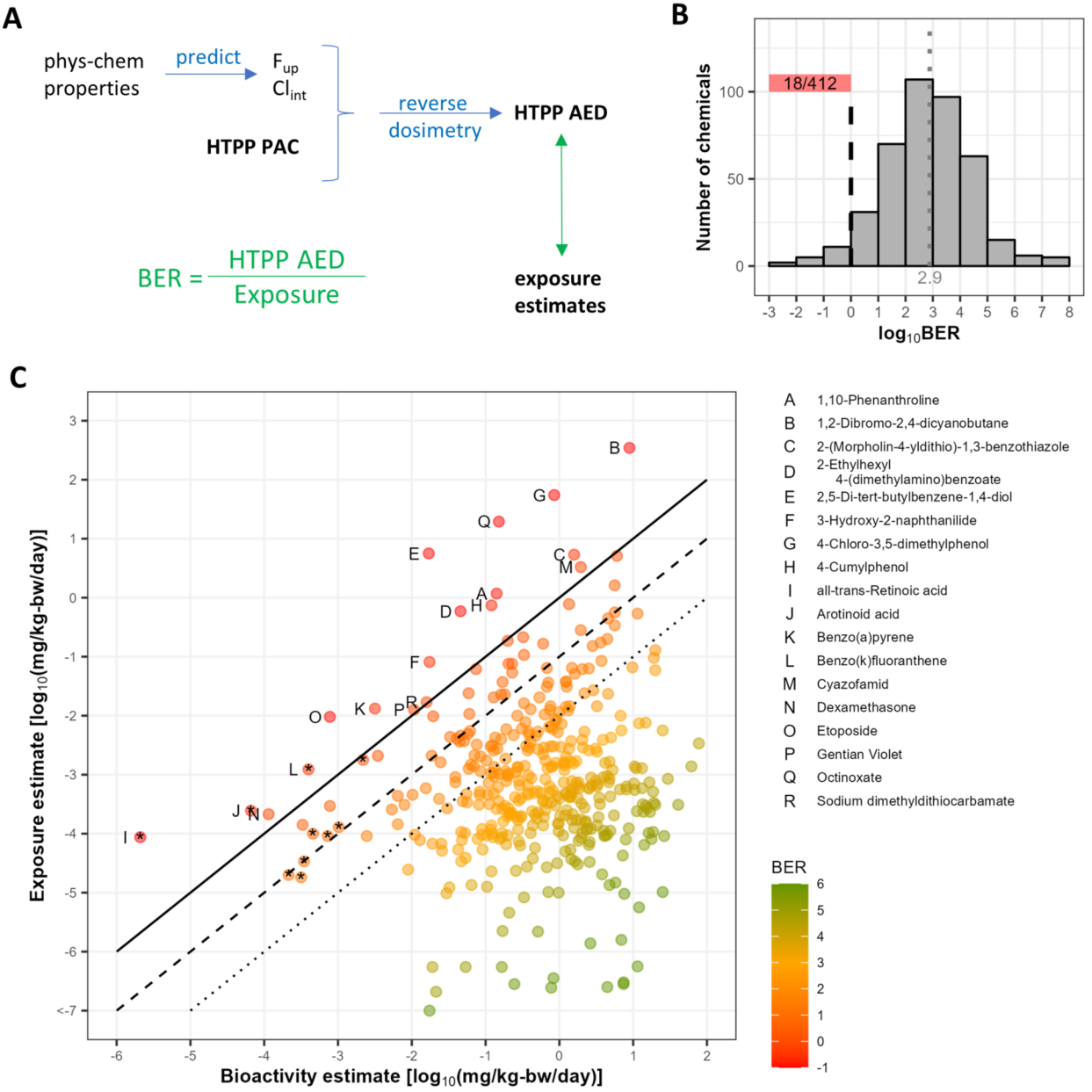
Willis, C., Nyffeler, J., Harrill, J.A., 2020. Phenotypic Profiling of Reference Chemicals Across Biologically Diverse Cell Types Using the Cell Painting Assay, SLAS Discovery, pp.

