FORUM

Screening for Developmental Neurotoxicity at the National Toxicology Program: The Future Is Here

Mamta Behl,[*,1] Kristen Ryan,[*] Jui-Hua Hsieh,[†] Frederick Parham,[*] Andrew J. Shapiro,[*] Bradley J. Collins,[*] Nisha S. Sipes,[*] Linda S. Birnbaum,[*] John R. Bucher,[*] Paul M. D. Foster,[*] Nigel J. Walker,[*] Richard S. Paules,[*] and Raymond R. Tice[‡]

[*]Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709; [†]Kelly Government Solutions, Durham, North Carolina 27709; and [‡]RTice Consulting, Hillsborough, North Carolina 27278

[1]To whom correspondence should be addressed at Division of the National Toxicology Program, 111 TW Alexander Drive, PO Box 12233, MD K2-12, Research Triangle Park, NC 27709. E-mail: mamta.behl@nih.gov.

ABSTRACT

The National Toxicology Program (NTP) receives requests to evaluate chemicals with potential to cause adverse health effects, including developmental neurotoxicity (DNT). Some recent requests have included classes of chemicals such as flame retardants, polycyclic aromatic compounds, perfluoroalkyl substances, and bisphenol A analogs with approximately 20–50 compounds per class, many of which include commercial mixtures. However, all the compounds within a class cannot be tested using traditional DNT animal testing guideline studies due to resource and time limitations. Hence, a rapid and biologically relevant screening approach is required to prioritize compounds for further in vivo testing. Because neurodevelopment is a complex process involving multiple distinct cellular processes, one assay will unlikely address the complexity. Hence, the NTP sought to characterize a battery of in vitro and alternative animal assays to quantify chemical effects on a variety of neurodevelopmental processes. A culmination of this effort resulted in a NTP-hosted collaborative project with approximately 40 participants spanning across domains of academia, industry, government, and regulatory agencies; collaborators presented data on cell-based assays and alternative animal models that was generated using a targeted set of compounds provided by the NTP. The NTP analyzed the assay results using benchmark concentration (BMC) modeling to be able to compare results across the divergent assays. The results were shared with the contributing researchers on a private web application during the workshop, and are now publicly available. This article highlights the overview and goals of the project, and describes the NTP's approach in creating the chemical library, development of NTPs data analysis strategy, and the structure of the web application. Finally, we discuss key issues with emphasis on the utility of this approach, and knowledge gaps that need to be addressed for its use in regulatory decision making.

Key words: developmental neurotoxicity; screening; in vitro; zebrafish.

This article is published as part of the NTP Neurotoxicology Screening Strategies Initiative.

BACKGROUND

The National Toxicology Program (NTP) receives requests for toxicological assessments on classes of chemicals such as flame retardants, bisphenol A (BPA) analogs, polycyclic aromatic compounds, and perfluorinated compounds. These nominations usually consist of 20–50 compounds per class and often include

commercial and isomeric mixtures. Among other toxicities, there is an increasing emphasis on examining the potential of these chemicals to cause developmental neurotoxicity (DNT) due to (i) their wide-spread exposure to sensitive populations including pregnant women and young children, and (ii) in some cases, their structural similarity to compounds known to be associated with DNT. An example of the latter is the similarity between the organophosphate (OP) flame retardants to some OP insecticides that are known DNTs (Makris, 2006). However, due to resource and time limitations, the NTP cannot evaluate the toxicity of most of the compounds within a class using traditional in vivo DNT guideline studies. Importantly, DNT guideline studies are usually conducted only when there is an a priori trigger, eg, clinical observations or histopathological changes in the brain noted from acute or subchronic studies, structural and/or use patterns of concern to known DNTs (such as extensive exposure of flame retardants with an OP backbone in children), or if there is suspected/known DNT. As a result, chemicals with unknown potential to cause DNT remain untested. Even in cases with in vivo DNT data, there are uncertainties in the current DNT test guidelines due to limitations with respect to sensitivity, reproducibility, and relevance to complex human diseases like autism or attention-deficit/hyperactivity disorder (ADHD) when extrapolating from rodent to humans (Bal-Price et al., 2015). This occurs primarily due to issues with respect to toxicokinetics, timing of exposure in brain development, use of functional tests that may not be as sensitive, and concerns that findings in rodents using guideline studies are not designed to capture many of the underlying biochemical or behavioral traits associated with these diseases.

Hence, there is a need to expand beyond traditional rodent studies to incorporate models that can screen for compounds rapidly and incorporate humanized cells/tissues. For a more efficient prioritization approach, the NTP has been evaluating a battery of medium-throughput, high-content assays that capture critical neurodevelopmental processes using in vitro cell-based and alternative animal models. The culmination of these efforts was presented in a recent NTP-hosted collaborative effort with the following goals: (i) evaluate emerging medium- to high-throughput and/or high-content cell-based in vitro and alternative animal (eg, zebrafish, planaria) model systems that screen for some aspect of DNT and might contribute to the development of a DNT screening battery, (ii) develop a data analysis approach to compare results from assays across diverse biologicals space, (iii) relate outcomes from these assays to current human exposure situations, and (iv) apply the battery approach in prioritizing NTP classes of compounds (and other chemicals) for further testing and/or in regulatory decision making.

The need for a DNT battery has been recognized for about a decade resulting in several important global efforts that have influenced progress in the field (Aschner et al., 2017; Bal-Price et al., 2012, 2018; Coecke et al., 2007; Crofton et al., 2011, 2012, 2014; Fritsche et al., 2018; Kadereit et al., 2012; Lein et al., 2007; Smirnova et al., 2014). The critical contribution of the NTP collaboration is that this was the first time that a unified data analysis pipeline has been applied across a common set of characterized chemicals that was provided to each investigator. Individual researchers tested the same set of chemicals in their respective cell-based or alternative animal assay that captured a unique aspect of neurodevelopment. The NTP then developed a data analysis pipeline that allowed for these assays to be combined to form a battery with improved biological coverage. The advantage of this approach is that the results from these assays can now be directly compared with each other to identify potential underlying brain development pathways that may be perturbed following chemical exposure. This information is now publicly available through NTP's, interactive web-application known as Developmental NeuroToxicity Data Integration and Visualization Enabling Resource (DNT-DIVER). This tool enables researchers to analyze, compare, and visualize results across multiple assays. Down-loadable data files are also available.

