



Establishment of a human cell-based in vitro battery to assess developmental neurotoxicity hazard of chemicals

Jonathan Blum^{a,1}, Stefan Masjosthusmann^{b,1}, Kristina Bartmann^b, Farina Bendt^b, Xenia Dolde^a, Arif Dönmez^b, Nils Förster^d, Anna-Katharina Holzer^a, Ulrike Hübenthal^b, Hagen Eike Keßel^b, Sadiye Kilic^a, Jördis Klose^b, Melanie Pahl^b, Lynn-Christin Stürzl^b, Iris Mangas^e, Andrea Terron^e, Kevin M. Crofton^f, Martin Scholze^g, Axel Mosig^d, Marcel Leist^{a,*}, Ellen Fritsche^{b,c,**}

^a *In Vitro Toxicology and Biomedicine, Dept Inaugurated By the Doerenkamp-Zbinden Foundation, University of Konstanz, 78457, Konstanz, Germany*

^b *IUF - Leibniz Research Institute for Environmental Medicine, 40225, Düsseldorf, Germany*

^c *Medical Faculty, Heinrich-Heine-University, 40225, Düsseldorf, Germany*

^d *Bioinformatics Group, Ruhr University Bochum, 44801, Bochum, Germany*

^e *European Food Safety Authority, PREV Unit, 43126, Parma, Italy*

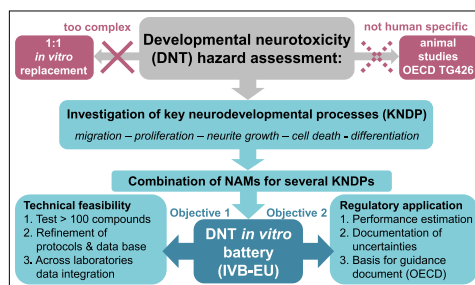
^f *R3Fellows LLC, Durham, NC, USA*

^g *Institute of Environment Health and Societies, Brunel University London, UK*

HIGHLIGHTS

- An in vitro testing battery (IVB) that allows screening of chemicals for developmental neurotoxicity (DNT) has been assembled.
- Performance estimates (>80% accuracy) have been obtained for the IVB, based on 45 negative/positive controls.
- Concentration-response data for altogether 120 compounds have been obtained for ten tests covering altogether 21 endpoints.
- Gaps of the IVB have been analyzed, and recommendations for the use of the IVB for regulatory testing have been put forward.

GRAPHICAL ABSTRACT



ARTICLE INFO

Handling Editor: Jian-Ying Hu

Keywords:

Testing battery
Stem cell
Brain development

ABSTRACT

Developmental neurotoxicity (DNT) is a major safety concern for all chemicals of the human exposome. However, DNT data from animal studies are available for only a small percentage of manufactured compounds. Test methods with a higher throughput than current regulatory guideline methods, and with improved human relevance are urgently needed. We therefore explored the feasibility of DNT hazard assessment based on new approach methods (NAMs). An in vitro battery (IVB) was assembled from ten individual NAMs that had been developed during the past years to probe effects of chemicals on various fundamental neurodevelopmental

* Corresponding author.

** Corresponding author. IUF - Leibniz Research Institute for Environmental Medicine, 40225, Düsseldorf, Germany.

E-mail addresses: jonathan.blum@uni-konstanz.de (J. Blum), marcel.leist@uni-konstanz.de (M. Leist), ellen.fritsche@iuf-duesseldorf.de (E. Fritsche).

¹ These authors contributed equally.

<https://doi.org/10.1016/j.chemosphere.2022.137035>

Received 8 July 2022; Received in revised form 20 October 2022; Accepted 24 October 2022

Available online 31 October 2022

0045-6535/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

In vitro testing
DNT

processes. All assays used human neural cells at different developmental stages. This allowed us to assess disturbances of: (i) proliferation of neural progenitor cells (NPC); (ii) migration of neural crest cells, radial glia cells, neurons and oligodendrocytes; (iii) differentiation of NPC into neurons and oligodendrocytes; and (iv) neurite outgrowth of peripheral and central neurons. In parallel, cytotoxicity measures were obtained. The feasibility of concentration-dependent screening and of a reliable biostatistical processing of the complex multi-dimensional data was explored with a set of 120 test compounds, containing subsets of pre-defined positive and negative DNT compounds. The battery provided alerts (hit or borderline) for 24 of 28 known toxicants (82% sensitivity), and for none of the 17 negative controls. Based on the results from this screen project, strategies were developed on how IVB data may be used in the context of risk assessment scenarios employing integrated approaches for testing and assessment (IATA).

1. Introduction

Screening of chemicals for a potential neurodevelopmental toxicity (DNT) hazard has been recognized as a pressing need by several large governmental and international organizations concerned with consumer safety. For instance, the US EPA and the European JRC took important roles in the organisation of a conference series (TestSmart) that was devoted to the development of a DNT test strategy useful in a regulatory context (Coecke et al., 2007; Lein et al., 2007; Crofton et al., 2011; Bal-Price et al., 2012). Also EFSA and the OECD embarked on similar efforts (Fritsche et al., 2017). In this context, several experimental programs were launched to probe novel approaches and to accelerate their implementation (Crofton et al., 2012; van Thriel et al., 2012; Krug et al., 2013b; Bal-Price et al., 2015; Baumann et al., 2016; Fritsche et al., 2018; Harrill et al., 2018; Behl et al., 2019; Lupu et al., 2020; Pistollato et al., 2021; Sachana et al., 2021; Vinken et al., 2021; Koch et al., 2022).

DNT is a field of toxicology concerned with effects of chemicals on the developing nervous system. Several experimental and epidemiological studies (on metals, pesticides and drugs) link compound exposure during early life phases (of the embryo, fetus or child) to functional alterations of the nervous system in adolescents or adults (Grandjean and Landrigan, 2014; Smirnova et al., 2014; Bennett et al., 2016). A particular concern is the possible role of DNT in the increased frequency of neurodevelopmental disorders, such as autism-spectrum disorders (Grandjean and Landrigan, 2006, 2014; Bellinger, 2012; Modafferi et al., 2021). The assessment is particularly challenging due to the multitude of potential toxicity manifestations (structural and functional). Moreover, there may be a time offset between toxicant exposure (before or after birth) and manifestation of effects (Grandjean et al., 2019).

The traditional methods to evaluate DNT hazard potential are based on animal studies following the OECD (OECD, 2007) or U.S. EPA (USEPA, 1998) test guidelines. To date only about 180 compounds world-wide have been tested using these guidelines (Crofton and Mundy, 2021). Several factors contribute to the limited availability of such studies: extensive time (e.g. 1–2 years) and resource requirement; limited triggered testing by chemical alerts; the need to reduce animal use; and the limited regulatory requirement for DNT testing as compared to some other test guidelines (e.g., carcinogenicity). The data available suffer from many uncertainties, and they require species extrapolation from rodents to humans. Moreover, they provide limited information on toxicity mechanisms. This can make them difficult to use in human risk assessments (Makris et al., 2009; Tsuji and Crofton, 2012; Tohyama, 2016; Paparella et al., 2020).

The strategic concepts of next generation risk assessment and of “toxicology for the 21st century” (Leist et al., 2008; Thomas et al., 2018; Pallocca et al., 2022a) suggest reductions in use of animal studies and development of new approach methods (NAMs) for toxicity assessment. The non-animal test methods should ideally be based on human-relevant test systems, reduce costs, allow a high throughput of test chemicals, and provide information on the toxicity mechanisms of toxicants. Many recent activities on scientific and regulatory levels have been undertaken to apply this strategy to the field of DNT (Sachana et al., 2019).

The establishment of DNT NAMs followed two major principles

(Bal-Price et al., 2015; Aschner et al., 2017). First, a concept was developed on how complex in vivo events and their disturbances could be modeled by simplified in vitro systems. It was found that the biological process of nervous system development can be broken down to less complex key neurodevelopmental processes (KNDP). Moreover, it was assumed that the disturbance of any KNDP may lead to DNT in humans. On this basis, NAMs were developed for most of the crucial KNDP. The second principle was that the performance and robustness of the NAMs should be at a high level, so that data could be used with high confidence. The concept of test readiness was developed to provide a measure of the NAM validation status (Bal-Price et al., 2018; Krebs et al., 2019, 2020b), and several assays were deemed ready and suitable for use in chemical screening. They include: proliferation, migration and differentiation assays based on neurospheres (NPC1-5 test methods); the neurite growth assays NeuroTox and PeriTox; the neural crest migration assay (cMINC); and an assays for neural network formation and synaptogenesis (Masjosthusmann et al., 2020; Crofton and Mundy, 2021; Carstens et al., 2022). Instead of a formal OECD-type validation (e.g. skin sensitization NAMs (OECD, 2021; Strickland et al., 2022)), the concept of a fit-for-purpose biological validation based on regulatory needs has been suggested (Leist et al., 2012; Hartung et al., 2013; Judson et al., 2013; Cote et al., 2016; Griesinger et al., 2016; Bal-Price et al., 2018; Andersen et al., 2019; Masjosthusmann et al., 2020). Its application to DNT NAM involved: understanding of all technologies related to test systems and endpoint assessment; a comparison of pivotal in vitro signaling pathways to those relevant in vivo; and an assessment of the cellular presence of toxicity targets known to play a role for human DNT (Aschner et al., 2017; Bal-Price et al., 2018; Koch et al., 2022).

No individual NAM covers all key aspects of neurodevelopmental biology. Thus no single test will detect effects on all KNDP. Therefore, a battery of assays is needed, to sufficiently cover all DNT toxicants. In 2016, participants of a meeting jointly organized by the European Food Safety Authority (EFSA) and the organisation for Economic Co-operation and Development (OECD) agreed that “an in vitro testing battery (based on available DNT NAM) could be used immediately to screen and prioritize chemicals” (Fritsche et al., 2017). A test run for such a battery was planned, in order to evaluate the technical feasibility, to identify potential gaps and to provide data and experience for setting up a draft guidance on how to run battery testing, and how to interpret data therefrom (Crofton and Mundy, 2021). The purpose of this manuscript is to describe the first test run of a DNT in vitro test battery based on methods available in European laboratories (IVB-EU). Extensive raw data and method documentations can be found in a report by EFSA (Masjosthusmann et al., 2020), and the experience and learnings from the IVB-EU have led to the preparation of the draft of an OECD guidance document, which is currently (July 2022) under revision in member countries (Crofton and Mundy, 2021). However, the data from 10 assays on 120 compounds (including 28 positive and 17 negative controls) have not been made available to academia and the interested public in a peer-reviewed publication. The same applies to the preliminary performance evaluation of the IVB-EU as a whole and the considerations concerning further use. The purpose of this manuscript is to make this important information available, and to provide a basis for further developments in academia, industry and by regulatory institutions

concerned with NAM-based DNT testing.

2. Materials and methods

2.1. Chemicals

A list of screen compounds ($n = 120$) was assembled by a working group, using the member's experience as members/employees at the US EPA, EFSA or in OECD working groups. Compounds were selected to be chemically and biologically somewhat diverse and to reflect groups of compounds with concern for a potential DNT hazard. For instance, flame retardants and pesticides were included, as some compounds in these groups are known for biological properties of relevance to DNT. One aspect of the selection process was also to allow for diversity of effects on different fundamental neurodevelopmental processes (and respective assays), and it was important to cover the full spectrum from compounds with no or low evidence for DNT liability to compounds with rich background data to allow for a wide spread of screen results. A subset of compounds ($n = 28$) were included as positive controls for DNT hazard, based on human data or robust animal data (Grandjean and Landrigan, 2006, 2014; Mundy et al., 2015; Ryan et al., 2016; Aschner et al., 2017) (Fig. S1). Another subset ($n = 17$) were compounds considered as negative controls. They were selected for their safe use during human pregnancy or because the available extensive data on their toxicity gave no evidence (by observation or mechanism) of any effects related to DNT (at the test concentrations used) (Fig. S2). A description of chemicals, including exact chemical identity and suppliers is found in the suppl. file 2 - sheet 1.

2.2. Test methods

All test methods used for screening were selected based on their high readiness level (Bal-Price et al., 2018), as well as a very comprehensive test description compatible with the OECD Guidance Document GD211 for in vitro test method descriptions. These ToxTemp files (Krebs et al., 2019) are included in suppl. file 1. Below, only brief descriptions are given for a quick overview. Notably, most assays had at least two endpoints, and some assays were run in more than one version, e.g. measurement after 72 and 120 h.

UKN2 Assay (cMINC): The assay, is based on neural crest cells differentiated from hiPSC (Nyffeler et al., 2017). Cells were seeded into 96-well plates around a stopper. The stopper was removed after 24 h to allow migration into the cell free area. Cells were exposed to the test compound for 24 h, and then stained with calcein-AM and Hoechst H-33342. The number of migrated double positive cells was quantified independent of an observer by high content imaging and image analysis (RingAssay software; <http://invitro-tox.uni-konstanz.de>). The cell viability was also determined by an automated imaging algorithm. Concentration-response curves from this test were based on six test compound concentrations (plus solvent control).

