Published in final edited form as:
Toxicol Sci. 2022 April 26; 187(1): 62–79. doi:10.1093/toxsci/kfac018.

Integrating data from in vitro New Approach Methodologies for Developmental Neurotoxicity

KE Carstens[1,2], **AF Carpenter**[1,2], **MM Martin**[1], **JA Harrill**[1], **TJ Shafer**[1], **K Paul Friedman**[1]

[1]Center for Computational Toxicology and Exposure, ORD, US EPA, RTP, NC 27711

[2]Oak Ridge Associated Universities, Oak Ridge, TN 37830

Abstract

In vivo developmental neurotoxicity (DNT) testing is resource intensive and lacks information on cellular processes affected by chemicals. To address this, DNT new approach methodologies (NAMs) are being evaluated, including: the microelectrode array neuronal network formation assay (NFA); and high-content imaging (HCI) to evaluate proliferation, apoptosis, neurite outgrowth, and synaptogenesis. This work addresses three hypotheses: (1) a broad screening battery provides a sensitive marker of DNT bioactivity; (2) selective bioactivity (occurring at non-cytotoxic concentrations) may indicate functional processes disrupted; and, (3) a subset of endpoints may optimally classify chemicals with *in vivo* evidence for DNT. The dataset was comprised of 92 chemicals screened in all 57 assay endpoints sourced from publicly available data, including a set of DNT NAM evaluation chemicals with putative positives (53) and negatives (13). The DNT NAM battery provides a sensitive marker of DNT bioactivity, particularly in cytotoxicity and network connectivity parameters. Hierarchical clustering suggested potency (including cytotoxicity) was important for classifying positive chemicals with high sensitivity (93%) but failed to distinguish patterns of disrupted functional processes. By contrast, clustering of selective values revealed informative patterns of differential activity but demonstrated lower sensitivity (74%). The false negatives were associated with several limitations, such as the maximal concentration tested or gaps in the biology captured by the current battery. This work demonstrates that this multi-dimensional assay suite provides a sensitive biomarker for DNT bioactivity, with selective activity providing possible insight into specific functional processes affected by chemical exposure and a basis for further research.

Introduction

Developmental neurotoxicity (DNT) refers to any adverse outcomes in the course of normal development of nervous system structures and/or function resultant to chemical exposure (EPA, 1998b). Many *in vivo* DNT studies involve exposure of pregnant rats from

Corresponding author: Katie Paul Friedman, 109 T.W. Alexander Drive, Mail Drop D143-02, Research Triangle Park, NC 27711, paul-friedman.katie@epa.gov, Tel. 919-541-0660, Fax. 919-541-1194.

implantation through the end of lactation, and include measurements of both functional and structural integrity of the developing nervous system, such as brain weights, morphometry, neuropathology assessment, and neurobehavior during critical windows of susceptibility (EPA, 1998a; OECD, 2007). However, *in vivo* DNT studies are extremely resource intensive (financial, labor, and animal use) (Crofton et al., 2011; Raffaele et al., 2008; Smirnova et al., 2014; Tsuji et al., 2012), preventing use in routine screening. Second, uncertainties regarding human relevance are apparent, including uncertainty regarding the disrupted functional processes and limited ability to assess complex adverse outcomes such as higher cognitive function. Third, DNT reference substances demonstrate low reproducibility in *in vivo* DNT study performance across different laboratories (Bal-Price et al., 2018). Fewer than 200 substances have been evaluated in the Office of Chemical Safety and Pollution Prevention Health Effect Series 870.3600 or Organisation for Economic Co-operation and Development Test Guideline 426 DNT guideline paradigms to date, leaving a potential knowledge gap. This has led to a 15 year internationally-coordinated research effort to build a battery of *in vitro* DNT NAMs (Fritsche et al., 2018; Lein et al., 2007; Sachana et al., 2019). The DNT NAM battery developed at the US EPA Office of Research and Development may complement the European Food Safety Authority-funded DNT testing battery recently evaluated by Masjosthusmann and colleagues (2020) and increases screening capacity to provide DNT-relevant bioactivity data that may be used in fit-for-purpose applications (Fritsche, et al., 2018; Lein, et al., 2007; Sachana, et al., 2019; Sachana et al., 2021).

The developing nervous system is known to be particularly sensitive to environmental insult during critical periods of development {Rice, 2000 #62;Grandjean, 2014 #123}, including evidence from exposure to ethanol {Laev, 1995 #63;Jones, 1973 #125}, lead {Alfano, 1983 #64;Canfield, 2003 #126}, methylmercury {Barone, 1998 #65;Grandjean, 1997 #127}, and chlorpyrifos {Whitney, 1995 #66;Das, 1999 #67;Rauh, 2006 #128}. It is generally accepted that no single *in vitro* screening assay can recapitulate all the critical process of neurodevelopment or clearly identify all substances that produce DNT due to the temporal and biological complexity of nervous system development (Bal-Price, et al., 2018). Some substances may affect a single neurodevelopmental process, whereas other substances may disrupt key neural processes early in development resulting in tangential or downstream effects on other processes (Harrill et al., 2018). Therefore, the integration of *in vitro* assays that represent diverse neurological processes across development will be critical for identifying DNT-relevant bioactivity (Fritsche, et al., 2018; Sachana, et al., 2019). Researchers at the US EPA have developed and implemented DNT NAMs that use human cell lines and primary rat cortical neurons to evaluate chemical effects on neural network formation, proliferation, apoptosis, synaptogenesis, neurite outgrowth and maturation (Brown et al., 2016; Frank et al., 2017; Harrill, et al., 2018; Shafer et al., 2019) using a high-content imaging (HCI) assay suite and a microelectrode array (MEA) network formation assay (NFA). The NFA measures neuronal activity, bursting, and connectivity (Brown, et al., 2016; Frank, et al., 2017; Shafer, et al., 2019). The HCI assay suite measures cell proliferation with a human neural progenitor cell line; apoptosis and viability assessment using a human neural progenitor cell line (Druwe et al., 2015); neurite outgrowth initiation with a human neuronal lineage cell line (Harrill et al., 2010); and, primary rat

cortical cell neurite outgrowth initiation, neurite maturation, and synaptogenesis (Harrill et al., 2013; Harrill et al., 2011b). These DNT NAMs detect chemical perturbations of a variety of critical neurodevelopmental processes *in vitro*.

The primary goal of this work is to assess the performance of existing DNT NAMs, revealing patterns of responses for the 92 substances screened to date in both MEA NFA and HCI assays relevant to DNT. No study to date has described how the NFA performs in conjunction with the battery of HCI assays to evaluate proliferation, apoptosis, neurite outgrowth, and synaptogenesis, for identifying positive and negative chemicals in a DNT NAM evaluation set of chemicals. Primarily, we wanted to investigate whether together these two assay technologies, composed of multiple assay endpoints evaluating different functional processes relevant to DNT, were capable of delineating positive and negative evaluation chemicals, and whether this *in vitro* activity tended to be free of cytotoxicity. Evaluating apparent patterns of response in these assays may help elucidate the *in vitro* biological effects of substances specific to DNT-relevant processes and function and inform the hypothesis that these DNT NAMs form an effective screening battery for putative DNT hazard. The novel computational analysis herein addresses three related hypotheses: (1) the DNT NAMs provide a sensitive indicator of disruptions in key neurodevelopmental processes; (2) evaluating bioactivity that may be selective, i.e. occurring at concentrations less than observed cytotoxicity, may provide a more specific indicator of the functional processes involved in putative DNT bioactivity; and, (3) the available DNT NAM screening battery may optimally classify putative positive and negative DNT NAM evaluation chemicals identified previously as helpful for assay development (Mundy et al., 2015). This work provides an analysis method for interpretation of these DNT NAMs for *in vitro* hazard identification and prioritization of chemicals for DNT bioactivity and further informs the need for multiple assay technologies for optimal identification of DNT-relevant bioactivity.

Methods

Overview

The overall workflow of this analysis is as follows: 92 chemicals were tested in a concentration-response screening which included 19 MEA parameters (including two cytotoxicity parameters) and 21 HCI parameters (including proliferation, apoptosis, neurite outgrowth (NOG) initiation, neurite maturation and synaptogenesis assays). The parameters were categorized into seven ‘activity types’ based on their associated neural processes. The HCI activity types include ‘neurite outgrowth’, ‘synaptogenesis and neurite maturation’, ‘proliferation’, and ‘apoptosis’, and the MEA activity types include ‘general activity’, ‘bursting’, ‘network connectivity’. See Table 1 and Supplemental Methods for detailed experimental design for each assay. The MEA NFA data was collected over a period of 12 days in vitro (DIV) with recordings occurring on DIV 5, 7, 9, and 12. Chemical exposure started on DIV 0, two hours after plating, and full media changes occurred on DIV 5 and 9 (two repeat doses). The length of the HCI assays ranged from 1-2 days with a single dose exposure occurring two hours after plating, with the exception of the synaptogenesis assay where the chemical exposure started on DIV 7 and data was collected on DIV 12. The chemical concentration-response data for each assay were normalized on a plate-by-plate

basis to the median of the vehicle control wells and the two lowest concentrations of test chemical wells and curve-fit using the ToxCast Pipeline (tcpl) to identify active or inactive chemicals (Filer et al., 2017). Seventeen MEA parameters (excluding the two cytotoxicity parameters) were fit in both the 'up' or 'down' activity direction. The HCI parameters were fit in the 'down' direction to capture a loss of bioactivity, with the exclusion of "Casp3_7_gain" which was fit in the 'up' direction. Examination of the fitted curves from the HCI parameters returned very little to no evidence suggesting increases in the assay parameters for functional processes. The concentration at 50% maximal activity (AC_{50}) values were analyzed across the 57 endpoints using hierarchical clustering to evaluate differential patterns of bioactivity among the 92 chemicals. Based on the observation that the MEA NFA results may have been associated with cytotoxicity, activity lower than the cytotoxicity AC_{50} was examined as an area under the curve (AUC), compressing efficacy and potency into a single metric of selective activity in order to identify groups of chemicals with similar selective DNT-relevant activity in the battery. In addition to evaluating patterns of activity, several other computational approaches were utilized: 1) a random forest regression analysis was used to evaluate the most informative parameters for predicting the minimum AC_{50} value, 2) *in vitro* to *in vivo* extrapolation (IVIVE) using high-throughput toxicokinetic (HTTK) modeling was applied to transform NAM-derived activity into administered equivalent doses (AEDs) for a subset of chemicals that were identified as false negatives; and, 3) DNT NAM *in vitro* potencies were compared to a dataset of predominantly non-neuronal cell models available in the ToxCast/ Tox21 database.

Software and source data

All data management, analysis, and figures were generated using the R statistical programming language using RStudio (version 3.6.3). All original code and source files are available on GitHub and at (https://github.com/USEPA/CompTox-Evaluation-DNTNAMs). The code and source data for these analyses are also available here (https://doi.org/10.23645/epacomptox.17152877). Assay data were compiled using the ToxCast pipeline (tcpl) R package (version 2.0.3) that is publicly available (https://cran.r-project.org/web/packages/tcpl/index.html) (Filer, et al., 2017) and these assay data are public in invitrodb version 3.4 (https://doi.org/10.23645/epacomptox.6062479.v6). Subsets of the source data were previously published in Brown et al., 2016, Frank et al., 2017, Harrill et al., 2018, and Shafer et al., 2019. Multi-concentration level 0, 3 and 5 data were downloaded from invitrodb (version 3.4) from the assay source names CCTE_Shafer_MEA_dev" and the "MUNDY_HCI". See Supplemental Methods Table 3 for a detailed description of the methods used to pipeline the NFA and HCI assay suite data in tcpl. Level 5 data were filtered with the following criteria: 1) chemicals with ≥3 flags were set to a hit-call of zero (flagging at multiple-concentration level 6 data in tcpl uses both the plate level concentration-response data and modeled parameters to flag for caution in interpreting the curve fit, e.g. noisy data, only highest concentration above baseline, or hit-call potentially confounded by overfitting), 2) chemicals with a fit category (fitc) of 36 or 45 (found in mc5 table in tcpl), indicating the curve top was less than or equal to 1.2 times the cutoff for a positive and the resultant AC_{50} was less than the concentration range screened, were set to a hit-call of zero, 3) any hit-call= −1, indicating the concentration series had fewer than 4 concentrations, was set to zero. It should be noted that the hit-calls presented here for MEA

results may differ slightly from those presented in Frank et al., 2017 and Shafer et al., 2019 due to the use of tcpl here.

