



Published in final edited form as:

ALTEX. ; 37(4): 623–638. doi:10.14573/altex.2002291.

Rapid Hazard Characterization of Environmental Chemicals Using a Compendium of Human Cell Lines from Different Organs

Zunwei Chen^{1,2}, Yizhong Liu^{1,2}, Fred A. Wright^{3,4}, Weihsueh A. Chiu^{1,2}, Ivan Rusyn^{1,2}

¹Interdisciplinary Faculty of Toxicology, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX, USA;

²Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX, USA;

³Bioinformatics Research Center, North Carolina State University, Raleigh, NC, USA

⁴Departments of Statistics and Biological Sciences, North Carolina State University, Raleigh, NC, USA

Abstract

The lack of adequate toxicity data for the vast majority of chemicals in the environment has spurred the development of new approach methodologies (NAMs). This study aimed to develop a practical high-throughput *in vitro* model for rapidly evaluating potential hazards of chemicals using a small number of human cells. Forty-two compounds were tested using human induced pluripotent stem cell (iPSC)-derived cells (hepatocytes, neurons, cardiomyocytes and endothelial cells), and a primary endothelial cell line. Both functional and cytotoxicity endpoints were evaluated using high-content imaging. Concentration-response was used to derive points-of-departure (POD). PODs were integrated with ToxPi and used as surrogate NAM-based PODs for risk characterization. We found chemical class-specific similarity among the chemicals tested; metal salts exhibited the highest overall bioactivity. We also observed cell type-specific patterns among classes of chemicals, indicating the ability of the proposed *in vitro* model to recognize effects on different cell types. Compared to available NAM datasets, such as ToxCast/Tox21 and chemical structure-based descriptors, we found that the data from the five-cell-type model was as good or even better in assigning compounds to chemical classes. Additionally, the PODs from this model performed well as a conservative surrogate for regulatory *in vivo* PODs and were less likely to underestimate *in vivo* potency and potential risk compared to other NAM-based PODs. In summary, we demonstrate the potential of this *in vitro* screening model to inform rapid risk-based decision-making through ranking, clustering, and assessment of both hazard and risks of diverse environmental chemicals.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.

Correspondence: Ivan Rusyn, MD, PhD, Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843, USA, (irusyn@tamu.edu).

Conflict of interest

All authors declare they have no competing interests.

1 Introduction

Most regulatory frameworks for evaluating the safety of drugs and chemicals include a requirement for studies in animals; however, because of the low throughput and high cost of these studies, considerable toxicological information gaps exist for most chemicals in commerce (Locke and Myers, 2011; Taylor et al., 2014; Kavlock et al., 2018). The development of novel non-animal models, both cell-based and computational approaches, to replace animals as the default option in chemical safety evaluation was stimulated by ethical and political pressures (Taylor, 2018), advances in biomedical research and technology, and the need to address the potential hazards from thousands of chemicals in commerce and the environment (NRC, 2007). In the United States and in the European Union, recent changes to the laws that govern the evaluation of commodity and environmental chemicals include provisions that encourage the use of alternative test methods for hazard and risk assessment applications, such as read-across, prioritization, and screening (ECHA, 2016; US EPA, 2018; Taylor et al., 2014). Novel analytical and *in vitro* data, now commonly referred to as new approach methodologies (NAMs), are being used in support of regulatory decisions (Kavlock et al., 2018; Paul Friedman et al., 2020); however, concerns about the limitations of NAMs in decision-making also have been voiced (Gocht et al., 2015; Berggren et al., 2015). The US Environmental Protection Agency (EPA) is developing a strategic plan to reduce the use of vertebrate animals in testing chemical substances and promote the development of alternative test methods; the goal is to eliminate animal testing from regulatory requirements for pesticides and industrial chemicals by 2035 (US EPA, 2019).

The efforts to expand the portfolio of NAMs and test their utility in decision-making are most prominent in the European Union (Daston et al., 2015; Berggren et al., 2015; Desprez et al., 2018; Escher et al., 2019) and the United States (Thomas et al., 2018; Judson et al., 2010a; Kavlock et al., 2018). Data on thousands of chemicals that have been tested in hundreds of lower organism, cell- or molecular-based assays (Kleinstreuer et al., 2014) are publicly available (Williams et al., 2017). These data are used to derive quantitative hazard predictions (Bell et al., 2018; Wetmore, 2015; Pearce et al., 2017; Wambaugh et al., 2015), to address potential data gaps (Chiu et al., 2018; Guyton et al., 2018), and to derive estimates of human health risk when combined with human exposure data or estimates (Sipes et al., 2017; Sirenko et al., 2017; Rotroff et al., 2010; Paul Friedman et al., 2020).

Notwithstanding recent advances in the development of NAMs and publications of a number of case studies on their use for decision-making, many stakeholders, both the industry and the regulators, remain unsure as to what assay(s) should be used to gather data on chemicals or mixtures not currently in ToxCast/Tox21 programs. A traditional approach to development of cell-based models for animal study replacement is to focus on one organ/tissue of concern to the toxicologists, such as the liver (Soldatow et al., 2013), central nervous system (Schmidt et al., 2017), kidney (Su et al., 2016), lung (Lee et al., 2018) or heart (Blanchette et al., 2019). Examples of a successful effort to create targeted sets of *in vitro* assays for a particular decision context are proposals to replace rat uterotrophic (Browne et al., 2015) and Hershberger (Kleinstreuer et al., 2018) assays. In addition, some decision contexts require rapid evaluation of potential chemical hazards in a limited number of assays, such as in response to chemical spills (Judson et al., 2010b; NTP, 2016). Still,

little consensus exists with respect to which assays are readily accessible, whether they are reproducible, and how the data shall be analyzed and interpreted.

It also has been reasoned that the pace of transition from animal data to NAMs will depend on the pace at which these new models are optimized to reflect the biology of humans, rather than that of animals (Herrmann et al., 2019). Cancer cell lines, primary cells isolated from non-transplant grade donor tissues, and induced pluripotent stem cell (iPSC)-derived cells are current options for studies of human biology *in vitro*. Of these choices, iPSC-derived organotypic cells are the most physiological and reproducible cell-based model for animal replacement (Anson et al., 2011); however, little toxicological data is available in iPSCs as they are not yet part of ToxCast/Tox21.

In this study, we aimed to conduct an initial test of the performance of a compendium of human *in vitro* models that comprise a small but diverse array of tissues of interest using a representative set of chemicals with known regulatory toxicity values that exemplify major distinct classes of contaminants found on Superfund sites. Specifically, we hypothesized that these cell-based assays can be used for rapid hazard evaluation and thus represent a sensible targeted set of alternative methods for NAM-enabled rapid risk assessment where timely decisions are needed but regulatory toxicity values are lacking. We show that the data from the five-cell-type model was as good or even better in assigning compounds to chemical classes, as compared to either data from large-scale chemical screening programs or chemical structure-based descriptors. In addition, the quantitative data from this model can serve as a conservative surrogate for regulatory decision-making in rapid hazard evaluation scenarios.

2 Materials and methods

Chemicals and biologicals

For our *in vitro* models, we selected four organ/tissue types from which iPSC-derived cells are available from a commercial vendor. iCell hepatocytes 2.0 (Catalogue # C1023), neurons (Catalogue # C1008), cardiomyocytes (Catalogue # CMC-100-010-001) and endothelial cells (Catalogue # C1023), including cell-specific media and supplements, were from Fujifilm Cellular Dynamics (Madison, WI). Pooled human umbilical vein endothelial cells (HUVECs) in EGM-2 medium (Catalogue # CC-2519A) and the EGMTM-2 BulletKitsTM (Catalogue # CC-3162) were from Lonza (Walkersville, MD). We selected these cell types because many of the chemicals have been shown to be associated with hepatotoxicity, neurotoxicity, cardiotoxicity, and vascular toxicity. Figure S1¹ shows the number of published reports for each type of toxicity as identified in a literature review (results are available through the Health Assessment Workspace Collaborative (Shapiro et al., 2018) web portal (see web links in the legend to Fig. S1¹)). The rationale for cell line selection, metabolic competency of the iCell hepatocyte model, and the justification for selected phenotypes in each cell type are detailed elsewhere (Grimm et al., 2015; Iwata et al., 2017; Sirenko et al., 2014a,b).

¹doi:10.14573/altex.2002291s

Additional reagents used were as follows: CellTiter-Glo[®] reagent was from Promega (Madison, WI, USA). EarlyTox[™] Cardiotoxicity Kits (Part# R8211) were from Molecular Devices (San Jose, CA, USA). RPMI 1640 medium, B-27 medium supplement, gentamicin (50 mg/mL), Calcein AM Green, MitoTracker Orange reagent, Hoechst 33342, human fibronectin, and Geltrex[™] LDEV-Free Reduced Growth Factor Basement Membrane were all from Life Technologies (Grand Island, NY, USA). Recombinant human VEGF was provided by R&D Systems (Minneapolis, MN, USA). Fetal bovine serum (FBS) and Medium 199 were purchased from Fisher Scientific (Waltham, MA, USA). Laminin (Catalogue #L2020-1MG, from Engelbreth-Holm-Swarm murine sarcoma basement membrane) was from Sigma-Aldrich (St. Louis, MO). The authors acknowledge that FBS-free or synthetic FBS-based culture conditions (van der Valk et al., 2018), as well as alternative synthetic basement membrane materials (Nguyen et al., 2017) should be utilized to replace animal-derived products, where appropriate.

The Agency for Toxic Substances and Disease Registry (ATSDR) maintains a priority list of hazardous substances/chemicals² that are frequently detected at the US National Priority List (NPL) sites, also known as “Superfund” sites, and are known human health hazards. From the list of over 300 compounds, we selected 42 chemicals (Tab. 1) based on the review of available information. These compounds represent several classes of pollutants that are ubiquitous in the environment, including polycyclic aromatic hydrocarbons (PAHs, n = 5), inorganic substances (n = 7), phthalates (n = 2), pesticides (n = 20), and other industrial chemicals (n = 8). ATSDR chemical classes are groupings that relate chemicals by similar features based on their structure, uses, physical properties, or other factors. Chemicals were selected for testing based on the following criteria: (i) is listed by ATSDR as priority chemical, (ii) has been evaluated by one or more government agencies and “safe exposure” levels have been established, (iii) was tested in ToxCast/Tox21, and (iv) reverse toxicokinetic and exposure data are publicly available through the EPA dashboard (Williams et al., 2017). Most chemicals were purchased from Sigma-Aldrich, except for heptachlor, heptachlor epoxide, 2,4,5-trichlorophenol, parathion, benzidine and o,p'-DDT, which were from ChemService (West Chester, PA).

Cell culture and chemical treatments

All cells were cultured in 384-well plates according to the manufacturer's (Fujifilm Cellular Dynamics or Lonza) recommendations with respect to cell culture media and supplements. Cell density and other cell culture conditions have been previously published for each of these cell types (Grimm et al., 2015; Iwata et al., 2017; Sirenko et al., 2014a,b) and details are included in Text S1¹. Cells were exposed to test chemicals in descending logarithmic order of concentrations (100, 10, 1, 0.1, and 0.01 μ M). Serial dilutions were originally prepared in 100% cell-culture grade DMSO and then further diluted 100-fold in corresponding cell culture medium to yield 4 \times working solutions in 1% DMSO. The final concentration of DMSO in assay wells following addition of test chemicals was 0.25% (v/v), an amount that was lower than in previous reports where it had no effects on each cell type-derived phenotype (Grimm et al., 2015; Iwata et al., 2017; Sirenko et al., 2014a,b).

