Research Article

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

Zunwei Chen1,2, Yizhong Liu1,2, Fred A. Wright3,4, Weihsueh A. Chiu1,2 and Ivan Rusyn1,2

1 Interdisciplinary Faculty of Toxicology and 2 Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX, USA; 3 Bioinformatics Research Center and 4 Departments 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.

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
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 EGM™-2 BulletKits™ (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 S1 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. S1)). 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).

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/chemicals 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 com-

---

1 doi:10.14573/altex.20022916
2 http://www.atsdr.cdc.gov/spl
### Tab. 1: Superfund priority chemicals used in this study

<table>
<thead>
<tr>
<th>ATSDR Chemical class</th>
<th>Chemical name</th>
<th>CAS number</th>
<th>Chemical formula</th>
<th>ATSDR rank (2017)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inorganic substances</td>
<td>Lead nitrate</td>
<td>10099-74-8</td>
<td>PbCl₂</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Mercuric chloride</td>
<td>7487-94-7</td>
<td>HgCl₂</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Cadmium chloride</td>
<td>10108-64-2</td>
<td>CdCl₂</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Potassium chromate (VI)</td>
<td>7789-00-6</td>
<td>K₂CrO₄</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Cobalt chloride</td>
<td>7646-79-9</td>
<td>CoCl₂</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Nickel chloride</td>
<td>7718-54-9</td>
<td>NiCl₂</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Zinc chloride</td>
<td>7646-85-7</td>
<td>ZnCl₂</td>
<td>75</td>
</tr>
<tr>
<td>Polycyclic aromatic hydrocarbons (PAHs)</td>
<td>Benzo(b)fluoranthene</td>
<td>205-99-2</td>
<td>C₂₀H₁₂</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Benzo(a)anthracene</td>
<td>56-55-3</td>
<td>C₁₈H₁₂</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Naphthalene</td>
<td>91-20-3</td>
<td>C₁₀H₈</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>Fluoranthene</td>
<td>206-44-0</td>
<td>C₁₈H₁₀</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td>Acenaphthen</td>
<td>83-32-9</td>
<td>C₁₂H₁₀</td>
<td>171</td>
</tr>
<tr>
<td>Pesticides</td>
<td>( p,p' )-DDT</td>
<td>50-29-3</td>
<td>C₁₄H₆Cl₅</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Dieldrin</td>
<td>60-57-1</td>
<td>C₁₂H₆Cl₅O</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Aldrin</td>
<td>309-00-2</td>
<td>C₁₂H₆Cl₆</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>( p,p' )-DDD</td>
<td>72-54-8</td>
<td>C₁₄H₁₀Cl₄</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Heptachlor</td>
<td>76-44-8</td>
<td>C₁₀H₅Cl₇</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Lindane</td>
<td>58-89-9</td>
<td>C₆H₆Cl₆</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Disulfoton</td>
<td>298-04-4</td>
<td>C₈H₉O₂PS₂</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Endrin</td>
<td>72-20-8</td>
<td>C₁₂H₆Cl₅O</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Diazinon</td>
<td>333-41-5</td>
<td>C₁₂H₁₁N₂O₃PS</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>Endosulfan</td>
<td>115-29-7</td>
<td>C₆H₆Cl₆O₂S</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Heptachlor epoxide</td>
<td>1024-57-3</td>
<td>C₁₀H₅Cl₇</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>( o,p' )-DDT</td>
<td>789-02-6</td>
<td>C₁₄H₆Cl₅</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Methoxychlor</td>
<td>72-43-5</td>
<td>C₁₈H₁₅Cl₅O₂</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Chlorpyrifos</td>
<td>2921-88-2</td>
<td>C₉H₁₁Cl₆NO₃PS</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>2,4-dinitrophenol</td>
<td>51-28-5</td>
<td>C₆H₄N₂O₅</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>Ethion</td>
<td>563-12-2</td>
<td>C₆H₂₂O₄P₂S₄</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>Azinphos-methyl</td>
<td>86-50-0</td>
<td>C₁₀H₁₂N₃O₃PS₂</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td>Dicofol</td>
<td>115-32-2</td>
<td>C₁₄H₆Cl₅O</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td>Parathion</td>
<td>56-38-2</td>
<td>C₁₀H₄NO₅PS</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>Trifluralin</td>
<td>1582-09-8</td>
<td>C₁₃H₁₆F₃N₂O₄</td>
<td>157</td>
</tr>
<tr>
<td>Other industrial chemicals</td>
<td>Benidine</td>
<td>92-87-5</td>
<td>C₁₂H₁₂N₂</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Pentachlorophenol</td>
<td>87-86-5</td>
<td>C₆Cl₃OH</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>2,4,6-trichlorophenol</td>
<td>88-06-2</td>
<td>C₆H₃Cl₃OH</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>2,4-dinitrotoluene</td>
<td>121-14-2</td>
<td>C₆H₆N₂O₄</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>2-Methyl-4,6-dinitrophenol</td>
<td>534-52-1</td>
<td>C₆H₆N₂O₆</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>1,2,3-Trichlorobenzene</td>
<td>87-61-6</td>
<td>C₆H₃Cl₃</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>2,4,5-Trichlorophenol</td>
<td>95-95-4</td>
<td>C₆H₃Cl₅OH</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>( p )-Cresol</td>
<td>106-44-5</td>
<td>C₆H₄O</td>
<td>175</td>
</tr>
<tr>
<td>Phthalates</td>
<td>Dibutyl phthalate</td>
<td>84-74-2</td>
<td>C₁₀H₂₂O₄</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Di(2-ethylhexyl) phthalate</td>
<td>117-81-7</td>
<td>C₂₄H₃₈O₄</td>
<td>77</td>
</tr>
</tbody>
</table>
pounds, 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, para-thion, 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 S1. Cells were exposed to test chemicals in descending logarithmic order of concentrations (100, 10, 1, 0.1, and 0.01 μM). 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× 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).