All Supplemental Files referenced in this work, containing additional results and explanations, can be found online here: https://doi.org/10.23645/epacomptox.17152964.

Chemical screening

A total of 92 chemicals were selected for this analysis that were tested in both the NFA and HCI methodologies, including 20 chemical repeats tested in the NFA. This includes 53 chemicals previously tested and evaluated in Harrill et al., 2018, Frank et al., 2017 and Shafer et al., 2019 that were selected based on evidence of possible *in vivo* DNT hazard (Mundy, et al., 2015), neurotoxic properties, as well as 13 putative negative DNT chemicals (Aschner et al., 2017; Harrill, et al., 2018). This set of 66 chemicals comprised a “DNT NAMs evaluation set” that can be used for the purpose of evaluating the sensitivity and specificity of the battery and includes a variety of chemical classes such as pyrethroids, organochlorines, organo metals, and pharmaceutical drugs. In addition, data for 26 organophosphate pesticides with known acute toxicities are also included here. Assay data from the OPs are available publicly via an Agency Issue Paper prepared for a Federal Insecticide, Fungicide, and Rodenticide Act Scientific Advisory Panel (EPA, 2020). A complete list of all chemicals screened in concentration response can be found in the supplemental materials (Supplemental Table 1). Eight additional chemicals were included as positive controls for the assay performance analysis that were not tested in all assay endpoints for the NFA and HCI assays (Supplemental Table 2). The chemicals presented in the comparison analysis with ToxCast/Tox21 dataset were derived from the ToxCast chemical library, which covers a broad chemical landscape and diversity of assays. All information on the chemicals and assays within the ToxCast chemical library are publicly available (https://comptox.epa.gov/dashboard).

Each chemical was screened with well replicates (n=3, where n is in reference to technical replicates) performed within the same experiment across multiple plates. For the vast majority of chemicals, screening concentrations ranged from 0.001 μM to 100 μM for the HCI and 0.03 μM to 30 μM for the NFA in half-log increments, with a few exceptions, such as sodium valproate which was tested at a maximal concentration of 3000 μM (see Supplemental Table 1 for detailed information on the minimum and maximal concentrations tested for each chemical in each assay technology). The lower concentration bound was reflective of practical limitations (e.g. the number of wells on the plate) and the upper concentration was reflective of DMSO limitations and typical concentration range within which activity is observed. The HCI assays were performed in a 96-well plate format using human neural progenitor cells (hNP1 cells derived from a neuroepithelial cell lineage of WA09 human embryonic stem cells, ArunA Biomedical (Athens, GA)) in the proliferation and apoptosis assays, primary rat cortical cells (dissected at P0 from Long Evans rat, (Charles River Laboratory, Raleigh, North Carolina)) in the synaptogenesis assay, or human neural cells (hN2 cells: immature neurons differentiated from the hNP1 cells described above) or primary rat cortical cells in the NOG assay. The NFA assay was performed in a 48-well microelectrode plate format using rat cortical cells.

Selectivity

Selective activity in this study is defined as activity below the threshold of cytotoxicity. This was determined with two methods: 1) by calculating the AUC of each concentration-response curve, applying an upper concentration cutoff at the cytotoxicity log10-AC_{50} for each assay, 2) by excluding log10-AC_{50} potency values above the cytotoxicity log10-AC_{50}. The first method was implemented in the 'patterns of selective activity' analysis. AUC was calculated by integration of each positive curve-fit using the model parameters from the winning model fit in tcpl. The upper limit of integration was defined as the cytotoxicity log10-AC_{50} for each chemical by assay (in the absence of any cytotoxicity, the AUC is based on the full curve-fit for the assay endpoint). The cytotoxicity endpoints included "MUNDY_HCI_Cortical_NOG_NeuronCount_loss" for the NOG (rat) assay, "MUNDY_HCI_hN2_NOG_NeuronCount_loss" for the NOG (hN2) assay, "MUNDY_HCI_hNP1_CellTiter_loss" for the proliferation assay. The MEA_NFA contained two cytotoxicity endpoints ("CCTE_Shafer_MEA_dev_LDH_dn", "CCTE_Shafer_MEA_dev_AB_dn") of which the minimum log10-AC_{50} was found between the two endpoints for each sample. Any assay endpoints that had a negative hit-call or a negative AUC from this calculation was set to zero in the analysis. The second method was implemented in the comparison of activity between early stage processes and the NFA as well as in the comparison of the DNT selective potency with ToxCast/ Tox21 data.

Patterns of activity in the DNT NAMs

Unsupervised hierarchical clustering was used to understand the differential patterns of activity across all 92 chemicals and 57 endpoints (log10-AC_{50} analysis) or 32 endpoints (selective activity analysis which excludes cytotoxicity endpoints and the NFA '_up' endpoints) as visualized in a heatmap. These analyses used Ward's D2 method for linkage (Ward, 1963) and the agglomerative clustering method as implemented in the R package 'gplots' version 3.0.4 (Warnes et al., 2016).The dendrogram shown on the heatmap was used as a guide for manual selection of cluster groupings. In the selectivity analysis, k-means clustering was also implemented as a secondary validation method to determine the optimal number of clusters. The elbow method was used to determine the optimal number of clusters (value of k using the 'kmeans' function in the 'stats' base R package), where the point of inflection of the curve (x= k, y= sum of squared distances from each point to its center) indicates the best model fit. Principal component analysis plots were also utilized to visualize the optimal number of clusters. The type II error rate of classifying putative positive and negative DNT NAM evaluation chemicals was computed based on active versus inactive clusters, respectively. An active cluster was defined as a cluster that captured chemicals with any activity in the selectivity heatmap. An inactive cluster was defined as a distinct cluster that demonstrated limited to no activity. A confusion matrix was computed, where a true positive (TP) was defined as an evaluation set positive that was captured by clusters that demonstrated selective activity and a true negative (TN) was defined as an evaluation set negative that was grouped in an inactive cluster. An evaluation set positive was considered a false negative (FN) if it was incorrectly captured in the 'negative' cluster and a false positive (FP) was any evaluation set negative that was captured in the active clusters. The sensitivity, specificity, and accuracy of the model was calculated: sensitivity = TP/(TP+FN), specificity = TN/(TN+FP), and accuracy = (TP + TN) / (TP + TN + FP + FN).

Important DNT NAM endpoints for chemical potency estimation

Random forest regression using the randomForest R package (version 4.6-14) was used to identify the most important endpoints for predicting the minimum log10-AC_{50} value for each chemical across 40 features (all endpoints in the DNT NAM battery excluding the NFA 'up' endpoints). The 92 chemicals were split 70/30 into training and testing datasets using the 'createDataPartition' in the R package "caret" (version 6.0-86) (Kuhn, 2008). The mean absolute error and the root residual mean square error were computed for the predicted values in the testing dataset. Feature importance was scored, ranked, and visualized based on two metrics: percent increase in mean square error and increase in node impurity.

In vitro to in vivo extrapolation (IVIVE) approach

For a subset of eleven chemicals in this dataset (seven that were identified as false negatives and four as true positives in the selectivity analysis), the bioactivity of the *in vitro* DNT NAM data was transformed into administered equivalent doses (AEDs) using high-throughput toxicokinetic (HTTK) data and models following the principles of reverse dosimetry (Bell et al., 2018; Sipes et al., 2017; Wambaugh et al., 2018; Wetmore et al., 2012). In total, seven of these 11 chemicals were considered "false negatives" using the DNT NAM assays and 4 of these 11 chemicals were selected to represent a diverse subset of "true positives" using the DNT NAM assays; insufficient data were available to estimate AEDs for all chemicals screened in the DNT NAM assays presented here. This methodological approach, referred to as *in vitro* to *in vivo* extrapolation (IVIVE), was used to compute an AED in units of milligram per kilogram bodyweight per day (mg/kg/day) from the NAM-derived concentration at 50% maximal activity (log10-AC50). The IVIVE approach relies on several high-level assumptions: 1) a nominal *in vitro* assay concentration approximates an *in vivo* plasma concentration that would correspond to a similar effect; 2) *in vivo* plasma concentration can be approximated based on steady-state kinetics; and, 3) a toxicokinetic model can be constructed using estimates of species-specific physiology and Phase I and Phase II enzyme-driven hepatic clearance. The HTTK information was built into the "httk" R package (version 2.0.3) (Pearce et al., 2017) which uses Monte Carlo simulation to incorporate population variability in the model. The 3-compartment steady-state model (3css) was selected based on the availability of hepatic intrinsic clearance (in units of $\mu L/min/10^6$ hepatocytes) values to predict both the 50^{th} or 95^{th} percentile in the population distribution (referred to as 3css50 or 3css95). Hepatic intrinsic clearance parameters were measured *in vitro* (Wambaugh, et al., 2018) for five chemicals and *in silico* for six chemicals (Sipes, et al., 2017) (Supplemental Table 3). Restrictive clearance was employed, i.e. hepatic clearance was assumed to be dependent on the fraction unbound in plasma. Due to the lack of hepatic clearance data in rat, the present analysis used human HTTK data and human physiology. Given this list of high-level assumptions, the NAM-derived AEDs were compared to *in vivo* lowest effect levels which were curated from a literature review of chemicals with *in vivo* DNT information (Mundy, et al., 2015). The lowest effect levels were not derived from DNT guideline studies therefore the dose does not necessarily reflect a true 'lowest effect' but instead the lowest dose found to demonstrate a DNT effect based on the available studies. The *in vivo* doses were transformed into human equivalent doses (HEDs) (Nair et al., 2016), where HED (mg/ kg) (see Table 1 of Nair and Jacob (2016) for equation).

All of the HTTK information and exposure citations used in this analysis are summarized in Supplemental Table 3.

Results

NFA and HCI assay performance

Assay endpoints—A total of 92 substances were tested in the DNT NAM battery, which included 4 HCI assays: proliferation (3 endpoints), neurite outgrowth (NOG) initiation (4 endpoints x 2 cell lines), neurite maturation and synaptogenesis (8 endpoints), and apoptosis (2 endpoints), and the MEA NFA, which included 17 endpoints measured in both the increased and decreased activity direction (34 endpoints) and viability (2 endpoints), for a total of 57 endpoints. The proliferation, NOG, and synaptogenesis assays each included a measure of cell viability (e.g. cell counts) and/or cytotoxicity (e.g. lactate dehydrogenase (LDH)); in this manuscript, these shall be collectively referred to as “cytotoxicity parameters”. Cytotoxicity parameters are relevant to DNT if neural cell death occurs *in vivo* during critical periods of nervous system development (Rice et al., 2000). Several critical processes of *in vivo* neurodevelopment and several cell models are encompassed in this suite of assays (Table 1).