²<http://www.atsdr.cdc.gov/spl>

Cytotoxicity assays

Cytotoxicity-related phenotypes in five tested cell types were assessed by high-content live cell imaging after a set exposure time (Tab. 2). Cells were stained with different fluorescent dyes (Hoechst 33342 for nuclei, Calcein AM Green for cytoplasm, and MitoTracker Orange for mitochondria) as detailed in (Grimm et al., 2015; Iwata et al., 2017; Sirenko et al., 2014a,b). Images of all cell culture plates were acquired with ImageXpress Micro Confocal High-Content Imaging System (Molecular Devices) using the DAPI (Hoechst 33342), FITC (Calcein AM Green), and TRITC (MitoTracker Orange) filters at 10× or 20× magnification. Acquired images were processed using the Multi-Wave-length Cell Scoring, Neurite Outgrowth, or Angiogenesis Tube Formation application modules in MetaXpress (Molecular Devices) image processing software, and quantitative data were extracted for concentration-response modeling (see below). In addition, ATP production of iCell neurons and HUVECs was evaluated using CellTiter-Glo assay as described in Text S2¹.

Physiologically-relevant phenotype assays

Physiologically-relevant phenotypes of each cell type were evaluated as detailed in Table 2 and reported previously (Grimm et al., 2015; Iwata et al., 2017; Sirenko et al., 2014a,b). Effects on the mitochondrial integrity and intensity of iCell hepatocytes, and neurite outgrowth of iCell neurons were measured using high-content imaging (ImageXpress Micro Confocal High-Content Imaging System, Molecular Devices). Calcium flux reflecting the contract beating of iCell cardiomyocytes was determined by a FLIPR tetra (Molecular Devices) instrument using EarlyTox™ Cardiotoxicity Kit as described in Text S3¹. Effects on angiogenesis of both iCell endothelial cells and HUVECs were measured by 3D cell culture using an extracellular gel matrix and followed by high content imaging as detailed in Text S4¹.

Assay quality controls and concentration-response profiling

The qualitative integrity of the screening assays in this study was evaluated using previously established conditions (Grimm et al., 2015). All responses were normalized to the vehicle control (0.25% DMSO-treated wells). Overall, quality control criteria were established to evaluate each cell-based assay based on five parameters (see Tab. S1, S2¹): (i) variance in replicate wells for two negative controls (vehicle-treated wells and cell medium only), (ii) the difference between two negative controls (vehicle vs cell culture media), (iii) intra- and (iv) inter-plate replicate correlation, and (v) EC₅₀ of the positive control chemicals/drugs that were specific for each cell type.

Vehicle control-scaled data for each treatment were fitted to a curve with a nonlinear logistic function to determine point-of-departure (POD) values, defined as the concentrations at which the fitted curve exceeds one standard deviation above or below the mean of vehicle-treated controls, using R software-based script as previously reported (Sirenko et al., 2013). The choice of one standard deviation “benchmark response” was based on the US EPA guidance for dose-response modeling and determination of the point-of-departure values (US EPA, 2012), as well as empirical testing of various thresholds as detailed in (Sirenko et al., 2013), which showed that a choice of one standard deviation generates consistently high classification accuracy.

Data integration in ToxPi

For data integration and visualization in Toxicological Priority Index Graphical User Interface (ToxPi GUI) (Marvel et al., 2018), we selected 48 phenotypes from all five cell types (Tab. 2). Following the standard ToxPi data protocol, POD values for each phenotype were inversely scaled on a 0–1 scale, with 0 representing the highest POD value in a given data set (i.e., the lowest observed bioactivity) and 1 representing the lowest measured POD value (i.e., the highest observed bioactivity). These scaled POD values were then used as quantitative inputs for bioactivity profiling in ToxPi.

Clustering and classification analyses

We used two approaches to grouping chemicals based on the biological profiling produced in this study, the bioactivity data from ToxCast/Tox21, and chemical structure-based Morgan fingerprint data. In an unsupervised analysis, chemicals were grouped based on the similarity between the biological/chemical profiling of the chemicals, without prior knowledge of chemical categories. To evaluate the outcome of such grouping, we include a quantitative metric into the unsupervised analysis workflow to assess the correspondence of the outcome to the original categories of each chemical. The details of the unsupervised analysis workflow are described elsewhere (Onel et al., 2019). The Fowlkes-Mallows (FM) index (Fowlkes and Mallows, 1983), a measure of similarity of two clusters, was calculated to enable quantitative comparative assessment between groupings achieved using each dataset to the known chemical categories. The higher the FM index, the more similar the grouping based on *in vitro* or chemical descriptor data was to the “perfect” grouping as shown in Table 1. The FM index ranges from 0.0 (no correspondence) to 1.0 (perfect correspondence). One-sided p-values for the FM index (using the null hypothesis of random assignment) were obtained using a standard z-statistic (Fowlkes and Mallows, 1983) that compares the observed value to the null expectation.

In the supervised analysis, assignments of chemicals to classes (Tab. 1) were used to build classification models, which were then used to predict the class for an unknown chemical. The term “supervised” is a statistical term (Kotsiantis, 2007) referring to models that are trained to perform automatic classification based on the available features, and using the classes as pre-defined groupings. In a supervised analysis, the intent is to identify the features that are best able to distinguish among the classes. For this purpose, the randomForest package in R v3.5 was used for class prediction, with 5-fold cross validation implemented in 50 random training/test data splits. The overall prediction accuracy from each database was calculated from cross-validation confusion matrices and the important distinguishing descriptors were further identified. A primary difference between unsupervised and supervised analysis is that the latter focuses on features that best distinguish among existing chemical categorizations.

Comparison to *in vivo* POD data and margin of exposure estimates

In vivo data are still the most commonly used PODs for use in regulatory decision-making, but recent analyses have suggested that NAM-based PODs may be useful as conservative surrogates for *in vivo* values (Paul Friedman et al., 2020). Thus, for the 42 chemicals in this study, we used the *in vivo* PODs from which the regulatory reference doses (RfDs) were

derived (POD_{RfD} values) as a benchmark. Specifically, we first compared the POD_{RfD} values to various NAM-based PODs, including the *in vitro* POD derived from this study using iPSC-derived cells and HUVECs, as well as two other *in vitro* data sets: the minimum of the distribution of 50% maximal activity concentration (AC_{50}) of high throughput *in vitro* assays in ToxCast database (i.e., most sensitive assay) and conservative POD_{NAM} values reported in (Paul Friedman et al., 2020). In addition, using ExpoCast exposure estimates, we compared margin of exposure (MoE) estimates based on POD_{RfD} values with those based on NAM data. Oral dose-based PODs or exposures were converted to C_{ss} (concentration at steady-state)-based values (or *vice versa*) using the high throughput toxicokinetic (httk) (Pearce et al., 2017) R package (v 1.10.1) at the upper 95th percentile for toxicokinetic variability. Due to the limitation of the availability of each data stream, only the chemicals shared in all the databases were taken into consideration for comparison (see details in Tab. S3¹).

3 Results

3.1 Screening assays and concentration-response profiling

In vitro effects of the test chemicals were evaluated for a wide range of functional and cytotoxicity phenotypes in five human cell types that represent four tissues (Tab. 2). POD values were derived from the concentration-response relationships for a total of 48 phenotypes (see quality control data for each phenotype in Tab. S1 and S2¹) and plotted (Fig. 1) separately for each cell type. Chemicals are grouped by their chemical class and ranked within each class from least to most bioactive based on the median response in iCell hepatocytes. Both for the individual chemicals and within a chemical class, there was a wide range of potency across all phenotypes. Each chemical had an effect in at least one cell type and no correlation in PODs was evident among cell types (Fig. S2¹), indicating that the chemicals elicited cell type-specific effects.

When the PODs were grouped by cell type (Fig. 2), the iCell cardiomyocytes clearly were, on average, the most sensitive to these chemicals. Across the 48 phenotypes included in the analysis, there was a wide range of effects for most of the evaluated chemicals. Not only were there chemicals that had effects at low concentrations, but there was a pronounced shift in the median and inter-quartile range, and for most of the phenotypes that were evaluated (Fig. 2, right panel). In other cell types, few chemicals had pronounced effects while most exhibited effects only at nominal test concentrations above 10 μ M. It is noteworthy that fewer effects were observed in metabolically-active iCell hepatocytes (Sirenko et al., 2014b) compared to other cell types. iCell endothelial cells were most resistant to the effects of chemicals tested in this study. In addition, functional effects had significantly lower PODs compared to cytotoxicity phenotypes, indicating higher sensitivity, in all *in vitro* data combined, and in data from iCell hepatocytes, cardiomyocytes and HUVECs (Fig. S3¹).

3.2 Ranking and clustering using ToxPi scores

To facilitate interpretation of the data from these experiments that involved five cell types and 48 phenotypes, we aggregated the concentration-response data and PODs derived from *in vitro* screening assays using the Toxicological Priority Index (ToxPi) (Marvel et al.,

2018). Each cell type was assigned an individual ToxPi “slice” (Fig. 3A). Specifically, PODs were converted into ToxPi scores as detailed in Section 2 and in Marvel et al. (2018). For each slice, the distance that the arc extends from the origin is proportional to its relative evidence of concern (e.g., longer = greater hazard because of lower POD), and the radial angle (width) indicates its weight in the overall model (in this analysis, data from each cell type were weighed equally). ToxPi scores were further combined into one pie chart to indicate the overall effect of each chemical on all five human cell types. ToxPi for three of the 42 tested chemicals are shown as examples in Figure 3B. Cadmium chloride showed the highest bioactivity (lowest PODs) in iCell hepatocytes compared to the other cell types, resulting in a large green slice in the ToxPi. Mercuric chloride and methoxychlor showed highest effects on iCell neurons and iCell cardiomyocytes, respectively.

The overall ToxPi scores for each chemical, reflecting the average of the normalized input scores for each slice of the respective bioactivity profile, were then used as a score to rank and cluster chemicals according to their overall bioactivity (Fig. 4A). ToxPi ranking using quantitative bioactivity data can be used for chemical prioritization (Reif et al., 2010). The 42 tested chemicals were ranked based on the summed effects in the five human cell lines. The three inorganic substances (mercuric chloride, cadmium chloride and potassium chromate) had the highest overall bioactivity score (Fig. 4B). When bioactivity profiles of the individual chemicals were combined into their respective classes, inorganic substances were on average most bioactive, followed by pesticides, phthalates, other industrial chemicals, and PAHs (Fig. 4C, Tab. 3). Furthermore, specific effects of different classes of chemicals on certain cell types were identified. While inorganic substances were bioactive in most cell types, pesticides had the highest bioactivity in iCell cardiomyocytes (Tab. 3, Fig. S4¹).

Chemicals were also clustered using ToxPi scores and bioactivity profiles (Fig. 4D). This visualization shows that while some compounds are clustered because of their relatively high potency (mercuric chloride, cadmium chloride and potassium chromate), other compounds have similar ToxPi profiles, indicating similarity in their effects on different cell types. For example, DDT-like organochlorine pesticides are clustered closely because of the similarity in both potency and effects across all five cell types. Similarly, other organochlorine pesticides cluster together because they showed the highest relative bioactivity in iCell cardiomyocytes. In addition, phenotype-specific effects of each chemical on each cell type were further identified by clustering chemicals using data on each cell type (Fig. S5¹). Cadmium chloride exhibited the most pronounced effects on iCell hepatocytes by affecting all phenotypes. Mercuric chloride dominated effects on iCell neurons. Pesticide methoxychlor was the most bioactive in iCell cardiomyocytes. iCell endothelial cells and HUVECs were most affected by potassium chromate.