UKN4 assay (NeuriTox): The assay is based on LUHMES cells that were cultured and handled as previously described (Lotharius et al., 2005; Scholz et al., 2011; Krug et al., 2013a). It assesses neurite outgrowth in central nervous system neurons (Delp et al., 2018). Cells were pre-differentiated for two days to commit them towards the neuronal fate. They were then re-seeded in 96-well plates and exposed to the chemical for 24 h. Viability and neurite area were determined by high-content imaging after staining with calcein-AM and H-33342. The neurite area was defined by a fully automated algorithm as the area of calcein-positive pixels minus the area of all cell soma (Stiegler et al., 2011). Concentration-response curves from this test were based on ten test compound concentrations (plus solvent control).

UKN5 Assay (PeriTox): The assay is based on immature sensory neurons differentiated from hiPSC as previously described (Hoelting et al., 2016; Holzer et al., 2022). The test measures neurite outgrowth in peripheral neurons. Frozen lots of peripheral neuron precursors were

thawed and seeded into 96-well plates. After 1 h, the cells were exposed to test chemicals for 24 h. Testing and endpoint measurements were exactly as for the UKN4 assay (despite 6 instead of 10 compound concentrations tested).

NPC1-5 Assays: The neurosphere assays (NPC1-5) are based on primary human neural progenitor cells (hNPCs; gestational week 16–19), that are grown as floating 3D neurospheres. Their growth and viability is assessed in the 3D neurospheres (NPC1). Alternatively, spheres can be plated onto a laminin-coated matrix, where the cells start migration and differentiation to form a secondary 3D co-culture. The latter approach allows the simultaneous assessment of radial glia migration (NPC2a), neuronal differentiation (NPC3), neuronal migration (NPC2b) and neurite outgrowth (NPC4) as well as oligodendrocyte differentiation (NPC5) and their migration (NPC2c) by fully automated high content imaging. Data were obtained and analyzed from recorded microscope images by a dedicated image processing software, trained on positive and negative control images, as described earlier in detail (Forster et al., 2022; Koch et al., 2022).

For the NPC1 assay, spheres (0.3 mm) were plated in 96-well plates (U-bottom; 1 sphere/well) and directly exposed to the test compound (in proliferation medium). DNA synthesis was assessed as functional endpoint after 3 days in vitro (DIV), using a luminescence-based bromodeoxyuridine (BrdU) ELISA (Nimtz et al., 2019). Cytotoxicity was assessed as a membrane integrity assay (CytoTox-ONE Assay) measuring the LDH release into the supernatant.

For the NPC2-5 assays, spheres (0.3 mm) were plated in poly-D-lysine/laminin-coated 96-well plates (F-bottom; 1 sphere/well) and directly exposed to the test compounds (in differentiation medium). Under control conditions, NPCs migrate radially out of the attached sphere and differentiate into radial glia, neurons and oligodendrocytes. Data were obtained after 72 h and 120 h. After 72 h (3 DIV), bright field images were taken of live cell cultures, and radial glia migration (NPC2a [72 h]) was assessed using ImageJ software. The medium was partially removed (50%) and used to assess cytotoxicity (CytoTox-ONE Assay). To continue the assay, the medium was replenished and cells were allowed to further differentiate and migrate for 48 h. At 5 DIV, cells were fixed and stained for TUBB3 (neuronal marker), O4 (oligodendrocyte marker) and Hoechst H-33258 (nuclear marker). The endpoint assessment was done by high content imaging followed by different image analysis algorithms. Neuronal and oligodendrocyte differentiation (NPC3 and NPC5) was assessed as the number of all TUBB3-positive and O4-positive cells in percent of the total number of nuclei in the migration area. Neurons and oligodendrocytes were automatically recognized by a machine learning software based on convolutional neural networks (Forster et al., 2022). The high-content image analysis software Omnishpero was used to determine radial glia migration (NPC2a [120 h]), neuronal migration (NPC2b) and oligodendrocyte migration (NPC2c) as well as neuronal morphology (NPC4a: neurite length; NPC4b: neurite area) (Schmuck et al., 2017). Cytotoxicity was assessed from samples of medium removed before the fixation by the CytoTox-ONE LDH Assay. Some additional cell viability data were obtained by using a resazurin reduction assay (CellTiter-Blue Assay). Concentration-response curves from all these tests were based on seven test compound concentrations.

2.3. Screen strategy

Most of the compounds ($n = 75$) were provided by EPA's ToxCast chemical contractor (Evotec, South San Francisco, CA) in v-bottom 96 well plates. Separate plates were provided for different assays, and volumes shipped ranged from 50 to 300 μ l as DMSO stock solutions (always 20 mM). Other compounds were obtained from commercial sources (indicated in the suppl. 2 Excel file). In some of these cases stock solution was higher than 20 mM and compounds were dissolved in water if they were highly water-soluble (e.g. valproic acid). The University of Konstanz robotics platform was used to either produce replicates of the

master plate for different screening runs and different assays (UKN assays) or to directly prepare the compound dilutions (1:3 steps) in the media in 96-well plates (NPC assays). Operators were blinded to the compound identity. For the UKN assays serial dilutions (1:3 steps) were prepared from the cloned master plates for each compound in DMSO on 96-well plates, and each of these stocks was transferred to a pre-dilution plate. On these plates compounds were diluted 1:3 in medium plus 1% DMSO to have constant levels of DMSO among all concentrations. Finally, pre-dilutions were transferred to assay plates with cells (e.g. 20 μ l transfer to 180 μ l cells corresponding to 1:10) in medium to a maximum DMSO level of 0.1% in each assay. Exact volumes and pre-dilutions were assay-dependent and are detailed in ToxTemps; suppl. file 1. Some compounds were tested in an adapted concentration range (e.g. it is known that valproic acid is a human teratogen and DNT toxicant at clinically used concentrations of 0.5–1 mM. Therefore, higher concentrations were also tested, and master stocks were prepared accordingly).

For some assays (e.g. UKN2), a pre-screening step was included, in which only 1–2 (highest) test compound concentrations were run. When they showed no effect, screening was ended. When there was an effect (at least 20% change of endpoint), a full concentration-response was obtained. Pre-screen and full concentration-response screen were performed three times independently for all assays. For the UKN assays this meant the use of different cell lots for each run, for the NPC assays it meant the use of cells from different donors and/or passages for each run. Each screen run contained 2–6 technical replicates (details in ToxTemps; suppl. file 1). In some cases, follow-up tests were run, when e.g. only the highest concentration showed a response. Then new stocks were produced, and the concentration range was extended to 60 or 100 μ M, depending on the solubility of the compound.

2.4. Data analysis

A fully automated data analysis workflow was implemented on the programming platform R (KeBel, 2022). Original code and source files are available on GitHub at (<https://github.com/iuf-duesseldorf/fritsche-lab-CRStats>). It included the following steps and outputs: (1) Pre-processing of data, where required by the definitions of the assay endpoints (see ToxTemps; suppl. file 1). For instance, the background signal was subtracted from all data points for the BrdU fluorescence readings. (2) Normalization of test compound data to the median of solvent controls. (3) Calculation of the median of the replicates for each experimental condition. (4) Concentration response fitting of the data for each compound. The best-fitting model (general logistic, 3-parameter log-logistic, 4-parameter log-logistic, 2-parameter exponential, 3-parameter exponential, 3-parameter Weibull, 4-parameter Weibull) was selected by the AKAIKE information criteria (Ritz et al., 2015; Jensen et al., 2020). (5) Re-normalization of the data, so that the upper asymptote of the selected curve fit was at 100% (Krebs et al., 2018; Kappenberg et al., 2020). (6) Calculation of the mean re-normalized values for each condition across independent test runs. (7) Concentration response fitting of the data for each compound. The best-fitting model (general logistic, 3-parameter log-logistic, 4-parameter log-logistic, 2-parameter exponential, 3-parameter exponential, 3-parameter Weibull, 4-parameter Weibull) was selected by the AKAIKE information criteria. (8) Determination of the benchmark concentration (BMC) as the point of the concentration-response curve that intersected with the benchmark response level (BMR). The BMR was determined and described for each assay (see ToxTemp; suppl. file 1), based on a biological and statistical rationale. It marked the extent of response considered to be statistically significant and toxicologically meaningful. It thus depended on the endpoint and on the base line noise. For most functional endpoints it was set at 75% (= 25% reduced normal function). For some assays it was set at 70% (higher baseline noise). For some viability measures it was set at 90% (a deviation of >10% was considered to potentially influence the functional endpoint). (9) After

determination of the BMC, the upper (BMCU) and lower limit (BMCL) of its 95% confidence interval were calculated (Krebs et al., 2020a).

2.5. Hit definitions and prediction models

The prediction models (Worth and Balls, 2001; Leist et al., 2010; Griesinger et al., 2016; Schmidt et al., 2017; Bal-Price et al., 2018; Krebs et al., 2020b) of the NAM used in the IVB-EU had been defined during the original test setup, as documented in the literature and the ToxTemp files. A key feature of all assays was that they had a specific functional endpoint (related to a KNDP) and an endpoint characterizing compound effects on cell viability. Within each NAM, a compound was considered a specific hit (toxicant), when it affected the functional endpoint at least at one concentration that did not affect viability (Fig. S3). Notably, this does not mean that specific cytotoxicity of a given cell population (e.g. neural crest cells) would not lead to DNT. However, specific toxicity to a subpopulation can only be determined across assays, not within one assay. At present, a procedure for such a cross-IVB interpretation has not been established. Within a given assay, cytotoxicity makes the interpretation of the functional endpoint difficult. Therefore, (i) functional endpoint data were only used for concentrations that were non-cytotoxic, and (ii) specific cytotoxicity to subpopulations was not considered in this first application of the IVB-EU. For the UKN assays, specific effects were determined by the ratio of benchmark concentrations for the functional endpoint (e.g. neurite growth in UKN4) and cytotoxicity (e.g. a 4-fold offset for UKN4). For the NPC assays, specific toxicity was assumed when the 95% confidence intervals of the functional endpoint and the viability endpoint did not overlap. As the separation between “hit” and “non-hit” leads to binary data with high uncertainties at the hit/non-hit boundary (Leontaridou et al., 2017; Delp et al., 2018), we introduced a borderline category for transition compounds (e.g. when confidence intervals in NPC assays overlapped by > 10%). Thus, a given compound was classified in each assay as “no hit”, “unspecific hit”, “specific hit” or “borderline hit” (Fig. S3).

2.6. Performance parameters

A set of 45 reference compounds (28 DNT positives; 17 DNT negatives) was used for a preliminary evaluation of the IVB-EU predictivity (more may be added in the future). Various hit definitions were used (e.g. only specific hits, or specific + borderline hits). If a positive control was a hit, it was considered true positive (TP), if it was not a hit, it was considered a false negative (FN). If a negative control was a hit, it was considered a false positive (FP) and if it was not a hit, it was considered a true negative (TN). Using these four numbers (FP, FN, TP, TN), the following performance parameters were defined:

$$\text{sensitivity} [\%] = \frac{TP}{(TP + FN)} * 100$$

$$\text{specificity} [\%] = \frac{TN}{(TN + FP)} * 100$$

$$\text{accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100$$

$$\text{balanced accuracy} = \frac{\text{sensitivity} + \text{specificity}}{2}$$

$$\text{positive predictive value (PPV)} = \frac{TP}{(TP + FP)} * 100$$

$$F1 \text{ score} = \frac{2}{\frac{1}{\text{sensitivity}} + \frac{1}{\text{PPV}}} = \frac{1}{2} * (\text{sensitivity} + \text{PPV})$$

$$\text{Matthews correlation coefficient (MCC)} = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