This article provides a brief description of NTP's chemical library, the data analysis approach, and salient features of the web application. It also addresses lessons learned with regards to data handling and design considerations, as well as knowledge gaps. Companion papers in this special issue highlight details on NTP's data analysis pipeline for alternative animal models (Hsieh et al.,2018), findings from collaborators on primary studies in in vitro alternative animal models (Dach et al., 2018; Quevedo et al., 2018; Zhang et al., 2018), how actives compare across some of these models (Hagstrom et al., 2018) and insight on how this approach may be used in regulatory decision making (Sachana et al., 2018).

OUR APPROACH

NTP's Chemical Library
Beginning in 2014, the NTP made available a diverse set of compounds for distribution via material transfer agreements to investigators conducting in vitro and alternative animal toxicological studies. The library included chemicals with published evidence of neurotoxicity (NT) or DNT, many of which were pesticides, drugs, and industrial chemicals. The library also contained compound classes of interest to the NTP (eg, flame retardants, polycyclic aromatic compounds, BPA analogs) and compounds reported to be negative/inactive in most developmental or acute NT assays. Additionally, it included several chemicals in replicate, to allow for the assessment of technical reproducibility within an experiment. The first NTP compound library contained 80 compounds (designated as NTP80); details on this list, including chemical name, source, purity, and literature supporting known chemical-specific NT, were published previously (Ryan et al., 2016). In 2016, 11 compounds (industrial chemicals, flame retardants, etc.) were added to the list to collectively generate the NTP91 compound library. The NTP91 library contains 87 unique compounds and 4 duplicate compounds to serve as technical replicates (hence sometimes also referred to NTP87). Some of the collaborators tested the NTP80 library while others the NTP91 library. General details regarding compound category, source, CASRN, and purity are provided for the NTP91 list in Supplementary Table 1. Supplementary Table 2 provides references for the 38 compounds of the list which are identified to have effects on DNT or general NT based on a literature review. Broad categories for the NTP91 compound library are presented in Figure 1A. Figure 1B provides the breadth of available in vivo data on the NTP91 compounds in terms of chronic, subchronic, carcinogenicity, acute, reproductive/developmental studies for comparison to (i) compound category or (ii) whether or not the compound is considered a DNT or NT as defined by the NTP. For all other categories, the NTP91 compound list was searched across the EPA Chemistry Dashboard (https://comptox.epa.gov) and the Leadscope Toxicity Database (Leadscope, Inc.) for evidence of whether the chemical has previously been tested for other toxicity domains.

Major features of the NTP library include: (i) all compounds were independently verified for identity and purity using
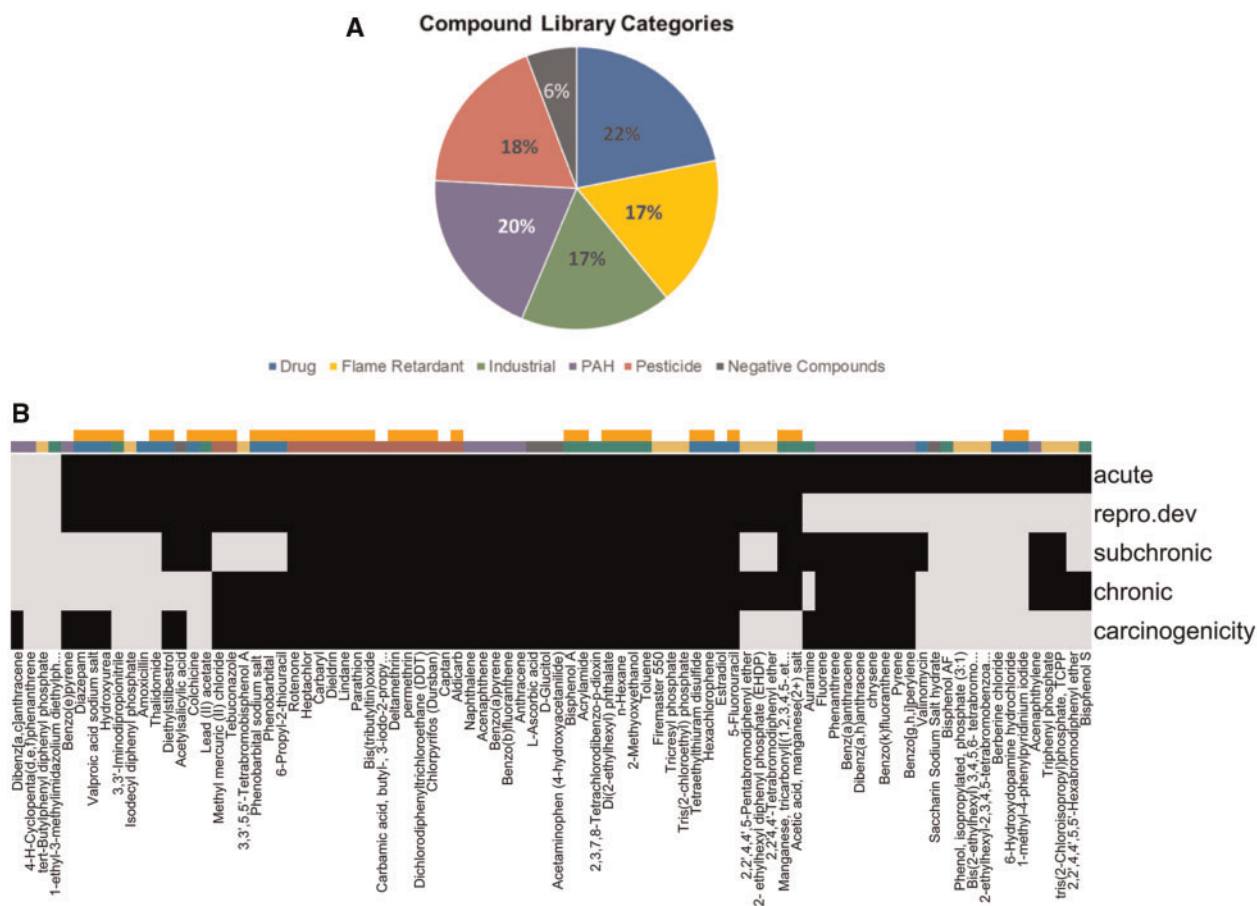
Figure 1. The NTP91 compound library. A, Pie chart of the NTP compound library, by use category (percent of total). As shown in the figure, we had a fairly uniform distribution of drugs (blue), flame retardants (yellow), industrial compounds (green), polycyclic aromatic hydrocarbons (PAH, purple), and pesticides (pink). B, Heat map identifying whether the compound has (black bars) or has not (light gray bars) been evaluated in a variety of systems toxicity assays. Data were derived from the EPA Chemistry Dashboard and Leadscope. Color bars indicate (i) if the compound has evidence of in vivo DNT or NT by NTP literature review (orange) and (ii) use category (eg, flame retardant [yellow]) repro. dev., reproductive and developmental.

standard chromatographic and mass spectrometry techniques prior to distribution; (ii) stock solutions at known concentrations were generated in bulk and shipped frozen in 96 well plates to the testing laboratories, ensuring that they received identical compound libraries; (iii) collaborators were provided the compounds in a blinded manner, and (iv) collaborators had access to a shared resource for questions they may have regarding handling and utilizing of compound plates.