3.3 Bioactivity-based class unsupervised grouping

Next, we tested how well the bioactivity data on the individual cell type, or in combination, can be used for grouping of tested chemicals into classes. A quantitative comparison of the unsupervised analysis was conducted using the Fowlkes-Mallows (FM) index (Fowlkes and Mallows, 1983; Onel et al., 2019). The results of the clustering were compared to the known

chemical groupings (Tab. 1) that were used as a reference. Figure 5A shows that clustering using the bioactivity profiles of the combination of all five cell types resulted in the highest FM index (FM = 0.56) and was highly significant compared to that expected under random permutation ($p < 0.001$). Among the individual cell types, iCell hepatocytes showed the highest FM index (FM = 0.41), albeit it was not significant. Data from HUVECs was least informative in this analysis. Because of the pronounced heterogeneity in the “value” of information from different cell types, we also evaluated whether even smaller sets of cell types may have clustering accuracy approaching the data on all five cell types. We found that a combination of the data from iCell cardiomyocytes and iCell neurons yielded an FM index that was as high as when the data from all five cell types was used (FM = 0.53, Fig. S6¹).

We also compared the ability of the targeted dataset obtained in this study to group chemicals into classes to that of a larger ToxCast/Tox21 *in vitro* dataset, or chemical structure-based descriptors (Morgan chemical fingerprints). Figure 5B shows that *in vitro* data on 48 phenotypes from five cell types obtained in this study has a higher FM index for grouping of 42 chemicals into five classes compared to other information that is available on these compounds. Figures 5C–E show the individual dendrograms for each of the comparisons in Figure 5B.

3.4 Bioactivity-based class supervised grouping

A different type of question that is often asked when using NAM data in decision-making is whether one can use the data obtained in the same set of assays as those for the compounds in a database to classify a new compound into a class. We conducted supervised analyses using a cross-validated random forest algorithm where every test compound was predicted using a classification model. In contrast to the unsupervised analysis, the supervised analysis attempts to train a model to identify the features that are most predictive of existing classification. Figure 6 shows the outcomes of the cross-validated classifications for each data type. Numbers on the top left to bottom right diagonals show correct class prediction, and the numbers off the diagonal show misclassifications and which class the compounds were misclassified into. Overall, the Morgan fingerprints-based classification was superior (81% accurate prediction) when compared to classifications based on either data from this study or ToxCast/Tox21 data (60% and 69%, respectively). It is also noteworthy that the *in vitro* data generated in this study can accurately classify most pesticides into the correct chemical class, whereas ToxCast/Tox21 data classified all inorganic substances correctly. The combination of the *in vitro* data and Morgan fingerprints, or combination of two *in vitro* datasets (Fig. S7A¹) did not improve prediction accuracy. The accuracy of classification with each type of data was significantly better than random assignment into classes (Fig. S7B¹). We emphasize that our prediction by supervised analysis was performed using cross-validation, which avoids overfitting inherent in fitting complex prediction models. If the chemical classes used had been truly meaningless, in the sense of “random,” then our reasonably high prediction accuracy values would not have been achieved. The prediction accuracy results suggest that the *a priori* classification is meaningful, and, in contrast to unsupervised analysis, highlight the specific measured biological features that are best able to discriminate among classes, as described below.

The supervised classification analysis, where every test compound was predicted using a classification model, can also be examined for information on the “most informative” features (i.e., features that are most predictive of existing classification) on which the models were developed. The top 10 most informative features from each dataset, i.e., phenotypes that contributed the most to the accuracy of the classification, are shown in Figure 7. Interestingly, for the *in vitro* data generated in this study, 5 of the top 10 most informative descriptors were functional phenotypes from iCell cardiomyocytes, followed by phenotypes from iCell neurons (Fig. 7A). For ToxCast/Tox21 data, the descriptors in the top 10 included largely disparate data from a wide range of models, i.e., from zebrafish, to cytotoxicity, to reporter assays (Fig. 7B). While Morgan fingerprints are difficult to interpret directly (Fig. 7C), a combination of bioactivity and chemical structure data showed that chemical descriptors do not dominate the list of informative features, and that *in vitro* data may be equally informative (Fig. 7D).

3.5 Comparison to *in vivo* POD data and margin of exposure estimates

It has recently been proposed that NAM-based PODs can serve as conservative surrogates for traditional *in vivo* PODs (Paul Friedman et al., 2020). Thus, we first compared various NAM-based PODs, including those based on our five cell types, to the regulatory PODs used as the basis for RfD toxicity values (POD_{RfD}). For our *in vitro*-based PODs, we used either the most sensitive POD for each cell type or the most sensitive POD across all cell types combined (Fig. 8). As shown in Figure 8A, only when all cell types are combined do our *in vitro* PODs represent a conservative surrogate for the POD_{RfD} , with only 25% of our *in vitro* PODs being higher than the corresponding POD_{RfD} , and those remaining 25% being within 10-fold of the *in vivo* value. In contrast, as shown in Figure 8C, only the approach of using the minimum (most sensitive) ToxCast AC_{50} has similarly conservative results, whereas cardiomyocytes alone and the POD_{NAM} from (Paul Friedman et al., 2020), which is a lower 5th percentile, had a substantial number of “unconservative” results. Note that these results appear to contrast with those reported by (Paul Friedman et al., 2020) because they used *in vivo* PODs from ToxRefDB, whereas we used the *in vivo* PODs that supported regulatory RfD toxicity values (Wignall et al., 2014).

A related comparison was with respect to the resulting screening-level risk characterization using a Margin of Exposure (MoE) approach. Specifically, we used a MoE benchmark of < 100 as an indication of “potential concern.” As shown in Figure 8B, more than half of the chemicals have implied MoEs less than a benchmark of 100 when using all cell types combined, with similar results for cardiomyocytes, but far fewer chemicals are suggested to be of “potential concern” when using other cell types. In Figure 8D, when restricting to chemicals common across different NAM-based approaches, we find that the POD_{RfD} -based “ground truth” suggests that only 2/16 chemicals are of “potential concern.” Using only iCell cardiomyocytes, or using all cell types, results in a more conservative estimate of 4 to 5/16 chemicals, with the median MoE being slightly more conservative than the *in vivo*-based MoE. In contrast, using the POD_{NAM} from (Paul Friedman et al., 2020) results in an “unconservative” estimate of only 1/16 chemicals of potential concern, with the median MoE being much higher (implying “safer”) than the *in vivo*-based MoE.

Overall, for this limited dataset, our PODs derived from high throughput *in vitro* data from five human cell types performed well as a conservative surrogate for regulatory *in vivo* PODs and were less likely to underestimate *in vivo* potency and potential risk compared to other NAM-based PODs.

4 Discussion

It is widely recognized that the future of regulatory toxicology lies in high-throughput *in vitro* assays and computational models based on human biology, rather than in continued testing in laboratory animals (NRC, 2007; National Academies of Sciences Engineering and Medicine, 2017). A wide array of both biological and computational tools is available to probe human function and disease at the molecular level through the transcriptome, epigenome, proteome and metabolome (Nielsen, 2017). Many thousands of immortalized cell lines collected from various tissues and individuals are now used in toxicological research (Chiu and Rusyn, 2018). There are large databases of publicly available biological data that can be explored to develop hypotheses about how chemicals, genes, and diseases may be connected (Miller, 2016; Davis et al., 2019; Williams et al., 2017). There are genetically diverse mammalian and non-mammalian models, *in vivo* and *in vitro*, that are used for toxicological research (Zeise et al., 2013). Complex human biology is being replicated in multicellular perfused microphysiological systems that mimic certain tissue functions (Marx et al., 2020). It appears that the field of regulatory science has finally overcome the long-lamented challenge of shortage of information for decisions on chemical safety (Lutter et al., 2013).

Alas, the quantity of the information now available is yet to be translated into actual examples of using these data in various decision contexts beyond now well-accepted screening-level, risk-based chemical prioritization (Harrill et al., 2019; Paul Friedman et al., 2020), or filling data gaps (Guyton et al., 2018). For new chemicals, complex substances, or mixtures, what is a sensible compendium of *in vitro* and *in silico* models that may satisfy the data requirements for a particular decision context? A number of examples have been published recently to address this question, especially in the context of grouping and read-across (De Abrew et al., 2019; Zhu et al., 2016; Escher et al., 2019). Indeed, it is critically important to establish both the strengths and limitations of cell-based *in vitro* screening methods, so that promising NAMs can be generated and used for decision-making in human and environmental health.

This study, even though primarily focused on an *in vitro* model that can be used for rapid hazard assessment, adds to the overall body of recent evidence on the topic of the utility of NAMs. We aimed to test performance of a small set of human *in vitro* models that represent a diverse array of tissues of interest to regulatory toxicologists. We took advantage of recently developed reproducible and physiologically-relevant human *in vitro* models derived from iPSCs (Li and Xia, 2019; Anson et al., 2011), models that are excellent replacements for animal tests and for which detailed methods and metrics of reproducibility have been established (Sirenko et al., 2013, 2014a,b; Grimm et al., 2018; Iwata et al., 2017; Klaren and Rusyn, 2018). We posited that commercially-available iPSC-derived cells are poised for wider use, replacement of animal studies, and inter-comparison of the outcomes in a

rigorous and reproducible manner (Anson et al., 2011). Presence of advanced cellular functions and absence of genetic drift because of repeated passaging, both problems of cancer cell lines, are advantages of iPSC-derived differentiated cells in toxicity testing (Kim et al., 2019). Our hypothesis was that these cell-based models, when probed for both physiological and toxicological effects of chemicals, can be used for rapid hazard evaluation and thus represent a sensible targeted set of alternative methods for NAM-enabled decisions, especially under conditions of rapid evaluations such as emergency response (Judson et al., 2010b).

Even though this study is not the first to attempt to probe the ability of a small dataset to group and classify diverse environmental chemicals, a number of important learnings have emerged. First, our comparison of cells representing various tissue types showed that iPSC-derived cardiomyocytes may be among the cell types that are most sensitive to effects across various chemical classes. This is noteworthy because iCell cardiomyocytes can be used as a highly reproducible *in vitro* model that faithfully replicates many *in vivo* cardiotoxic phenotypes (Grimm et al., 2018). Our previous studies showed that environmental chemicals have adverse effects on cardiomyocytes, similar to many known cardiotoxic drugs (Sirenko et al., 2017; Burnett et al., 2019; Blanchette et al., 2019); however, it is noteworthy that this metabolism-limited cell type was most affected by the diverse set of Superfund priority chemicals from different classes.

Second, the fact that the chemicals tested in this study showed very divergent effects across multiple cell types, leading to distinct class-specific bioactivity profiles that can be used to group substances, also strongly supports the need for tissue diversity of *in vitro* models. Moreover, when used for NAM-based risk characterization, multiple cell types together performed better than any individual cell type for ensuring that the risk is not underestimated. These findings suggest that when testing is not meant to be mechanism- or effect-based, inclusion of cells from multiple tissues should be a design principle for *in vitro* test batteries that are to be used as NAMs. Such tissue-diverse data should also increase confidence in the “biological coverage” of *in vitro* NAMs.

Third, we observed that *in vitro* bioactivity data may be as good as or, in some cases better than, chemical descriptors for grouping of chemical substances into classes. In addition, important synergies are realized when biological and chemical descriptors are combined. These findings are in line with previous observations that chemical-biological data are most powerful for grouping (Low et al., 2011, 2013, 2014), as well as that they are most interpretable by the decision-makers (Zhu et al., 2016).