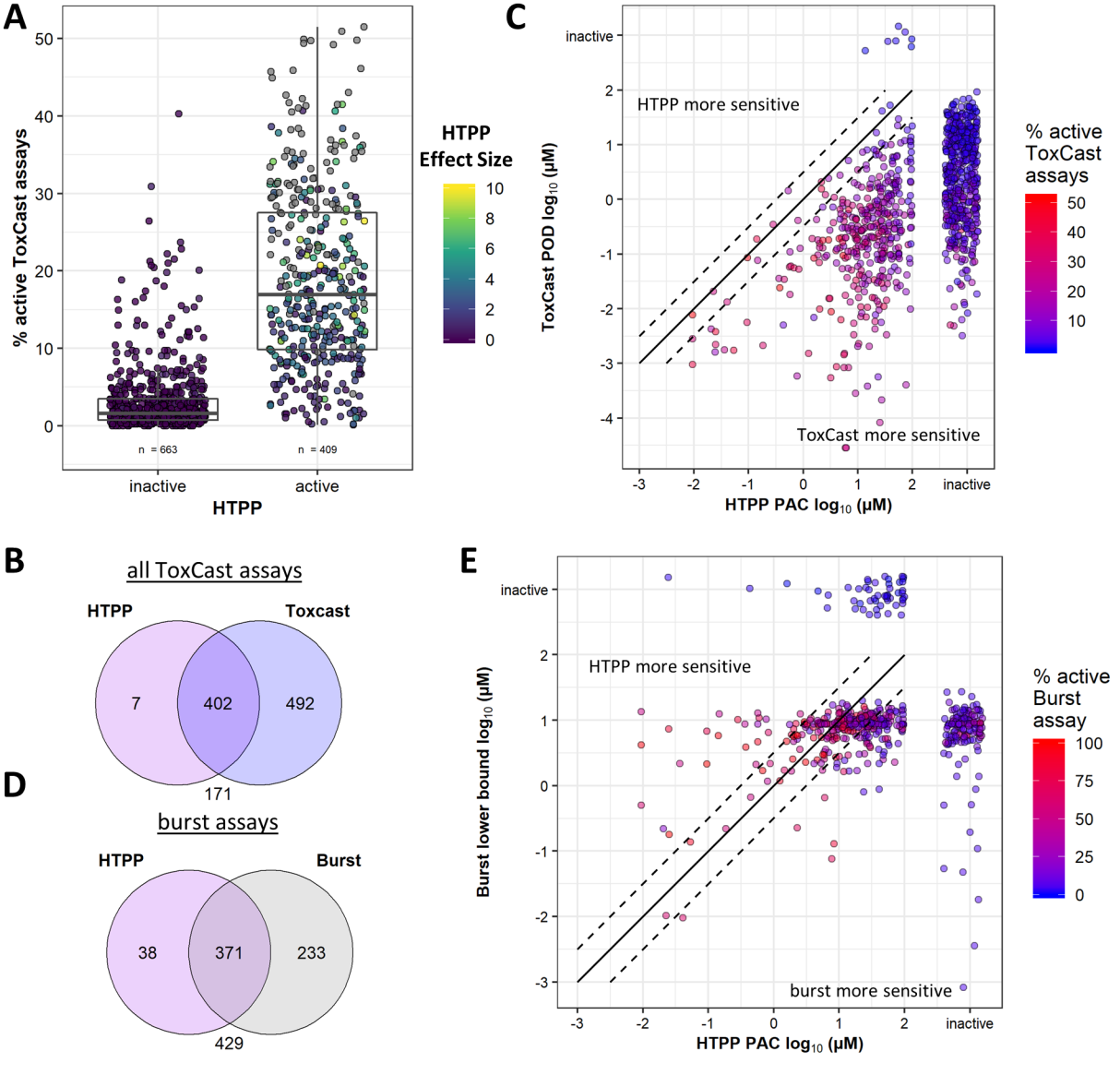
Woehrmann, M.H., Bray, W.M., Durbin, J.K., Nisam, S.C., Michael, A.K., Glassey, E., Stuart, J.M., Lokey, R.S., 2013. Large-scale cytological profiling for functional analysis of bioactive compounds. Molecular BioSystems **9**, 2604-2617.