Assay reproducibility

Each of these assays were developed following the principles outlined in Crofton et al., 2011 and have been demonstrated to perform well with assay positive controls (Brown, et al., 2016; Druwe, et al., 2015; Harrill, et al., 2018; Harrill et al., 2011a; Harrill, et al., 2013; Harrill, et al., 2011b). To further demonstrate the reproducibility and robustness of these assays,, we evaluated several additional quantitative indicators of assay reproducibility.. First, baseline variability was evaluated as the median and mean coefficient of variation (CV) of the vehicle control wells for: the NFA parameters, NFA-associated cytotoxicity assays, and the HCI assays, the mean and median CVs for DMSO were 18-21%, 7-8%, and 8.4 and 8.7%, respectively. Baseline variability indicates the level of changes required to see a difference from background, with the MEA NFA demonstrating greater baseline variability relative to the HCI assay suite (Supplemental Table 4). This difference is perhaps due to the greater biological complexity inherent to functional networks in contrast to measurement of specific cellular morphological or biochemical changes in the HCI assay suite. Second, qualitative and quantitative reproducibility was evaluated in the NFA assay, wherein 20 chemicals were tested as at least two unique chemical samples with differences in sourcing (Supplemental Table 5). Each chemical was given a reproducibility score based on three response categories: 1) ‘strong’ (replicates were consistently positive with >3 hits or consistency negative with 0 hits), 2) ‘equivocal’ (1 replicate was between 1 and ≤3 hits and 1 replicate was negative), or 3) ‘weak’ (1 replicate was positive and 1 was replicate negative or equivocal) (Supplemental Results 1A). Overall, the replicability of the chemical repeats in the NFA was 75% (15/20 chemicals were scored as strong or equivocal), despite the use of separately sourced chemical samples. The variability in potency across repeats (measured by computing the average standard deviation of bioactivity at each endpoint) was less than 0.6 log10-μM for each chemical, except for 2,2’,4,4’-tetrabromodiphenyl ether and di(2-ethylhexyl) phthalate where there was more uncertainty in the potency values (Supplemental

Table 4), suggesting generally that AC_{50} values in the NFA might reasonably vary by ± 0.5 log10-μM, or approximately ± 1 standard deviation. There are several reasons why some chemicals demonstrated 'weak' concordance (e.g. 6-propyl-2-thiouracil, chlorpyrifos oxon, di-(2-ehtylhexyl)phthalate, diazinon, methamidophos) such as chemical sample stability, biological variability in culture preparations, or differences in laboratory equipment and personnel (Supplemental Results 1A).

Assay performance was also assessed with positive controls that are known to cause effects on neurodevelopmental processes measured in these assays ("assay positive controls") or to be inactive ("assay negative controls"; Supplemental Table 2). Overall, the activity of the controls performed as expected with positive controls demonstrating selective activity in the NFA and HCI assays and the negative control (acetaminophen) demonstrating little to no activity in the battery (Supplemental Figure 2, Supplemental Results 1B). For example, the positive control, lithium chloride, a known inhibitor of neurite outgrowth (Harrill, et al., 2011b), decreased all 4 endpoints across the NOG initiation assay in both the rat cortical cells and the human hN2 neural cells (Supplemental Figure 2A). Aphidicolin, a known inhibitor of cell proliferation (Kohno et al., 2006; Mundy et al., 2010) decreased all 3 endpoints in the neuroprogenitor proliferation assay. Bisindolylmaleimide I, a known inhibitor of an ontogeny of network activity (Harrill, et al., 2011b; Robinette et al., 2011), was active in 16 NFA endpoints measuring decreased network activity which was reproducible across 4 chemical repeats (Supplemental Figure 2B).

Finally, the Z' factor, a measure of how reliably a signal can be distinguished from noise, was computed for the assay-matched positive controls. A Z' of 0.5-1.0 indicates an assay with sufficiently high signal-to-background (noise) distinction (Paul Friedman et al., 2016; Zhang et al., 1999). The Z' across all NFA endpoints ranged from 0.55-0.8, with bisindolylmaleimide I demonstrating the highest effect size of the positive controls, and overall suggesting reproducible, robust results in the NFA with a substance known to perturb network formation (Supplemental Table 6, Supplemental Results 1C). For the HCI assays, the Z' was more variable, ranging from 0-0.8, indicating positive controls with high efficacy and high reproducibility for the viability and NOG assays and perhaps a future need for better positive controls in the synaptogenesis and hNP1 proliferation assays (Supplemental Table 7). For example, staurosporine, known to induce apoptosis (Bertrand et al., 1994; Druwe, et al., 2015; Feng et al., 2002), had a robust Z' value of 0.8 and large effect sizes as indicated by the strictly standardized mean difference (Bray et al., 2004) in the endpoint measuring apoptosis in hNP1 neuroprogenitors, indicating an optimal positive control. A known inhibitor of proliferation, aphidicolin, produced a small effect size and more variability as indicated by a low signal-to-noise (another metric that indicates the positive control response from the baseline response) and Z' (Z' value of 0.1), making aphidicolin a less optimal control but one that clearly affected hNP1 proliferation. Together, the CV of vehicle control wells, the concordance of NFA chemical repeats, the activity and Z' factor of known positive and negative controls indicates that this battery of broad phenotypic endpoints has sufficiently low baseline variability and high signal-to-noise to detect chemical perturbations of these neurodevelopmental processes *in vitro*.

Patterns of general DNT-relevant activity using the log10-AC50 metric (μM)

Here we evaluated the patterns of activity across the DNT NAM battery (Figure 1, Supplemental Table 8). We hypothesized that the integration of all assays would reveal informative differential patterns of bioactivity that may ultimately be useful for identifying putative DNT activity from a set of distinct cellular processes. The 92 chemicals were tested in a concentration window of 0.001 to 100 μM with a narrow potency range (median potency range ±SD of 8.30 ± 11.3 μM). Hierarchical clustering of activity and potency using the log10-AC_{50} (μM) in the combined DNT battery analysis revealed three main chemical groups (rows/ y-axis dendrogram) and a general 'on/off' activity pattern, particularly for the NFA. Cluster 1 grouped chemicals that were very active in the NFA and moderately active to not active in the HCI assays (≤26/57 hits), cluster 2 grouped high activity (with activity seen in NFA and some or all of the HCI assays) and high potency chemicals (≥23/57 hits), and cluster 3 captured very low activity to negative chemicals (≤6/57 hits). Hierarchical clustering of activity type (column labels/ x-axis dendrogram) revealed that the four NFA activity types measuring decreased activity ('down') cluster together (bursting, network, general activity, cytotoxicity). Notably, three synaptogenesis endpoints (dark orange bar) associated with the NFA endpoints 'down' endpoints: 1) neurite length, 2) pre-synaptic puncta in the neurite compartment, and 3) cytotoxicity in the same assay. All other endpoints measured with the HCI technology clustered together in a separate main effect group, and the NFA endpoints measuring increased activity clustered together, demonstrating little to no activity. A current limitation to contextualizing this result is the lack of assay positive controls in the NFA that are known to increase NFA parameters. The two NOG initiation assays (tested in either human hN2 neural cells (yellow bar) or rat primary cortical cultures (dark blue bar)) were not closely associated, suggesting that some chemicals may differentiate by cell-type rather than neurological process. Given the strong clustering by assay technology and the lack of differential patterns across the NFA 'down' endpoints, we evaluated the correlation between endpoints using a correlation matrix (Supplemental Figure 3 and Supplemental Results 2). The NFA 'down' endpoints were strongly correlated (mean correlation coefficient ± SD of 0.9 ± 0.02), suggesting that log10-AC_{50} for all 17 parameters may not be needed for indicating activity in the NFA.

Informative endpoints in identifying the most sensitive endpoint

We next evaluated assay endpoint sensitivity, to understand which endpoints lead to minimum AC_{50} value estimates, as well as the hypotheses regarding how selective bioactivity in the DNT NAM battery may allow for greater differentiation of *in vitro* phenotype. We examined potential trends in the most sensitive assay endpoints by determining the endpoint with the minimum log10-AC_{50} response per chemical across the DNT battery. In comparing the NFA and HCI assays, the minimum log10-AC_{50} value was determined by the NFA for 76.3% of chemicals, with the activity-type 'network connectivity' defining the minimum log10-AC_{50} value for nearly 50% of the 92 chemicals screened (Figure 2A). The NFA endpoint measuring decreased mutual information, a measure of network connectivity, was the most sensitive endpoint overall (Figure 2B), followed by the NFA cytotoxicity endpoint using LDH. Four out of the top eight most sensitive endpoint were measures of cytotoxicity (black bars) in different assays and cell-types (rat cortical, hNP1, and hN2 cells). The normalized mutual information endpoint

is a derived endpoint that was designed to be a sensitive indicator of perturbations in a developing neural network (Ball et al., 2017). Importantly, approximately 90% of chemicals with more than one hit in any assay were within one order of magnitude difference from the closest log10-AC_{50} response. Indeed, all potencies for the 92 chemicals fell within a median range ± SD of 0.919 ±1.05 log10-μM. Moreover, the minimum log10-AC_{50} value as defined by the NFA versus the HCI assay suite demonstrated a low coefficient of determination (R^2) of 0.43, indicating variability in assay sensitivity between the two assay technologies (Supplemental Figure 4). These findings were supported by a random forest regression analysis which predicted the minimum log10-AC_{50} value with a mean absolute error of 0.572 log10-μM and a root residual mean square error of 1.083 log10-μM. Ranking of feature importance for predicting the minimum log10-AC_{50} value identified cytotoxicity and network connectivity endpoints as the most informative endpoints in predicting the minimum log10-AC_{50} value across the battery (Supplemental Figure 5, Supplemental Table 9, Supplemental Results 3). Overall, these results reveal an important role for cytotoxicity and network connectivity parameters in determining the minimum log10-AC_{50} value and indicates that no single endpoint is sufficient to define a most sensitive DNT effect in the DNT NAM battery for all chemicals. Moreover, endpoints measuring other neurodevelopmental processes (e.g. loss in NOG branch points and loss in neurite length in the synaptogenesis assay) are also sensitive in identifying neuronal perturbations, supporting the application of multiple endpoints and key neurodevelopmental processes in the battery.

Patterns of selective activity

To explore the hypothesis that selective bioactivity in the DNT NAMs may allow for greater differentiation of *in vitro* phenotypes, we evaluated patterns of DNT-relevant activity at concentrations below the AC_{50} for cytotoxicity. Patterns of selective activity excluded NFA ‘up’ endpoints due to limitations on interpreting these endpoints, including few instances of increases in these endpoints and a lack of positive controls that reliably increase these endpoints. These patterns of selective activity were evaluated using a hierarchical clustering approach (Figure 3 and Table 2, Supplemental Table 10). The selectivity analysis importantly reveals that there is a narrow concentration window (median range ± SD of 0.101 ± 0.732 log10-μM) to observe activity that is not likely to be confounded by cytotoxicity in the DNT NAM battery (the median difference between the cytotoxicity AC_{50} and the mean selective potency ± SD was 0.109 ± 0.334 log10-μM for the NFA and 0.086 ± 0.447 log10-μM for the HCI assay suite, Supplemental Tables 11 and 12). Hierarchical clustering by activity type and chemical revealed five main chemical clusters demonstrating differential patterns of activity that were not evident when considering bioactivity that co-occurred with cytotoxicity, with support for three observations: (1) active chemicals cluster into four groups based on the magnitude of the sum of the selective AUC for all endpoints by chemical; (2) a subset of endpoints measuring general activity and/or network connectivity in the NFA, NOG (hN2), and synaptogenesis/ neurite maturation appear to differentiate chemicals with high selective potency and/or efficacy (mean AUC sum ± SD of 1140 ± 441.5); and, (3) the association of the selective AUC data across assay endpoints suggests cell-type and species-specific effects (Supplemental Results 4). Cluster 1, with a mean AUC sum ±SD of 1230 ± 373.2, captured chemicals that were selectively active in synaptogenesis and/or neurite maturation

assays, particularly in endpoints measuring decreased number of synapses, number of neurite-associated puncta, neurite length loss, and neurite branch points per cell. The 14 chemicals grouped in cluster 1 consisted of metals (e.g. triethyltin bromide, cadmium, bis(tributyltin)oxide, and methylmercury), organochlorines (heptachlor, hexachlorophene, and dieldrin), and neuroactive drugs (fluoxetine, chlorpromazine, and chlordiazepoxide). Colchicine, an alkaloid gout therapeutic that causes cell death via interference with microtubules and prevents tubulin-dependent neurite outgrowth in neurons (Gorman et al., 1999; Uppuluri et al., 1993), was captured in cluster 1 and was notably dissimilar with an AUC sum of 2043. Cluster 2, with a mean AUC sum ± SD of 1046 ± 510.3, captured chemicals that were selectively active in general activity and network connectivity, particularly in endpoints measuring decreased mutual information between electrodes, number of spikes in a network spike, standard deviation of network spike duration, number of bursts per minute, and mean firing rate. The chemicals haloperidol and deltamethrin were particularly closely associated in the hierarchical clustering (and in the PCA of the k-means analysis in Supplemental Figure 6D), demonstrating strong selective activity across all network, general and bursting activity endpoints. Cluster 3, with a mean AUC sum ± SD of 305.4 ± 156.4, captured low efficacy chemicals with moderate activity across multiple neurodevelopmental processes. Several subclusters demonstrated trends in selective activity, such as 6-aminopyridine-3-carboxamide, paraquat dichloride, and z-tetrachlorvinphos (blue arrows), which were selective in all three endpoints measuring NOG (rat) but not NOG (hN2), indicating possible cell-type/species-specific effects. Cluster 4, with a mean AUC sum ± SD of 80.54 ± 58.83, captured low efficacy chemicals with disparate (or highly selective) activity. Cluster 5 was manually defined based on inactivity. The selectivity analysis demonstrates activity with greater differentiation of *in vitro* phenotypes than non-selective activity (Figure 1) and identifies groups of chemicals that, in the absence of cytotoxic effects, appear to target distinct neurodevelopmental processes.