Finally, we found that a limited set of *in vitro* data may be equally or even more informative than the much larger datasets from large-scale chemical screening programs (Thomas et al., 2018). Overabundance of NAM data is not necessarily a recipe for more accurate prediction, as has been shown for various types of biological (Kreutz et al., 2013) and chemical (Fourches et al., 2015) data. One approach to dealing with such “big data” problems is to apply variable selection (Knudsen et al., 2013) or deep learning (Grapov et al., 2018) algorithms to uncover meaningful “signals” in large datasets. Regretfully, these exercises seldom have resulted in selection of a reasonably small set of assays/endpoints that are

reasonably accurate for prediction and do not require extensive and lengthy experimentation. Only recently, influential examples have emerged of how a small set of assays can be used to replace a specific animal test (Kleinstreuer et al., 2018; Browne et al., 2015). On the other hand, the data from our study performed at least as well, if not better, than larger NAM datasets, not only for grouping of chemicals into classes, but also in serving as surrogate NAM-based PODs for rapid risk characterization. Additional confidence in these results could be obtained by evaluating a larger set of ToxCast/Tox21 chemicals.

Notwithstanding the need for diverse high-throughput *in vitro* data streams to rapidly inform hazard identification and to fill the knowledge gap for chemicals with minimum toxicity data, challenges remain about their use in prioritization and screening level assessment strategies as well as tradeoffs between speed and uncertainty (Paul Friedman et al., 2020). For instance, while high throughput screening data could play key roles in decision-making for emergency response, there are many limitations with respect to predicting chemical fate and effects in the environment, challenges that might lead to potentially missed hazards (Ginsberg et al., 2019). Furthermore, there is also uncertainty in the extrapolation from *in vitro* bioactivity to *in vivo* toxicity (Bell et al., 2018), and gaps exist in the cell-based *in vitro* screening and potential effects on human health since most cell assay endpoints are still related to cytotoxicity and non-specific effects (Judson et al., 2016). Overall, however, our findings support the notion that the field of *in vitro* toxicology and NAM implementation would be well served by agreeing on a reasonably small subset of differentiated, human cell-based models with both cytotoxicity-based and functional readouts that can be used in different decision contexts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors wish to thank Drs Lora Yanagisawa, Burcu Beycal, and Nan-Hung Hsieh (Texas A&M University) for technical assistance with some of the experiments and data processing, and Drs David Reif (North Carolina State University), Grace Patlewicz (U.S. EPA) and Oksana Sirenko (Molecular Devices, Inc.) for useful discussion. This work was supported, in part, by a grant from the National Institutes of Health (P42 ES027704) and a cooperative agreement with the United States Environmental Protection Agency (STAR RD83580201). The views expressed in this manuscript do not reflect those of the funding agencies. The use of specific commercial products in this work does not constitute endorsement by the funding agencies.

References

- Anson BD, Kolaja KL and Kamp TJ (2011). Opportunities for use of human iPS cells in predictive toxicology. *Clin Pharmacol Ther* 89, 754–758. doi:10.1038/clpt.2011.9 [PubMed: 21430658]
- Bell SM, Chang X, Wambaugh JF et al. (2018). In vitro to in vivo extrapolation for high throughput prioritization and decision making. *Toxicol In Vitro* 47, 213–227. doi:10.1016/j.tiv.2017.11.016 [PubMed: 29203341]
- Berggren E, Amcoff P, Benigni R et al. (2015). Chemical safety assessment using read-across: Assessing the use of novel testing methods to strengthen the evidence base for decision making. *Environ Health Perspect* 123, 1232–1240. doi: 10.1289/ehp.1409342 [PubMed: 25956009]
- Blanchette AD, Grimm FA, Dalajamts C et al. (2019). Thorough QT/QTc in a dish: An in vitro human model that accurately predicts clinical concentration-QTc relationships. *Clin Pharmacol Ther* 105, 1175–1186. doi:10.1002/cpt.1259 [PubMed: 30346629]

- Browne P, Judson RS, Casey WM et al. (2015). Screening chemicals for estrogen receptor bioactivity using a computational model. *Environ Sci Technol* 49, 8804–8814. doi:10.1021/acs.est.5b02641 [PubMed: 26066997]
- Burnett SD, Blanchette AD, Grimm FA et al. (2019). Population-based toxicity screening in human induced pluripotent stem cell-derived cardiomyocytes. *Toxicol Appl Pharmacol* 381, 114711. doi:10.1016/j.taap.2019.114711 [PubMed: 31425687]
- Chiu WA and Rusyn I (2018). Advancing chemical risk assessment decision-making with population variability data: Challenges and opportunities. *Mamm Genome* 29, 182–189. doi:10.1007/s00335-017-9731-6 [PubMed: 29299621]
- Chiu WA, Guyton KZ, Martin MT et al. (2018). Use of high-throughput in vitro toxicity screening data in cancer hazard evaluations by IARC monograph working groups. *ALTEX* 35, 51–64. doi:10.14573/altex.1703231 [PubMed: 28738424]
- Daston G, Knight DJ, Schwarz M et al. (2015). SEURAT: Safety evaluation ultimately replacing animal testing – Recommendations for future research in the field of predictive toxicology. *Arch Toxicol* 89, 15–23. doi:10.1007/s00204-014-1421-5 [PubMed: 25433540]
- Davis AP, Grondin CJ, Johnson RJ et al. (2019). The comparative toxicogenomics database: Update 2019. *Nucleic Acids Res* 47, D948–D954. doi:10.1093/nar/gky868 [PubMed: 30247620]
- De Abrew KN, Shan YK, Wang X et al. (2019). Use of connectivity mapping to support read across: A deeper dive using data from 186 chemicals, 19 cell lines and 2 case studies. *Toxicology* 423, 84–94. doi:10.1016/j.tox.2019.05.008 [PubMed: 31125584]
- Desprez B, Dent M, Keller D et al. (2018). A strategy for systemic toxicity assessment based on non-animal approaches: The cosmetics Europe long range science strategy programme. *Toxicol In Vitro* 50, 137–146. doi:10.1016/j.tiv.2018.02.017 [PubMed: 29499337]
- ECHA – European Chemicals Agency (2016). New Approach Methodologies in Regulatory Science – Proceedings of a scientific workshop Helsinki, 19–20 4 2016. doi:10.2823/543644
- Escher SE, Kamp H, Bennekou SH et al. (2019). Towards grouping concepts based on new approach methodologies in chemical hazard assessment: The read-across approach of the EU-ToxRisk project. *Arch Toxicol* 93, 3643–3667. doi:10.1007/s00204-019-02591-7 [PubMed: 31781791]
- Fourches D, Muratov E and Tropsha A (2015). Curation of chemogenomics data. *Nat Chem Biol* 11, 535. doi:10.1038/nchembio.1881 [PubMed: 26196763]
- Fowlkes EB and Mallows CL (1983). A method for comparing two hierarchical clusterings. *J Am Stat Assoc* 78, 553–569. doi:10.1080/01621459.1983.10478008
- Ginsberg GL, Pullen Fedinick K, Solomon GM et al. (2019). New toxicology tools and the emerging paradigm shift in environmental health decision-making. *Environ Health Perspect* 127, 125002. doi:10.1289/EHP4745 [PubMed: 31834829]
- Gocht T, Berggren E, Ahr HJ et al. (2015). The SEURAT-1 approach towards animal free human safety assessment. *ALTEX* 32, 9–24. doi:10.14573/altex.1408041 [PubMed: 25372315]
- Grapov D, Fahrman J, Wanichthanarak K et al. (2018). Rise of deep learning for genomic, proteomic, and metabolomic data integration in precision medicine. *OMICS* 22, 630–636. doi:10.1089/omi.2018.0097 [PubMed: 30124358]
- Grimm FA, Iwata Y, Sirenko O et al. (2015). High-content assay multiplexing for toxicity screening in induced pluripotent stem cell-derived cardiomyocytes and hepatocytes. *Assay Drug Dev Technol* 13, 529–546. doi:10.1089/adt.2015.659 [PubMed: 26539751]
- Grimm FA, Blanchette A, House JS et al. (2018). A human population-based organotypic in vitro model for cardiotoxicity screening. *ALTEX* 35, 441–452. doi:10.14573/altex.1805301 [PubMed: 29999168]
- Guyton KZ, Rusyn I, Chiu WA et al. (2018). Application of the key characteristics of carcinogens in cancer hazard identification. *Carcinogenesis* 39, 614–622. doi:10.1093/carcin/bgy031 [PubMed: 29562322]
- Harrill J, Shah I, Setzer RW et al. (2019). Considerations for strategic use of high-throughput transcriptomics chemical screening data in regulatory decisions. *Curr Opin Toxicol* 15, 64–75. doi:10.1016/j.cotox.2019.05.004 [PubMed: 31501805]

- Herrmann K, Pistollato F and Stephens ML (2019). Beyond the 3Rs: Expanding the use of human-relevant replacement methods in biomedical research. *ALTEX* 36, 343–352. doi:10.14573/altex.1907031 [PubMed: 31329258]
- Iwata Y, Klaren WD, Lebakken CS et al. (2017). High-content assay multiplexing for vascular toxicity screening in induced pluripotent stem cell-derived endothelial cells and human umbilical vein endothelial cells. *Assay Drug Dev Technol* 15, 267–279. doi:10.1089/adt.2017.786 [PubMed: 28771372]
- Judson RS, Houck KA, Kavlock RJ et al. (2010a). In vitro screening of environmental chemicals for targeted testing prioritization: The ToxCast project. *Environ Health Perspect* 118, 485–492. doi:10.1289/ehp.0901392 [PubMed: 20368123]
- Judson RS, Martin MT, Reif DM et al. (2010b). Analysis of eight oil spill dispersants using rapid, in vitro tests for endocrine and other biological activity. *Environ Sci Technol* 44, 5979–5985. doi:10.1021/es102150z [PubMed: 20602530]
- Judson R, Houck K, Martin M et al. (2016). Analysis of the effects of cell stress and cytotoxicity on in vitro assay activity across a diverse chemical and assay space. *Toxicol Sci* 153, 409. doi:10.1093/toxsci/kfw148 [PubMed: 27605417]
- Kavlock RJ, Bahadori T, Barton-Maclaren TS et al. (2018). Accelerating the pace of chemical risk assessment. *Chem Res Toxicol* 31, 287–290. doi:10.1021/acs.chemrestox.7b00339 [PubMed: 29600706]
- Kim TW, Che JH and Yun JW (2019). Use of stem cells as alternative methods to animal experimentation in predictive toxicology. *Regul Toxicol Pharmacol* 105, 15–29. doi:10.1016/j.yrtph.2019.03.016 [PubMed: 30930176]
- Klaren WD and Rusyn I (2018). High-content assay multiplexing for muscle toxicity screening in human-induced pluripotent stem cell-derived skeletal myoblasts. *Assay Drug Dev Technol* 16, 333–342. doi:10.1089/adt.2018.860 [PubMed: 30070899]
- Kleinstreuer NC, Yang J, Berg EL et al. (2014). Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. *Nat Biotechnol* 32, 583–591. doi:10.1038/nbt.2914 [PubMed: 24837663]
- Kleinstreuer NC, Browne P, Chang X et al. (2018). Evaluation of androgen assay results using a curated Hershberger database. *Reprod Toxicol* 81, 272–280. doi:10.1016/j.reprotox.2018.08.017 [PubMed: 30205137]
- Knudsen T, Martin M, Chandler K et al. (2013). Predictive models and computational toxicology. *Methods Mol Biol* 947, 343–374. doi:10.1007/978-1-62703-131-8_26 [PubMed: 23138916]
- Kotsiantis SB (2007). Supervised machine learning: A review of classification techniques. *Proceedings of the 2007 conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*. IOS Press, 3–24.
- Kreutz C, Raue A, Kaschek D et al. (2013). Profile likelihood in systems biology. *FEBS J* 280, 2564–2571. doi:10.1111/febs.12276 [PubMed: 23581573]
- Lee JJ, Miller JA, Basu S et al. (2018). Building predictive in vitro pulmonary toxicity assays using high-throughput imaging and artificial intelligence. *Arch Toxicol* 92, 2055–2075. doi:10.1007/s00204-018-2213-0 [PubMed: 29705884]
- Li S and Xia M (2019). Review of high-content screening applications in toxicology. *Arch Toxicol* 93, 3387–3396. doi:10.1007/s00204-019-02593-5 [PubMed: 31664499]
- Locke PA and Myers DB Jr. (2011). A replacement-first approach to toxicity testing is necessary to successfully reauthorize TSCA. *ALTEX* 28, 266–272. doi:10.14573/altex.2011.4.266 [PubMed: 22130480]
- Low Y, Uehara T, Minowa Y et al. (2011). Predicting drug-induced hepatotoxicity using QSAR and toxicogenomics approaches. *Chem Res Toxicol* 24, 1251–1262. doi:10.1021/tx200148a [PubMed: 21699217]
- Low Y, Sedykh A, Fourches D et al. (2013). Integrative chemical-biological read-across approach for chemical hazard classification. *Chem Res Toxicol* 26, 1199–1208. doi:10.1021/tx400110f [PubMed: 23848138]