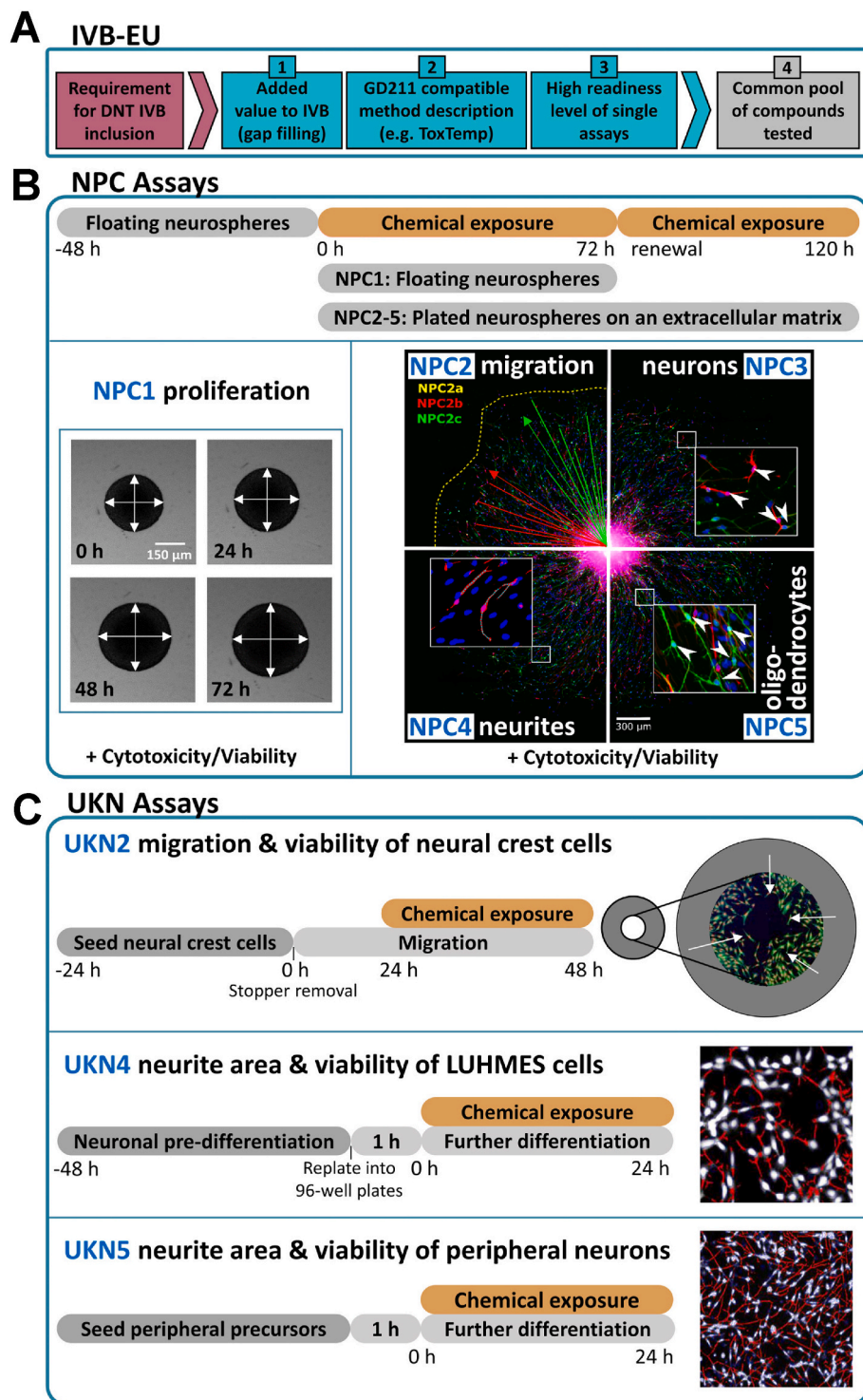


Fig. 1. Requirements and composition of the IVB-EU. (A) Criteria for assays to be included in the DNT test battery designated here IVB-EU. Criteria 1–3 were applied to this study. Criterion 4 was fulfilled in the course of this study and is suggested to be considered for future battery expansion. GD211 = OECD guidance document 211 on documentation of in vitro methods. (B) Schematic representation of the assays based on human neural progenitor cells (NPC) and their progeny. The general test system generation and exposure scheme is indicated on top. For the NPC1 test, floating neurospheres were exposed to toxicants for 72 h, and bromodeoxyuridine (BrdU) incorporation was used as endpoint for proliferation of NPC. For the NPC2-5 assays, neurospheres were plated and allowed to form secondary co-cultures of various cell types. Endpoints related to migration (NPC2), neuronal differentiation (NPC3), neurite growth (NPC4) and oligodendrocyte formation (NPC5) were assessed after 120 h by immunostaining and high content imaging. (C) Schematic representation of UKN assays. Cell types used and exposure schemes are indicated. Viability and migration of the cells in all assays were determined simultaneously by automated high content imaging after staining of the cell cultures with calcein-AM and Hoechst H-33342. The UKN2 assay evaluated the migration of neural crest cells into an empty circular area. The UKN4/UKN5 assays evaluated neural outgrowth of central nervous system and peripheral nervous system immature neurons. Detailed descriptions of NPC and UKN assays are given in the ToxTemps.

2.7. Data accessibility

The full raw data set from the IVB-EU has been entered into the ToxCast data base and is available in a machine-readable format used by many computational toxicologists after the fall 2022 ToxCast release (US EPA ORD, 2022).

3. Results and discussion

3.1. The DNT in vitro battery (IVB)

A large panel of assays with direct or indirect relevance to DNT can be found in the literature. Criteria needed to be developed to select a prototype battery of assays that was large enough for the main objective of this study, i.e. providing a basis for preparation of a general technical guidance document on battery testing for regulatory applications. At the same time, reasons of feasibility and limited resources called for keeping the number of NAMs included in the test run low. Experts with a regulatory background (from the US and Europe) were involved in the selection. The overall plan was to start testing in some European laboratories on a core battery (IVB-EU) of fully ready NAMs, and then to combine data on the same set of compounds with tests established at the US EPA. The three main selection criteria for the DNT NAMs were: (i) complementarity, (ii) documentation, and (iii) the readiness level (Fig. 1A). The first point meant that the assays were selected in a way to fill gaps of knowledge and to cover many KNDPs. It was also considered here to use assays for overlapping biological functions to learn about their orthogonality for later designs of tiered testing and sub-batteries. The second point referred to the availability of method documentations useful at a regulatory level (i.e. defined by OECD guidance document GD211) for the use of NAMs. Linked to this was the third criterion which referred to the technical performance of the NAMs, and the level of confidence into their predictivity and relevance. These issues are in some legislations referred to as validation state (Leist et al., 2012; Hartung et al., 2013; Judson et al., 2013; Cote et al., 2016; Griesinger et al., 2016; Bal-Price et al., 2018; Andersen et al., 2019; Masjosthusmann et al., 2020). In the selection of assays for the IVB-EU, we used a more flexible definition, termed “readiness” (Krebs et al., 2020b; Patterson et al., 2021). The assays used here all had undergone such an evaluation (Bal-Price et al., 2018; Klose et al., 2021a; Koch et al., 2022).

An additional criterion important for development of additional assays, now recommended in the draft OECD DNT-IVB test guideline is use of a common pool of test compounds (Fig. 1A). Ten assays fulfilled all criteria, and they were considered to be suitable for forming the IVB-EU. In addition to the above points, all selected assays use human cells, cover four major KNDP, reflect seven different brain cell types and represent different neurodevelopmental stages (Fig. 1B and C; Fig. 2).

To obtain an overview of test battery relevance and predictivity, a gap analysis was performed. Comparison of the included tests with the known neurodevelopmental processes showed that some KNDP are currently not covered by the IVB-EU. These include very early developmental processes such as stem cell differentiation into neural progenitor cells and subsequent neural tube construction, as well as processes necessary for neuronal circuit building, like formation, maturation and function of neuronal networks. As such gaps may reduce the sensitivity of DNT predictions, we explored the availability of assays that fulfill the IVB-EU inclusion criteria and could become part of an expanded full battery (Fig. 2). Many assays for network formation have indeed already shown to be at high readiness, yet these are based on rat cortical cells (Carstens et al., 2022) calling for human cell-based neuronal network formation assays. The early embryonal stages of neural development may be covered by the UKN1 assay (Dresler et al., 2020; Meisig et al., 2020). Some functional endpoints related to non-neuronal cells are also desirable for the IVB, as these cells (astrocytes, microglia, myelinating oligodendrocytes, microvascular endothelial cells) do not only have support and immune function, but rather participate in multiple neurodevelopmental processes (Allen and Lyons, 2018). Several 3D systems have been described to include the necessary cell types (Brull et al., 2020; Chesnut et al., 2021; Nunes et al., 2022), but still need some development to meet basic inclusion criteria (set up of test methods, throughput, documentation) for the IVB. The same applies to dedicated assays to investigate neurotransmitter systems (e.g. glutamate and acetylcholine signaling) (Klima et al., 2021; Loser et al., 2021b). However, a large part of signaling systems is covered already by the recent development of neural network formation assays (Frank et al., 2017; Nimtz et al., 2020). An interesting endpoint to comprehensively capture neuronal differentiation is transcriptome profiling (Pallocca et al., 2016; Shinde et al., 2017; Simon et al., 2019; Dresler et al., 2020; Meisig et al., 2020; Hu et al., 2022). This was exemplified here by the UKN1 assay. Modern high throughput sequencing techniques (Simon

KNDPs	Precursor proliferation	Migration	Differentiation	Neurite outgrowth	Neural network formation and function	Cell activation & stimulation
	Death of specific cell populations					
In vitro methods	NPC1 NPCs	NPC2a radial glia	NPC3 neuron	NPC4 CNS neuron	MEA based assays *	transport activity
	hiPSC based NPCs *	NPC2b neuronal	NPC5 oligo-dendrocytes	UKN4 CNS neuron	maturation & synaptogenesis	neuro-transmitter signals *
	radial glia	NPC2c oligo-dendrocytes	(NPC6) oligo-dendrocyte maturation *	UKN5 PNS neuron	myelination	inflammation (glia activation)
		UKN2 neural crest	NEPs (UKN1 & RoFA) *			signal transduction gaps
			astrocytes & radial glia *			
					Covered in IVB-EU	
					Potential gap	

Fig. 2. Key neurodevelopmental processes (KNDP) covered by IVB-EU. Categories of KNDPs, according to Bal-Price et al., 2018 are listed on top. Specific cell death in a neurodevelopmental sub-population may either be considered a KNDP or an adverse effect. As it is measured as endpoint in all assays of other KNDP, it was considered to be broadly covered by the IVB-EU without a dedicated own assay. The lower part of the figure indicates NAM (designated here: in vitro methods) that are related to the respective KNDP on top of each column. The coverage of KNDPs by assays that are part of the current IVB-EU is shown (bold). For some KNDPs, more than one test was available. The reason was that several distinct subpopulations e. g. migrate (radial glia, neurons, oligodendrocytes and neural crest cells) or grow neurites (different types of CNS and PNS neurons). Potential gaps of the current IVB-EU are shown as assays in the non-bold in vitro method boxes. Assays that have already been established in the co-authors’ labs are indicated by asterisks. They may be included in an extended version of the IVB, once they fulfill all inclusion criteria (Fig. 1). CNS: central nervous system; hiPSC: human induced pluripotent stem cells; NEP: neuroepithelial precursor; NPC: neural progenitor cell; MEA: microelectrode array; PNS: peripheral nervous system; RoFA: rosette formation assay.

et al., 2019; Jaklin et al., 2022; Spreng et al., 2022) now allow sufficient throughput for screening applications and it is likely that such assays will add additional information to the IVB in the future.

3.2. Readiness overview

The readiness of the assays of the DNT IVB was assessed on two tiers: first, the readiness of individual assays, as assessed earlier in individual publications, was an inclusion criterion (Fig. 1) of the IVB-EU. Second, the readiness of the overall battery and the performance of the assays under screening conditions was evaluated.

Concerning the first point, the underlying considerations are briefly re-iterated here, as they impinge on the interpretation and on the overall confidence into data from the NAMs of the IVB-EU. As for all toxicological assays, relevance, predictivity and reliability/robustness were considered. A major focus was put on the latter point, as suggested earlier (Leist et al., 2014; Krebs et al., 2019; Pallocca et al., 2022b). Earlier publications (summarized in Masjosthusmann et al. (2020)), and the ToxTemp (suppl. file 1) give more background information. One aspect helping to keep typical sources of variability low is that the selected IVB-EU assays all used a fully automated data capturing and evaluation procedure. However, the ultimate proof of the pudding for robustness, a blinded inter-lab comparison study, still has to be done for the assays.

When simple methods for 1:1 replacement of acute toxicity endpoints were evaluated, relevance and predictivity have been defined as separate aspects of NAMs. However, this concept has been modified for complex endpoints and batteries. In such more complex cases, the predictivity of a single NAM (for a given regulatory endpoint derived from animal studies) cannot be calculated, and the aspects of predictivity and relevance are strongly intertwined (Escher et al., 2022). In such cases, a scientific validation process is suggested that builds on two pillars: (i) comparison of the biological basis of the test system to that of the modeled human biology, and (ii) comparison of pathway modulations that lead to endpoint changes in the NAM to pathway changes known to be relevant to the respective human pathophysiology (Hartung, 2007; Leist et al., 2012; Hartung et al., 2013; Bal-Price et al., 2018; Piersma et al., 2018; Patterson et al., 2021). For the NAMs included in the IVB-EU, the test systems have been extensively documented and compared to the respective human developing nervous system counterparts. This involved the levels of cell morphology, cell function, and cell markers (see ToxTemps; suppl. file 1). Moreover, the relevant systems were profiled for their respective transcriptomes (Krug et al., 2014; Hoelting et al., 2016; Pallocca et al., 2017; Gutbier et al., 2018; Masjosthusmann et al., 2018; Klose et al., 2021a, 2021b, 2022). Also, the responses of the NAMs to modulation of signaling pathways relevant for brain development have been investigated by the use of compounds known to specifically affect signaling pathways (for overview: Klose et al. (2021b); Koch et al. (2022); Krebs et al. (2020b); Masjosthusmann et al. (2020)). A high-level summary of the responses to such “mechanistic tool compounds” is summarized in Fig. S4. One example is the Notch pathway, which determines a crucial switch between neurogenesis and oligodendrogenesis in vivo. By using the Notch pathway inhibitor DAPT, we can mimic this differentiation switch also in vivo with the NPC3/5 tests (Koch et al., 2022). Another illustrative example is the Rho pathway, which is involved in neurite growth in vivo. Activation of the RhoA kinase by narciclasine decreases neurite outgrowth in the NPC4, UKN4 and UKN5 assays. This successful characterization of neurodevelopmentally-relevant signaling in the IVB-EU assays is considered as the physiological basis and qualitative evidence for relevance and predictivity.