The selective AUC heatmap also indicates which chemicals were contained in an *in vivo* DNT NAM evaluation set, including putative “positives” (53 chemicals, black row annotation in Figure 3) and “negatives” (13 chemicals, gray row annotation in Figure 3) (Supplemental Table 1, as proposed by previous publications (Aschner, et al., 2017; Harrill, et al., 2018; Mundy, et al., 2015)). A confusion matrix was computed for the classification of putative *in vivo* DNT positives in the active clusters (clusters 1-4) and the putative *in vivo* DNT negatives in the inactive cluster (cluster 5), resulting in an accuracy of 77%, a sensitivity of 74%, and a specificity of 92%. Based on selective activity, the DNT battery performed with a high specificity rate, identifying 12/13 putative negatives in the inactive cluster, and a moderate sensitivity rate, incorrectly classifying 14/53 putative positives in the inactive cluster. In contrast, non-selective activity (including cytotoxicity and MEA NFA ‘up’ endpoints, Figure 1) classified DNT NAM evaluation chemicals with a sensitivity of 93%, a specificity of 69%, and an accuracy of 88% (active clusters were defined as ≥1 positive hit-call). The higher sensitivity can be explained by capturing chemicals that were cytotoxic without any other DNT NAM-relevant activity at concentrations below the AC_{50} for cytotoxicity (5,5-diphenylhydantoin, acrylamide, caffeine, cyclophosphamide, amphetamine, fosthiazate) or that demonstrated disparate activity in the NFA ‘up’ endpoint(s) (maneb, phenol, diethylene glycol, nicotine, acephate).

Only four chemicals were identified as false negatives: carbamazepine, naloxone, sodium fluoride, and thalidomide. One possible explanation for the misclassification of putative positives in these analyses may be that the screening concentration was not sufficiently high to cause effects *in vitro*. To enable comparison of the maximum screened concentrations to doses known to cause adversity *in vivo,* IVIVE was performed using a high-throughput toxicokinetic (HTTK) model and data, as operationalized by the R library "httk," to approximate NAM-derived administered equivalent doses (AEDs) in units of mg/kg/day (Pearce, et al., 2017; Wambaugh, et al., 2018). Only seven of 14 putative false negatives had sufficient information to estimate AEDs and compare them to *in vivo* points-of-departure (Mundy, et al., 2015). Based on reported DNT-relevant *in vivo* rodent doses and converted human equivalent doses (HED), three of these seven (caffeine, maneb, and 5,5'-diphenylhydantoin) may have not been screened at a high enough concentration (Figure 4, Supplemental Table 13). The AED based on the maximum screened concentration for sodium valproate approached doses associated with *in vivo* adversity, and cyclophosphamide monohydrate, acrylamide, and nicotine appeared to have been screened at sufficient concentrations. Four true positives from the DNT NAM evaluation set were included for comparison, demonstrating *in vivo* points-of-departure that were several orders of magnitude below the AEDs based on the maximum screened concentration.

We further explored trends in the perturbations of the key cellular processes, evaluating whether disruptions in early stage processes (e.g. synaptogenesis and/or neurite maturation, NOG) correspond with disruptions in a later-stage process (e.g. network formation). In this analysis, the unity of activity between the early stage assays (NOG and synaptogenesis) and the NFA were presented in Venn diagrams (Figure 5). A chemical was considered active if it was a hit in at least one assay endpoint, including cytotoxicity. In the NOG assays, all chemicals that disrupted NOG (rat) also disrupted the NFA, while 35/37 chemicals that disrupted the NOG (hN2) assay also disrupted the NFA. In the synaptogenesis assay, 42/45 chemicals that disrupted the synaptogenesis assay also disrupted the NFA. These data indicate that screening simply for early stage processes, NOG (in rat or hN2 cells) or synaptogenesis and/or neurite maturation effects (in rat cells), would not have fully identified all hits in the NFA. When examining selective hits, there was a lower ratio of chemicals that were active in the early stage process and the NFA, indicating that some of the overlap between the early stage processes and the NFA can be explained by cytotoxicity (Supplemental Figure 7). Together, these results indicate that many chemicals that perturbed early stage processes also disrupted network formation (selective or non-selective), however there was a subset of chemicals that were active in the NFA that were not active in any early stage process (Supplemental Results 5), indicating that: 1) network formation activity cannot necessarily predict perturbations in early stage processes, and 2) there are additional biological activities captured by network formation that are not captured by NOG/synaptogenesis.

Comparison of DNT NAMs to ToxCast/Tox21 data

Lastly, we compared DNT NAM activity to the activity observed in the broader ToxCast/Tox21 screening program, with the goal of identifying substances with neural-specific effects. The ToxCast/Tox21 database (as of invitrodb version 3.3, 2020 release) contains

1,569 assay endpoints tested in a variety of cell types and assay technologies (https://doi.org/10.23645/epacomptox.6062479.v5). This aim of this comparison was to reveal substances with selective bioactivity in the DNT NAM battery. Thirty-three chemicals that were highly cytotoxic in the DNT battery, as defined by positive hit-calls in at least 3 of 8 cytotoxicity assay endpoints in the DNT NAM battery, were identified. The minimum cytotoxicity log10-AC_{50} value from the DNT NAM battery was compared to the cytotoxicity burst threshold potency in ToxCast assays, a lower bound estimate for a concentration that might cause cell stress or cytotoxicity *in vitro* (Richard et al., 2016). In general, substances that were cytotoxic in the DNT battery were also cytotoxic in other ToxCast assays, with the exception of phosmet, lead(II) acetate trihydrate, ethoprop, and cytarabine (red arrows), which were not cytotoxic in available ToxCast assays (Figure 6A) as defined by the burst. Cytotoxicity was observed within < 1 log10-μM between the DNT NAM battery and ToxCast assays for 30/37 chemicals, suggesting that these highly cytotoxic chemicals in developing neurons are similarly cytotoxic in other cell types and/or assay platforms. The minimum DNT NAM cytotoxicity was more potent than the 5^{th} percentile ToxCast AC_{50} for seven chemicals (blue arrows: fluoxetine hydrochloride, heptachlor, methylmercury, dichlorvos, tebuconazole, terbufos, and bis(tributyltin)oxide) indicating that these chemicals are cytotoxic to neurons at concentrations that demonstrate little or no bioactivity in other cell-types and/or assay platforms. Moreover, the interquartile range and median for the DNT NAM log10-AC_{50} values (Figure 6A) indicate that some, if not all, DNT bioactivity can be explained by cytotoxic effects for these 37 cytotoxic chemicals. The minimum *selective* bioactive concentrations (log10-AC_{50} lower than cytotoxicity log10-AC_{50} values) in the DNT NAM battery were compared to the range and lower 5^{th} percentile of ToxCast log10-AC_{50} values (Figure 6B). Four chemicals (blue arrows: fluoxetine, isoniazid, methylmercury, and maneb) appeared to be more selectively potent in the DNT NAM battery compared to the 5^{th} percentile ToxCast AC_{50} (> 1 log10-μM), suggesting that these chemicals have strong selective effects on neuronal cell types in the absence of cytotoxicity. Moreover, 47/58 of the chemicals with selective DNT effects had a minimum DNT NAM AC_{50} value that was lower than the median ToxCast AC_{50} values, demonstrating that the DNT NAM battery captures potent, sensitive effects relative to general ToxCast activity.

Discussion

This work presented a novel integrated analysis of concentration-response data for 92 chemicals screened in a DNT NAM battery that evaluates key neurodevelopmental processes: proliferation, apoptosis, synaptogenesis, neurite outgrowth and maturation, and neural network formation and function (Brown, et al., 2016; Frank, et al., 2017; Harrill, et al., 2018; Shafer, et al., 2019). The DNT NAM assay suite together successfully functioned as a broad phenotypic screen of neurodevelopmental processes *in vitro*, provided insight into complex differential patterns of disrupted functional biology not captured by each assay alone, and demonstrated the need for multiple assay endpoints evaluating distinct neurodevelopmental functional processes to delineate positive and negative evaluation chemicals. Clustering analyses suggest the minimum log10-AC_{50} value in the DNT NAM battery may provide a sensitive marker of DNT bioactivity, and selective bioactivity may reveal specific functional processes impeded *in vitro*. The DNT NAM selectivity analysis

demonstrated high specificity (92%) and lower sensitivity (74%) in classifying DNT NAM evaluation chemicals, while non-selective activity demonstrated a lower specificity (69%) and a high sensitivity (93%). Finally, we identified chemicals that demonstrate more potent bioactivity in neuronal cell types compared to the broader ToxCast/Tox21 dataset, which includes a variety of non-neuronal cell types and technologies, potentially informing the relative importance of *in vitro* DNT bioactivity for chemicals with many bioactivities *in vitro*.

The DNT NAM battery demonstrated reliability, based on reproducibility with assay performance controls known to disrupt specific functional processes relevant to neurodevelopment, such as proliferation by aphidicolin, NOG initiation by lithium chloride, and network activity by bisindolylmaleimide I. The NFA demonstrated 75% qualitative concordance (repeated observation of active, equivocal, or negative behavior) for screening of replicate samples for 20 chemicals. Further, baseline sampling variability and qualitative concordance suggest signal can be distinguished from noise, and that positive, equivocal, or negative behavior in the DNT NAM battery assays is likely to be reproducible despite experimental changes and chemical sourcing differences over time.

We identified measures of network connectivity (mutual information) and cytotoxicity as the most sensitive assay endpoints in the DNT NAM battery on a potency basis. The mutual information parameter, a calculated measure of dependence between any two random electrodes in a well (Ball, et al., 2017), is by design a sensitive readout and its ranking as the overall most sensitive endpoint is unsurprising. However, potency for the DNT NAM battery occurs within a narrow concentration range (median ± SD of 0.919 ±1.05 log10-μM across the 92 chemicals). Feature importance from a random forest regression analysis for minimum potency ranked cytotoxicity, measures of decreased network connectivity and general activity, and a measure of neurite length loss in the NOG (rat) assay as most informative of minimum potency by chemical. This analysis suggests that cytotoxicity in neuronal cells is a sensitive marker of potentially DNT-relevant bioactivity and that different cell-types and species may inform minimum bioactive concentration in DNT NAMs.

Differences in methodology between assays may have affected assay sensitivity, including exposure duration and cell type: the MEA NFA used rat primary cortical neurons and involved a 12-day chemical exposure period with two doses, in contrast to most of the HCI assays which involved a single chemical exposure of 48 hours or less, with the exception of a 5 day exposure in the synaptogenesis assay. Harrill and colleagues (2011a, b) compared sensitivity of human hN2 neural cells and rat cortical neurons in the NOG assays and found that the human hN2 cells were more sensitive to NOG inhibitors. Moreover, the six cytotoxicity parameters in the DNT NAM battery measure different aspects of cellular health, such as number of nuclei, number of neurons, or lactate dehydrogenase. Given that the battery includes different cell-types and species, it will be critical to investigate the impact of cell-type sensitivity in future studies.