The Screening Battery
Over the past several years, the NTP has been developing rapid approaches for screening large numbers of compounds for biological/toxicological activity, as part of the Tox21 federal partnership. In parallel, the NTP responds to requests for toxicological evaluation of classes of chemicals such as a group of organophosphorus flame retardants for which DNT, developmental toxicity (DT), and/or acute NT are considered a concern (Behl *et al.*, 2015, 2016; Glazer *et al.*, 2017; Jarema *et al.*, 2015).To expand on existing efforts, we developed a library containing a larger, diverse set of chemicals (Supplementary Tables 1 and 2) with assays representing varying levels of biological complexity covering domains of DT, DNT, and NT (Figure 2). Additional details on findings from individual assays with NTP's compound library can be found in recent publications (Delp *et al.*, 2018, Nyffeler *et al.*, 2017; Ryan *et al.*, 2016) and companion

manuscripts in this issue (Dach *et al.*, 2018; Hagstrom *et al.*, 2018; Sirenko *et al.*, 2018; Zhang *et al.*, 2018).

Briefly, the NTP evaluated a set of cell-based and alternative animal models that capture critical neurodevelopmental processes for potential incorporation into a screening battery. The cell-based assays included measures of neuronal proliferation, differentiation, neurite outgrowth (Ryan *et al.*, 2016), migration (Delp *et al.*, 2018; Nyffeler *et al.*, 2017), and neuronal network formation (Brown *et al.*, 2016; Frank *et al.*, 2017). All of these assays incorporated a metric that evaluated compound effects on cell viability in addition to effects on specific neurodevelopmental processes of interest. These assays have previously been shown to be in vitro models for critical stages of neurodevelopment and are perturbed by compounds that produce DNT in animals and humans (Aschner *et al.*, 2017; Harrill *et al.*, 2018; Mundy *et al.*, 2015; Nyffeler *et al.*, 2017). The alternative animal model assays including zebrafish (Geier *et al.*, 2018; Nishimura *et al.*, 2015; Raftery *et al.*, 2014) and planaria (Hagstrom *et al.*, 2015) were used to screen for DNT and NT, measuring general early development, motor activity (measurement of behavior), and mortality.

Data Analysis
Although the assays share a common goal of detecting chemicals with a potential to cause DNT, they are from 11
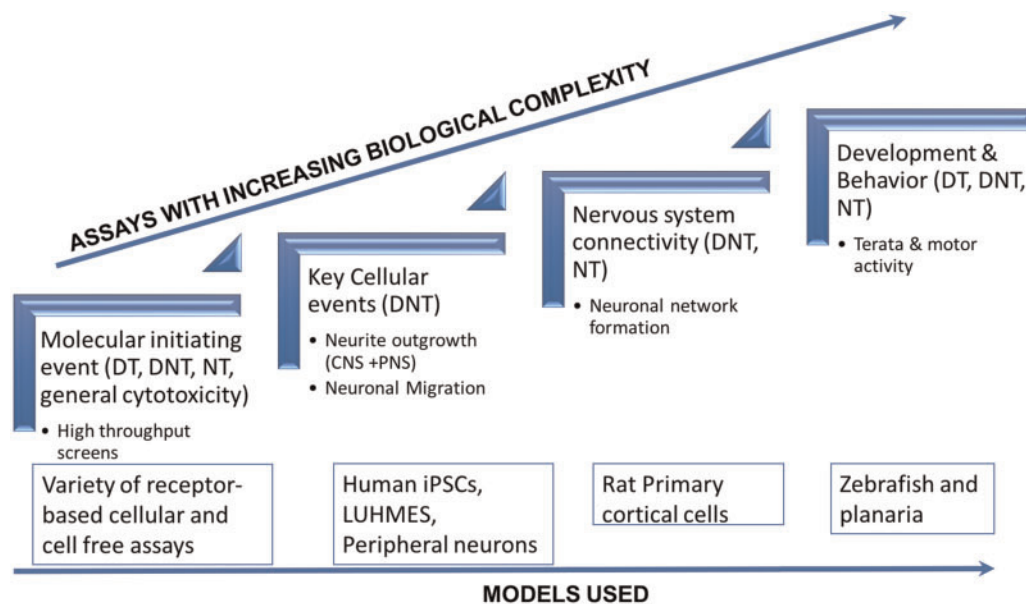
Figure 2. Schematic of assays included in the battery by increasing biological complexity. The figure shows the primary assays that were covered in the initial collaboration from left to right in increasing level of biological complexity. A total of 80 assays were evaluated as part of the Tox21 effort that covered receptor-based cellular assays including mitochondrial activity, stress response pathways, and general cytotoxicity. Additionally, 137 assays were assessed in cell-free and Novascreen models based on ToxCast data that included target genes related to genes of axon guidance or other axon parameter. As we move toward the right, the models covered functional aspects of key events that included neuronal differentiation, outgrowth and neural network formation in human-derived iPSC cells, immortalized cell lines, and rat primary cultures. Finally, to incorporate whole organisms in screening, we measured complex behavior (locomotor activity) and terata in zebrafish and planaria. DT, developmental toxicity; DNT, developmental neurotoxicity, and NT, neurotoxicity.

laboratories, collecting a variety of different data measurements, and each of which has a preferred data storage format suitable for in-house data analysis. Thus, the first challenge was to make the data directly comparable by incorporating a single unified structure. To accomplish this goal, we created a data ingest pipeline for "extraction, transformation, and loading" (ETL) of the data from the different sources. For each data source, intermediate files were created that captured one row per plate in one file, and one row per well in another file. Then, a series of Jupyter notebooks (ie, the pipeline) were written to process intermediate files and establish a relational database (REF). A Jupyter notebook is a free and open-source interactive file format which contains code and documentation, and is widely used in data science and reproducible science (http://jupyter.org). The resulting database contained over 100 000 distinct wells from plates and nearly 1 million individual responses.