While the above-mentioned steps were important for the selection of NAMs and for giving confidence into their individual function within the IVB-EU, we also engaged in an effort to obtain information on the validity of the entire IVB-EU, as a battery. We considered the key parameters robustness, predictivity and relevance (Hartung et al., 2004;

Pallocca and Leist, 2022). Concerning relevance, it was mainly considered how many cell types and how many signaling pathways important for brain development were covered. A gap analysis showed that there was a need for few additional cells (e.g. microglia) and for some additional functions (e.g. neuronal network formation, astrocyte function). Moreover, more coverage of signaling (e.g. BDNF pathway and nicotinic signaling pathway) would be desirable. However, most relevant cell types were already represented, and many pathways known to be affected by toxicants were shown to be identifiable by at least one assay

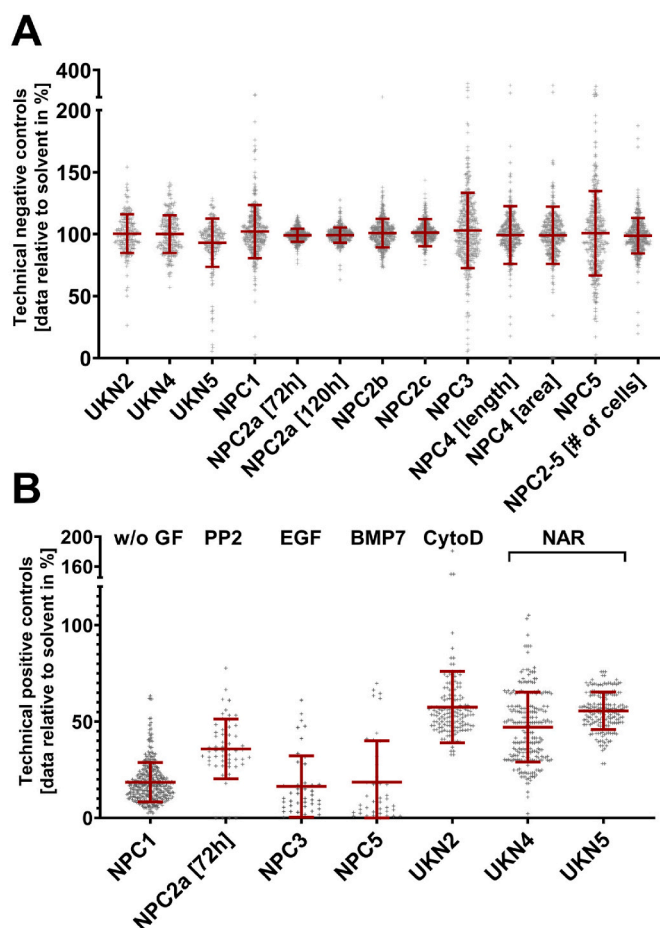


Fig. 3. Baseline noise and signal variation of acceptance controls in the IVB-EU assays. All tests were performed in a way so that each assay plate or experimental run contained wells with (i) negative controls, and at least one (ii) positive control. The reading of (ii) vs. (i) was used as acceptance criterion of the respective plate for UKN2, 4 and 5. If the positive control was not in a pre-specified range, the plate data were not included in screen results and measurements were repeated. Depending on the assay, plates contained different numbers of compounds. For some tests, the different concentrations of a given compound were on different plates. Thus, some plates contained the (iii) lowest concentration of a compound, and some did not. (A) To obtain a measure of inter-plate and intra-experimental variability of the baseline signal, the lowest concentration of each test compound (iii) was compared to the solvent control (i) on each plate. Altogether >200 data points were obtained for each IVB-EU endpoint from the testing campaign. For easier overview, the means \pm SD are indicated on top of the data points. (B) For each plate, the reading of the positive controls (ii) was compared to that of the negative controls (i) and normalized to negative control readings. The means \pm SD of data for positive controls are given for the IVB-EU endpoints. The compounds used to set acceptance criteria were as follows: w/o GF: without growth factor (omission of normally present growth factors in the positive control well); PP-2: SRC-kinase inhibitor; EGF: epidermal growth factor; BMP7: bone morphogenetic protein 7; CytoD: cytochalasin D; NAR: narciclasine. Details on concentrations are found in the ToxTemps (suppl. file 1).

(Fig. 2; Fig. S4).

One estimate for the robustness of screening results from the test battery is the baseline noise level of the NAM. As the results of all assays are normalized to solvent control data (which are set to 100%, and therefore do not vary by default), we used a surrogate baseline data set: from each concentration-response curve of the screen compounds, we selected the lowest concentration and assumed that this was in most cases a no-effect concentration. This assumption was consistent with the average of all these data points being about 100% for all assays. With this approach it was possible to visualize the baseline noise (as standard deviation around the average signal, Fig. 3A). From such data, we also calculated the assay-specific coefficients of variation (CoVs, see ToxTemp; suppl. file 1). As a second measure of robustness, we evaluated the responses of each test to the concurrent positive technical controls, which were run along on each plate/for every experiment during the screen (Fig. 3B). The positive controls were also used to determine acceptability of the respective plates/experiments for further evaluation. The plates/experiments, for which the acceptance criteria (see ToxTemp; suppl. file 1) were not met (<10% for all tests), were discarded.

3.3. Performance analysis

The predictivity of the IVB as a whole is a key feature of its regulatory applicability. This was examined as follows: First, all of the above discussed aspects of mechanistic validation were considered: the biology and pathophysiology covered by the entirety of assays of the IVB-EU suggested a high, but not perfect, biological applicability domain. This pointed at a sufficient predictivity for many purposes.

In a second step, we evaluated the capacity of the IVB-EU to correctly identify negative and positive controls. A list of 45 such calibration compounds was assembled from various literature references (Kadereit et al., 2012; Grandjean and Landrigan, 2014; Mundy et al., 2015; Aschner et al., 2017; Paparella et al., 2020; Crofton and Mundy, 2021). The challenges and shortcomings of this approach have been widely discussed (see above references), but our compound selection appeared to be a good compromise based on the present state of knowledge (Fig. 4A and B).

Prediction models for test batteries are an active field of research, and many possibilities exist (tiered approaches, Bayesian models, Boolean rules and decision trees). The difficulty to agree on the defined approaches for the small (3 NAM) battery used to predict dermal sensitization exemplifies these difficulties (Strickland et al., 2022). Here, we used a simple Boolean rule to define a battery hit as any compound that was a hit in one of the included DNT IVB-EU NAMs. A negative was defined as a compound not being a hit in any of the assays. This rule allows for a high transparency and simplicity. For statistical reasons, this battery prediction model may be associated with a high false discovery rate (testing for multiple endpoints considered to be independent). This was considered to be acceptable for screening and prioritization use. Moreover, the use of full concentration-response curves (instead of single data points) for definition of all positive hits reduced this problem. The false discovery rate was further reduced by our use of data from three independent experiments.

The 28 positive controls were used to obtain a preliminary measure of assay sensitivity (to be refined with time and the addition of more control compounds). We used different stringencies of hit definitions to obtain an estimate of the IVB-EU performance with respect to detection of DNT toxicants. When only the specific hits (compounds causing functional impairment at non-cytotoxic concentrations) were counted, the sensitivity of the IVB-EU was 68%. When borderline hits were included, this went up to 82%. When also cytotoxic compounds were included in the "hits", a further increase was observed. However, interpretation of cytotoxic compounds is presently not part of the IVB prediction model (Fig. 4A,C).

The 17 negative controls were used to obtain data on specificity.

When specific and borderline hits were counted, a value of 100% was obtained. Specificity dropped to 94%, when also cytotoxic effects were counted as "hit" (Fig. 4B,C).

Altogether, these preliminary performance estimates indicate that a balanced accuracy of about 80% or higher can be reached with the present IVB-EU. Based on the set of positive/negative control compounds, several additional performance measures were calculated (Fig. 4C) and it is particularly noteworthy that the IVB-EU had a high positive predictive value (PPV). This supports the conclusion that compounds identified as a hit should be prioritized for further evaluation of potential human hazard. Such data would also suggest that such chemicals better be excluded at early stages from further development (e.g. as a drug).

Nicotine serves as a good example for gaps in the IVB-EU, identified by the performance evaluation. It was identified as a false negative in the battery, and thus is indicative of a shortcoming with respect to sensitivity. The major action of nicotine is the stimulation of ionotropic acetylcholine receptors, and the IVB-EU does not (yet) include NAMs that would cover this biological function. This information is important when it comes to the interpretation of data from compounds that target nicotinic receptors, like neonicotinoid insecticides (Sheets et al., 2016; Loser et al., 2021a). Assays that fill these gaps are already under development (Fig. 2), and inclusion of assays based on zebra fish embryos and other model organisms (e.g. *C. elegans*) are considered an additional approach to close battery gaps (Atzei et al., 2021; Dasgupta et al., 2022).

Another limitation of the DNT IVB-EU is hard to overcome: the number of control compounds with clearly documented human effects is very limited, and also the compounds having been tested in DNT guideline studies in animals is small (Aschner et al., 2017). For this reason, performance metrics on the basis of currently-available control-compound predictivity will remain superficial. A way forward is to focus more on mechanistic validation approaches (Leist et al., 2012; Judson et al., 2013; Cote et al., 2016; Griesinger et al., 2016; Bal-Price et al., 2018; Andersen et al., 2019; Masjosthusmann et al., 2020) to gain further confidence into the predictivity of the battery for human adversities.

A final, but very important, consideration on predictivity is that this concept is highly context-dependent. In each sharply-defined use domain, it seems important to ask how far the battery is fit-for-purpose. Four issues need to be specified: (i) what regulatory problem is to be addressed (e.g. risk assessment of a new chemical, or prioritization of compounds for further testing); (ii) is there a focus on high positive predictivity or high negative predictivity; (iii) which type of chemicals is being examined (predictivity may be very high within certain compound groups, while it may be low for some compound classes); (iv) which types of biology (targets, pathways) play a role. It is likely that some adverse outcome pathways (AOP) are covered well, while others not at all. For example, acetylcholine esterase inhibitors may not be detected easily by the current IVB-EU, but this gap would be easily filled by an additional enzymatic assay (Li et al., 2017).