Of major interest was the similarity of chemicals thought to act on related neuronal targets with other chemicals that may demonstrate hierarchical cluster associations. The cluster with the highest selective activity values contained several metals, organochlorines, and

neurological drugs that perturbed the NFA and synaptogenesis and NOG. Excessive metal concentrations in the developing nervous system have been associated with impaired cell function and neurological diseases (Wright et al., 2007) (Malecki, 2001; Sadiq et al., 2012). Methylmercury exposure has been associated with deficits in neuronal migration, proliferation, and growth (Antunes Dos Santos et al., 2016; Choi, 1989; Radonjic et al., 2013), consistent with effects observed on NOG in human hN2 neural cells. Several drugs also demonstrated selectivity for synaptogenesis and NOG (hN2) perturbations, such as fluoxetine, an antidepressant drug that inhibits the selective reuptake of serotonin, which modulates differentiation, migration, axonal guidance, synaptogenesis and dendritic pruning in the developing brain (Gaspar et al., 2003; Haydon et al., 1984) (Getz et al., 2011; Xu et al., 2010). Fluoxetine was closely associated with another drug that targets serotonin, chlorpromazine, which also demonstrated strong selectivity for synaptogenesis and NOG (hN2). This association suggests the possibility of a relationship between the molecular target and DNT NAM bioactivity; however, several other chemicals in the dataset share similar receptor targets and were not closely associated. Haloperidol, an antipsychotic that targets the D2 receptor (Scalzo et al., 1989), decreased network formation activity and bursting activity, and was most strongly associated with deltamethrin, a pyrethroid that disrupts voltage-gated sodium channels (Soderlund et al., 2002). Using a limited dataset, a clear relationship between molecular target and DNT NAM selective bioactivity pattern for functional processes was not revealed.

A deeper dive into “mis-classified” chemicals using selective activity suggests the need for more experimentation on chemicals that can increase NFA signal; possible biological gaps in the DNT NAM battery; and, possible differences in the concentrations achievable *in vitro* versus *in vivo*. Only one putative negative was incorrectly identified as a positive in the selectivity analysis: isoniazid. However, with evidence that isoniazid disrupts neurotransmission *in vitro* (Carta et al., 2008; Hamada et al., 2020), isoniazid may not be an informative “true” negative. The selectivity analysis incorrectly classified 14 false negatives in the inactive cluster, while the non-selective activity analysis classified four false negatives. In total, five of the 14 false negatives in the selectivity analysis were chemicals active only at cytotoxic concentrations. Indeed, cytotoxicity has been identified as the most highly integrated ‘key event’ in a derived network of adverse outcome pathways for neurotoxicity (Spinu et al., 2019). We also examined whether any false negatives may be explained by activity in the MEA NFA ‘up’ endpoints which were excluded from the selectivity analysis, but only borderline and/or suspected noise were observed MEA NFA in the “up” direction for these false negative chemicals (see examples of this behavior in Supplemental Figure 9). Evidence with the chemicals screened to date in the MEA NFA failed to demonstrate that increases in the parameters measured can be observed; for example, chemicals that antagonize γ-Aminobutyric acid type A ($GABA_A$) receptors have been reported to increase mean firing rate in acute MEAs (Kosnik et al., 2020; Strickland et al., 2018{Cao, 2012 #334)} but do not increase mean firing rate in the MEA NFA, possibly due to longer exposures during network development.

Observed false negatives may have been due to gaps in model biology. Adenosine receptors may not be sufficiently expressed in these cell models to correctly detect caffeine (Johansson et al., 1997). Although naloxone, an antagonist of the μ-type opioid receptor (Wang et al.,

2007), was inactive in the DNT NAM battery, the receptor is likely present because another μ-type opioid receptor antagonist, loperamide, was highly active. Masjosthusmann and colleagues (2020) reported on DNT assays evaluating neural crest cell migration, neuronal differentiation, and oligodendrocyte differentiation that also identified nicotine as a false negative. Another false negative in the analysis herein, maneb, was active in the neural crest cell migration assay, identifying a possible gap in the key neurodevelopmental processes covered by the DNT NAM battery described here.

Additional explanations of the false negatives in this analysis include methodological limitations, such as chemical bioavailability *in vitro* and maximum screened concentrations. Most chemicals were screened at concentrations from 0.001 to 100 μM in half-log increments in the HCI assays and 0.03 to 30 μM in the NFA (Supplemental Table 1). The AEDs based on the maximum concentrations screened for sodium valproate, caffeine, maneb, and 5,5'-diphenylhydantoin were lower than or approached the doses associated with *in vivo* effects in studies of DNT, suggesting higher concentrations may have been required to observe DNT-relevant effects. Conversely, the doses needed to perturb DNT-relevant functional processes might be lower than those necessary to impact apical measures *in vivo*. Future work should include improved prediction of brain and serum concentrations based on *in vitro* assay concentrations to better contextualize these bioactive concentrations. Additionally, many chemicals screened in this work lacked the information needed for high-throughput toxicokinetic models to predict AEDs; in the future, with the availability of more data or structure-activity relationships to predict the needed modeling parameters, quantitative IVIVE approaches could be used before or after screening experiments to indicate whether the appropriate concentration range had been screened when *in vivo* data are available for comparison.

Comparing DNT NAM potency values to ToxCast/Tox21 data identified substances with selective DNT in neuronal cell types and possible gaps in the currently available targeted screening assays. Generally, highly cytotoxic substances in the DNT NAM battery were also cytotoxic in available assays in the ToxCast database. Chemicals with cytotoxicity in the DNT NAM assays (phosmet, lead(II) acetate trihydrate, ethoprop, and cytarabine), but not in available cytotoxicity assay data in ToxCast, may be chemicals of greater interest for putative DNT-related bioactivity. Moreover, the chemicals that demonstrated more potent 'selective' activity compared to ToxCast/Tox21 assays may be indicative of a targeted mode of action that was not previously captured.

Taken together, this novel computational analysis of the DNT NAM battery provides insight into the complex biology underlying differential patterns of DNT-relevant bioactivity and further informs the need for an integrated-multi-dimensional assay suite for DNT hazard identification. Importantly, this analysis demonstrated that the NFA and HCI assay technologies together, comprised of multiple assay endpoints evaluating key distinct functional processes in neuronal development, are necessary for the optimal classification of positive and negative DNT evaluation chemicals. Addition of DNT NAM assays to an integrated screening paradigm (Thomas et al., 2019) that uses neuronal cells, and is expanded to non-neuronal cell-types (e.g. microglia, astrocytes) representing key neurodevelopmental processes that are not currently included in this DNT NAM battery

(e.g. neural crest cell migration, myelination), may further improve classification of DNT NAM evaluation chemicals. Although selective activity in the DNT NAM battery may implicate specific functional processes perturbed, cytotoxicity in these neuronal cell lines was identified as a sensitive indicator of putative DNT-related activity, suggesting that relative cytotoxicity in neuronal cell types may help inform putative DNT hazard and derive sensitive points of departures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to acknowledge Drs. John Cowden and Joseph Bundy at the US Environmental Protection Agency and Dr. William Mundy (retired) for insightful comments on previous versions of this manuscript.

Funding

K.E.C. was supported by appointment to the Research Participation Program of the U.S. Environmental Protection Agency, Office of Research and Development, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. EPA. A.F.C was supported via an agreement administered by the Oak Ridge Associated Universities.

Abbreviations

AB	Alamar Blue
AC_{50}	concentration at 50% maximal activity
AED	administered equivalent dose
AUC	area under the curve
CV	coefficient of variation
DIV	day *in vitro*
DMSO	dimethyl sulfoxide
DNT	Developmental neurotoxicity
EPA	[United States] Environmental Protection Agency
FN	false negative
FP	false positive
$GABA_A$	γ-Aminobutyric acid type A
HCI	high content imaging
HED	human equivalent dose

hNP1	human neural progenitor cells derived from a neuroepithelial cell lineage of WA09 human embryonic stem cells, ArunA Biomedical (Athens, GA)
hN2	human neural cells (immature neurons differentiated from the hNP1 cells)
HTTK	high-throughput toxicokinetics
IVIVE	in vitro to in vivo extrapolation
LDH	lactate dehydrogenase
MEA	microelectrode array
NAM	new approach methodology
NOG	neurite outgrowth
NFA	network formation assay
Tcpl	ToxCast Pipeline
TN	true negative
TP	true positive

References:

Alfano DP, Petit TL, and LeBoutillier JC (1983). Development and plasticity of the hippocampal-cholinergic system in normal and early lead exposed rats. Brain Res 312(1), 117–24. [PubMed: 6686078]

Antunes Dos Santos A, Appel Hort M, Culbreth M, López-Granero C, Farina M, Rocha JB, and Aschner M (2016). Methylmercury and brain development: A review of recent literature. J Trace Elem Med Biol 38, 99–107. [PubMed: 26987277]

Aschner M, Ceccatelli S, Daneshian M, Fritsche E, Hasiwa N, Hartung T, Hogberg HT, Leist M, Li A, Mundi WR, et al. (2017). Reference compounds for alternative test methods to indicate developmental neurotoxicity (DNT) potential of chemicals: example lists and criteria for their selection and use. Altex 34(1), 49–74. [PubMed: 27452664]

Bal-Price A, Hogberg HT, Crofton KM, Daneshian M, FitzGerald RE, Fritsche E, Heinonen T, Hougaard Bennekou S, Klima S, Piersma AH, et al. (2018). Recommendation on test readiness criteria for new approach methods in toxicology: Exemplified for developmental neurotoxicity. Altex 35(3), 306–352. [PubMed: 29485663]

Ball KR, Grant C, Mundy WR, and Shafer TJ (2017). A multivariate extension of mutual information for growing neural networks. Neural Netw 95, 29–43. [PubMed: 28850900]

Barone S Jr., Haykal-Coates N, Parran DK, and Tilson HA (1998). Gestational exposure to methylmercury alters the developmental pattern of trk-like immunoreactivity in the rat brain and results in cortical dysmorphology. Brain Res Dev Brain Res 109(1), 13–31. [PubMed: 9706388]

Bell SM, Chang X, Wambaugh JF, Allen DG, Bartels M, Brouwer KLR, Casey WM, Choksi N, Ferguson SS, Fraczkiewicz G, et al. (2018). In vitro to in vivo extrapolation for high throughput prioritization and decision making. Toxicol In Vitro 47, 213–227. [PubMed: 29203341]

Bertrand R, Solary E, O'Connor P, Kohn KW, and Pommier Y (1994). Induction of a common pathway of apoptosis by staurosporine. Exp Cell Res 211(2), 314–21. [PubMed: 8143779]

Bray MA, Carpenter A, Imaging Platform, B. I. o. M. I. T., and Harvard (2004). Advanced Assay Development Guidelines for Image-Based High Content Screening and Analysis. In Assay Guidance Manual (Markossian S, Grossman A, Brimacombe K, Arkin M, Auld D, Austin CP, Baell J, Chung TDY, Coussens NP, Dahlin JL, et al., Eds.). Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda (MD).

Brown JP, Hall D, Frank CL, Wallace K, Mundy WR, and Shafer TJ (2016). Editor's Highlight: Evaluation of a Microelectrode Array-Based Assay for Neural Network Ontogeny Using Training Set Chemicals. Toxicol Sci 154(1), 126–139. [PubMed: 27492221]

Canfield RL, Henderson CR, Cory-Slechta DA, Cox C, Jusko TA, and Lanphear BP (2003). Intellectual impairment in children with blood lead concentrations below 10 microg per deciliter. N. Eng. J. Med 348, 1517–1526.