Since the data share a standard format, the database allows for the upload of additional assays in the future. The results of the analysis can be traced back to the original raw data, providing much needed efficiency and transparency when analyzing diverse datasets. In addition, results from these analyses were saved in the database as well, building a system architecture which was subsequently used for the interactive web application.

A second challenge for the integrated analysis was designing a flexible yet comparable method to identify active chemicals and to quantify their potency in each assay. Current popular strategies to quantify the potency of actives use either an $EC_{X\%}$ (effective EC_x [concentration at X% of chemical's maximal effect] value), based on an analysis of concentration response data, or a LOAEL (lowest-observed-adverse-effect level value), based on pairwise statistical tests between responses at each concentration and responses of the control data. The $EC_{X\%}$ is

not desirable for comparing activity of chemicals which elicit varying maximal effect and the LOAEL is limited to the actual concentrations tested and does not provide an uncertainty factor for the reported potency. Therefore, instead of adopting either of these two strategies, we applied the benchmark concentration (BMC) modeling strategy, which was designed to tackle shortcomings of the LOAEL/$EC_{X\%}$ approach and has previously been used for modeling in vivo animal data for quantitative risk assessment (eg, benchmark dose [BMD] modeling for in vivo data). Both BMC approach and $EC_{X\%}$ approach consider the concentration-response relationship, but the meaning of the calculated potency is different. The $EC_{X\%}$ typically represents the potency at the response relative to the maximal response that the chemical elicits, whereas a BMC is usually a potency at which the response is equivalent to a user-defined benchmark response (BMR) relative to the background response rate. This concentration-response modeling-based strategy enables an objective and direct comparison of results between data sources generating a value that can serve as a point of departure (POD) in risk assessment.

Data Reporting and Visualization

The purpose of the collaboration was to enable comparisons across assays while allowing researchers to examine their assay-specific data. Enabling this level of flexibility would require either prohibitively large reports or highly summarized findings based on numerous assumptions. Therefore, we designed an interactive web application for data exploration and visualization, using the database created for data analysis (Figs. 3A–H). The key goals of the web application included being able to:

1. Compare the different assay parameters (such as testing concentration range and time points), which is useful when
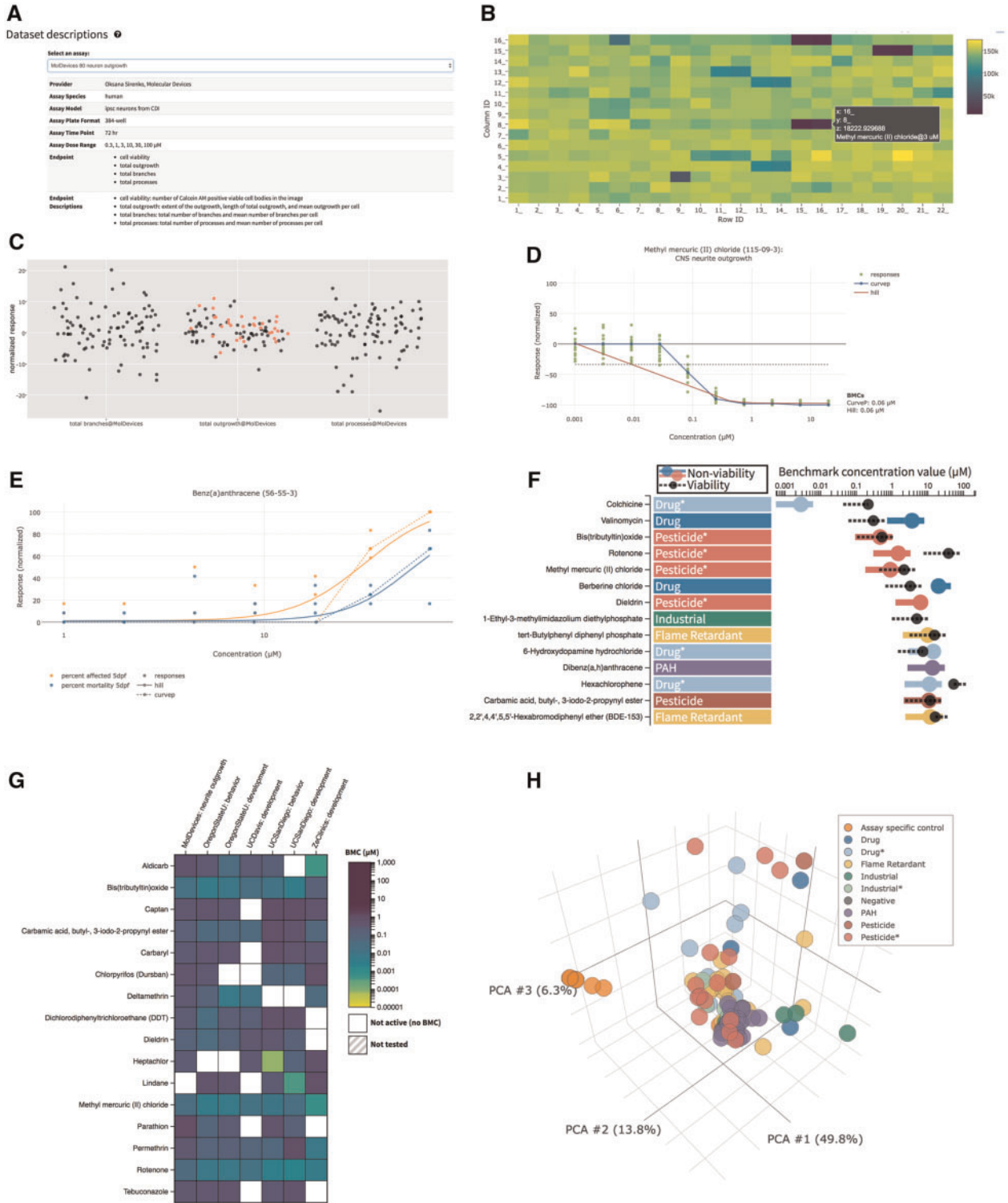
Figure 3. Salient features of web application DNT-DIVER. A, Description of each assay in the database, including model system, doses-tested, and endpoints measured. B, Plate map and response for each well on the plate. C, Exploration of vehicle control response variability; used for better understanding normalization techniques. D, Individual dose-response dataset and BMC curve fits. E, Summary of curve-fit data for multiple endpoints (here development and mortality endpoint are shown). F, Activity summary showing BMC estimate for both specific effect (colored point) and viability (black point). G, BMC heat map comparing chemicals and readouts from multiple laboratories; H, Principal component analysis (PCA) showing each chemical, colored by chemical class, in biological endpoint readout space (79 readouts reduced to 3 dimensions).