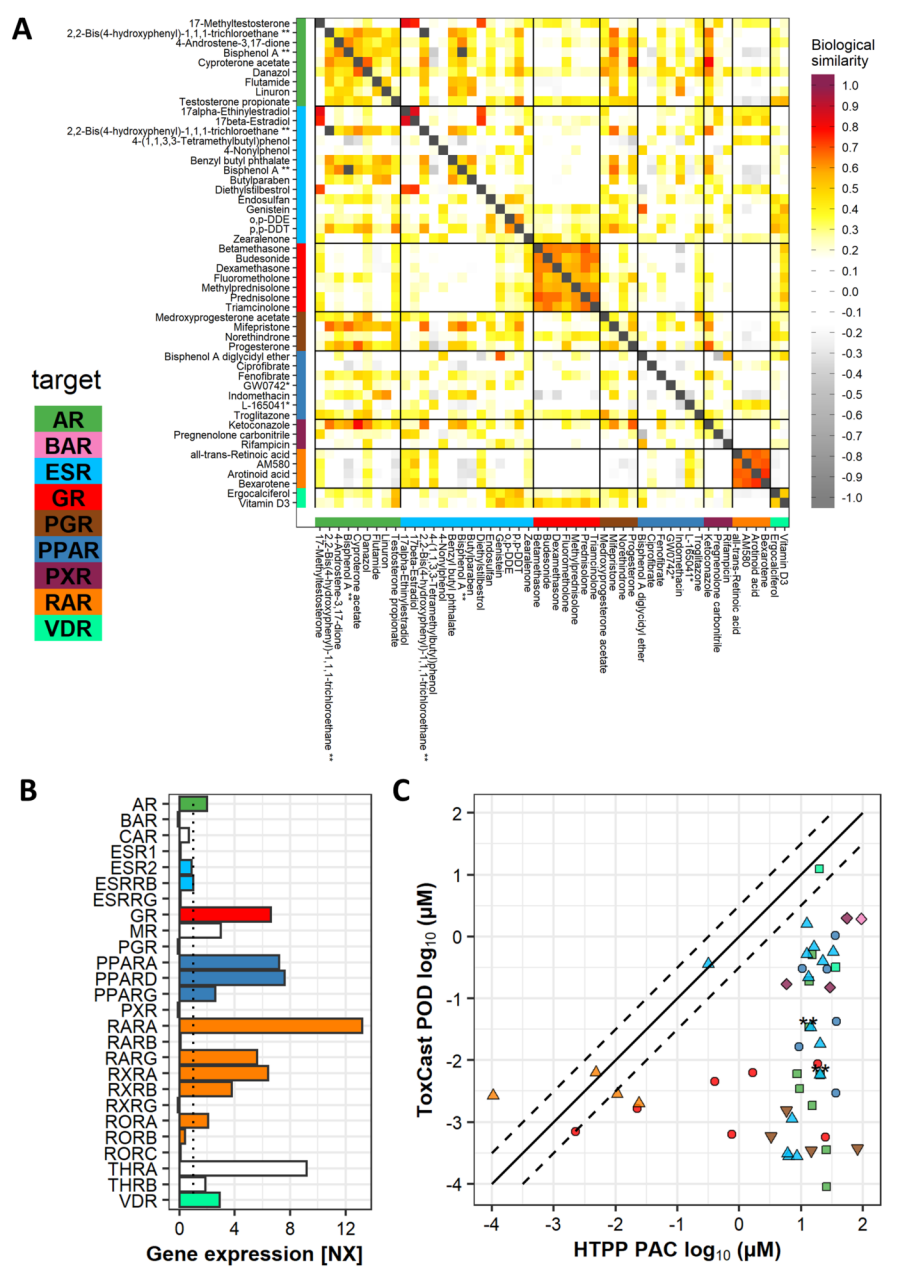
Yang, C.H., Tarkhov, A., Marusczyk, J., Bienfait, B., Gasteiger, J., Kleinoeder, T., Magdziarz, T., Sacher, O., Schwab, C.H., Schwoebel, J., Terfloth, L., Arvidson, K., Richard, A., Worth, A., Rathman, J., 2015. New Publicly Available Chemical Query Language, CSRML, To Support Chemotype Representations for Application to Data Mining and Modeling. J Chem Inf Model **55**, 510-528.