Cao Z, Hammock BD, McCoy M, Rogawski MA, Lein PJ, and Pessah IN (2012). Tetramethylenedisulfotetramine alters Ca2þ dynamics in cultured hippocampal neurons: mitigation by NMDA receptor blockade and GABA(A) receptorpositive modulation. Toxicol. Sci 130, 362–372. [PubMed: 22889812]

Carta M, Murru L, Barabino E, Talani G, Sanna E, and Biggio G (2008). Isoniazid-induced reduction in GABAergic neurotransmission alters the function of the cerebellar cortical circuit. Neuroscience 154(2), 710–9. [PubMed: 18456415]

Choi BH (1989). The effects of methylmercury on the developing brain. Prog Neurobiol 32(6), 447–70. [PubMed: 2664880]

Crofton KM, Mundy WR, Lein PJ, Bal-Price A, Coecke S, Seiler AE, Knaut H, Buzanska L, and Goldberg A (2011). Developmental neurotoxicity testing: recommendations for developing alternative methods for the screening and prioritization of chemicals. Altex 28(1), 9–15. [PubMed: 21311847]

Das KP, and Barone S Jr. (1999). Neuronal differentiation in PC12 cells is inhibited by chlorpyrifos and its metabolites: is acetylcholinesterase inhibition the site of action? Toxicol Appl Pharmacol 160(3), 217–30. [PubMed: 10544056]

Druwe I, Freudenrich TM, Wallace K, Shafer TJ, and Mundy WR (2015). Sensitivity of neuroprogenitor cells to chemical-induced apoptosis using a multiplexed assay suitable for high-throughput screening. Toxicology 333, 14–24. [PubMed: 25841707]

Feng G, and Kaplowitz N (2002). Mechanism of staurosporine-induced apoptosis in murine hepatocytes. Am J Physiol Gastrointest Liver Physiol 282(5), G825–34. [PubMed: 11960779]

Filer DL, Kothiya P, Setzer RW, Judson RS, and Martin MT (2017). tcpl: the ToxCast pipeline for high-throughput screening data. Bioinformatics 33(4), 618–620. [PubMed: 27797781]

Frank CL, Brown JP, Wallace K, Mundy WR, and Shafer TJ (2017). From the Cover: Developmental Neurotoxicants Disrupt Activity in Cortical Networks on Microelectrode Arrays: Results of Screening 86 Compounds During Neural Network Formation. Toxicol Sci 160(1), 121–135. [PubMed: 28973552]

Fritsche E, Grandjean P, Crofton KM, Aschner M, Goldberg A, Heinonen T, Hessel EVS, Hogberg HT, Bennekou SH, Lein PJ, et al. (2018). Consensus statement on the need for innovation, transition and implementation of developmental neurotoxicity (DNT) testing for regulatory purposes. Toxicol Appl Pharmacol 354, 3–6. [PubMed: 29447839]

Gaspar P, Cases O, and Maroteaux L (2003). The developmental role of serotonin: news from mouse molecular genetics. Nat Rev Neurosci 4(12), 1002–12. [PubMed: 14618156]

Getz A, Xu F, Zaidi W, and Syed NI (2011). The antidepressant fluoxetine but not citalopram suppresses synapse formation and synaptic transmission between Lymnaea neurons by perturbing presynaptic and postsynaptic machinery. Eur J Neurosci 34(2), 221–34. [PubMed: 21722207]

Gorman AM, Bonfoco E, Zhivotovsky B, Orrenius S, and Ceccatelli S (1999). Cytochrome c release and caspase-3 activation during colchicine-induced apoptosis of cerebellar granule cells. Eur J Neurosci 11(3), 1067–72. [PubMed: 10103099]

Grandjean P, and Landrigan PJ (2014). Neurobehavioural effects of developmental toxicity. Lancet. Neurol 13, 330–338. [PubMed: 24556010]

Grandjean P, Weihe P, White RF, Debes F, Araki S, Yokoyama K,Murata K, Sørensen N, Dahl R, and Jørgensen PJ (1997). Cognitive deficit in 7-year-old children with prenatal exposure to methylmercury. Neurotoxicol. Teratol 19, 417–428. [PubMed: 9392777]

Hamada Y, Figueroa C, Martín-Sánchez M, Falzon D, and Kanchar A (2020). The safety of isoniazid tuberculosis preventive treatment in pregnant and postpartum women: systematic review and meta-analysis. Eur Respir J 55(3).

Harrill JA, Freudenrich T, Wallace K, Ball K, Shafer TJ, and Mundy WR (2018). Testing for developmental neurotoxicity using a battery of in vitro assays for key cellular events in neurodevelopment. Toxicol Appl Pharmacol 354, 24–39. [PubMed: 29626487]

Harrill JA, Freudenrich TM, Machacek DW, Stice SL, and Mundy WR (2010). Quantitative assessment of neurite outgrowth in human embryonic stem cell-derived hN2 cells using automated high-content image analysis. Neurotoxicology 31(3), 277–90. [PubMed: 20188755]

Harrill JA, Freudenrich TM, Robinette BL, and Mundy WR (2011a). Comparative sensitivity of human and rat neural cultures to chemical-induced inhibition of neurite outgrowth. Toxicol Appl Pharmacol 256(3), 268–80. [PubMed: 21354195]

Harrill JA, Robinette BL, and Mundy WR (2011b). Use of high content image analysis to detect chemical-induced changes in synaptogenesis in vitro. Toxicol In Vitro 25(1), 368–87. [PubMed: 20969947]

Harrill JA, Robinette BL, Freudenrich T, and Mundy WR (2013). Use of high content image analyses to detect chemical-mediated effects on neurite sub-populations in primary rat cortical neurons. Neurotoxicology 34, 61–73. [PubMed: 23142577]

Haydon PG, McCobb DP, and Kater SB (1984). Serotonin selectively inhibits growth cone motility and synaptogenesis of specific identified neurons. Science 226(4674), 561–4. [PubMed: 6093252]

Johansson B, Georgiev V, Lindström K, and Fredholm BB (1997). A1 and A2A adenosine receptors and A1 mRNA in mouse brain: effect of long-term caffeine treatment. Brain Res 762(1-2), 153–64. [PubMed: 9262169]

Jones KL, Smith DW, Ulleland CN, and Streissguth P (1973). Pattern of malformation in offspring of chronic alcoholic mothers. Lancet. 1, 1267–1271. [PubMed: 4126070]

Kohno R, Ikeda Y, Yonemitsu Y, Hisatomi T, Yamaguchi M, Miyazaki M, Takeshita H, Ishibashi T, and Sueishi K (2006). Sphere formation of ocular epithelial cells in the ciliary body is a reprogramming system for neural differentiation. Brain Res 1093(1), 54–70. [PubMed: 16697356]

Kosnik MB, Strickland JD, Marvel SW, Wallis DJ, Wallace K, Richard AM, Reif DM, and Shafer TJ (2020). Concentration-response evaluation of ToxCast compounds for multivariate activity patterns of neural network function. Arch. Toxicol 94, 469–484. [PubMed: 31822930]

Kuhn M (2008). Building predictive models in R using the caret package. Journal of Statistical Software 28(5), 1–26. [PubMed: 27774042]

Laev H, Karpiak SE, Gokhale VS, and Hungund BL (1995). In utero ethanol exposure retards growth and alters morphology of cortical cultures: GM1 reverses effects. Alcohol Clin Exp Res 19(5), 1226–33. [PubMed: 8561295]

Lein P, Locke P, and Goldberg A (2007). Meeting report: alternatives for developmental neurotoxicity testing. Environ Health Perspect 115(5), 764–8. [PubMed: 17520065]

Malecki EA (2001). Manganese toxicity is associated with mitochondrial dysfunction and DNA fragmentation in rat primary striatal neurons. Brain Res Bull 55(2), 225–8. [PubMed: 11470319]

Masjosthusmann S, Blum J, Bartmann K, Dolde X, Holzer A-K, Stürzl L-C, Keßel EH, Förster N, Dönmez A, Klose J, et al. (2020). Establishment of an a priori protocol for the implementation and interpretation of an in-vitro testing battery for the assessment of developmental neurotoxicity. EFSA Supporting Publications 17(10), 1938E.

Mundy WR, Padilla S, Breier JM, Crofton KM, Gilbert ME, Herr DW, Jensen KF, Radio NM, Raffaele KC, Schumacher K, et al. (2015). Expanding the test set: Chemicals with potential to disrupt mammalian brain development. Neurotoxicol Teratol 52(Pt A), 25–35. [PubMed: 26476195]

Mundy WR, Radio NM, and Freudenrich TM (2010). Neuronal models for evaluation of proliferation in vitro using high content screening. Toxicology 270(2-3), 121–30. [PubMed: 20149836]

Nair AB, and Jacob S (2016). A simple practice guide for dose conversion between animals and human. J Basic Clin Pharm 7(2), 27–31. [PubMed: 27057123]

OECD (2007). Test No. 426: Developmental Neurotoxicity Study, OECD Guidelines for the Testing of Chemicals. Publishing, Paris 4, S., (Ed.).

Paul Friedman K, Watt ED, Hornung MW, Hedge JM, Judson RS, Crofton KM, Houck KA, and Simmons SO (2016). Tiered High-Throughput Screening Approach to Identify Thyroperoxidase Inhibitors Within the ToxCast Phase I and II Chemical Libraries. Toxicol Sci 151(1), 160–80. [PubMed: 26884060]

Pearce RG, Setzer RW, Strope CL, Wambaugh JF, and Sipes NS (2017). httk: R Package for High-Throughput Toxicokinetics. J Stat Softw 79(4), 1–26. [PubMed: 30220889]

Radonjic M, Cappaert NL, de Vries EF, de Esch CE, Kuper FC, van Waarde A, Dierckx RA, Wadman WJ, Wolterbeek AP, Stierum RH, et al. (2013). Delay and impairment in brain development and function in rat offspring after maternal exposure to methylmercury. Toxicol Sci 133(1), 112–24. [PubMed: 23457123]

Raffaele KC, Fisher JE Jr., Hancock S, Hazelden K, and Sobrian SK (2008). Determining normal variability in a developmental neurotoxicity test: a report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. Neurotoxicol Teratol 30(4), 288–325. [PubMed: 18280700]

Rauh VA, Garfinkel R, Perera FP, Andrews HF, Hoepner L, Barr DB, Whitehead R, Tang D, and Whyatt RW (2006). Impact of prenatal chlorpyrifos exposure on neurodevelopment in the first 3 years of life among inner-city children. Pediatrics. 118, e1845–e1859. [PubMed: 17116700]

Rice D, and Barone S Jr. (2000). Critical periods of vulnerability for the developing nervous system: evidence from humans and animal models. Environ Health Perspect 108 Suppl 3(Suppl 3), 511–33. [PubMed: 10852851]

Richard AM, Judson RS, Houck KA, Grulke CM, Volarath P, Thillainadarajah I, Yang C, Rathman J, Martin MT, Wambaugh JF, et al. (2016). ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. Chem Res Toxicol 29(8), 1225–51. [PubMed: 27367298]

Robinette BL, Harrill JA, Mundy WR, and Shafer TJ (2011). In vitro assessment of developmental neurotoxicity: use of microelectrode arrays to measure functional changes in neuronal network ontogeny. Front Neuroeng 4, 1. [PubMed: 21270946]

Sachana M, Bal-Price A, Crofton KM, Bennekou SH, Shafer TJ, Behl M, and Terron A (2019). International Regulatory and Scientific Effort for Improved Developmental Neurotoxicity Testing. Toxicol Sci 167(1), 45–57. [PubMed: 30476307]

Sachana M, Shafer TJ, and Terron A (2021). Toward a Better Testing Paradigm for Developmental Neurotoxicity: OECD Efforts and Regulatory Considerations. Biology (Basel) 10(2).