interpreting results across end-points and assays (Figure 3A).
2. Visualize vehicle response variability of each end-point in each data-set that was used to assign benchmark thresholds (Figs. 3B and 3C).
3. Visualize individual concentration-response curves (Figs. 3D and 3E).
4. Rank order chemicals for activity (as defined by a BMC being calculated for any effect including cell death) and selectivity (defined by a specific DNT effect that occurs at a concentration lower than that at which cytotoxicity is induced) (Figure 3F).
5. Compare effects of a specific chemical across assays or to compare the performance of classes of chemicals in a particular assay with options of a heat map view, boxplot view, or a principal component analysis view (Figs. 3G and 3H).
6. Download original data and BMC analysis via flat file formats or application programming interfaces (API).

The web application DNT-DIVER is now publicly available at https://sandbox.ntp.niehs.nih.gov/neurotox/ (last accessed December 14, 2018), and will be available for at least two-years post-publication of this article. The web application was designed to separate data (eg, assay results) from views (eg, tables, charts). To allow easy interactions with the underlying data in the database, a Representational State Transfer Application Programming Interface (REST API) was created using the Python Django web framework (Fielding, 2000). Views were implemented in two technologies: R-Shiny and JavaScript using d3.js and plotly (Bostock et al., 2011). By loosely coupling data from views, future assays or datasets can be integrated easily into the application, or the visuals can be repurposed for other projects. Components were containerized in Docker for reproducible deployment (Boettiger, 2015).

KNOWLEDGE GAPS AND LESSONS LEARNED

Based on our experience with different datasets, we have identified several areas for consideration while creating a battery of diverse assays, some of which have been reported previously (Bal-Price et al., 2018; Crofton et al., 2011, 2012; Fritsche et al., 2018; Kadereit et al., 2012).

A Systematic Approach to ETL and Data Analysis
Although our data analysis pipeline was designed for DNT assays, the fundamental problem that it attempts to solve is common in the field of toxicology: how to compare and contrast diverse data in biological and chemical space using a systematic framework. This is particularly difficult in a research environment where datasets are updated frequently, analyzes are modified, and final results and visualizations must be re-generated as upstream changes propagate downstream. The approach we took applied software-development principles, rather than a "one size fits all" solution which had a number of benefits:

1. The use of standardized intermediate file formats provide flexibility for data analysts to convert raw formats to re-usable formats.
2. The use of a series of Jupyter notebooks to combine the intermediate files to SQL-based data makes it easier to modify steps (ie, swap a Jupyter notebook) as the pipeline evolves.
3. The use of a RESTful API to confine the database data to essential building blocks for web application development allows more modular and re-usable visualizations.

The critical step for handling diverse datasets is to create a common schema for database importing. We believe our database schema can be used (with some modifications) for many other similar projects.

Application of BMC Modeling
There are a wide variety of approaches to data analysis from in vitro and alternative assays; however, there is no consensus on which of these approaches is most informative to decision makers. The BMC modeling approach is being more extensively used in in vitro datasets because it offers the advantage of a user-defined BMR relative to the background response thereby allowing for an objective comparison of results across data sources (Hsieh et al., 2015; Huang et al., 2011). Although BMC modeling provides a unified data analysis approach for comparing across datasets, there are still many details to be addressed including guidance for the selection of best-fitting model and/or model averaging, model parameterization, etc. We found that by normalizing the chemical responses using vehicle control responses (per plate) and linearly shifting baseline response to 0, it facilitated the streamline BMC modeling processes (see Hsieh et al., 2018 for further details). Normalization of chemical responses using vehicle control responses per plate is a common practice in in vitro medium- and high-throughput assays. However, it is oftentimes not used when analyzing data generated using alternative animal models. After the normalization procedure, the responses can be linearly shifted in order to set baseline response as 0. Setting baseline response to 0 has several advantages in BMC data analysis: first, it makes it easier to interpret the responses, where value > 0 (< 0) represents increased (decreased) effect relative to the vehicle control; second, it makes the responses in vehicle control comparable across datasets; third, it allows for a standardized BMR identification approach. Analyzing the response variation of vehicle control across plates (eg, standard deviation, SD) allows for a comparison of the background variation between assays. Setting a proper BMR is known as the critical step in vivo BMD modeling. Despite guidelines [BMD manual], we found that these may not be directly applicable to our in vitro DNT studies (Ryan et al., 2016). Therefore, as in Ryan et al., we adopted 3× SD of the vehicle control response as an activity threshold, a method that has previously been used in in vitro high throughput screening efforts (Hsieh et al., 2015; Huang et al., 2011). However, we also found that the 3× SD BMR approach is not applicable to the alternative animal model data, probably due to the nonnormal distribution of vehicle control responses. For these data, we applied a new approach that is currently under development for BMR identification. Details of this approach are discussed in our companion paper by Hsieh et al. (2018) using a sample zebrafish dataset. This approach can further be applied to in vitro data, an effort that the NTP is currently working on that will be discussed in a future publication.

The selectivity dilemma. For DNT studies, it is not only important to identify actives but also to prioritize actives for their primary DNT effect, separate from secondary outcomes (eg, cell death). These prioritized actives are considered to have higher "selectivity" than the remaining actives. In a previous related study on neurite outgrowth (Ryan et al., 2016), we created a selectivity score by using the ratio of the BMC values of the actives in both primary outcome (ie, neurite outgrowth) and secondary outcome (ie, cell viability) and applied a pre-defined cut-off to

identify selective actives. The cut-off was based on the testing concentration interval generally used in the assays. Although the choice is sensible, this approach does not take into account assay or chemical data variability. Therefore, instead of providing a pre-defined cut-off, we have allowed the user to filter the actives based on a user-defined selectivity cut-off. Currently, we are still exploring methods to define selectivity statistically (Hsieh *et al.*, 2018).