**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, West Chester, PA).

**Tab. 2: In vitro toxicity phenotypes evaluated in this study**

See Table S4 for detailed description of each phenotype.

<table>
<thead>
<tr>
<th>Cell type&lt;sup&gt;a&lt;/sup&gt;</th>
<th>iCell hepatocytes</th>
<th>iCell neurons</th>
<th>iCell cardiomyocytes&lt;sup&gt;b&lt;/sup&gt;</th>
<th>iCell endothelial cells&lt;sup&gt;c&lt;/sup&gt;</th>
<th>HUVEC&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalog #</td>
<td>C1023</td>
<td>C1008</td>
<td>CMC-100-010-001</td>
<td>C1114</td>
<td>CC-2519A</td>
</tr>
<tr>
<td>Time point</td>
<td>24 h</td>
<td>72 h</td>
<td>15 or 90 min</td>
<td>18 or 24 h</td>
<td>18 or 24 h</td>
</tr>
<tr>
<td><strong>Functional phenotypes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Mitochondrial integrity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Mitochondrial intensity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Total outgrowth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Mean outgrowth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Total process</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Total branches</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Cells with significant growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Beats per minute</td>
<td>- Mean tube length</td>
<td>- Total tube length</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Peak amplitude</td>
<td>- Mean tube length</td>
<td>- Mean tube length</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Peak spacing</td>
<td>- Total tube area</td>
<td>- Total tube area</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Peak width</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Peak rise time</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Peak decay time</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Decay to rise ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cytotoxicity phenotypes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Cell number</td>
<td>- Cell number</td>
<td>- Cell number</td>
<td>- Cell number</td>
<td>- Cell number</td>
</tr>
<tr>
<td></td>
<td>- Nuclei intensity</td>
<td>- Mitochondrial integrity</td>
<td>- Mitochondrial integrity</td>
<td>- Mitochondrial integrity</td>
<td>- Mitochondrial integrity</td>
</tr>
<tr>
<td></td>
<td>- All cell mean area</td>
<td>- Cytoplasmic integrity</td>
<td>- Cytoplasmic integrity</td>
<td>- Cytoplasmic integrity</td>
<td>- Cytoplasmic integrity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Total cells body area</td>
<td>- Nuclei mean area</td>
<td>- Nuclei mean area</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- ATP&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td>- ATP&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> iCell lines were purchased from FujiFilm Cellular Dynamics, HUVEC cell line was purchased from Lonza.<br><sup>b</sup> Cytotoxicity phenotypes were measured in iCell cardiomyocytes at 90 min.<br><sup>c</sup> Cytotoxicity phenotypes were measured in iCell endothelial cells and HUVEC at 24 h.<br><sup>d</sup> CellTiter-Glo® assay.
es) using the DAPI (Hoechst 33422), FITC (Calcein AM Green), and TRITC (MitoTracker Orange) filters at 10× or 20× magnification. Acquired images were processed using the Multi-Wavelength 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 Early-Tox™ 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 concentration, 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 Early-Tox™ 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.