Sadiq S, Ghazala Z, Chowdhury A, and Büsselberg D (2012). Metal toxicity at the synapse: presynaptic, postsynaptic, and long-term effects. J Toxicol 2012, 132671. [PubMed: 22287959]

Scalzo FM, Ali SF, and Holson RR (1989). Behavioral effects of prenatal haloperidol exposure. Pharmacol Biochem Behav 34(4), 727–31. [PubMed: 2623028]

Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K, and Friedman KP (2019). Evaluation of Chemical Effects on Network Formation in Cortical Neurons Grown on Microelectrode Arrays. Toxicol Sci 169(2), 436–455. [PubMed: 30816951]

Sipes NS, Wambaugh JF, Pearce R, Auerbach SS, Wetmore BA, Hsieh JH, Shapiro AJ, Svoboda D, DeVito MJ, and Ferguson SS (2017). An Intuitive Approach for Predicting Potential Human Health Risk with the Tox21 10k Library. Environ Sci Technol 51(18), 10786–10796. [PubMed: 28809115]

Smirnova L, Hogberg HT, Leist M, and Hartung T (2014). Developmental neurotoxicity - challenges in the 21st century and in vitro opportunities. Altex 31(2), 129–56. [PubMed: 24687333]

Soderlund DM, Clark JM, Sheets LP, Mullin LS, Piccirillo VJ, Sargent D, Stevens JT, and Weiner ML (2002). Mechanisms of pyrethroid neurotoxicity: implications for cumulative risk assessment. Toxicology 171(1), 3–59. [PubMed: 11812616]

Spinu N, Bal-Price A, Cronin MTD, Enoch SJ, Madden JC, and Worth AP (2019). Development and analysis of an adverse outcome pathway network for human neurotoxicity. Arch Toxicol 93(10), 2759–2772. [PubMed: 31444508]

Strickland JD, Martin MT, Richard AM, Houck KA, and Shafer TJ (2018). Screening the ToxCast phase II libraries for alterations in network function using cortical neurons grown on multi-well microelectrode array (mwMEA) plates. Arch. Toxicol 92, 487–500. [PubMed: 28766123]

Thomas RS, Bahadori T, Buckley TJ, Cowden J, Deisenroth C, Dionisio KL, Frithsen JB, Grulke CM, Gwinn MR, Harrill JA, et al. (2019). The Next Generation Blueprint of Computational Toxicology at the U.S. Environmental Protection Agency. Toxicol Sci 169(2), 317–332. [PubMed: 30835285]

Tsuji R, and Crofton KM (2012). Developmental neurotoxicity guideline study: issues with methodology, evaluation and regulation. Congenit Anom (Kyoto) 52(3), 122–8. [PubMed: 22925212]

Uppuluri S, Knipling L, Sackett DL, and Wolff J (1993). Localization of the colchicine-binding site of tubulin. Proc Natl Acad Sci U S A 90(24), 11598–602. [PubMed: 8265596]

U.S. EPA. (1998a). Health Effects Test Guidelines OCSPP 870.6300 Developmental Neurotoxicity Study. Washington, DC.

U.S. EPA. (1998b). U.S. Environmental Protection Agency Guidelines for Neurotoxicity Risk Assessment. EPA/630/R-95/001F EPA/630/R-95/001F. Washington, DC.

U.S. EPA. (2020). Use of New Approach Methodologies to Derive Extrapolation Factors and Evaluate Developmental Neurotoxicity for Human Health Risk Assessment. Washington, DC.

Wambaugh JF, Hughes MF, Ring CL, MacMillan DK, Ford J, Fennell TR, Black SR, Snyder RW, Sipes NS, Wetmore BA, et al. (2018). Evaluating In Vitro-In Vivo Extrapolation of Toxicokinetics. Toxicol Sci 163(1), 152–169. [PubMed: 29385628]

Wang D, Sun X, and Sadee W (2007). Different effects of opioid antagonists on mu-, delta-, and kappa-opioid receptors with and without agonist pretreatment. J Pharmacol Exp Ther 321(2), 544–52. [PubMed: 17267582]

Ward JH (1963). Hierarchical Grouping to Optimize an Objective Function. J. Am. Stat. Assoc 58, 236–244.

Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, and Lumley T (2016). Package ‘gplots’: various r programming tools for plotting data. R package version 3.0. 1(Computer Software).

Wetmore BA, Wambaugh JF, Ferguson SS, Sochaski MA, Rotroff DM, Freeman K, Clewell HJ 3rd, Dix DJ, Andersen ME, Houck KA, et al. (2012). Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol Sci 125(1), 157–74. [PubMed: 21948869]

Whitney KD, Seidler FJ, and Slotkin TA (1995). Developmental neurotoxicity of chlorpyrifos: cellular mechanisms. Toxicol Appl Pharmacol 134(1), 53–62. [PubMed: 7545834]

Wright RO, and Baccarelli A (2007). Metals and neurotoxicology. J Nutr 137(12), 2809–13. [PubMed: 18029504]

Xu F, Luk C, Richard MP, Zaidi W, Farkas S, Getz A, Lee A, van Minnen J, and Syed NI (2010). Antidepressant fluoxetine suppresses neuronal growth from both vertebrate and invertebrate neurons and perturbs synapse formation between Lymnaea neurons. Eur J Neurosci 31(6), 994–1005. [PubMed: 20377614]

Zhang JH, Chung TD, and Oldenburg KR (1999). A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. J Biomol Screen 4(2), 67–73. [PubMed: 10838414]

Figure 1. Hierarchical clustering of log10-AC_{50} values for the DNT NAM battery.

Rows of the heatmap indicate chemical activity across activity type associated with the DNT NAM assay endpoints illustrated by a color annotation bar. Within the heatmap, more potent chemicals are indicated by increasingly dark blue, and yellow indicates a negative hit-call in the assay endpoint represented. Three main clusters are identified.

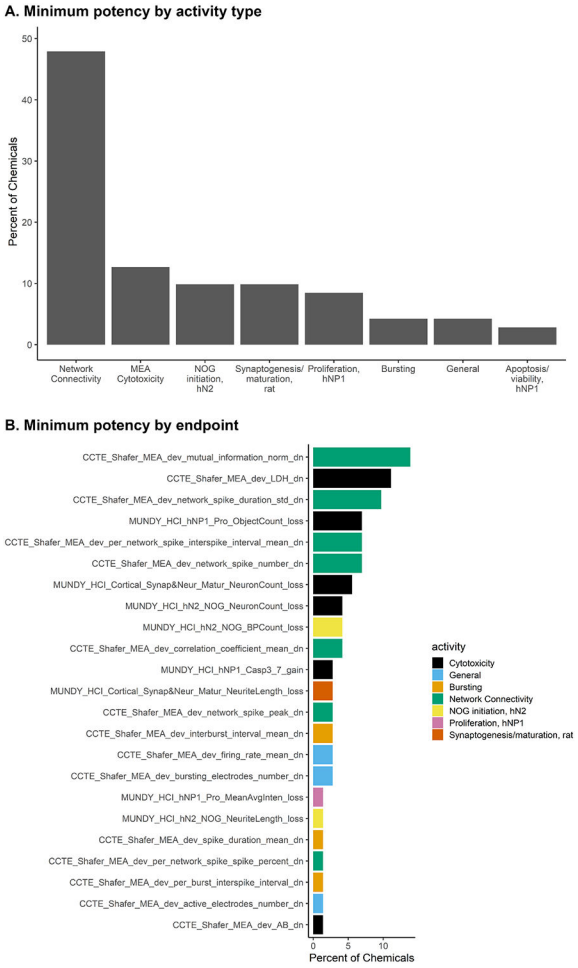

Figure 2. Sensitive endpoints in the DNT NAM battery.

(A) Bar plot indicates the percentage of chemicals for which the most sensitive endpoint (lowest AC_{50} value) corresponded to each activity type.

(B) Bar plot indicates the percentage of chemicals for which each assay endpoint was the most sensitive (lowest AC_{50} value).

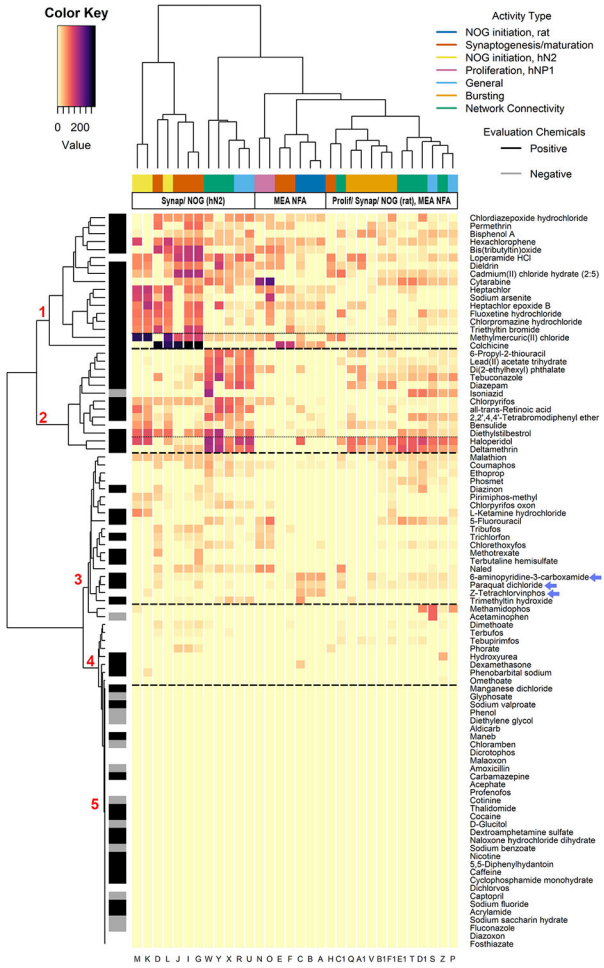

Figure 3. Selective activity in the DNT NAM battery.

Rows of the heatmap indicate chemical activity across activity type associated with the DNT NAM assay endpoints illustrated by a color annotation bar. Within the heatmap, yellow indicates a negative or non-selective AUC, whereas increasing pink to black color indicates increasing selective AUC values (more selective activity observed).

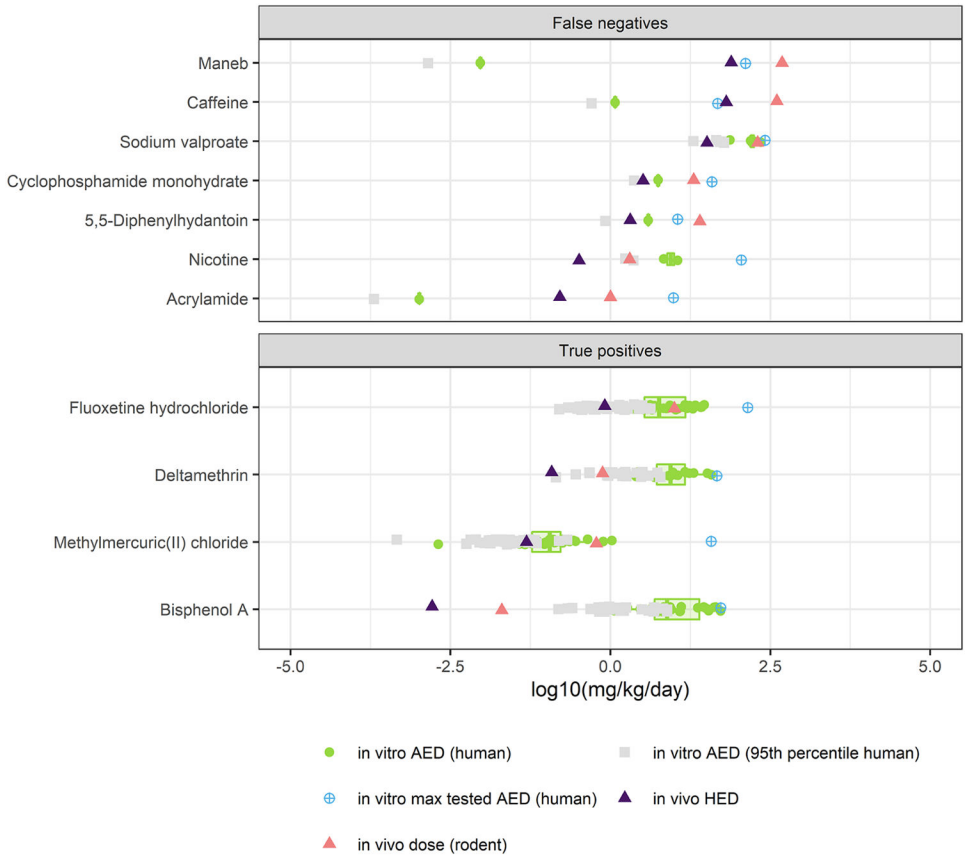

Figure 4. Comparison of in *vitro* AEDs to *in vivo* human equivalent doses of the false negatives

The bioactivity of the *in vitro* DNT NAM data was transformed to AEDs using HTTK data and models following the principles of reverse dosimetry. The *in vivo* LEL in rodents was transformed to a human equivalent dose (HED) using allometric scaling. The chemicals include a subset of seven false negative chemicals and four true positive chemicals with available HTTK data. The x-axis units are in log10 mg/kg/day.