- Low YS, Sedykh AY, Rusyn I et al. (2014). Integrative approaches for predicting in vivo effects of chemicals from their structural descriptors and the results of short-term biological assays. *Curr Top Med Chem* 14, 1356–1364. doi:10.2174/1568026614666140506121116 [PubMed: 24805064]
- Lutter R, Barrow C, Borgert CJ et al. (2013). Data disclosure for chemical evaluations. *Environ Health Perspect* 121, 145–148. doi:10.1289/ehp.1204942 [PubMed: 23228957]
- Marvel SW, To K, Grimm FA et al. (2018). ToxPi graphical user interface 2.0: Dynamic exploration, visualization, and sharing of integrated data models. *BMC Bioinformatics* 19, 80. doi:10.1186/s12859-018-2089-2 [PubMed: 29506467]
- Marx U, Akabane T, Andersson TB et al. (2020). Biology-inspired microphysiological systems to advance patient benefit and animal welfare in drug development. *ALTEX* 37, 364–394. doi:10.14573/altex.2001241
- Miller GW (2016). Making data accessible: The Dryad experience. *Toxicol Sci* 149, 2–3. doi:10.1093/toxsci/kfv238 [PubMed: 26728592]
- National Academies of Sciences Engineering and Medicine (2017). *Using 21st Century Science to Improve Risk-Related Evaluations*. Washington, DC, USA: National Academies Press.
- Nguyen EH, Daly WT, Le NNT et al. (2017). Versatile synthetic alternatives to Matrigel for vascular toxicity screening and stem cell expansion. *Nat Biomed Eng* 1, 0096. doi:10.1038/s41551-017-0096 [PubMed: 29104816]
- Nielsen J (2017). Systems biology of metabolism. *Annu Rev Biochem* 86, 245–275. doi:10.1146/annurev-biochem-061516-044757 [PubMed: 28301739]
- NRC – National Research Council (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC, USA: National Academies Press.
- NTP – National Toxicology Program (2016). NTP Research Program on Chemicals Spilled into the Elk River in West Virginia: Final Update. N. T. Program https://ntp.niehs.nih.gov/ntp/research/areas/wvspill/wv_finalupdate_july2016_508.pdf
- Onel M, Beykal B, Ferguson K et al. (2019). Grouping of complex substances using analytical chemistry data: A framework for quantitative evaluation and visualization. *PLoS One* 14, e0223517. doi:10.1371/journal.pone.0223517 [PubMed: 31600275]
- Paul Friedman K, Gagne M, Loo LH et al. (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. doi:10.1093/toxsci/kfz201 [PubMed: 31532525]
- Pearce RG, Setzer RW, Strobe CL et al. (2017). htk: R package for high-throughput toxicokinetics. *J Stat Softw* 79, 1–26. doi:10.18637/jss.v079.i04 [PubMed: 30220889]
- Reif DM, Martin MT, Tan SW et al. (2010). Endocrine profiling and prioritization of environmental chemicals using ToxCast data. *Environ Health Perspect* 118, 1714–1720. doi:10.1289/ehp.1002180 [PubMed: 20826373]
- Rotroff DM, Wetmore BA, Dix DJ et al. (2010). Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. *Toxicol Sci* 117, 348–358. [PubMed: 20639261]
- Schmidt BZ, Lehmann M, Gutbier S et al. (2017). In vitro acute and developmental neurotoxicity screening: An overview of cellular platforms and high-throughput technical possibilities. *Arch Toxicol* 91, 1–33. doi:10.1007/s00204-016-1805-9 [PubMed: 27492622]
- Shapiro AJ, Antoni S, Guyton KZ et al. (2018). Software tools to facilitate systematic review used for cancer hazard identification. *Environ Health Perspect* 126, 104501. doi:10.1289/EHP4224 [PubMed: 30392397]
- Sipes NS, Wambaugh JF, Pearce R et al. (2017). An intuitive approach for predicting potential human health risk with the Tox21 10k library. *Environ Sci Technol* 51, 10786–10796. doi:10.1021/acs.est.7b00650 [PubMed: 28809115]
- Sirenko O, Cromwell EF, Crittenden C et al. (2013). Assessment of beating parameters in human induced pluripotent stem cells enables quantitative in vitro screening for cardiotoxicity. *Toxicol Appl Pharmacol* 273, 500–507. doi:10.1016/j.taap.2013.09.017 [PubMed: 24095675]
- Sirenko O, Hesley J, Rusyn I et al. (2014a). High-content high-throughput assays for characterizing the viability and morphology of human iPSC-derived neuronal cultures. *Assay Drug Dev Technol* 12, 536–547. doi:10.1089/adt.2014.592 [PubMed: 25506803]

- Sirenko O, Hesley J, Rusyn I et al. (2014b). High-content assays for hepatotoxicity using induced pluripotent stem cell-derived cells. *Assay Drug Dev Technol* 12, 43–54. doi:10.1089/adt.2013.520 [PubMed: 24229356]
- Sirenko O, Grimm FA, Ryan KR et al. (2017). In vitro cardiotoxicity assessment of environmental chemicals using an organotypic human induced pluripotent stem cell-derived model. *Toxicol Appl Pharmacol* 322, 60–74. doi:10.1016/j.taap.2017.02.020 [PubMed: 28259702]
- Soldatow VY, LeCluyse EL, Griffith LG et al. (2013). In vitro models for liver toxicity testing. *Toxicol Res* 2, 23–39. doi:10.1039/c2tx20051a
- Su R, Xiong S, Zink D et al. (2016). High-throughput imaging-based nephrotoxicity prediction for xenobiotics with diverse chemical structures. *Arch Toxicol* 90, 2793–2808. doi: 10.1007/s00204-015-1638-y [PubMed: 26612367]
- Taylor K, Stengel W, Casalegno C et al. (2014). Experiences of the REACH testing proposals system to reduce animal testing. *ALTEX* 31, 107–128. doi:10.14573/altex.1311151 [PubMed: 24609452]
- Taylor K (2018). Ten years of REACH – An animal protection perspective. *Altern Lab Anim* 46, 347–373. doi:10.1177/026119291804600610 [PubMed: 30657330]
- Thomas RS, Paules RS, Simeonov A et al. (2018). The US Federal Tox21 program: A strategic and operational plan for continued leadership. *ALTEX* 35, 163–168. doi:10.14573/altex.1803011 [PubMed: 29529324]
- US EPA – Environmental Protection Agency (2012). Benchmark Dose Technical Guidance. Risk Assessment Forum, US EPA. EPA/100/R-12/001. https://www.epa.gov/sites/production/files/2015-01/documents/benchmark_dose_guidance.pdf
- US EPA (2018). Strategic Plan to Promote the Development and Implementation of Alternative Test Methods within the TSCA Program. Office of Chemical Safety and Pollution Prevention. https://www.epa.gov/sites/production/files/2018-06/documents/epa_alt_strat_plan_6-20-18_clean_final.pdf
- US EPA (2019). Directive to Prioritize Efforts to Reduce Animal Testing. <https://www.epa.gov/sites/production/files/2019-09/documents/image2019-09-09-231249.pdf>
- van der Valk J, Bieback K, Buta C et al. (2018). Fetal bovine serum (FBS): Past – Present – Future. *ALTEX* 35, 99–118. doi:10.14573/altex.1705101 [PubMed: 28800376]
- Wambaugh JF, Wetmore BA, Pearce R et al. (2015). Toxicokinetic triage for environmental chemicals. *Toxicol Sci* 147, 55–67. doi:10.1093/toxsci/kfv118 [PubMed: 26085347]
- Wetmore BA (2015). Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment. *Toxicology* 332, 94–101. doi:10.1016/j.tox.2014.05.012 [PubMed: 24907440]
- Wignall JA, Shapiro AJ, Wright FA et al. (2014). Standardizing benchmark dose calculations to improve science-based decisions in human health assessments. *Environ Health Perspect* 122, 499–505. doi:10.1289/ehp.1307539 [PubMed: 24569956]
- Williams AJ, Grulke CM, Edwards J et al. (2017). The CompTox chemistry dashboard: A community data resource for environmental chemistry. *J Cheminform* 9, 61. doi:10.1186/s13321-017-0247-6 [PubMed: 29185060]
- Zeise L, Bois FY, Chiu WA et al. (2013). Addressing human variability in next-generation human health risk assessments of environmental chemicals. *Environ Health Perspect* 121, 23–31. doi:10.1289/ehp.1205687 [PubMed: 23086705]
- Zhu H, Bouhifd M, Donley E et al. (2016). Supporting read-across using biological data. *ALTEX* 33, 167–182. doi:10.14573/altex.1601252 [PubMed: 26863516]