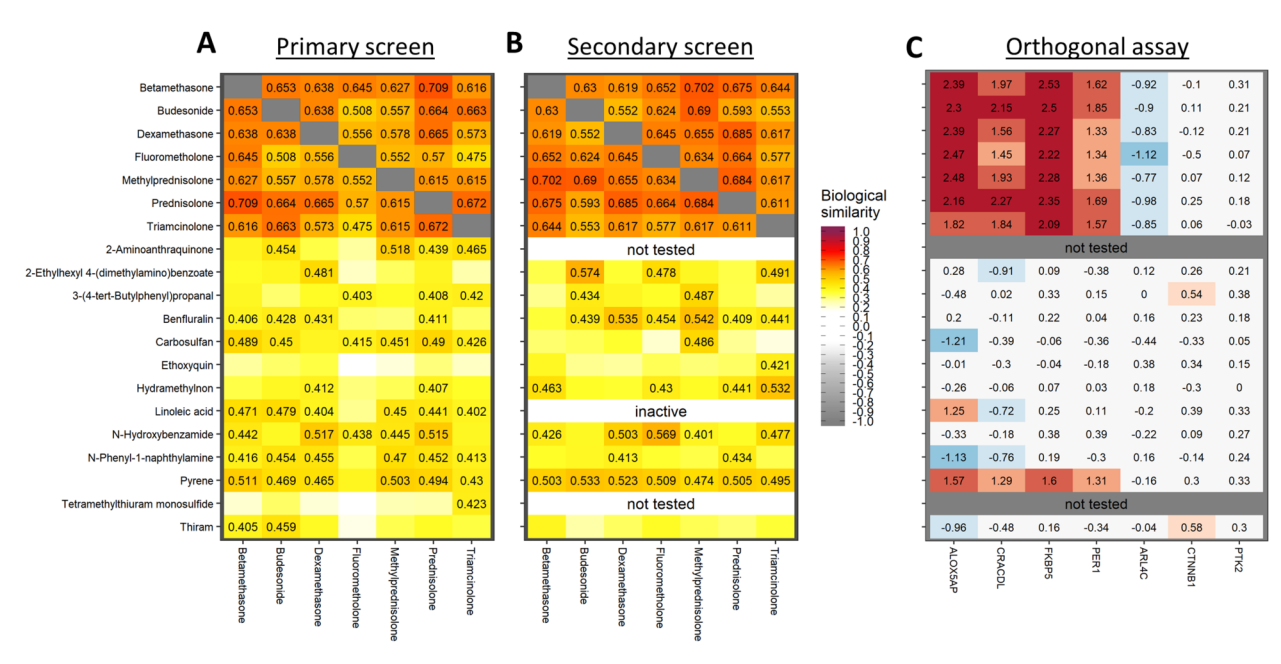
Young, D.W., Bender, A., Hoyt, J., McWhinnie, E., Chirn, G.-W., Tao, C.Y., Tallarico, J.A., Labow, M., Jenkins, J.L., Mitchison, T.J., Feng, Y., 2008. Integrating high-content screening and ligand-target prediction to identify mechanism of action. Nature Chemical Biology **4**, 59-68.