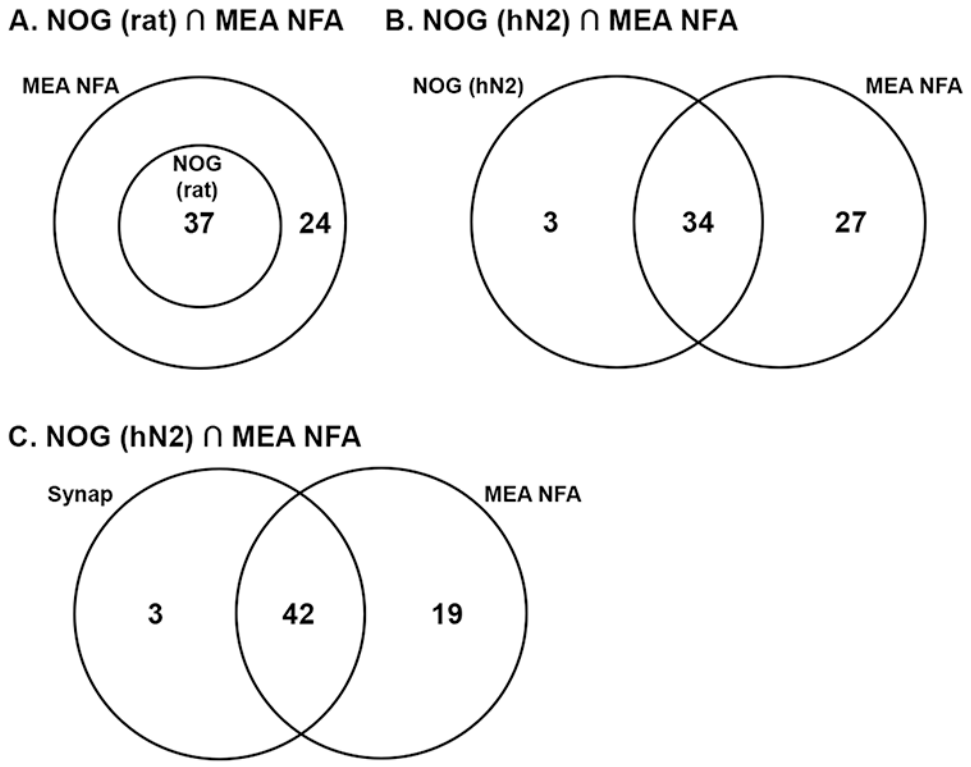

Figure 5. The intersect of activity in early stage processes and network formation

Venn diagrams indicate the intersect between chemicals that were active in the early stage process and in the NFA. A chemical was considered active if it was a hit in at least one endpoint in the assay, including cytotoxicity. The early stage processes include: A) NOG (in rat cortical neurons), B) NOG (in human neural cells), or C) synaptogenesis and neurite maturation (in rat cortical cells).

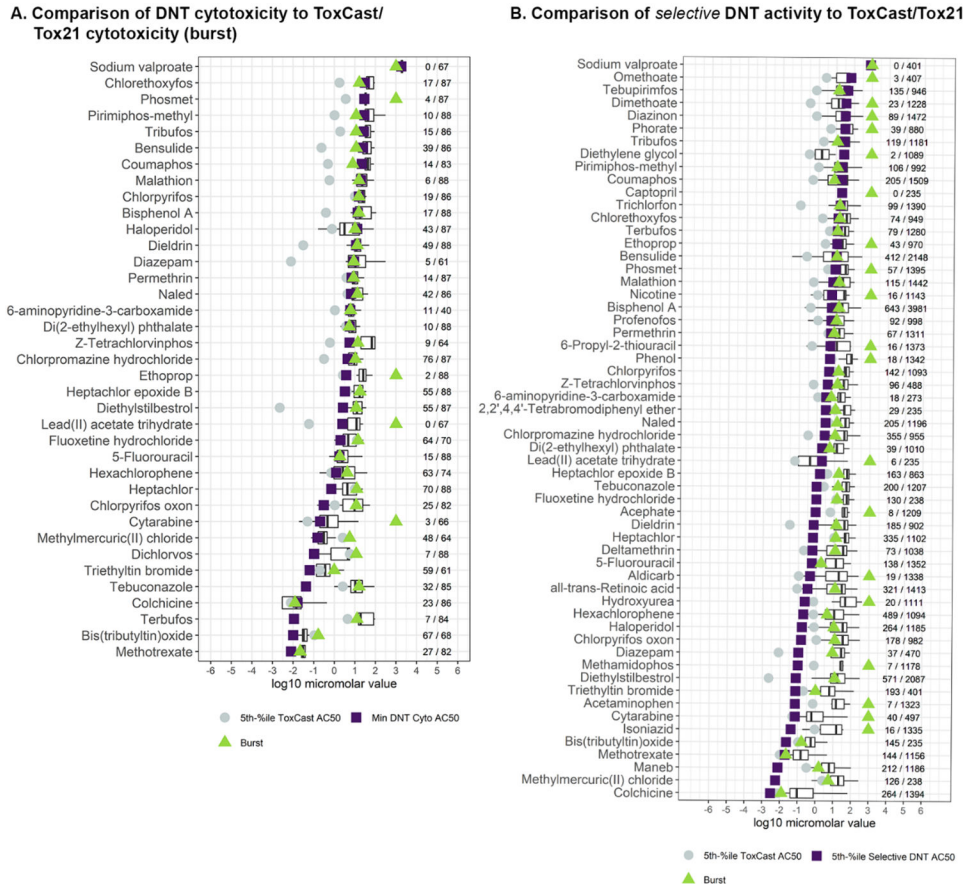

Figure 6. DNT NAM potency and ToxCast/Tox21 potency

(A) The 5th percentile AC_{50} value from ToxCast/Tox21 (gray circle), the minimum DNT cytotoxicity log10-AC_{50} (purple square), the cytotoxicity burst threshold (green triangle), and the interquartile range and median for the AC_{50} values in the DNT battery (boxplot). The number of positive burst assay endpoints over the total number of burst assay endpoints screened is provided to the right to indicate cytotoxicity hit-rate. The x-axis units are in log10-micromolar.

(B) The 5th percentile log10-AC_{50} value from ToxCast/Tox21 (gray circle), the selective (not including cytotoxicity hits) DNT log10-AC_{50} (purple square), the cytotoxicity burst threshold (green triangle), and the interquartile range and median for the log10-AC_{50} values in the ToxCast/Tox21 assay endpoints (boxplot). The number of positive ToxCast/Tox21 assay endpoints over the total number of ToxCast/Tox21 assay endpoints screened is provided to the right to indicate hit-rate. The x-axis units are in log10-micromolar.

Table 1.
Overview of Assay Endpoints Grouped by Activity Type.

The assay technology type (HCI or MEA), activity type, assay short name, cell type, functional process and endpoints evaluated, and then the expected possible assay response direction are briefly summarized.

Assay technology type	Activity Type	Assay short name	Cell type	Overview of functional processes and endpoints evaluated	Direction
HCI	Proliferation	MUNDY_HCI_hNP1_Pro	Human hNP1	Proliferation based on BrdU labeling, the intensity of this response, and number of nuclei per well (cytotoxicity).	Down
	Apoptosis	MUNDY_HCI_hNP1	Human hNP1	Apoptosis based on luminescent signal produced by detection of caspase 3/7 cleavage. The number of cells were measured based on the amount of ATP present in each well (cell viability).	Both
	Neurite outgrowth, hN2	MUNDY_HCI_hN2_NOG	Human hN2	Neurite outgrowth based on morphology of βII-tubulin labeled neurons. Measurements of neurite length, number of neurites, number of neurite branch points, and number of neurons per well (cytotoxicity).	Down
	Neurite outgrowth, rat	MUNDY_HCI_Cortical_NOG	Rat cortical		
	Synaptogenesis and neurite maturation	MUNDY_HCI_Cortical_Synap&Neur_Matur_	Rat cortical	Synaptogenesis based on morphology of MAP2 and synapsin labeled neurons. Measurements of neurite length, number of neurites, number of neurite branch points, the number of pre-synaptic puncta in the cell body compartment or the neurite compartment, the total number of synapses, the number of neurite-associated puncta, and number of neurons pre well (cytotoxicity).	Down
MEA	Network formation activity: **General**	CCTE_Shafer_MEA_dev	Rat cortical	General activity based on measurements: mean firing rate, burst rate, number of active electrodes, and number of bursting electrodes.	Both
	Network formation activity: **Bursting**	CCTE_Shafer_MEA_dev	Rat cortical	Bursting activity based on measurements: interspike interval within a burst, number of spikes within a burst, mean duration of a burst, and mean interval between bursts.	Both
	Network formation activity: **Network**	CCTE_Shafer_MEA_dev	Rat cortical	Network activity based on measurements: number of spikes in a network spike, number of electrodes active at peak of network spike, mean spike duration, SD of network spike duration, mean interspike interval for spikes in network spikes, number of spikes in network spike, percent of spikes in network spike (out of total spikes), mean correlation between all electrodes, and	Both

Assay technology type	Activity Type	Assay short name	Cell type	Overview of functional processes and endpoints evaluated	Direction
				mutual information between all electrodes.	
	MEA **Cytotoxicity**	CCTE_Shafer_MEA_dev	Rat cortical	Two measurements of cytotoxicity: alamarBlue ('_AB) which detects metabolically active cells and lactate dehydrogenase activity assay ('_LDH') which detects LDH released into the cell culture medium upon damage to the plasma membrane. Due to the half-life of LDH and the length of time between media changes, total cellular LDH was measured as an indicator of remaining cells at DIV12.	Down

Table 2.

Key for interpreting patterns of selective response in Figure 3.

The column label from Figure 3 and corresponding assay endpoint name are provided.

Column Label in Figure 3	Assay endpoint name
A	MUNDY_HCI_Cortical_NOG_BPCount_loss
B	MUNDY_HCI_Cortical_NOG_NeuriteCount_loss
C	MUNDY_HCI_Cortical_NOG_NeuriteLength_loss
D	MUNDY_HCI_Cortical_Synap&Neur_Matur_BPCount_loss
E	MUNDY_HCI_Cortical_Synap&Neur_Matur_CellBodySpotCount_loss
F	MUNDY_HCI_Cortical_Synap&Neur_Matur_NeuriteCount_loss
G	MUNDY_HCI_Cortical_Synap&Neur_Matur_NeuriteLength_loss
H	MUNDY_HCI_Cortical_Synap&Neur_Matur_NeuriteSpotCountPerNeuriteLength_loss
I	MUNDY_HCI_Cortical_Synap&Neur_Matur_NeuriteSpotCountPerNeuron_loss
J	MUNDY_HCI_Cortical_Synap&Neur_Matur_SynapseCount_loss
K	MUNDY_HCI_hN2_NOG_BPCount_loss
L	MUNDY_HCI_hN2_NOG_NeuriteCount_loss
M	MUNDY_HCI_hN2_NOG_NeuriteLength_loss
N	MUNDY_HCI_hNP1_Pro_MeanAvgInten_loss
O	MUNDY_HCI_hNP1_Pro_ResponderAvgInten_loss
P	CCTE_Shafer_MEA_dev_active_electrodes_number_dn
Q	CCTE_Shafer_MEA_dev_burst_duration_mean_dn
R	CCTE_Shafer_MEA_dev_burst_rate_dn
S	CCTE_Shafer_MEA_dev_bursting_electrodes_number_dn
T	CCTE_Shafer_MEA_dev_correlation_coefficient_mean_dn
U	CCTE_Shafer_MEA_dev_firing_rate_mean_dn
V	CCTE_Shafer_MEA_dev_interburst_interval_mean_dn
W	CCTE_Shafer_MEA_dev_mutual_information_norm_dn
X	CCTE_Shafer_MEA_dev_network_spike_duration_std_dn
Y	CCTE_Shafer_MEA_dev_network_spike_number_dn
Z	CCTE_Shafer_MEA_dev_network_spike_peak_dn
A1	CCTE_Shafer_MEA_dev_per_burst_interspike_interval_dn
B1	CCTE_Shafer_MEA_dev_per_burst_spike_percent_dn
C1	CCTE_Shafer_MEA_dev_per_network_spike_interspike_interval_mean_dn
D1	CCTE_Shafer_MEA_dev_per_network_spike_spike_number_mean_dn
E1	CCTE_Shafer_MEA_dev_per_network_spike_spike_percent_dn
F1	CCTE_Shafer_MEA_dev_spike_duration_mean_dn