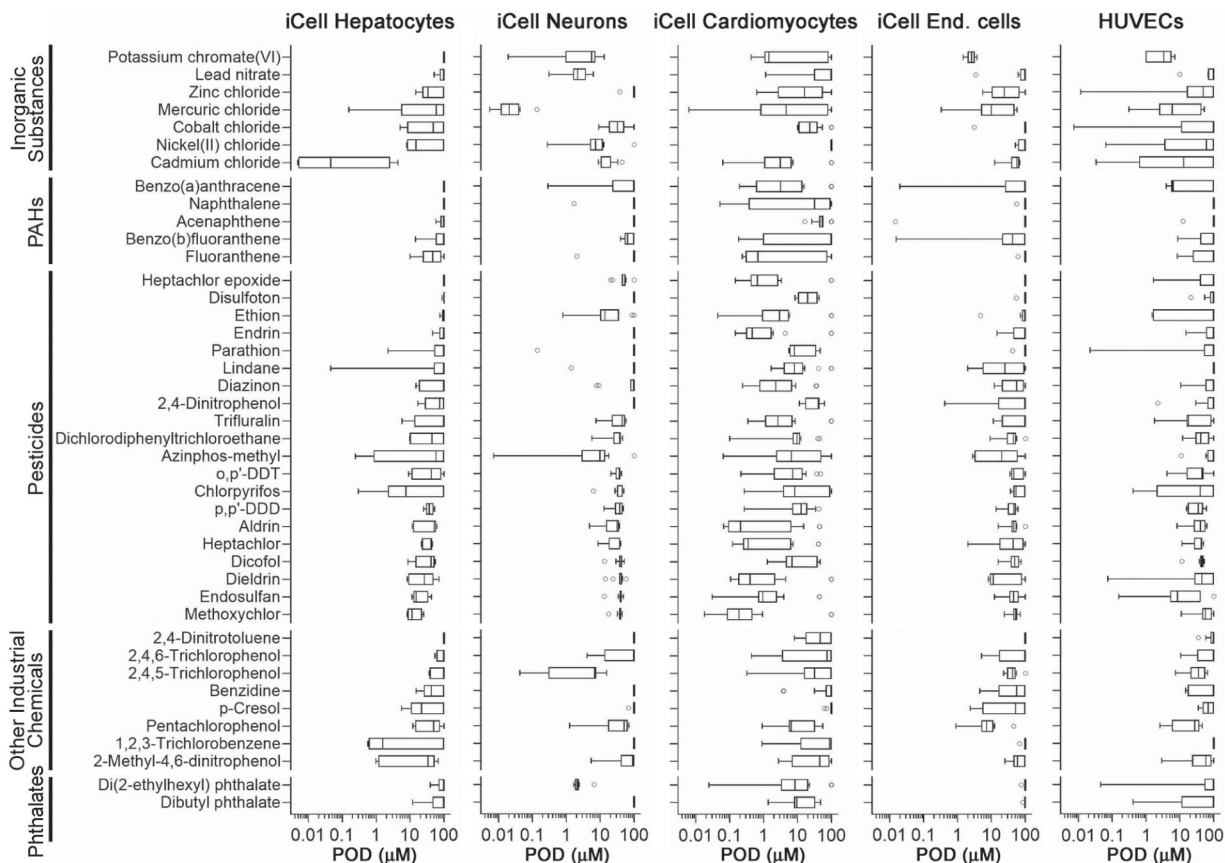


Fig. 1: Quantitative analysis of chemical-specific effects in five cell types

Box (inter-quartile range and median) and whiskers (min to max) plots show the range of PODs (one standard deviation of vehicle-treated wells) across 48 phenotypes in five cell types (Tab. 2) for each of the 42 Superfund priority list chemicals (Tab. 1). Chemicals were grouped into classes (Tab. 1) and then sorted within a class based on the mean POD values of the phenotypes in iCell hepatocytes.

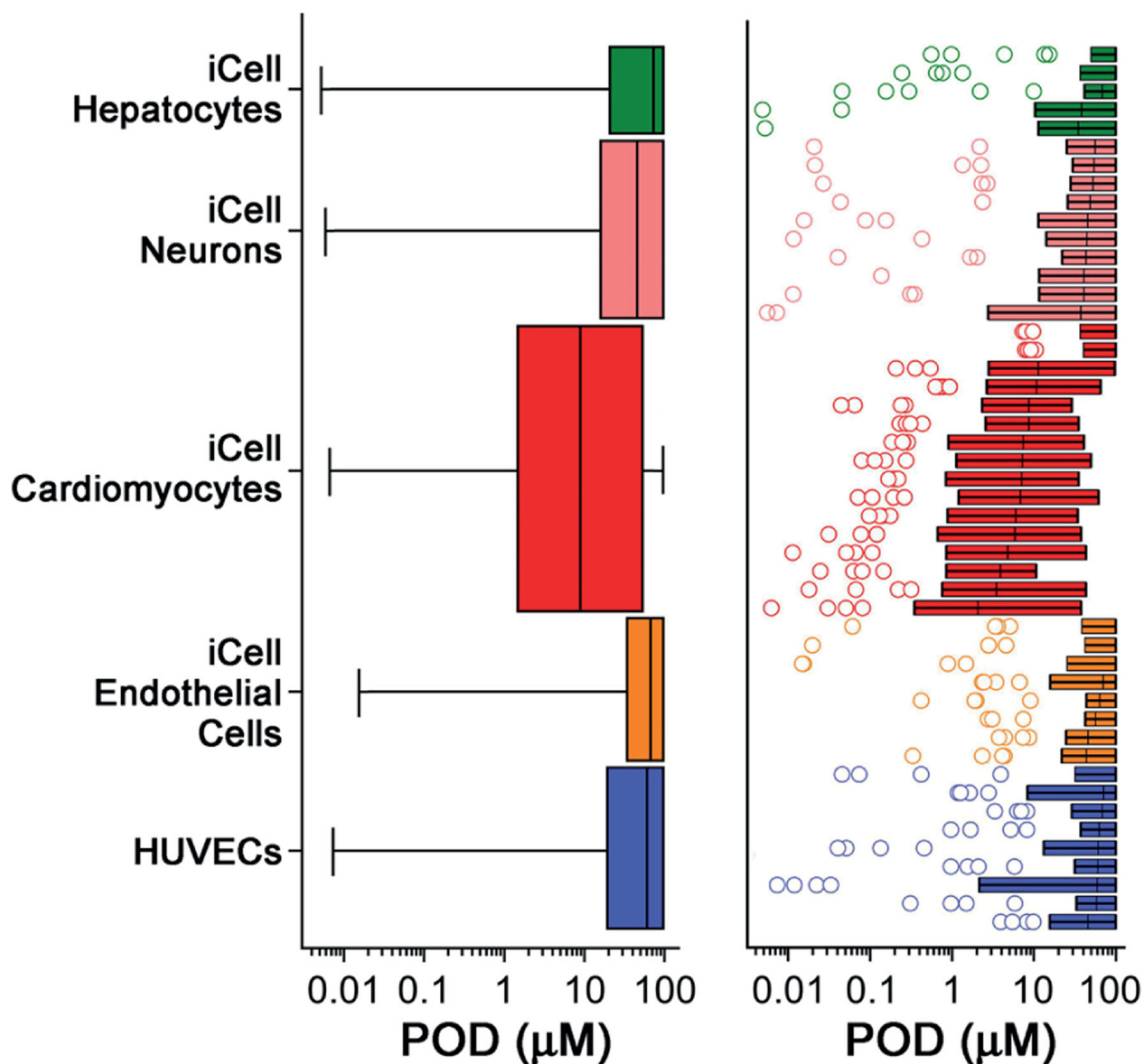


Fig. 2: Quantitative analysis of cell-specific effects of the 42 Superfund priority list chemicals
 The left panel shows box (inter-quartile range and median) and whiskers (min to max) plots of PODs (one standard deviation of vehicle-treated wells) for all 42 tested chemicals (Tab. 1) in each cell type. The size of each box and whiskers plot is proportional to the number of phenotypes evaluated in each cell type (Tab. 2). The right panel shows box (inter-quartile range and median) and whiskers (Tukey) plots of PODs (one standard deviation of vehicle-treated wells) for all 42 tested chemicals (Tab. 1) in each phenotype. Phenotypes are grouped based on the cell type (Tab. 2). Outlier chemicals are shown as circles.

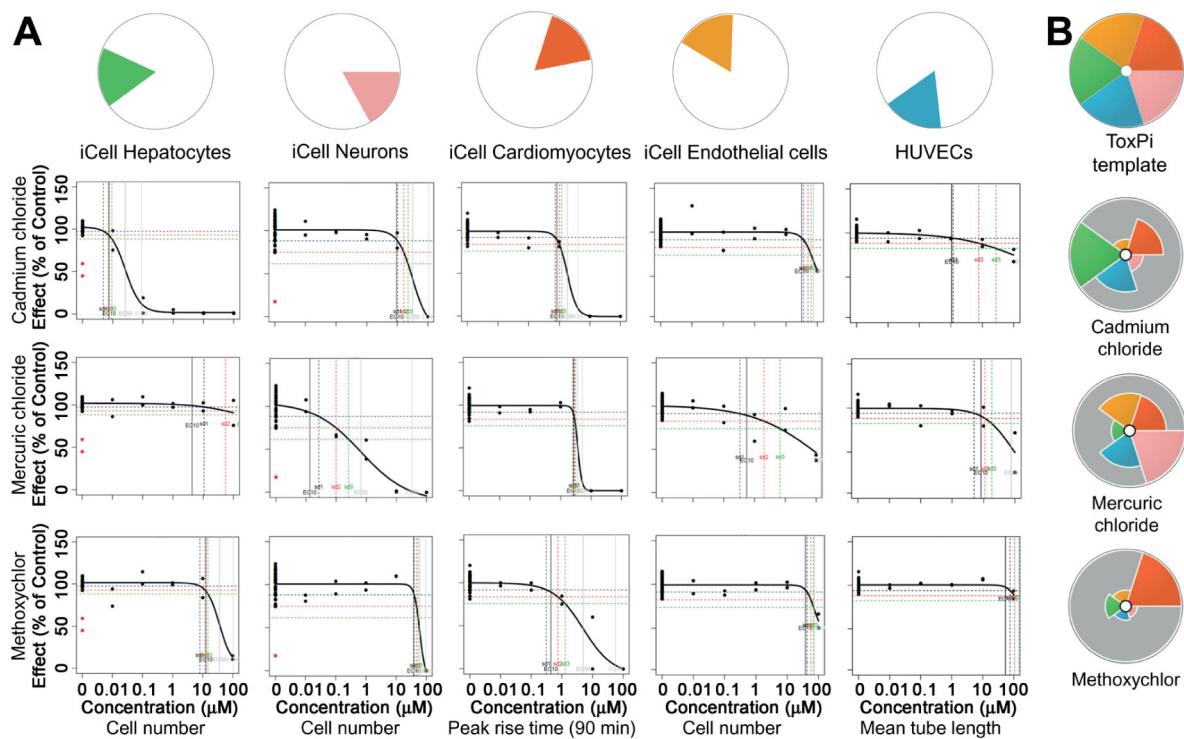


Fig. 3: Data integration from concentration-response modeling for each chemical and phenotype using Toxicological Prioritization Index (ToxPi) approach

(A) Representative examples of concentration-response fits (lines) to the data (dots) are shown for three chemicals (rows) and five cell-specific phenotypes (columns). Pie chart slices are colored to distinguish effects in each cell type. (B) Examples of ToxPi images for three selected chemicals.