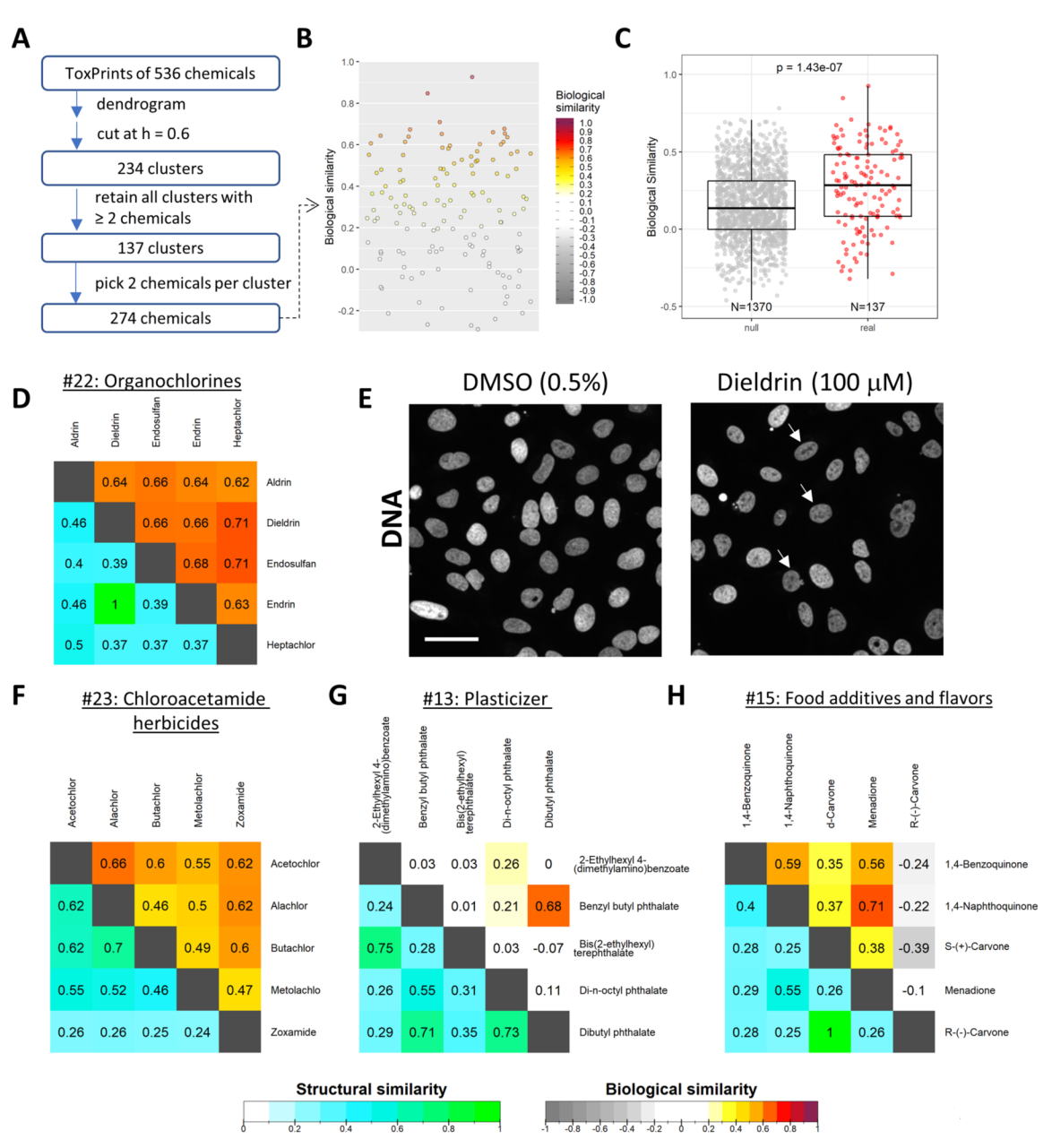

A
0
0
1
3
1
12
31
1151
1
2
3
4
5
6
7
8
Number of non-cytostatic concentrations
B
Global Mahalanobis
Category-level Mahalanobis
test chemical
null chemical
7 / 290
8 / 290
-2
-1
0
1
2
PAC log10 (µM)
C
inactive
2
1
0
-1
-2
-3
PAC log10 (µM) (Global Mahalanobis)
Category-level Mahalanobis more sensitive
Global Mahalanobis more sensitive
Global Mahalanobis
Category-level Mahalanobis
42
450
98
626
PAC log10 (µM) (Category-level Mahalanobis)
D
pharmaceuticals (ZINC15PHARMA)
pesticide active ingredients (PESTACTIVES)
inert, non-food chemicals (INERTNONFOOD)
p=5.0e-02
p=3.4e-07
p=4.0e-14
food contact chemicals (FOODCONTACTSDB)
substances added to foods (FDAFOODSUBS)
Safer Chemical Ingredients List (SCIL)
p=2.6e-14
p=2.4e-17
p=2.5e-08
>=20
15
10
5
0
n=284
n=228
n=379
n=568
n=267
n=111
HTPP Effect Size (top_over_cutoff)
-
Y
Presence in list
E
13 phys-chem properties
average mass
atmospheric hydroxylation rate
bioconcentration factor
biodegradation half-life
boiling point
Henry's law
OPERA KM days
octanol air partition
soil adsorption coefficient
octanol water partition
melting point
vapor pressure
water solubility
Predicted
Inactive
active
HTPP
Inactive
529
110
Active
160
365
76% Balanced Accuracy