Expanding the Chemical Library

Although ours are the first chemical libraries to be distributed internationally among DNT research groups, they contain a relatively limited number of compounds, especially with respect to compounds that have known DNT/NT potential. The goal of this effort was to establish proof of concept using a combination of compounds that included known DNTs/NTs, NTP chemical classes of interest, and negative controls. The known DNTs/NTs were not further specifically categorized as DNT or NT because we were focusing on a more generalized prioritization strategy with emphasis on comparing data across assays. However, moving forward, it would be valuable to expand the current NTP91 compound library to include more DNT compounds to better characterize the sensitivity and specificity of these assays. The NTP is working with collaborators to identify additional appropriate positive and negative DNT chemicals for various assays, and creating a more comprehensive set of compounds for data generation (Sachana *et al.*, 2018).

Biological Coverage and Relevance

In addition to evaluating DNT, we also included a limited number of assays spanning various domains of DT and NT in this test battery. In this preliminary effort, we did not distinguish between compounds which specifically caused DNT (vs NT) because our main goal was to prioritize compounds for further in vivo testing; chemicals identified as actives in these screens will likely be tested in NTP's modified one generation rodent study, during which specific effects on DT versus DNT and NT can be further evaluated (Foster, 2014).

There was a consensus among collaborators that while identifying complete coverage representing all possible pathways by which DNT may occur is not necessary (or possible) for an initial screening evaluation, a thorough assessment of how biologically representative the test battery is would help build confidence in outcomes. For example, there are several DNT assays such as 2D and 3D mixed cell cultures including "brain on a chip" that incorporate microfluidics (Koo *et al.*, 2018) were not represented in this battery due to throughput limitations, and/or are still in developmental or optimization stages. The NTP is currently working to evaluate these model systems and are anticipated to be powerful second-tier screens, due to their ability to inform on the underlying biology and/or adverse outcome pathways of a chemical. Other examples of such models include in vitro models that evaluate the role of blood brain barrier, metabolic capability, and reflect genetic diversity.

For alternative animal models, there was a high degree of regulatory interest and emphasis on expanding beyond early activity in zebrafish embryos to developing assays that more closely mimic critical behavioral aspects of DNT guideline studies (eg, motor activity, motor and sensory response, learning, and memory). To incorporate such models in future screens, efforts at the NTP are already underway where zebrafish are exposed to chemicals during early development and evaluated for neurobehavioral outcomes such as motor activity, startle, learning and memory, and anxiety later in life (Glazer *et al.*, 2017,

2018). Another topic of interest under current investigation is the need to understand performance in multiple strains of fish to be able to improve confidence in the data generated. Finally, there was consensus for the need for a global harmonization effort to ensure that standardized terminology is used across researchers to define and describe their findings. To further address some of these concerns, the NTP has a program-wide ongoing effort on the Systematic Evaluation of the Application of Zebrafish in Toxicology (SEAZIT).

Finally, we recognize that in addition to covering key events of neurodevelopment, it is important to define the biological plausibility and relevance (eg, metabolism), and to recognize limitations of the in vitro and alternative animal assays because there are no direct links to endpoints in humans. Some of these questions are being addressed using adverse outcome pathway approaches (Bal-Price and Meek, 2017).

Need for better communication and collaboration. One of the key, yet often overlooked, factors critical to the success of advancing the DNT/NT (or any) field is the ability to communicate and collaborate across disciplines. During this effort, there was extensive interaction between researchers with different scientific backgrounds, data scientists, software developers, toxicologists, and statisticians. Although this diversity of expertise resulted in a steep learning curve for most of the participants, the willingness of the participants to be patient, to learn, and to contribute made for a much richer and more successful endeavor. Additionally, NTP's DNT-DIVER website provides transparent and effective mode of communication for researchers to compare and contrast their data globally.

CLOSING THOUGHTS: THE FUTURE IS HERE

Over the past decade, there has been increased recognition for the need to develop more and better rapid screening methods to identify compounds with a potential for DNT/NT (Aschner *et al.*, 2017; Bal-Price *et al.*, 2012, 2018; Coecke *et al.*, 2007; Crofton *et al.*, 2008, 2011, 2012, 2014; Fritsche *et al.*, 2018; Kadereit *et al.*, 2012; Lein *et al.*, 2007; Smirnova *et al.*, 2014). There is a pressing need to demonstrate the utility, and limitations of these approaches so that they can be put into effect as soon as possible, especially in the context of prioritizing chemicals for further in-depth evaluation to help protect public health. Currently, it can take over a decade for regulations to be put into effect from the time a compound is identified as potentially neurotoxic due to the rigor required for decision making. In the interim, there is a global rise in neurodevelopmental disorders and susceptible populations continue to be exposed (EPA, 2015), which drives an increased public concern about the potential contributions of chemicals to these increases that cannot be addressed in the absence of data. Although we need to constantly challenge ourselves to improve our understanding of knowledge gaps to define the DNT space more comprehensively, we strongly believe, that in the interim, batteries such as these can be put into effect to provide useful information on hazard identification, prioritization, and regulatory—decision making for a more timely protection of public health. In fact, at the NTP, we have already started implementing such a battery to prioritize a class of flame retardants for further in vivo testing (Behl *et al.*, 2015). Regulatory and scientific efforts are underway to discuss how batteries such as these can be used by OECD and regulatory agencies by providing case examples of Integrated Approaches to Testing and Assessment (IATAs) based on readiness criteria (Fritsche, 2016; Sachana *et al.*, 2018).

DATA AVAILABILITY

To access the normalized data and BMCs descried in this paper and visualized on NTP's interactive web-application, DNT-DIVER, go to: https://doi.org/10.22427/NTP-DATA-002-00062-0001-0000-1. Last accessed December 14, 2018.

SUPPLEMENTARY DATA

Supplementary data are available at *Toxicological Sciences* online.

ACKNOWLEDGMENTS

FUNDING

REFERENCES

Aschner, M., Ceccatelli, S., Daneshian, M., Fritsche, E., Hasiwa, N., Hartung, T., Hogberg, H. T., Leist, M., Li, A., Mundi, W. R., *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**, 49–74.