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 PODRfD values to various NAM-based PODs, including the 

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 predefined 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.

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.
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 S2) 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 of the 48 phenotypes (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.

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

Box-plot (inter-quartile range and median) and whiskers (min to max) plots show the range of PODs for 42 chemicals across 48 phenotypes in five cell types (Tab. 1). Each cell type represented one of the four tissue types (Tab. 1). Chemicals were grouped into classes (Tab. 1) and then sorted within a class based on the mean POD values of the phenotypes.
one cell type and no correlation in PODs was evident among cell types (Fig. S2\(^1\)), 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 \(\mu\)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. S3\(^1\)).

### 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. S4\(^1\)).

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, cadmium chloride, and potassium chromate) had the highest overall bioactivity score.
each chemical on each cell type were further identified by clustering chemicals using data on each cell type (Fig. S5). 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.

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. S5). 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.
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-
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 bi-
ological features that are best able to discriminate among classes,
as described below.

The supervised classification analysis, where every test com-
pound was predicted using a classification model, can also be ex-
amined 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
descriptors 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 mod-
els, i.e., from zebrafish, to cytotoxicity, to reporter assays (Fig.
7B). While Morgan fingerprints are difficult to interpret direct-

---

**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.
trast, as shown in Figure 8C, only the approach of using the minimum (most sensitive) ToxCast AC50 has similarly conservative results, whereas cardiomyocytes alone and the PODNAM 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 \textit{in vivo} PODs from ToxRefDB, whereas we used the \textit{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 PODRfD-based

3.5 Comparison to \textit{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 \textit{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 (PODRfD). For our \textit{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 \textit{in vitro} PODs represent a conservative surrogate for the PODRfD, with only 25% of our \textit{in vitro} PODs being higher than the corresponding PODRfD, and those remaining 25% being within 10-fold of the \textit{in vivo} value. In contrast, as shown in Figure 8C, only the approach of using the minimum (most sensitive) ToxCast AC50 has similarly conservative results, whereas cardiomyocytes alone and the PODNAM 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 \textit{in vivo} PODs from ToxRefDB, whereas we used the \textit{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 PODRfD-based

Fig. 6: Confusion matrices for chemical classification into five classes using \textit{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 \textit{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) \textit{In vitro} data from this study. (B) ToxCast/Tox21 data. (C) Morgan fingerprints. (D) Morgan fingerprint combined with \textit{in vitro} data from this study.
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 "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\textsubscript{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 that 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. Regrettably, 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.
References
Kavlock, R. J., Bahadori, T., Barton-Maclaren, T. S. et al. (2018).


Nguyen, E. H., Daly, W. T., Le, N. N. T. 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


Reif, D. M., Martin, M. T., Tan, S. W. et al. (2010). Endocrine profiling and prioritization of environmental chemicals using ToxCast data. Environ Health Perspect 118, 1714-1720. doi:10.1289/ehp.1002180


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


Conflict of interest
All authors declare they have no competing interests.

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.