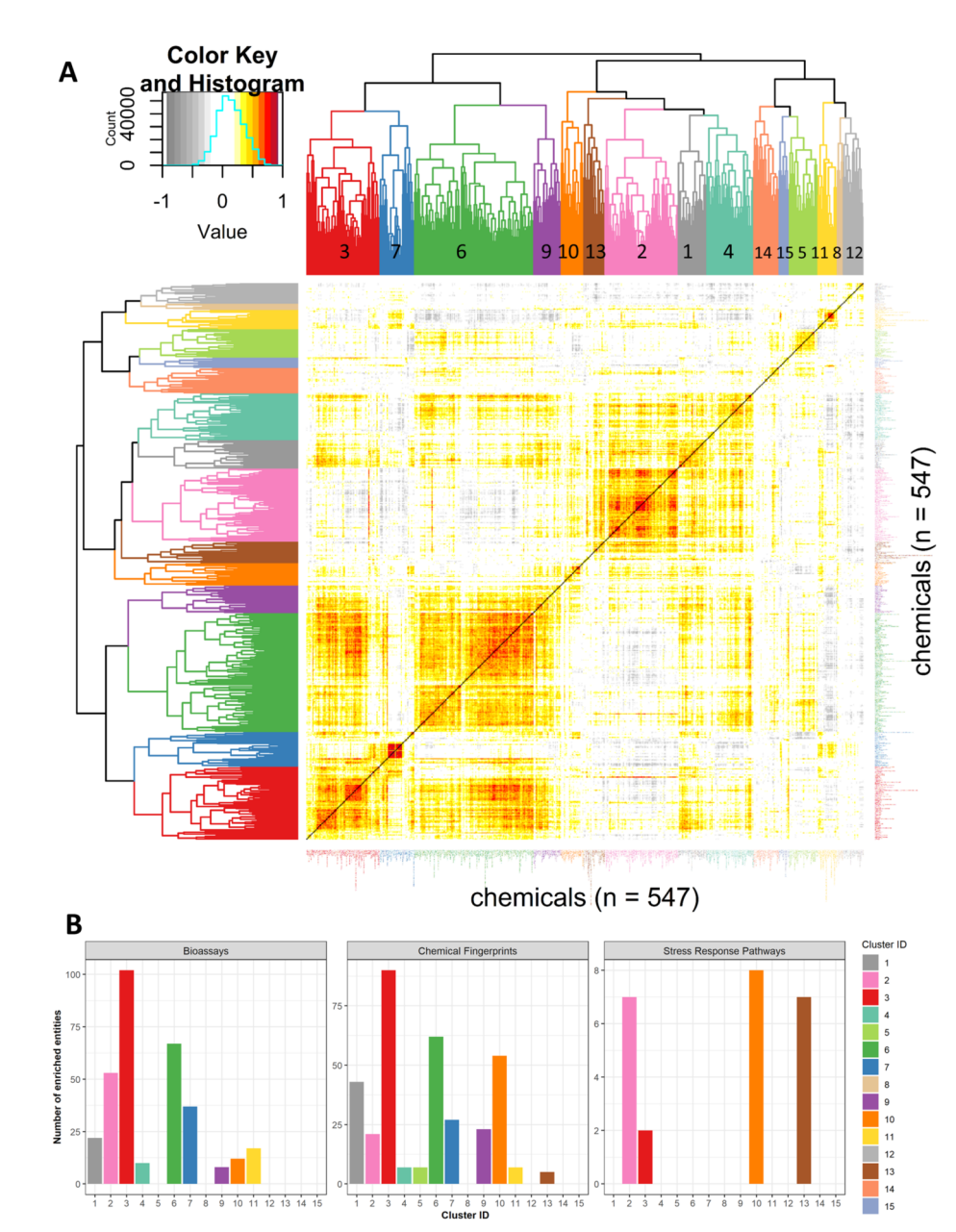

A
phys-chem properties
predict
F_{up}
Cl_{int}
HTPP PAC
reverse dosimetry
HTPP AED
exposure estimates
BER = HTPP AED / Exposure
B
18/412
Number of chemicals
100
50
0
2.9
-3 -2 -1 0 1 2 3 4 5 6 7 8
log_{10}BER
C
Exposure estimate [log_{10}(mg/kg-bw/day)]
Bioactivity estimate [log_{10}(mg/kg-bw/day)]
3 2 1 0 -1 -2 -3 -4 -5 -6 <-7
-6 -5 -4 -3 -2 -1 0 1 2
A 1,10-Phenanthroline
B 1,2-Dibromo-2,4-dicyanobutane
C 2-(Morpholin-4-yldithio)-1,3-benzothiazole
D 2-Ethylhexyl 4-(dimethylamino)benzoate
E 2,5-Di-tert-butylbenzene-1,4-diol
F 3-Hydroxy-2-naphthanilide
G 4-Chloro-3,5-dimethylphenol
H 4-Cumylphenol
I all-trans-Retinoic acid
J Arotinoid acid
K Benzo(a)pyrene
L Benzo(k)fluoranthene
M Cyazofamid
N Dexamethasone
O Etoposide
P Gentian Violet
Q Octinoxate
R Sodium dimethyldithiocarbamate
BER
6 5 4 3 2 1 0 -1