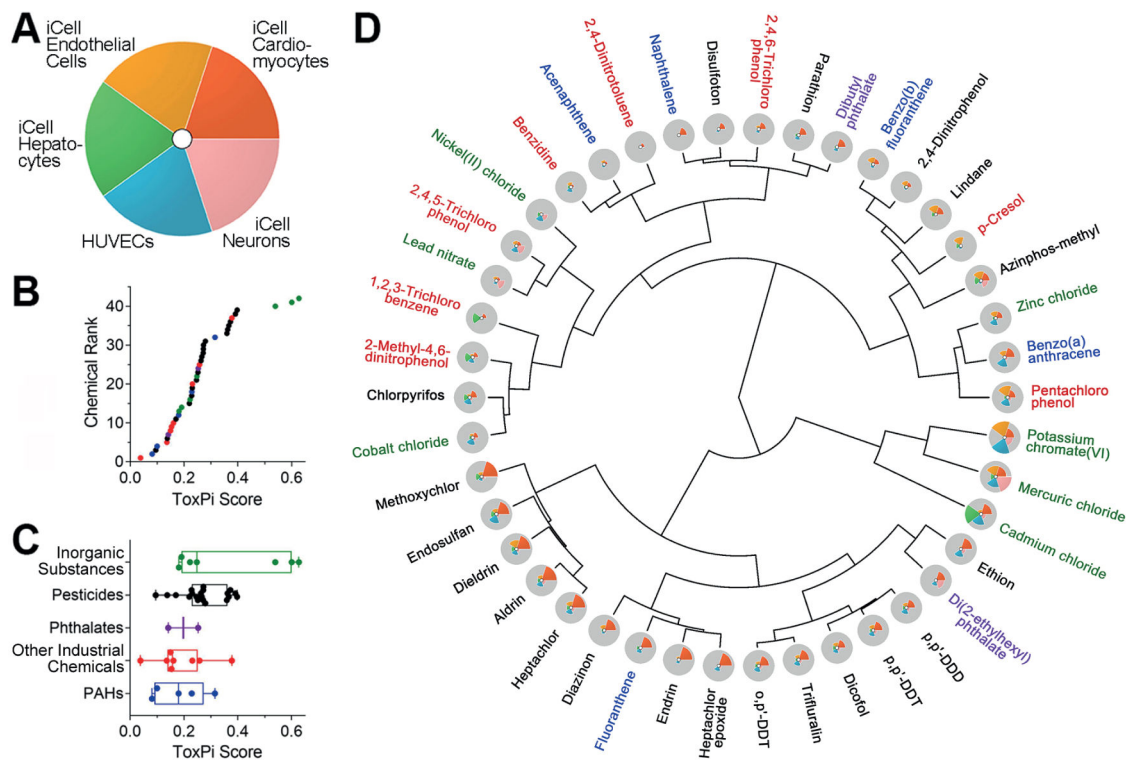


Fig. 4: ToxPi analysis-based ranking and clustering of 42 Superfund priority list chemicals based on the effects in five cell types

(A) Legend to the ToxPi visualization of the effects on five cell types. (B) Ranking of the tested chemicals based on the overall ToxPi scores. Chemicals are colored based on chemical class. Table S5¹ contains the data from the ToxPi analysis. (C) Box (inter-quartile range and median) and whiskers (min to max) plots show the range of ToxPi scores for each chemical (dots) for each class. Chemical classes (Tab. 1) were ranked based on the median value. (D) Clustering (Ward's D method) of 42 Superfund priority list chemicals using ToxPi scores. Chemical names are colored based on chemical class as in panel C.

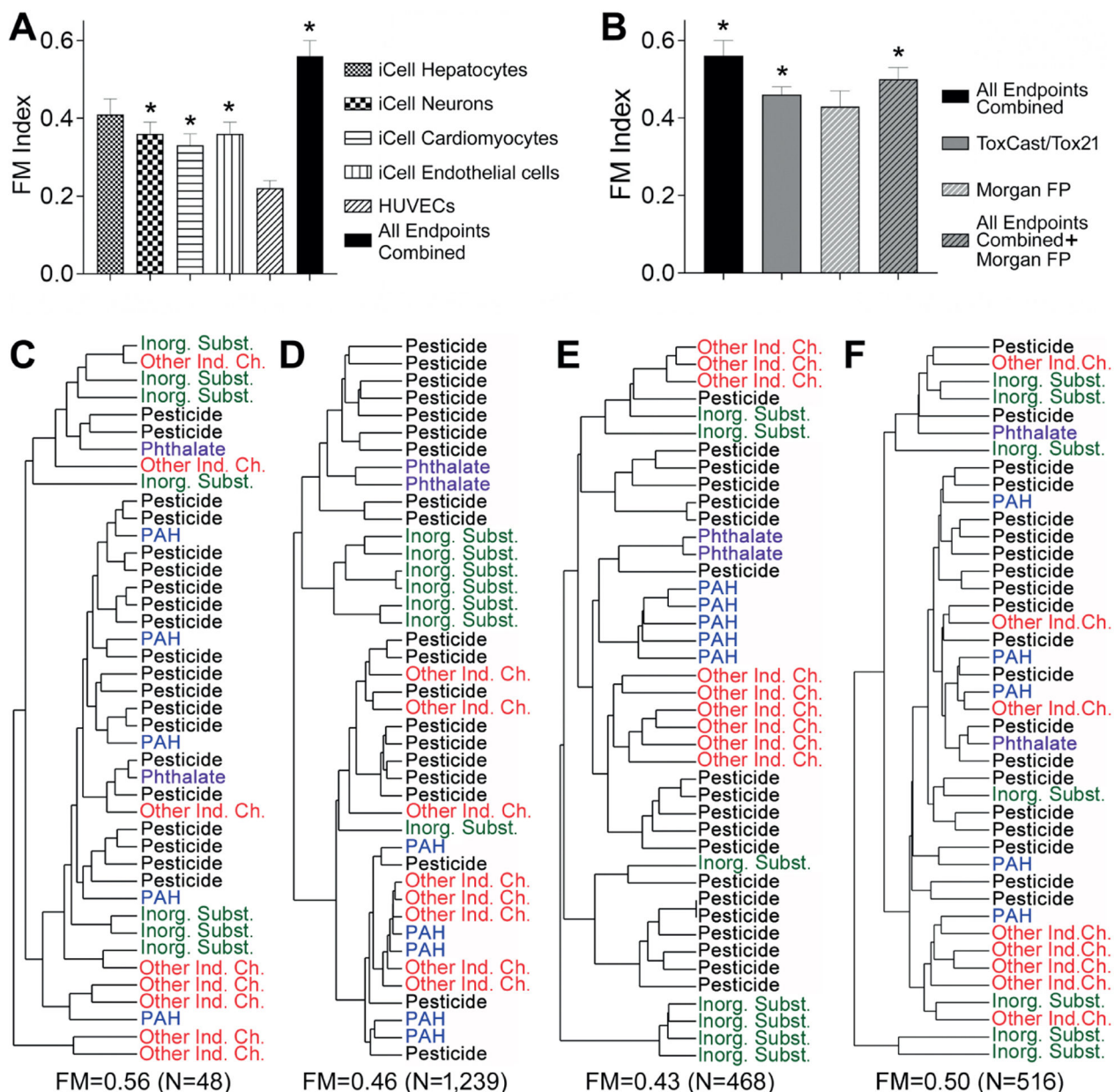


Fig. 5: Quantitative analysis of the grouping of 42 Superfund priority list chemicals with various data streams

(A) Fowlkes-Mallows (FM) index for clustering of chemicals into five classes (Tab. 1) using *in vitro* data from each cell type, or all data combined. (B) FM index for clustering of chemicals using data in this study (black bar), or other publicly available *in vitro* or chemical descriptors (e.g., Morgan fingerprints [FP]), or a combination thereof. Asterisks (*) indicate that one-sided p-values were < 0.05 for the observed FM index value compared to the null expectation. (C-F) Clustering dendrograms (average Pearson correlation method) for each data stream shown in (B). FM index and the number of variables included in each comparison are shown below each plot. (C) *In vitro* data from this study, all endpoints combined. (D) ToxCast/Tox21 data (as of November 2019). (E) Morgan fingerprints. (F)

Morgan fingerprints combined with *in vitro* data from this study. Identity of each chemical in each clustering diagram is listed in Table S6¹.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

All Endpoints Combined (60%)	Pesticides	Other Ind. Ch.	PAHs	Phthalates	Inorg. Subst.
Pesticides	19	3	4	1	3
Other Ind. Ch.		3	1		1
PAHs		1			
Phthalates					
Inorganic Subst.	1	1		1	3

ToxCast/Tox21 (69%)	Pesticides	Other Ind. Ch.	PAHs	Phthalates	Inorg. Subst.
Pesticides	18	3	1	2	1
Other Ind. Ch.	2	5	4		
PAHs					
Phthalates					
Inorganic Subst.					6

Morgan FP (81%)	Pesticides	Other Ind. Ch.	PAHs	Phthalates	Inorg. Subst.
Pesticides	18	3		1	
Other Ind. Ch.	1	3			
PAHs			5		
Phthalates				1	
Inorganic Subst.	1	2			7

All Endpoints Combined+ Morgan FP (66%)	Pesticides	Other Ind. Ch.	PAHs	Phthalates	Inorg. Subst.
Pesticides	19	4	2	1	3
Other Ind. Ch.	1	3	1		
PAHs			2		
Phthalates					
Inorg. Subst.		1		1	4

Fig. 6: Confusion matrices for chemical classification into five classes using *in vitro* and/or chemical descriptors

Known (columns) chemical assignment into each of five classes (Tab. 1) is compared to predicted (rows) class assignment using random forest algorithm with 5-fold cross validation as detailed in Section 2. Classification outcomes for the analyses using data from all phenotypes in this study (top left), ToxCast/Tox21 data (top right), Morgan fingerprints [FP] (bottom left), or data from this study and Morgan FP combined (bottom right) are shown. Accuracy of classification for each dataset is shown in the top left corner of each matrix. Numbers in the cells filled with green (on diagonal) and light pink (off diagonal) indicate the number of chemicals that were classified correctly or misclassified, respectively.

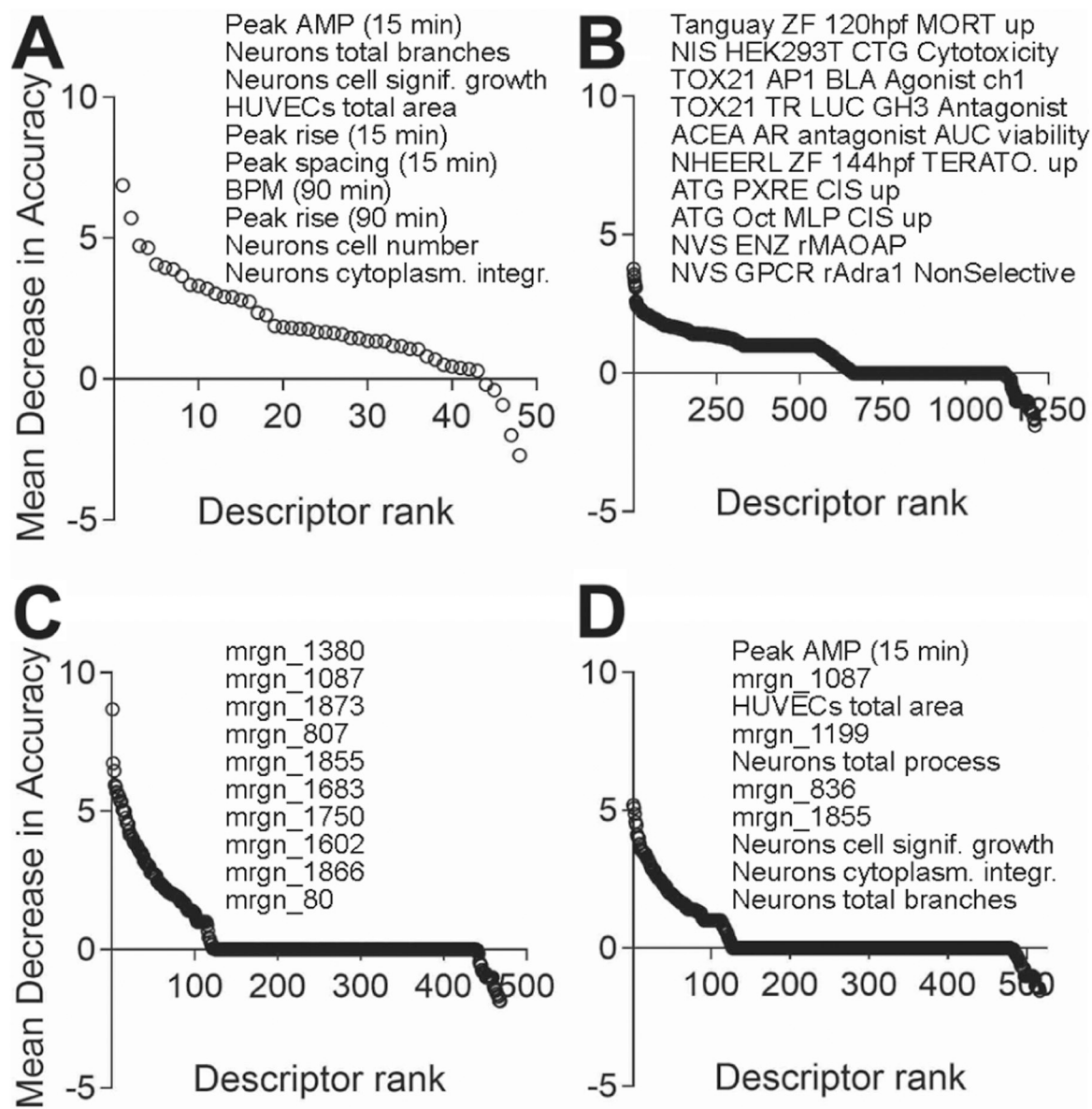


Fig. 7: Classification accuracy-contributing phenotypes

Importance of the *in vitro* or chemical structure descriptors contributing to the classification accuracy from different data streams (Fig. 6) was analyzed as detailed in Section 2. Top 10 features are listed. (A) *In vitro* data from this study. (B) ToxCast/Tox21 data. (C) Morgan fingerprints. (D) Morgan fingerprint combined with *in vitro* data from this study.