3.4. Compound testing and hit identification

In addition to the 45 compounds tested for the IVB-EU performance analyses, all 10 assays were challenged with additional 75 test compounds, so that the total screen comprised 120 chemicals (suppl. file 2). The result of the screen were benchmark concentrations (BMC) of effect (or no effect data within the used concentration range) for 120 compounds on ten functional and six viability endpoints, i.e. 1920 concentration response curves. A matrix including 405 BMCs for the IVB hits (with measures of uncertainty) was generated. To allow a better overview and focus, all compounds were compiled that affected at least one functional endpoint at a non-cytotoxic concentration (n = 59). To better visualize the activity profile of compounds, the endpoints for which toxicants had the highest potency (most sensitive endpoint(s)) were

A

Positive controls	specific + brdl. + cytotox	specific + brdl.	specific
Cadmium chloride	TP	TP	TP
Chlorpyrifos	TP	TP	FN
Dexamethasone	TP	TP	TP
Hexachlorophene	TP	TP	TP
Lead (II) acetate trihydrate	TP	TP	TP
Manganese (II) chloride	TP	TP	TP
Methylmercury chloride	TP	TP	TP
PBDE 47	TP	TP	TP
PBDE 99	TP	TP	FN
(±) Ketamine hydrochloride	FN	FN	FN
5,5-Diphenylhydantoin	FN	FN	FN
Acrylamide	TP	TP	TP
all-trans-Retinoic acid	TP	TP	TP
Chlorpromazine hydrochloride	TP	TP	TP
Deltamethrin	TP	TP	TP
Domoic acid	FN	FN	FN
Haloperidol	TP	TP	TP
Maneb	TP	TP	FN
Methylazoxymethanol acetate	TP	TP	TP
Nicotine	FN	FN	FN
Paraquat dichloride hydrate	TP	TP	TP
PFOA	TP	FN	FN
PFOSK	TP	TP	TP
Sodium valproate	TP	TP	TP
Tebuconazole	TP	TP	TP
Tributyltin chloride	TP	TP	TP
Trichlorfon	TP	TP	TP
Triethyl-tin bromide	TP	TP	FN

B

Negative controls	Acetaminophen	TN	TN	TN
	Amoxicillin	TN	TN	TN
	Aspirin	TN	TN	TN
	Buspirone	FP	TN	TN
	Chlorpheniramine maleate	TN	TN	TN
	D-Glucitol	TN	TN	TN
	Diethylene glycol	TN	TN	TN
	D-Mannitol	TN	TN	TN
	Doxylamine succinate	TN	TN	TN
	Famotidine	TN	TN	TN
	Ibuprofen	TN	TN	TN
	Metformin	TN	TN	TN
	Metoprolol	TN	TN	TN
	Penicillin	TN	TN	TN
	Saccharin	TN	TN	TN
	Sodium benzoate	TN	TN	TN
Warfarin	TN	TN	TN	

C

Performance [%]	Sensitivity	86	82	68
	Specificity	94	100	100
	Accuracy	89	89	80
	Balanced accuracy	90	91	84
	PPV	96	100	100
	F1 score	91	91	84
	MCC	78	80	67

(caption on next column)

Fig. 4. Performance overview of the test battery (IVB-EU). A set of predefined negative ($n = 17$) and positive ($n = 28$) control compounds was included in the set of screening compounds ($n = 120$). The rationale for their selection is given in Fig. S1 and S2. Note that the controls were randomly included in the overall screening workflow without being given any preferences or special treatment. This means that the standard prediction models of the assays were applied to them, so that they were classified as “no hit”, “cytotoxic”, “borderline (brdl)” or “specific hit” in individual NAM (see Fig. S3). A reference compound was considered to be a “positive” on the level of the overall IVB-EU, when it was an “alert” in at least one of the individual assays. The tabular display of the figure uses three definitions for an alert: anything that is not a “no hit” (first column), anything that was a specific hit or brdl (second column) or only specific hits (third column). (A) Alerts were considered true positives (TP), non-alerts were considered false negatives (FN). (B) Non-alerts were considered true negatives (TN), alerts were considered false positives (FP). (C) Performance parameters of the current DNT IVB-EU in percent. All parameters were calculated based on the TP, FN, TN, FP as indicated in (A) and (B). PPV: positive predictive value; MCC: Matthews correlation coefficient.

highlighted (Fig. 5). Compounds were considered to be about equally potent across test endpoints, when their activity did not differ by more than a factor of three. This is due to technical issues (the test concentrations were separated by a factor of three in the concentration-response curves), but also due to statistical considerations (the confidence intervals of BMCs separated by factor 3 overlapped in 85% of all cases).

Besides the 59 compounds that produced at least one specific hit (comprising 23 positive controls and 36 other compounds), there were also 61 compounds that had no specific hit in any of the 10 functional endpoints. Ten of these compounds were cytotoxic to one or more cell populations (Fig. S5A), while 51 compounds (including 16 negative controls) had no effect at all (Fig. S5B). This finding of 35 fully negatives (excluding the known negative controls) extends observations from the preliminary predictivity evaluation (using known negative control compounds) that showed that the IVB-EU, despite its large number of tests and endpoints, is not highly unspecific.

3.5. Hit patterns in the DNT IVB screen

Concerning the further analysis of battery hits, several strategies were followed. One approach was to select some individual hit compounds or groups of compounds for further toxicological evaluation. For instance, an expert group of EFSA and the OECD used IVB-EU data on deltamethrin and flufenacet for a case study within the OECD IATA program (EFSA PPR Panel, 2021). Another example is the group of flame retardants, for which the battery data were used to support a comprehensive hazard assessment (Klose et al., 2021a). Such specific toxicological follow-ups were beyond the scope of the present study. Instead, we analyzed general hit patterns of the screen to learn more about the relationship (complementarity/necessity) of the various assays and endpoints.

The first question was, how functional endpoints and specific hits related to the viability endpoints and cytotoxicity hits. To understand the overall data structure, we generated an overview, comparing for each specific hit compound the potency for the most sensitive functional endpoint in the battery (MSE) with the potencies for all cytotoxic effects across the battery test systems (cytotoxicity hits). There were 57 specific hits, plus two compounds (maneb and clorpyrifos), which were classified as borderline hits, and are being included here in the group of functional hits. Altogether 17 of the 59 compounds (29%) did not affect any of the battery’s viability endpoints. For this subgroup, the functional endpoint provided a definite gain in sensitivity, compared to cytotoxicity assays. It is also very likely that the functional endpoint was directly affected by the test compounds, i.e. it was not an indirect effect of unspecific cytotoxicity.

As an alternative approach to understand the role of cytotoxicity, we

<table border="1"> <tr> <td>M</td> <td>within</td> <td>outside</td> </tr> <tr> <td>S</td> <td>3-fold</td> <td>3-fold</td> </tr> <tr> <td>E</td> <td>MSE range</td> <td>range</td> </tr> </table>			M	within	outside	S	3-fold	3-fold	E	MSE range	range	UKN2	UKN4	UKN5	NPC1	NPC2a [72h]	NPC2a [120h]	NPC2b	NPC2c	NPC3	NPC4 [length]	NPC4 [area]	NPC5	NPC2-5 [# of cells]
M	within	outside																						
S	3-fold	3-fold																						
E	MSE range	range																						
Compound / Test method ¹																								
positive control compounds	Dexamethasone (DEX)				7.3				5.7															
	Tributyltin chloride (TTC)	6.9			7.2							6.8												
	Hexachlorophene (HCP)	6.4			4.9	6.9					7.1													
	Methylmercury(II) chloride (MMC)		7.0	6.4		6.2	6.3		6.3	6.7	6.3		6.1											
	Chlorpromazine hydrochloride (CPH)	5.2		5.5	6.0					5.4	5.3	6.4												
	Deltamethrin (DM)	4.8				6.0		5.4			4.9	6.4												
	Cadmium chloride (CCL)	6.3			5.3	5.8	5.6						5.5											
	Triethyltin bromide (TETB)			6.2																				
	all-trans-Retinoic acid (RA)	5.3			5.7																			
	Tebuconazole (TBN)												5.6											
	Haloperidol (HPD)		4.9	4.9		4.9							5.6											
	PBDE-47 ²	4.3											5.4											
	Trichlorfon (TCF)										5.3		4.9											
	PFOSK ³	4.5							5.0															
	Maneb (MAB)	5.0																						
	PBDE 99 (BDE)		4.9																					
	Paraquat dichloride hydrate (PQ)	4.0	4.0	3.8		4.9																		
	Manganese(II) chloride (MC)												4.7											
	Lead(II) acetate trihydrate (LAT)	4.6																						
	Chlorpyrifos (CPF)	4.4	4.0																					
	Methylazoxymethanol acetate (MAM)				3.8																			
	Sodium valproate (VPA)				3.3				3.3				3.3											
	Acrylamide (AAM)			2.9																				
all other hits	Narciclasine (NAR)	7.3	7.9	7.7	7.8			7.2	8.3	7.9	8.0	7.8	7.6											
	Rotenone (RTN)	7.5	7.0	7.2		6.8	7.0		6.3	6.3	6.2		6.7											
	Glufosinate-ammonium (GFA)								7.1															
	Alpha-Endosulfan (AES)	5.1					4.8						6.4											
	Metaflumizone (MFM)												6.3											
	Endosulfan sulfate (ESS)	5.4											6.3											
	Indoxacarb (IXC)	5.0			6.3								5.8											
	3,3',5,5'-Tetrabromobisphenol A (TBBPA)	5.9											6.3											
	gamma-Cyhalothrin (CHT)												6.1											
	Emamectin benzoate (EMB)	6.0			6.0																			
	2-Ethylhexyl diphenyl phosphate (EDP)	5.3			6.0								5.2											
	Beta-Cypermethrin (BCPM)												6.0											
	Malathion (MLT)												5.9											
	Endosulfan (EDS)						4.9						5.7											
	tau-Fluvalinate (TFV)												5.7											
	tert-Butylphenyl diphenyl phosphate (BDP)	5.3					5.1						5.7											
	MPP ⁴		5.7		5.6																			
	Tris(1,3-dichloro-2-propyl) phosphate (TDCPP)						4.9						5.6											
	Flubendiamide (FBD)												5.5											
	Fipronil (FPN)	4.8	5.0			4.9			4.9	4.9	4.9	4.9	5.5	4.8										
	Triphenyl phosphates isopropylated (TPI)	5.2											5.4											
	Tri-o-tolyl phosphate (TOTP)	5.4					5.3						5.2	4.7										
	Isodecyl diphenyl phosphate (IDDP)	5.0											5.4											
	Penthiopyrad (PP)					4.8							5.3											
	Chlorpyrifos-oxon (CPFO)				5.2		5.0				4.9													
	Tris(2-butoxyethyl)phosphate (TBOEP)												5.2											
	beta-Cyfluthrin (BCFT)												5.2											
	Carbaryl (CBR)	5.0	4.8			5.1	5.2				5.1	5.0												
	Acibenzolar-S-methyl (ABM)				5.1																			
	Tris(methylphenyl) phosphate (TTP)	5.1																						
	Clothianidin (CTN)									5.0														
	Triphenylphosphate (TPHP)	5.0											4.9											
	Bisphenol A (BPA)	4.9											4.9											
	Thiacloprid (TC)												4.9											
	1-Naphthol (NT)	4.7				4.9																		
	Azinphos-methyl (APM)					4.8	4.8																	

¹all concentrations are given in -logM; ²2,2',4,4'-Tetrabromodiphenyl ether;

³Heptadecafluorooctanesulfonic acid potassium salt; ⁴1-Methyl-4-phenylpyridinium iodide,

Fig. 5. Hit summary of the IVB-EU screen. Overall, 120 compounds were screened in the current DNT IVB-EU. Screened substances were considered as “hits” when they were classified as a “specific hit” or a “borderline compound” in at least one assay of the battery (assays indicated on top of the columns). The upper section of the table shows all 23 hits amongst the 28 positive controls used in the screen (the remaining five positive controls were no hits). The lower section shows all additional 36 hits amongst the screened compounds. Within the groups, the compounds are ranked based on potency (indicated in units of - log [M]). The table includes all hits of the screen. For each compound, the most sensitive endpoint (MSE) is highlighted. In addition, hits of the respective chemical in other assays, which were of similar potency as in the MSE assay (within a 3-fold range), are also highlighted. The compounds that affected only viability endpoints in the IVB-EU are listed in Fig. S5A. The compounds that affected no endpoint at all are listed in Fig. S5B. Exact and complete screen data (including the uncertainties assessed as 95% confidence interval) are included in a suppl. file 2 – sheet 2 & 3.