Bal-Price, A., and Meek, M. E. B. (2017). Adverse outcome pathways: Application to enhance mechanistic understanding of neurotoxicity. *Pharmacol. Ther.* **179**, 84–95.

Bal-Price, A., Coecke, S., Costa, L., Crofton, K. M., Fritsche, E., Goldberg, A., Grandjean, P., Lein, P. J., Li, A., Lucchini, R., *et al.* (2012). Advancing the science of developmental neurotoxicity (DNT): Testing for better safety evaluation. *ALTEX* **29**, 202–215.

Bal-Price, A., Crofton, K. M., Leist, M., Allen, S., Arand, M., Buetler, T., Delrue, N., FitzGerald, R. E., Hartung, T., Heinonen, T., *et al.* (2015). International STakeholder NETwork (ISTNET): Creating a developmental neurotoxicity (DNT) testing road map for regulatory purposes. *Arch. Toxicol.* **89**, 269.

Bal-Price, A., Hogberg, H. T., Crofton, K. M., Daneshian, M., FitzGerald, R. E., Fritsche, E., Heinonen, T., Hougaard Bennekou, S., Klima, S., Piersma, A. H., *et al.* (2018). Recommendation on test readiness criteria for new approach methods in toxicology: Exemplified for developmental neurotoxicity. *ALTEX* 35(3):306–352.

Behl, M., Hsieh, J.-H., Shafer, T. J., Mundy, W. R., Rice, J. R., Boyd, W. A., Freedman, J. H., Hunter, E. S., Jarema, K. A., Padilla, S., *et al.* (2015). Use of alternative assays to identify and prioritize organophosphorus flame retardants for potential developmental and neurotoxicity. *Neurotoxicol. Teratol.* **52**, 181–193.

Behl, M., Rice, J. R., Smith, M. V., Co, C. A., Bridge, M. F., Hsieh, J. H., Freedman, J. H., and Boyd, W. A. (2016). Editor's highlight: Comparative toxicity of organophosphate flame retardants and polybrominated diphenyl ethers to *Caenorhabditis elegans*. *Toxicol. Sci.* **154**, 241–252.

Boettiger, C. (2015). An introduction to Docker for reproducible research. *ACM SIGOPS Operat. Syst. Rev.* **49**, 71–79.

Bostock, M., Ogievetsky, V., and Heer, J. (2011). D3: Data-driven documents. *IEEE Trans. Vis. Comput. Graph* **17**, 2301–2309.

Brown, J. P., Hall, D., Frank, C. L., Wallace, K., Mundy, W. R., and Shafer, T. J. (2016). Editor's highlight: Evaluation of a micro-electrode array-based assay for neural network ontogeny using training set chemicals. *Toxicol. Sci.* **154**, 126–139.

Coecke, S., Goldberg, A. M., Allen, S., Buzanska, L., Calamandrei, G., Crofton, K., Hareng, L., Hartung, T., Knaut, H., Honegger, P., *et al.* (2007). Workgroup report: Incorporating in vitro alternative methods for developmental neurotoxicity into international hazard and risk assessment strategies. *Environ. Health Perspect.* **115**, 924–931.

Crofton, K. M., Mundy, W. R., and Shafer, T. J. (2012). Developmental neurotoxicity testing: A path forward. *Congenit. Anom.* **52**, 140–146.

Crofton, K. M., Mundy, W. R., Lein, P. J., Bal-Price, A., Coecke, S., Seller, A. E., 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**, 9–15.

Crofton, K., Fritsche, E., Ylikomi, T., *et al.* (2014). International STakeholder NETwork (ISTNET) for creating a developmental neurotoxicity testing (DNT) roadmap for regulatory purposes. *ALTEX* **31**, 223–224.

Dach, K., Yaghoobi, B., Schmuck, M.R, Carty, D. R., Morales, K. K., Harvey, D. J., and Lein, P. J. (2018) Teratological and Behavioral Screening of the National Toxicology Program 91-Compound Library in Zebrafish (*Danio rerio*). *Toxicol. Sci.* **167**, 77–91.

Delp, J., Gutbier, S., Klima, S., Hoelting, L., Pinto-Gil, K., Hsieh, J. H., Aichem, M., Klein, K., Schreiber, F., Tice, R. R., *et al.* (2018). A high-throughput approach to identify specific neurotoxicants/developmental toxicants in human neuronal cell function assays. *ALTEX* 35(2):235–253.

EPA. (2015). Neurodevelopmental disorders. https://www.epa.gov/sites/production/files/2015-10/documents/ace3_neurodevelopmental.pdf. Accessed December 13, 2018.

Fielding, R. T. (2000). Architectural styles and the design of network-based software architectures. Doctoral dissertation, University of California, Irvine.

Foster, P. M. (2014). Regulatory forum opinion piece: New testing paradigms for reproductive and developmental toxicity – The NTP Modified One generation study and OECD 443. *Toxicol. Pathol.* **42**, 1165–1167.

Frank, C. L., Brown, J. P., Wallace, K., Mundy, W. R., and Shafer, T. J. (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**, 121–135.

Fritsche, E., Crofton, K.M., Hernandez, A.F., Hougaard Bennekou, S., Leist, M., Bal-Price, A., Reaves, E., Wilks, M.F., Terron, A., Solecki, R., Sachana, M., Gourmelon, A. (2017). OECD/EFSA workshop on developmental neurotoxicity (DNT): The use of non-animal test methods for regulatory purposes. *ALTEX* 34(2):311–315.

Fritsche, E., Grandjean, P., Crofton, K. M., Aschner, M., Goldberg, A., Heinonen, T., Hessel, E. V. S., Hogberg, H. T., Bennekou, S. H., Lein, P. J., *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.

Geier, M. C., James Minick, D., Truong, L., Tilton, S., Pande, P., Anderson, K. A., Teeguardan, J., and Tanguay, R. L. (2018). Systematic developmental neurotoxicity assessment of a representative PAH Superfund mixture using zebrafish. *Toxicol. Appl. Pharmacol.* **18**, 30121–30122.

Glazer, L., Wells, C. N., Drastal, M., Odamah, K. A., Galat, R. E., Behl, M., and Levin, E. D. (2017). Developmental exposure to low concentrations of two brominated flame retardants, BDE-47 and BDE-99, causes life-long behavioral alterations in zebrafish. *Neurotoxicology* Advance Access published September 19, 2017. pii: S0161-813X(17)30196-1. doi: 10.1016/j.neuro.2017.09.007.