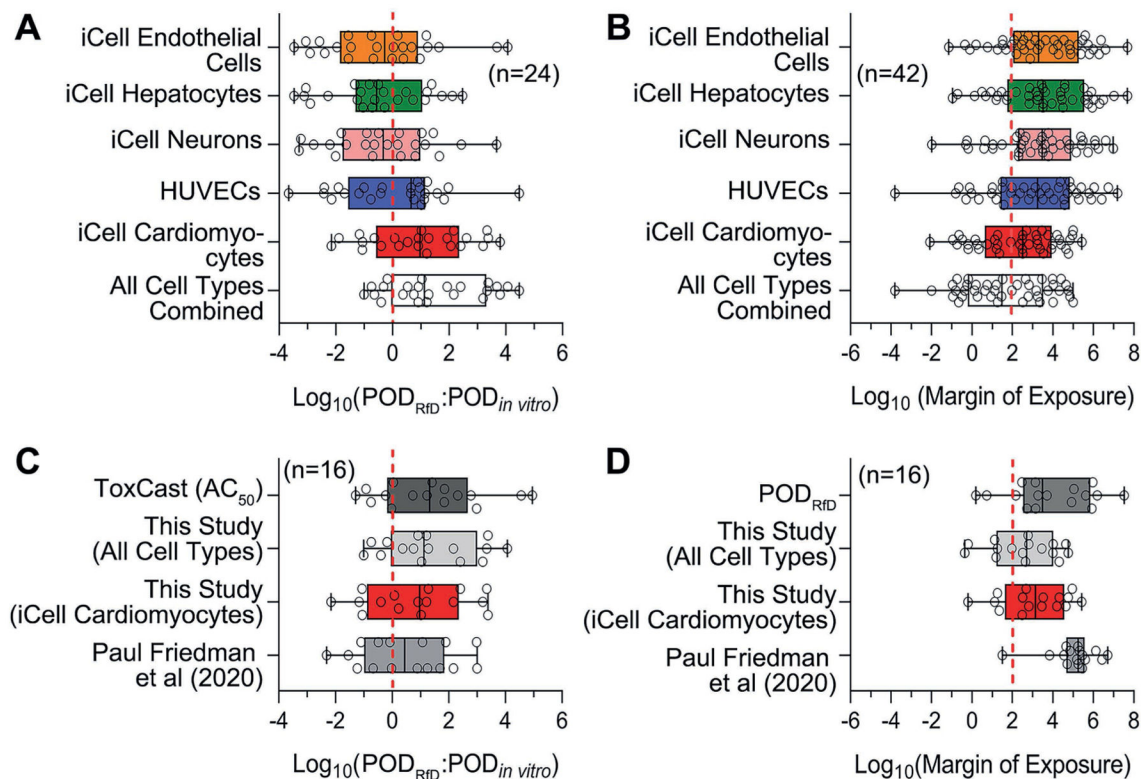


Fig. 8: POD data comparison across different *in vitro* and *in vivo* datasets and margin of exposure estimates

Minimum of *in vitro* PODs generated from each cell type and all cell types combined in this study were compared to *in vivo* POD derived from Reference dose (A). Margins of exposure were calculated based on *in vitro* PODs from this study and the estimated exposure levels (B). The ratio between *in vivo* and *in vitro* (C), and the margins of exposure (D) were further compared across different datasets. All of the ratio outputs were log transformed for comparison; *n* represents the number of chemicals from 42 Superfund priority list chemicals covered by different datasets for comparison and detailed in Table S3¹.

Tab. 1:

Superfund priority chemicals used in this study

ATSDR Chemical class	Chemical name	CAS number	Chemical formula	ATSDR rank (2017)
Inorganic substances	Lead nitrate	10099-74-8	PbCl ₂	2
	Mercuric chloride	7487-94-7	HgCl ₂	3
	Cadmium chloride	10108-64-2	CdCl ₂	7
	Potassium chromate (VI)	7789-00-6	K ₂ CrO ₄	17
	Cobalt chloride	7646-79-9	CoCl ₂	51
	Nickel chloride	7718-54-9	NiCl ₂	57
	Zinc chloride	7646-85-7	ZnCl ₂	75
Polycyclic aromatic hydrocarbons (PAHs)	Benzo(b)fluoranthene	205-99-2	C ₂₀ H ₁₂	10
	Benzo(a)anthracene	56-55-3	C ₁₈ H ₁₂	38
	Naphthalene	91-20-3	C ₁₀ H ₈	81
	Fluoranthene	206-44-0	C ₁₆ H ₁₀	138
	Acenaphthene	83-32-9	C ₁₂ H ₁₀	171
Pesticides	<i>p,p'</i> -DDT	50-29-3	C ₁₄ H ₉ Cl ₅	13
	Dieldrin	60-57-1	C ₁₂ H ₈ Cl ₆ O	18
	Aldrin	309-00-2	C ₁₂ H ₈ Cl ₆	25
	<i>p,p'</i> -DDD	72-54-8	C ₁₄ H ₁₀ Cl ₄	26
	Heptachlor	76-44-8	C ₁₀ H ₅ Cl ₇	28
	Lindane	58-89-9	C ₆ H ₆ Cl ₆	34
	Disulfoton	298-04-4	C ₈ H ₁₉ O ₂ PS ₃	37
	Endrin	72-20-8	C ₁₂ H ₈ Cl ₆ O	40
	Diazinon	333-41-5	C ₁₂ H ₂₁ N ₂ O ₃ PS	41
	Endosulfan	115-29-7	C ₉ H ₆ Cl ₆ O ₃ S	44
	Heptachlor epoxide	1024-57-3	C ₁₀ H ₅ Cl ₇ O	47
	<i>o,p'</i> -DDT	789-02-6	C ₁₄ H ₉ Cl ₅	53
	Methoxychlor	72-43-5	C ₁₆ H ₁₅ Cl ₃ O ₂	55

ATSDR Chemical class	Chemical name	CAS number	Chemical formula	ATSDR rank (2017)
	Chlorpyrifos	2921-88-2	C ₉ H ₁₁ Cl ₃ NO ₃ PS	64
	2,4-dinitrophenol	51-28-5	C ₆ H ₄ N ₂ O ₅	89
	Ethion	563-12-2	C ₉ H ₂₂ O ₄ P ₂ S ₄	99
	Azinphos-methyl	86-50-0	C ₁₀ H ₁₂ N ₃ O ₃ PS ₂	131
	Dicofol	115-32-2	C ₁₄ H ₉ Cl ₅ O	145
	Parathion	56-38-2	C ₁₀ H ₁₄ NO ₃ PS	148
	Trifluralin	1582-09-8	C ₁₃ H ₁₆ F ₃ N ₃ O ₄	157
	Benzidine	92-87-5	C ₁₂ H ₁₂ N ₂	30
	Pentachlorophenol	87-86-5	C ₆ Cl ₅ OH	54
	2,4,6-trichlorophenol	88-06-2	C ₆ H ₂ Cl ₃ OH	85
	2,4-dinitrotoluene	121-14-2	C ₇ H ₆ N ₂ O ₄	98
	2-Methyl-4,6-dinitrophenol	534-52-1	C ₇ H ₆ N ₂ O ₅	100
	1,2,3-Trichlorobenzene	87-61-6	C ₆ H ₃ Cl ₃	137
	2,4,5-Trichlorophenol	95-95-4	C ₆ H ₂ Cl ₃ OH	142
	<i>p</i> -Cresol	106-44-5	C ₇ H ₈ O	175
	Dibutyl phthalate	84-74-2	C ₁₆ H ₂₂ O ₄	58
	Di(2-ethylhexyl) phthalate	117-81-7	C ₂₄ H ₃₈ O ₄	77
Other industrial chemicals				
Phthalates				

Tab. 2:

In vitro toxicity phenotypes evaluated in this study

Cell type ^a	iCell hepatocytes	iCell neurons	iCell cardiomyocytes ^b	iCell endothelial cells ^c	HUVEC ^c
Catalog #	C1023	C1008	CMC-100-010-001	C1114	CC-2519A
Time point	24 h	72 h	15 or 90 min	18 or 24 h	18 or 24 h
Functional phenotypes	<ul style="list-style-type: none"> - Mitochondrial integrity - Mitochondrial intensity 	<ul style="list-style-type: none"> - Total outgrowth - Mean outgrowth - Total process - Total branches - Cells with significant growth 	<ul style="list-style-type: none"> - Beats per minute - Peak amplitude - Peak spacing - Peak width - Peak rise time - Peak decay time - Decay to rise ratio 	<ul style="list-style-type: none"> - Total tube length - Mean tube length - Total tube area 	<ul style="list-style-type: none"> - Total tube length - Mean tube length - Total tube area
Cytotoxicity phenotypes	<ul style="list-style-type: none"> - Cell number - Nuclei intensity - All cell mean area 	<ul style="list-style-type: none"> - Cell number - Mitochondrial integrity - Cytoplasmic integrity - Total cells body area - ATP^d 	<ul style="list-style-type: none"> - Cell number - Mitochondrial integrity 	<ul style="list-style-type: none"> - Cell number - Mitochondrial integrity - Mitochondrial intensity - Cytoplasmic integrity - Nuclei mean area 	<ul style="list-style-type: none"> - Cell number - Mitochondrial integrity - Mitochondrial intensity - Cytoplasmic integrity - Nuclei mean area - ATP^d

See Table S41 for detailed description of each phenotype.

^a iCell lines were purchased from FujiFilm Cellular Dynamics, HUVEC cell line was purchased from Lonza.

^b Cytotoxicity phenotypes were measured in iCell cardiomyocytes at 90 min.

^c Cytotoxicity phenotypes were measured in iCell endothelial cells and HUVEC at 24 h.

^d CellTiter-Glo[®] assay.

Tab. 3:

Ranges in ToxPi scores for each chemical class and cell type

Cell type	PAHs	Pesticides	Inorganic substances	Other industrial chemicals	Phthalates
iCell hepatocytos	0-0.14	0-0.32	0-0.88	0-0.45	0.026-0.03
iCell neurons	0-0.11	0-0.37	0.01-1	0-0.47	0-0.46
iCell cardiomyocytes	0.10-0.55	0.18-0.78	0-0.50	0.01-0.34	0.37-0.42
iCell endothelial cells	0.02-0.27	0-0.38	0.04-0.72	0-0.49	0.005-0.009
HUVECs	0-0.41	0-0.38	0.10-0.75	0-0.36	0.13-0.18
Overall (combination of all phenotypes)	0.08-0.32	0.10-0.39	0.18-0.63	0.04-0.38	0.14-0.25