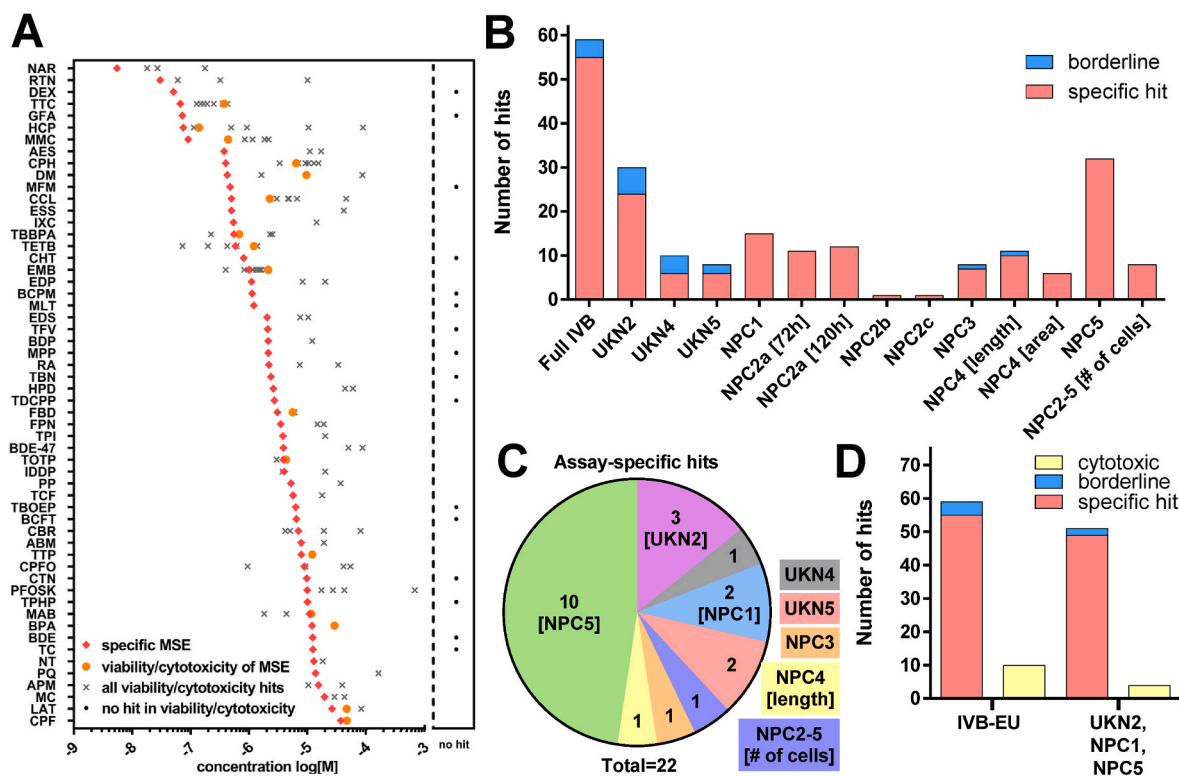


Fig. 6. Contribution of individual NAM to the overall IVB-EU. The screen was performed, hits were identified and the most sensitive endpoint (MSE) was defined for each compound as detailed in Fig. 5 (A). A potency overview of all hit compounds (see Fig. 5 for abbreviation) is displayed: The compounds are sorted according to the potency of their MSE. Note that all MSE data refer to a specific test endpoint (i.e. migration, differentiation, proliferation, neurite growth). In addition, the concentrations at which compounds were detected to be cytotoxic are indicated. Compounds that were not cytotoxic in any assay are indicated by a dot right of the dashed line. The cytotoxic concentration measured in the same assay as the MSE is given a separate symbol (filled circle) to allow an easy overview. Note that for many compounds, no cytotoxicity was measured in the assay that produced the MSE. For design reasons, three low potency compounds were not included in the figure: MAM (MSE = -3.8) orange point at x , 3 additional cytotoxic hits; VPA (MSE = -3.3) orange point at -2.7 , four other cytotoxicity hits; AAM (MSE = -2.9) no other cytotoxic hit. All data are given in log(M). (B) The number of hits (out of 120 screen compounds) is indicated for each assay of the battery, and for the total IVB-EU (most leftward bar). The number of specific hits and of borderline hits can both be seen within one bar. The respective set of data for cytotoxic compounds in visualized in Fig. S7. (C) The number of compounds that were a hit in only one assay is displayed for all assays, e.g. 10 compounds were detected only in NPC5, but no other assay; one compound was detected only in UKN4 and no other assay. (D) The number of hits (separated in specific hits, borderline hits and cytotoxic-only compounds) was compared for the full IVB-EU and a hypothetical mini-battery consisting of 3 assays (UKN2, NPC1, NPC5). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

asked, how the MSE concentration related to the cytotoxic potency in the same or in any other assay. There were only five compounds (8%) for which a cytotoxic endpoint was observed at higher (\geq factor 2) potency than the functional MSE (Fig. 6A). One example is carbaryl (CBR), which specifically inhibited neurite growth in the UKN4 assay (functional endpoint). It was particularly potent as cytotoxicant for peripheral neurons and mixed NPC cultures. This may indicate that CBR exerts a cell type-specific cytotoxicity for such neural cell populations. Such viability effects may be relevant for neurodevelopment, but further investigations would be required to allow clear conclusions.

We used a comparison to published data as one preliminary approach to test whether cytotoxicity hits of the IVB-EU are specific for neurodevelopmental cell types. We hypothesized that we may see a difference between cytotoxic potencies on conventional cell lines (HepG2, HEK293, etc.) and on the test systems used here, if a compound shows a developmental-stage specific cytotoxicity. Information on unspecific toxicity (called: cytotoxicity lower bound) was obtained from the ToxCast data base (Judson et al., 2016). For the 41 compounds, for which sufficient data was available, we found that cytotoxicity hit potency in the IVB-EU was at least 10-fold below the cytotoxicity lower bound for 7 compounds; 34 compounds showed no particular sensitivity in IVB-EU test systems compared to cell lines used for ToxCast screening (Fig. S6A). This may indicate that some, but not all cytotoxicity hits may be specific for neurodevelopmental cell types. To complete this

comparison, we also checked how the functional hits of the IVB-EU compared to the cytotoxicity lower bound. In general, the cytotoxicity threshold in ToxCast was often in the range of 5–20 μ M. Thus, the 17 IVB screen hits with MSEs <1 μ M (for which the cytotoxicity lower bound was available), seemed to separate clearly from general cytotoxicity except for TETB. The situation is complex for compounds with higher MSE potency in the IVB-EU. The data set is too small and compound behaviour is very heterogeneous. However, it is plausible, that specificity may be reduced (or lost) at higher screen concentrations (>20 μ M). It has been shown that unspecific baseline toxicity increases from this threshold on, due to membrane incorporation and alterations of protein conformations (Escher et al., 2019; Lee et al., 2021, 2022). Therefore, hits in a higher concentration range (e.g. MAM, VPA, AAM) need good justifications (e.g. clinically-observed plasma levels at hit concentration levels) and/or a detailed mechanistic follow-up providing a rationale for specific functional effects in the observed concentration range (Fig. S6B).

All these potency comparisons have an important caveat: the data we obtained are based on nominal concentrations, and these might differ from the free effective concentrations in the medium, and especially at the target sites (Kisitu et al., 2020). Especially, for comparisons to assays with tumor cell lines, it needs to be considered, that such systems usually use serum supplements containing protein and lipids, while most stem cell culture media used here had a low protein and lipid content. Under

the conditions used for the IVB-EU, the free concentrations are very close to the total concentrations in medium (Krebs et al., 2020b), while this is not necessarily the case for serum-containing media.

The second question we asked was, how the hits distributed over the different assays of the battery. Altogether 67 compounds affected at least one test endpoint: 57 specific, 2 borderline, 10 cytotoxic and 51 compounds affected no endpoint at concentrations up to 20 μM (Fig. 6B, Fig. S5&S7). All cytotoxic compounds had potencies of ≥8 μM (Fig. S5A). The number of hits obtained in each assay was also compiled. For instance, the NPC5 assay (examining the KNDP oligodendrocyte differentiation) identified the highest number (n = 34) of specific hits (Fig. 6B). Moreover, 10 compounds were hits only in this assay and would have been missed as potential toxicants without the NPC5 test as part of the IVB-EU (Fig. 6C). The second highest hit rate (n = 30) was found for the UKN2 assay (represents the KNDP of neural crest cell migration). Three compounds were unique hits in this test, i.e. not identified by another endpoint. Most other assays (UKN4, UKN5, NPC1, NPC2a, NPC3 and NPC4) identified 8–15 specific hits, and each of the assay identified at least one test compound that would have been missed by the other tests of the battery (Fig. 6C). This illustrates that the cell

types and endpoints assembled in the IVB-EU all differ in the pattern of toxicity pathways and targets they represent. This analysis also showed that the test methods are not redundant, even with this small number (n = 120) of screened chemicals. We anticipate that the broad coverage of cell types, developmental stages and endpoints of the IVB-EU will be even more required to ensure maximal sensitivity, when the chemical space is enlarged by broader test campaigns and a more-wide spread use of the battery.

A third question we asked dealt with resource optimization. Some assays, such as NPC2b/c (migration of neurons and oligodendrocytes) or UKN4 (neurite outgrowth) contributed relatively little to the overall hit rate, and one may consider them to be deleted from the battery or replaced. This would be a step towards a faster, more economical “mini-battery”, which would be expected to have a slightly reduced sensitivity, but not greatly reduced overall performance (accuracy; Matthews coefficient). However, in case of the neurosphere assay, individual read-outs are multiplexed, meaning that omission of one endpoint will not lead to saving resources, e.g. NPC2b/c are automatically assessed when NPC3/5 are evaluated. As NPC3 is multiplexed with NPC2 and 5, also this assay adds negligible extra time and costs to the overall assays

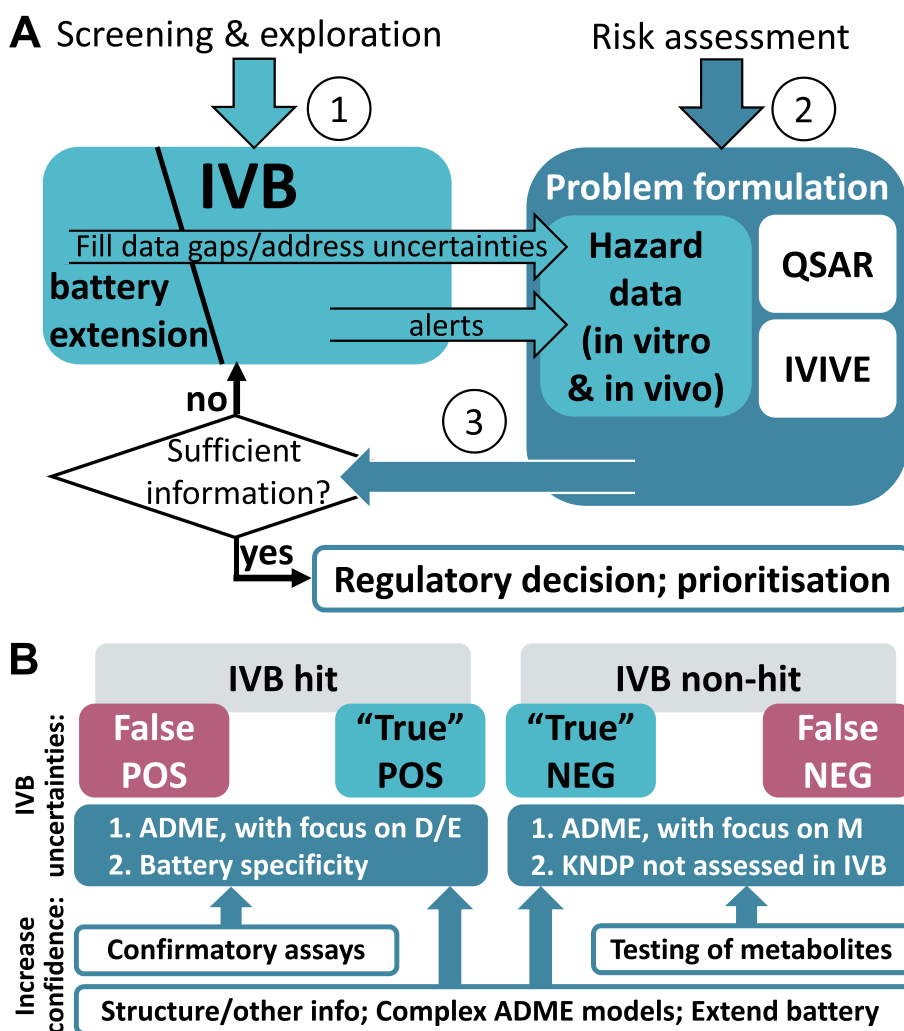


Fig. 7. Outlook on further uses and extensions of the IVB. (A) Incorporation of the IVB into an integrated approach to testing and assessment (IATA): Two different scenarios are depicted. In the first (1) the IVB will be used for screening of compound groups to generate hazard alerts (IVB hits). One way to follow up on these would be in the context of an IATA. In the second scenario (2), risk assessment of single chemicals would be performed in an IATA. This approach starts with a problem formulation (considering or not considering particular exposure situations). In this context all available data on hazard identification and characterization are collected. These may be extended via data of scenario (1). Quantitative structure activity relationships (QSAR) and in vitro-to-in vivo extrapolation (IVIVE) are shown as exemplary elements of the IATA framework. Further elements could include absorption, distribution, metabolism and excretion data (ADME) or an exposure assessment. If the hazard data of the assessed compound are considered not sufficient to derive a robust point of departure (PoD), further information could be obtained from the IVB. (3) In some cases, IVB extensions would be needed to fill data gaps and to reduce uncertainties, until sufficient information is available for regulatory action. (B) Each test method or battery has some uncertainties. The level of uncertainties that can be accepted depends on the problem formulation. For IVB hits and non-hits, one needs to consider that these may be either false positives/negatives, or compounds with a correctly identified hazard (“true” positives/negatives). One potential reason for misidentification is a lack of ADME features represented in the in vitro test systems. For example in vivo distribution and elimination (D/E) features may be misrepresented in the in vitro system. As a result, a compound never reaching the fetal brain because of the placental barrier may show effects on neurons in vitro. In contrast, some false negatives can be explained by a lack of metabolism (M) i.e. in vivo toxic metabolites which are not present in the IVB. Another reason is that a toxicant affects a key neurodevelopmental process (KNDP) that is not included in the IVB. In order to reduce the level of uncertainties and gain confidence into the results, further information can be added (low, white boxes). This includes information transfer

across tested compounds (grouping and readacross (RAX)), complex ADME models, confirmatory assays (battery extension), and direct testing of potential metabolites. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

NPC2-5. Hence, a mini-battery should only omit assays that practically save resources, i.e. individual assays. If one continues this line of thought, a minimal DNT IVB may consist of NPC1 (NPC proliferation), NPC2-5 and UKN2 (NCC migration) test methods (Fig. 6D). In our screen, this mini-battery would have identified 52 compounds (88% of all specific and borderline hits) of the 59 hits covered by the whole IVB-EU. Such a reduced approach may be used e.g. for quick/inexpensive pre-screens, e.g. in situations where sensitivity is of low importance, but compounds are to be ranked according to their priority for further testing. However, one may also consider adding an assay to a mini-battery that is not yet included in the IVB-EU. The gap analysis (Fig. 2) suggested that some biological domains are still poorly covered, and that an important gap would be filled by a neural network formation assay (Carstens et al., 2022). Thus, future batteries would need to consider the assays presented here, in addition to other established and emerging DNT NAM.