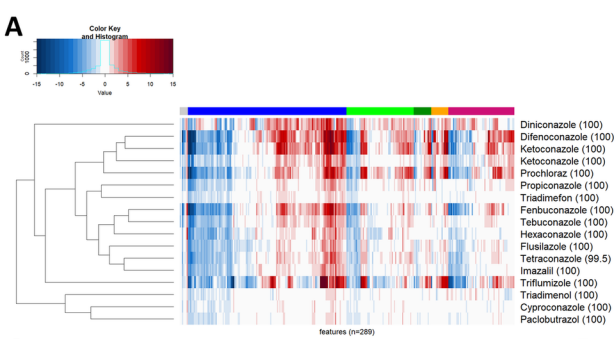

A
% active ToxCast assays
HTPP
Effect Size
n = 663
n = 409
inactive
active
HTPP
B
all ToxCast assays
HTPP
Toxcast
7
402
492
171
D
burst assays
HTPP
Burst
38
371
233
429
C
inactive
HTPP more sensitive
ToxCast more sensitive
ToxCast POD log10 (μM)
HTPP PAC log10 (μM)
% active ToxCast assays
E
Burst lower bound log10 (μM)
HTPP more sensitive
burst more sensitive
HTPP PAC log10 (μM)
% active Burst assay

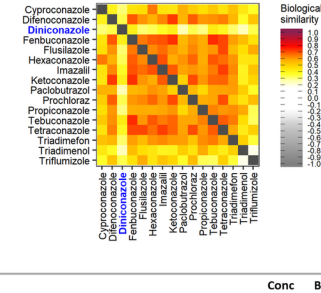

A
17-Methyltestosterone
2,2-Bis(4-hydroxyphenyl)-1,1,1-trichloroethane **
4-Androstene-3,17-dione
Bisphenol A **
Cyproterone acetate
Danazol
Flutamide
Linuron
Testosterone propionate
17alpha-Ethinylestradiol
17beta-Estradiol
2,2-Bis(4-hydroxyphenyl)-1,1,1-trichloroethane **
4-(1,1,3,3-Tetramethylbutyl)phenol
4-Nonylphenol
Benzyl butyl phthalate
Bisphenol A **
Butylparaben
Diethylstilbestrol
Endosulfan
Genistein
o,p-DDE
p,p-DDT
Zearalenone
Betamethasone
Budesonide
Dexamethasone
Fluorometholone
Methylprednisolone
Prednisolone
Triamcinolone
Medroxyprogesterone acetate
Mifepristone
Norethindrone
Progesterone
Bisphenol A diglycidyl ether
Ciprofibrate
Fenofibrate
GW0742*
Indomethacin
L-165041*
Troglitazone
Ketoconazole
Pregnenolone carbonitrile
Rifampicin
all-trans-Retinoic acid
AM580
Arotinoid acid
Bexarotene
Ergocalciferol
Vitamin D3
Biological similarity
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1.0
target
AR
BAR
ESR
GR
PGR
PPAR
PXR
RAR
VDR
B
AR
BAR
CAR
ESR1
ESR2
ESRRB
ESRRG
GR
MR
PGR
PPARA
PPARD
PPARG
PXR
RARA
RARB
RARG
RXRA
RXRB
RXRG
RORA
RORB
RORC
THRA
THRB
VDR
0
4
8
12
Gene expression [NX]
C
ToxCast POD log10 (μM)
HTPP PAC log10 (μM)
2
1
0
-1
-2
-3
-4
-4
-3
-2
-1
0
1
2
**
**