Glazer, L., Wells, C. N., Drastal, M., Odamah, K. A., Galat, R. E., Behl, M., and Levin, E. D. (2018). Developmental exposure to low concentrations of two brominated flame retardants, BDE-47 and BDE-99, causes life-long behavioral alterations in zebrafish. *Neurotoxicology* **66**, 221–232.

Grandjean, P., and Landrigan, P. J. (2014). Neurobehavioural effects of developmental toxicity. *Lancet Neurol.* **13**, 330–338.

Hagstrom, D., Cochet-Escartin, O., Zhang, S., Khuu, C., and Collins, E. M. (2015). Freshwater planarians as an alternative animal model for neurotoxicology. *Toxicol. Sci.* **147**, 270–285.

Hagstrom, D., Truong, L., Zhang, S., Tanguay, R., and Collins, E. M. (2018). Comparative analysis of zebrafish and planarian model systems for developmental neurotoxicity screens using an 87-compound library, 1–11.

Harrill, J. A., Freudenrich, T., Wallace, K., Ball, K., Shafer, T. J., and Mundy, W. R. (2018). Testing for developmental neurotoxicity using a battery of in vitro assays for key cellular events in neurodevelopment. *Toxicol. Appl. Pharmacol.* 354:24–39.

Hsieh, J.-H., Ryan, K., Sedykh, A., Lin, J.-A., Shapiro, A. J., Parham, F., and Behl, M. (2018). Application of benchmark concentration (BMC) analysis on zebrafish data – A new perspective for quantifying toxicity in alternative animal models. *Toxicol. Sci.* **167**, 92–104.

Hsieh, J. H., Sedykh, A., Huang, R., Xia, M., and Tice, R. R. (2015). A data analysis pipeline accounting for artifacts in Tox21 quantitative high-throughput screening assays. *J. Biomol. Screening* **20**, 887–897.

Huang, R., Xia, M., Cho, M.-H., Sakamuru, S., Shinn, P., Houck, K. A., Dix, D. J., Judson, R. S., Witt, K. L., Kavlock, R. J., *et al.* (2011). Chemical genomics profiling of environmental chemical modulation of human nuclear receptors. *Environ. Health Perspect.* **119**, 1142–1148.

Jarema, K. A., Hunter, D. L., Shaffer, R. M., Behl, M., and Padilla, S. (2015). Acute and developmental behavioral effects of flame retardants and related chemicals in zebrafish. *Neurotoxicol. Teratol.* **52**, 194–209.

Kadereit, S., Zimmer, B., van Thriel, C., *et al.* (2012). Compound selection for in vitro modeling of developmental neurotoxicity. *Front. Biosci.* **17**, 2442–2460.

Koo, Y., Hawkins, B. T., and Yun, Y. (2018). Three-dimensional (3D) tetra-culture brain on chip platform for organophosphate toxicity screening. *Sci. Rep.* **8**, 2841.

Lein, P., Locke, P., and Goldberg, A. (2007). Meeting report: Alternatives for developmental neurotoxicity testing – Test-Smart developmental neurotoxicology. *Environ. Health Perspect.* **115**, 764–768.

Makris, S. (2006). Regulatory considerations in developmental neurotoxicity of organophosphorus and carbamate pesticides. In *Toxicology of Organophosphate & Carbamate Compounds*, pp. 633–641. doi: 10.1016/B978-012088523-7/50044-2. Accessed December 13, 2018.

Mundy, W. R., Padilla, S., Breier, J. M., Crofton, K. M., Gilbert, M. E., Herr, D. W., Jensen, K. F., Radio, N. M., Raffaele, K. C., Schumacher, K., *et al.* (2015). Expanding the test set: Chemicals with potential to disrupt mammalian brain development. *Neurotoxicol. Teratol.* **52**, 25–35.

Nishimura, Y., Murakami, S., Ashikawa, Y., Sasagawa, S., Umemoto, N., Shimada, Y., and Tanaka, T. (2015). Zebrafish as a systems toxicology model for developmental neurotoxicity testing. *Congenit. Anom.* **55**, 1–16.

Nyffeler, J., Dolde, X., Krebs, A., Pinto-Gil, K., Pastor, M., Behl, M., Waldmann, T., and Leist, M. (2017). Combination of multiple neural crest migration assays to identify environmental toxicants from a proof-of-concept chemical library. *Arch. Toxicol.* **91**, 3613–3632.

Quevedo, C., Behl, M., Ryan, K., Alday, A., Muriana, M., and Alzualde, A. (2018). Detection and prioritization of developmentally neurotoxic and/or neurotoxic compounds using zebrafish. *Toxicol. Sci.* doi: 10.1093/toxsci/kfy291.

Raftery, T. D., Isales, G. M., Yozzo, K. L., and Volz, D. C. (2014). High-content screening assay for identification of chemicals impacting spontaneous activity in zebrafish embryos. *Environ. Sci. Technol.* **48**, 804–810.

Ryan, K. R., Sirenko, O., Parham, F., Hsieh, J. H., Cromwell, E. F., Tice, R. R., and Behl, M. (2016). Neurite outgrowth in human induced pluripotent stem cell-derived neurons as a high-throughput screen for developmental neurotoxicity or neurotoxicity. *Neurotoxicology* **53**, 271–281.

Sachana, M., Bal-Price, A., Crofton, K. M., Bennekou, S. H., Shafer, T. J., Behl, M., and Terron, A. (2018) International regulatory and scientific effort for improved developmental neurotoxicity testing. *Toxicol. Sci.* **167**, 45–57.

Sirenko, O., Parham, F., Dea, S., Sodhi, N., Biesmans, S., Mora, S., Ryan, K., Behl, M., Chandy, G., Crittenden, C., *et al.* (2018) Functional and mechanistic neurotoxicity profiling using human iPSC–derived neuronal 3D cultures. *Toxicol. Sci.* **167**, 58–76.

Smirnova, L., Hogberg, H. T., Leist, M., and Hartung, T. (2014). Developmental neurotoxicity – Challenges in the 21st century and in vitro opportunities. *ALTEX* **31**, 129–156.

Zhang, S., Hagstrom, D., Hayes, P., Graham, A., and Collins, E. M. (2018). Multi-behavioral endpoint testing of an 87-chemical compound library in freshwater planarians. *Toxicol. Sci.* **167**, 26–44.