4. Conclusions and outlook

We have demonstrated here how NAMs with endpoints related to KNDP can be selected and assembled to an in vitro battery to screen for DNT hazard of chemicals. The technical feasibility and the implementation of solid reporting standards have been demonstrated by the use of 120 test compounds in a battery test-run that produced close to 2000 BMCs. These were used to provide battery performance estimates and to classify test compounds as specific hits, cytotoxicants or non-hits. The pattern of results was used to discuss the contribution of the assays and their endpoints to the overall IVB-EU and to define gaps still to be filled.

Pivotal questions for the future are (i) how battery hits would be further used and (ii) how the IVB-EU (or its future expanded version = IVB) could be implemented in a regulatory context (Fig. 7A and B). We anticipate that the first application of the IVB will be for screening of data-poor compounds to explore their DNT liabilities. As the overwhelming majority of chemicals lacks data on DNT hazard, compounds of particular concern (because of high exposure or structural alerts) may be screened first. The IVB would produce alerts for further testing. The underlying toxicological rationale is that disturbance of any KNDP covered by the IVB has the potential to lead to DNT. In a regulatory environment, the IVB data would provide a hazard characterization, and could be used as point-of-departure for further steps. In this context, physiology-based kinetic modelling (PBK) followed by in vitro-to-in vivo extrapolations (IVIVE) could be applied to convert the BMCs to estimated adverse doses (AEDs). These would be used to perform a risk assessment.

With growing experience and confidence into the IVB, its output could become a pivotal element of DNT risk assessment. Such a development is supported by the guidance document on the generation and use of the NAM-based DNT data (Crofton and Mundy, 2021). In a risk assessment situation with a defined problem formulation (e.g. for pesticide marketing re-approval in the EU, or during registration of a chemical in Japan) the compound to be evaluated would be run through the battery to provide hazard data. These might be clear and unambiguous. Or they may need to be complemented by additional rounds of testing in battery extensions. Together with the use of ADME data or other information (such as QSAR) and an IVIVE procedure, sufficient information for risk assessment would be generated (Fig. 7A).

One important aspect of using the battery data as hazard characterization is the interpretation and follow-up of hits. It is at present unclear, whether the number of positive battery endpoints correlates with the strength of DNT hazard. Hence, in the hazard characterization scenario one would be equally concerned if a compound produced one or several hits. However, the BMCs producing the hits have to be considered as multiple hits in the same order of magnitude suggest a higher concern than hits that only produce one low BMC. In the screening and prioritization scenario concern could be based on a

combination of BMC magnitude and number of hits similar to the approach practiced in Klose et al. (2021a) in the flame retardant case study. However, singleton-hit chemicals can be of high concern as exemplified by the illustrative example lead, which is one of the best-proven human DNT toxicants and only affected one functional endpoint of the IVB-EU.

For each battery hit, there is always the uncertainty, that it is either a true positive, i.e. that the battery results reflect real DNT hazard for humans, or that it is a false positive (FP). A reason for the latter scenario may be toxicokinetic (ADME) properties. E.g. a compound may never reach the foetal or child brain because of barrier functions, but there is no such barrier in vitro. Some FP will also arise from test classification uncertainties (alpha error) and the IVB false discovery rate (FDR) due to the combination of a large number of assays. Fortunately, there are also ways to build confidence into the hit pattern and to reduce the uncertainty of a hit being a FP. The assays and their prediction models can be trimmed for high specificity (multiple test runs, full concentration-response curves, conservative thresholds for hit definition). Another powerful approach is to functionally group hit compounds and to use information on one compound to read across to others. This way, consistency and plausibility can be established and/or strengthened.

For some applications, also non-hits play an important role, e.g. for providing confidence to consumers on the safety of food constituents or contaminants. Non-hits may either be true (no hazard) or be false negatives (FN), i.e. have non-discovered toxic properties. The main sources of uncertainty on negatives are the gaps in the battery (KNDP or specific signaling pathway not covered) and toxicokinetic aspects. For instance, a tested parent compound may not be toxic, but a metabolite generated only in vivo may be a DNT toxicant. Fortunately, there are also strategies available to increase confidence in negative hits. If this is of particular importance, the sensitivity of assays can be increased by running a higher number of replicates. Also, a less conservative prediction model may be applied. This strategy is demonstrated here by the introduction of a borderline category, to capture toxic compounds that would otherwise have dropped out of the hit definition. Another major approach is the extension of the battery, e.g. by combination with the US EPA assays (Carstens et al., 2022). Last, but not least, grouping, and other information from data bases and the literature could be used for further evaluation of negative hits and decisions on potential extended testing (Fig. 7A).

Author contributions

All authors read, commented, and approved the manuscript. Jonathan Blum: study conception, investigation, data analysis, supervision, figure design, writing of article. Stefan Masjosthusmann: study conception, data analysis, supervision, figure design, writing of article. Kristina Bartmann: investigation. Farina Bendt: investigation. Xenia Dolde: investigation, data analysis. Anna-Katharina Holzer: investigation, data analysis. Ulrike Hübenthal: investigation. Sadiye Kilic: investigation. Jödis Klose: investigation. Melanie Pahl: investigation. Lynn-Christin Stürzl: investigation. Arif Dönmez: software development, data analysis. Nils Förster: software development, data analysis. Hagen Eike Keßel: software development, data analysis. Martin Scholze: software development, data analysis. Axel Mosig: software development, supervision. Iris Mangas: editing of article. Andrea Terron: editing of article. Kevin Crofton: editing of article. Marcel Leist: study conception, supervision, funding acquisition, project administration, figure design, writing of article. Ellen Fritsche: study conception, supervision, funding acquisition, project administration, figure design, writing of article.

Funding

This work was supported by the European Food Safety Authority (EFSA-Q-2018-00308), the Danish Environmental Protection Agency (EPA), Denmark, under the grant number MST-667-00205, the State

Ministry of Baden-Wuerttemberg, Germany, for Economic Affairs, Labour and Tourism (NAM-Accept), the project CERST (Center for Alternatives to Animal Testing) of the Ministry for culture and science of the State of North-Rhine Westphalia, Germany (file number 233-1.08.03.03-121972/131-1.08.03.03-121972), the European Chemical Industry Council Long-Range Research Initiative (Cefic LRI) under the project name AIMT11 and the BMBF (NeuroTool). It has also received funding from the European Union's Horizon 2020 research and innovation program under grant agreements No. 964537 (RISK-HUNT3R), No. 964518 (ToxFree), No. 101057014 (PARC) and No. 825759 (ENDpoiNTs).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ellen Fritsche, Kristina Bartmann, Arif Dönmez and Axel Mosig are shareholders of the company DNTOX that provides DNT-IVB assay services. The authors declare no potential conflicts of interest with respect to the research in this article. All other authors have no conflict of interest to declare.

Data availability

Data will be made available on request.

Acknowledgements

The authors are very grateful to Tanja Waldmann as well as Tim Shafer, Katie Paul Friedman and Kelly Carstens (all involved in compound selection, study design, data interpretation and transfer of data to the ToxCast data base). We would also like to thank R. Göttler for the experimental and J. Kapr for layout support. Furthermore, we thank T. Mayer, S. Müller, and the screening centre of the University of Konstanz for the experimental and technical support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chemosphere.2022.137035>.

Abbreviations

AOP	– adverse outcome pathway
BMC	– benchmark concentration
BMCL	– lower limit of 95% confidence interval of BMC
BMCU	– upper limit of 95% confidence interval of BMC
DIV	– days <i>in vitro</i>
DNT	– developmental neurotoxicity
EFSA	– European Food Safety Authority
FDR	– false discovery rate
hNPC	– human neural progenitor cell
hiPSC	– human induced pluripotent stem cell
IVB	– <i>in vitro</i> battery
IVB-EU	– DNT IVB based on methods available in European laboratories
IVIVE	– <i>in vitro</i> to <i>in vivo</i> extrapolation
KNDP	– key neurodevelopmental process
MSE	– most sensitive endpoint
NAM	– new approach methods
PPV	– positive predictive value
IATA	– integrated approaches for testing and assessment
OECD	– Organisation for Economic Co-operation and Development
TN	– true negative
TP	– true positive
UKN	– University of Konstanz

US EPA – United States Environmental Protection Agency

References

- Allen, N.J., Lyons, D.A., 2018. Glia as architects of central nervous system formation and function. *Science* 362, 181–185.
- Andersen, M.E., McMullen, P.D., Phillips, M.B., Yoon, M., Pendse, S.N., Clewell, H.J., Hartman, J.K., Moreau, M., Becker, R.A., Clewell, R.A., 2019. Developing context appropriate toxicity testing approaches using new alternative methods (NAMs). *ALTEX* 36, 523–534.
- Aschner, M., Ceccatelli, S., Daneshian, M., Fritsche, E., Hasiwa, N., Hartung, T., Hogberg, H.T., Leist, M., Li, A., Mundi, W.R., Padilla, S., Piersma, A.H., Bal-Price, A., Seiler, A., Westerink, R.H., Zimmer, B., Lein, P.J., 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.
- Atzei, A., Jense, I., Zwart, E.P., Legradi, J., Venhuis, B.J., van der Ven, L.T.M., Heusinkveld, H.J., Hessel, E.V.S., 2021. Developmental neurotoxicity of environmentally relevant pharmaceuticals and mixtures thereof in a zebrafish embryo behavioural test. *Int. J. Environ. Res. Publ. Health* 18.
- Bal-Price, A., Crofton, K.M., Leist, M., Allen, S., Arand, M., Buetler, T., Delrue, N., FitzGerald, R.E., Hartung, T., Heinonen, T., Hogberg, H., Benekou, S.H., Lichtensteiger, W., Oggier, D., Paparella, M., Axelstad, M., Piersma, A., Rached, E., Schilter, B., Schmuck, G., Stoppini, L., Tongiorgi, E., Tiramani, M., Monnet-Tschudi, F., Wilks, M.F., Ylikomi, T., Fritsche, E., 2015. International Stakeholder Network (ISTNET): creating a developmental neurotoxicity (DNT) testing road map for regulatory purposes. *Arch. Toxicol.* 89, 269–287.
- Bal-Price, A., Hogberg, H.T., Crofton, K.M., Daneshian, M., FitzGerald, R.E., Fritsche, E., Heinonen, T., Hougaard Benekou, S., Klima, S., Piersma, A.H., Sachana, M., Shafer, T.J., Terron, A., Monnet-Tschudi, F., Viviani, B., Waldmann, T., Westerink, R.H.S., Wilks, M.F., Witters, H., Zurich, M.G., Leist, M., 2018. Recommendation on test readiness criteria for new approach methods in toxicology: exemplified for developmental neurotoxicity. *ALTEX* 35, 306–352.
- Bal-Price, A.K., Coecke, S., Costa, L., Crofton, K.M., Fritsche, E., Goldberg, A., Grandjean, P., Lein, P.J., Li, A., Lucchini, R., Mundy, W.R., Padilla, S., Persico, A.M., Seiler, A.E., Kreysa, J., 2012. Advancing the science of developmental neurotoxicity (DNT): testing for better safety evaluation. *ALTEX* 29, 202–215.
- Baumann, J., Gassmann, K., Masjosthusmann, S., DeBoer, D., Bendt, F., Giersiefer, S., Fritsche, E., 2016. Comparative human and rat neurospheres reveal species differences in chemical effects on neurodevelopmental key events. *Arch. Toxicol.* 90, 1415–1427.
- Behl, M., Ryan, K., Hsieh, J.H., Parham, F., Shapiro, A.J., Collins, B.J., Sipes, N.S., Birnbaum, L.S., Bucher, J.R., Foster, P.M.D., Walker, N.J., Paules, R.S., Tice, R.R., 2019. Screening for developmental neurotoxicity at the national toxicology program: the future is here. *Toxicol. Sci.* 167, 6–14.
- Bellinger, D.C., 2012. A strategy for comparing the contributions of environmental chemicals and other risk factors to neurodevelopment of children. *Environ. Health Perspect.* 120, 501–507.
- Bennett, D., Bellinger, D.C., Birnbaum, L.S., Bradman, A., Chen, A., Cory-Slechta, D.A., Engel, S.M., Fallin, M.D., Halladay, A., Hauser, R., Hertz-Picciotto, I., Kwiatkowski, C.F., Lanphear, B.P., Marquez, E., Marty, M., McPartland, J., Newschaffer, C.J., Payne-Sturges, D., Patisaul, H.B., Perera, F.P., Ritz, B., Sass, J., Schantz, S.L., Webster, T.F., Whyatt, R.M., Woodruff, T.J., Zoeller, R.T., Anderko, L., Campbell, C., Conry, J.A., DeNicola, N., Gould, R.M., Hirtz, D., Huffing, K., Landrigan, P.J., Lavin, A., Miller, M., Mitchell, M.A., Rubin, L., Schettler, T., Tran, H. L., Acosta, A., Brody, C., Miller, E., Miller, P., Swanson, M., Witherspoon, N.O., American College of O., , Gynecologists, Child Neurology, S., Endocrine, S., International Neurotoxicology, A., International Society for Children's H., the E., International Society for Environmental, E., National, 2016. Council of asian pacific islander, P., national hispanic medical, A., national medical, A., In: Project TENDR: Targeting Environmental Neuro-Developmental Risks the TENDR Consensus Statement. *Environ Health Perspect.* vol. 124, pp. A118–A122.
- Brull, M., Spreng, A.S., Gutbier, S., Loser, D., Krebs, A., Reich, M., Kraushaar, U., Britschgi, M., Patsch, C., Leist, M., 2020. Incorporation of stem cell-derived astrocytes into neuronal organoids to allow neuro-glial interactions in toxicological studies. *ALTEX* 37, 409–428.
- Carstens, K.E., Carpenter, A.F., Martin, M.M., Harrill, J.A., Shafer, T.J., Paul Friedman, K., 2022. Integrating data from *in vitro* new approach methodologies for developmental neurotoxicity. *Toxicol. Sci.* 187, 62–79.
- Chesnut, M., Paschoud, H., Repond, C., Smirnova, L., Hartung, T., Zurich, M.G., Hogberg, H.T., Pamies, D., 2021. Human iPSC-derived model to study myelin disruption. *Int. J. Mol. Sci.* 22.
- Coecke, S., Goldberg, A.M., Allen, S., Buzanska, L., Calamandrei, G., Crofton, K., Hareng, L., Hartung, T., Knaut, H., Honegger, P., Jacobs, M., Lein, P., Li, A., Mundy, W., Owen, D., Schneider, S., Silbergeld, E., Reum, T., Trnovec, T., Monnet-Tschudi, F., Bal-Price, A., 2007. Workgroup report: incorporating *in vitro* alternative methods for developmental neurotoxicity into international hazard and risk assessment strategies. *Environ. Health Perspect.* 115, 924–931.
- Cote, I., Andersen, M.E., Ankley, G.T., Barone, S., Birnbaum, L.S., Boelkeheide, K., Bois, F.Y., Burgeon, L.D., Chiu, W.A., Crawford-Brown, D., Crofton, K.M., DeVito, M., Devlin, R.B., Edwards, S.W., Guyton, K.Z., Hattis, D., Judson, R.S., Knight, D., Krewski, D., Lambert, J., Maull, E.A., Mendrick, D., Paoli, G.M., Patel, C. J., Perkins, E.J., Poje, G., Portier, C.J., Rusyn, I., Schulte, P.A., Simeonov, A., Smith, M.T., Thayer, K.A., Thomas, R.S., Thomas, R., Tice, R.R., Vandenberg, J.J., Villeneuve, D.L., Wesselkamper, S., Whelan, M., Whittaker, C., White, R., Xia, M., Yauk, C., Zeise, L., Zhao, J., DeWoskin, R.S., 2016. The next generation of risk