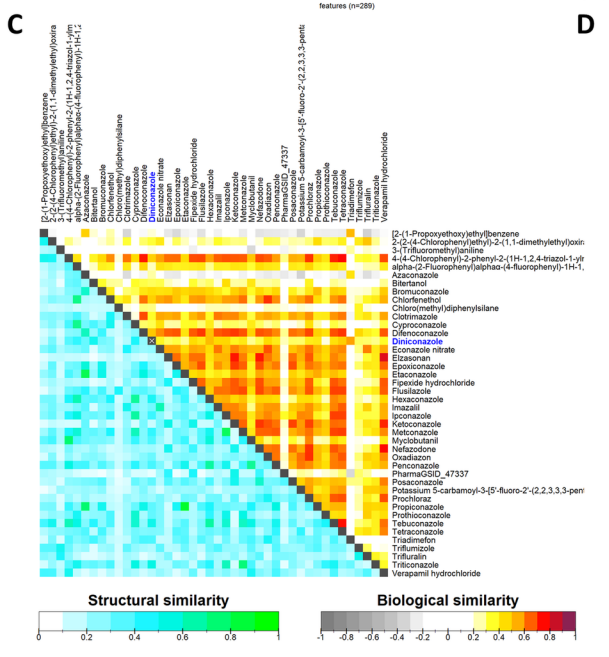

A
Primary screen
B
Secondary screen
C
Orthogonal assay
Biological similarity
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1.0
Betamethasone
Budesonide
Dexamethasone
Fluorometholone
Methylprednisolone
Prednisolone
Triamcinolone
2-Aminoanthraquinone
2-Ethylhexyl 4-(dimethylamino)benzoate
3-(4-tert-Butylphenyl)propanal
Benfluralin
Carbosulfan
Ethoxyquin
Hydramethylnon
Linoleic acid
N-Hydroxybenzamide
N-Phenyl-1-naphthylamine
Pyrene
Tetramethylthiuram monosulfide
Thiram
not tested
inactive
not tested
not tested
ALOX5AP
CRACDL
FKBP5
PER1
ARL4C
CTNNB1
PTK2


A
ToxPrints of 536 chemicals
dendrogram
cut at h = 0.6
234 clusters
retain all clusters with ≥ 2 chemicals
137 clusters
pick 2 chemicals per cluster
274 chemicals
B
Biological similarity
Biological similarity
C
p = 1.43e-07
Biological Similarity
N=1370
N=137
null
real
D
#22: Organochlorines
Aldrin
Dieldrin
Endosulfan
Endrin
Heptachlor
E
DMSO (0.5%)
Dieldrin (100 μM)
DNA
F
#23: Chloroacetamide herbicides
Acetochlor
Alachlor
Butachlor
Metolachlor
Zoxamide
G
#13: Plasticizer
2-Ethylhexyl 4-(dimethylamino)benzoate
Benzyl butyl phthalate
Bis(2-ethylhexyl) terephthalate
Di-n-octyl phthalate
Dibutyl phthalate
H
#15: Food additives and flavors
1,4-Benzoquinone
1,4-Naphthoquinone
d-Carvone
Menadione
R-(-)-Carvone
S-(+)-Carvone
Structural similarity
Biological similarity


A
Color Key and Histogram
Count
0
40000
-1
0
1
Value
3
7
6
9
10
13
2
1
4
14
15
5
11
8
12
chemicals (n = 547)
chemicals (n = 547)
B
Bioassays
Chemical Fingerprints
Stress Response Pathways
Number of enriched entities
Cluster ID
Cluster ID
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15







D

	DTXSID	Chemical Name	Conc [uM]	Biol Sim
1	DTXSID0021125	Phenolphthalein	100	0.550
2	DTXSID2032398	Fludioxonil	30.1	0.547
3	DTXSID2032550	Fenpyroximate (Z,E)	100	0.546
4	DTXSID8046666	Methyl Violet	3	0.535
5	DTXSID2037712	4,4',4''-Ethane-1,1,1-triyltriphenol	100	0.516
6	DTXSID1024338	Thiophanate-methyl	100	0.511
7	DTXSID4022442	Bisphenol B	100	0.509
8	DTXSID5020576	17alpha-Ethinylestradiol	100	0.503
9	DTXSID6032192	Tri-o-cresyl phosphate	100	0.491
10	DTXSID9020114	Auramine hydrochloride	10	0.482
11	DTXSID2023224	Loratadine	100	0.480
12	DTXSID6021248	Rotenone	1	0.478
13	DTXSID3020465	Diethylstilbestrol	30	0.471
14	DTXSID1021243	Rhodamine 6G	10	0.458
15	DTXSID8029868	Atorvastatin	94.5	0.456
16	DTXSID9032379	Dithiopyr	99.6	0.455
16	DTXSID4020870	2,2'-Methylenebis(4-methyl-6-tert-butylphenol)	30	0.455
16	DTXSID1034181	Raloxifene hydrochloride	30	0.455
19	DTXSID8041248	2,5-Di-tert-butylbenzene-1,4-diol	100	0.451
20	DTXSID0020573	17beta-Estradiol	100	0.449
21	DTXSID5032525	Bifenazate	100	0.448
21	DTXSID7024081	Endothal	100	0.448
23	DTXSID1026900	Dodecyltrimethylammonium chloride	30	0.443
24	DTXSID1033664	17-Methyltestosterone	100	0.436
25	DTXSID9032113	Tebuconazole	100	0.435
26	DTXSID6020226	Cadmium chloride	100	0.429
27	DTXSID9032329	Bensulide	100	0.423
28	DTXSID8032548	Fenbuconazole	100	0.422
29	DTXSID5023871	Amitraz	100	0.420
30	DTXSID7029879	Ketoconazole	100	0.417