- assessment multi-year study-highlights of findings, applications to risk assessment, and future directions. *Environ. Health Perspect.* 124, 1671–1682.
- Crofton, K.M., Mundy, W.R., 2021. External scientific report on the interpretation of data from the developmental neurotoxicity in vitro testing assays for use in integrated approaches for testing and assessment. *EFSA Support. Pub.* 18, 6924E.
- Crofton, K.M., Mundy, W.R., Lein, P.J., Bal-Price, A., Coecke, S., Seiler, A.E., Knaut, H., Buzanska, L., Goldberg, A., 2011. Developmental neurotoxicity testing: recommendations for developing alternative methods for the screening and prioritization of chemicals. *ALTEX* 28, 9–15.
- Crofton, K.M., Mundy, W.R., Shafer, T.J., 2012. Developmental neurotoxicity testing: a path forward. *Congenital. Anom.* 52, 140–146.
- Dasgupta, S., Simonich, M.T., Tanguay, R.L., 2022. Zebrafish behavioral assays in toxicology. *Methods Mol. Biol.* 2474, 109–122.
- Delp, J., Gutbier, S., Klima, S., Hoelting, L., Pinto-Gil, K., Hsieh, J.H., Aichem, M., Klein, K., Schreiber, F., Tice, R.R., Pastor, M., Behl, M., Leist, M., 2018. A high-throughput approach to identify specific neurotoxicants/developmental toxicants in human neuronal cell function assays. *ALTEX* 35, 235–253.
- Dresler, N., Madjar, K., Holzer, A.K., Kapitza, M., Scholz, C., Kranaster, P., Gutbier, S., Klima, S., Kolb, D., Dietz, C., Trefzer, T., Meisig, J., van Thriel, C., Henry, M., Berthold, M.R., Bluthgen, N., Sachinidis, A., Rahnenfuhrer, J., Hengstler, J.G., Waldmann, T., Leist, M., 2020. Development of a neural rosette formation assay (RoFA) to identify neurodevelopmental toxicants and to characterize their transcriptome disturbances. *Arch. Toxicol.* 94, 151–171.
- Escher, B.I., Glauch, L., Konig, M., Mayer, P., Schlichting, R., 2019. Baseline toxicity and volatility cutoff in reporter gene assays used for high-throughput screening. *Chem. Res. Toxicol.* 32, 1646–1655.
- Escher, S.E., Partosch, F., Konzok, S., Jennings, P., Luijten, M., Kienhuis, A., de Leeuw, V., Reuss, R., Lindemann, K.-M., Bennekou, S.H., 2022. Development of a roadmap for action on new approach methodologies in risk assessment. *EFSA Support. Pub.* 19, 7341E.
- Forster, N., Butke, J., Kessel, H.E., Bendt, F., Pahl, M., Li, L., Fan, X., Leung, P.C., Klose, J., Masjosthusmann, S., Fritsche, E., Mosig, A., 2022. Reliable identification and quantification of neural cells in microscopic images of neurospheres. *Cytometry* 101, 411–422.
- Frank, C.L., Brown, J.P., Wallace, K., Mundy, W.R., 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., Goumelon, A., 2017. OECD/EFSA workshop on developmental neurotoxicity (DNT): the use of non-animal test methods for regulatory purposes. *ALTEX* 34, 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., Leist, M., Mundy, W.R., Paparella, M., Piersma, A.H., Sachana, M., Schmuck, G., Solecki, R., Terron, A., Monnet-Tschudi, F., Wilks, M.F., Witters, H., Zurich, M.G., Bal-Price, A., 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.
- Grandjean, P., Abdennebi-Najar, L., Barouki, R., Cranor, C.F., Etzel, R.A., Gee, D., Heindel, J.J., Hougaard, K.S., Hunt, P., Nawrot, T.S., Prins, G.S., Ritz, B., Soffritti, M., Sunyer, J., Weihe, P., 2019. Timescales of developmental toxicity impacting on research and needs for intervention. *Basic Clin. Pharmacol. Toxicol.* 125 (Suppl. 3), 70–80.
- Grandjean, P., Landrigan, P.J., 2006. Developmental neurotoxicity of industrial chemicals. *Lancet* 368, 2167–2178.
- Grandjean, P., Landrigan, P.J., 2014. Neurobehavioural effects of developmental toxicity. *Lancet Neurol.* 13, 330–338.
- Griesinger, C., Desprez, B., Coecke, S., Casey, W., Zuang, V., 2016. Validation of alternative in vitro methods to animal testing: concepts, challenges, processes and tools. *Adv. Exp. Med. Biol.* 856, 65–132.
- Gutbier, S., May, P., Berthelot, S., Krishna, A., Trefzer, T., Behbehani, M., Efreanova, L., Delp, J., Gstraunthaler, G., Waldmann, T., Leist, M., 2018. Major changes of cell function and toxicant sensitivity in cultured cells undergoing mild, quasi-natural genetic drift. *Arch. Toxicol.* 92, 3487–3503.
- Harrill, J.A., Freudenrich, T., Wallace, K., Ball, K., Shafer, T.J., 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.
- Hartung, T., 2007. Food for thought ... on validation. *ALTEX* 24, 67–80.
- Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., Gribaldo, L., Halder, M., Hoffmann, S., Roi, A.J., Prieto, P., Sabbioni, E., Scott, L., Worth, A., Zuang, V., 2004. A modular approach to the ECVAM principles on test validity. *Altern. Lab Anim* 32, 467–472.
- Hartung, T., Hoffmann, S., Stephens, M., 2013. Mechanistic validation. *ALTEX* 30, 119–130.
- Hoelting, L., Klima, S., Karreman, C., Grinberg, M., Meisig, J., Henry, M., Rotshteyn, T., Rahnenfuhrer, J., Bluthgen, N., Sachinidis, A., Waldmann, T., Leist, M., 2016. Stem cell-derived immature human dorsal root ganglia neurons to identify peripheral neurotoxicants. *Stem Cells Transl Med* 5, 476–487.
- Holzer, A.K., Suciu, I., Karreman, C., Goy, T., Leist, M., 2022. Specific attenuation of purinergic signaling during bortezomib-induced peripheral neuropathy in vitro. *Int. J. Mol. Sci.* 23.
- Hu, W., Liu, C.W., Jimenez, J.A., McCoy, E.S., Hsiao, Y.C., Lin, W., Engel, S.M., Lu, K., Zylka, M.J., 2022. Detection of azoxystrobin fungicide and metabolite azoxystrobin-acid in pregnant women and children, estimation of daily intake, and evaluation of placental and lactational transfer in mice. *Environ. Health Perspect.* 130, 27013.
- Jaklin, M., Zhang, J.D., Schafer, N., Clemann, N., Barrow, P., Kung, E., Sach-Peltason, L., McGinnis, C., Leist, M., Kustermann, S., 2022. Optimization of the TeraTox assay for preclinical teratogenicity assessment. *Toxicol. Sci.* 188, 17–33.
- Jensen, S.M., Kluxen, F.M., Streibig, J.C., Cedergreen, N., Ritz, C., 2020. bmd: an R package for benchmark dose estimation. *PeerJ* 8, e10557.
- Judson, R., Houck, K., Martin, M., Richard, A.M., Knudsen, T.B., Shah, I., Little, S., Wambaugh, J., Woodrow Setzer, R., Kothiy, P., Phuong, J., Filer, D., Smith, D., Reif, D., Rotroff, D., Kleinstreuer, N., Sipes, N., Xia, M., Huang, R., Crofton, K., Thomas, R.S., 2016. Editor's highlight: analysis of the effects of cell stress and cytotoxicity on in vitro assay activity across a diverse chemical and assay space. *Toxicol. Sci.* 152, 323–339.
- Judson, R., Kavlock, R., Martin, M., Reif, D., Houck, K., Knudsen, T., Richard, A., Tice, R., Whelan, M., Xia, M., Huang, R., Austin, C., Daston, G., Hartung, T., Fowle 3rd, J., Wooge, W., Tong, W., Dix, D., 2013. Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. *ALTEX* 30, 51–56.
- Kadereit, S., Zimmer, B., van Thriel, C., Hengstler, J.G., Leist, M., 2012. Compound selection for in vitro modeling of developmental neurotoxicity. *Front Biosci (Landmark Ed)* 17, 2442–2460.
- Kappenberg, F., Brecklinghaus, T., Albrecht, W., Blum, J., van der Wurp, C., Leist, M., Hengstler, J.G., Rahnenfuhrer, J., 2020. Handling deviating control values in concentration-response curves. *Arch. Toxicol.* 94, 3787–3798.
- Kebel, 2022. Biostatistics and its impact on hazard characterization using in vitro developmental neurotoxicity assays. *ALTEX*. <https://doi.org/10.1101/2022.10.18.512648>. Submitted for publication.
- Kisitu, J., Hollert, H., Fisher, C., Leist, M., 2020. Chemical concentrations in cell culture compartments (C5) - free concentrations. *ALTEX* 37, 693–708.
- Klima, S., Brull, M., Spreng, A.S., Suciu, I., Falt, T., Schwamborn, J.C., Waldmann, T., Karreman, C., Leist, M., 2021. A human stem cell-derived test system for agents modifying neuronal N-methyl-D-aspartate-type glutamate receptor Ca(2+)-signalling. *Arch. Toxicol.* 95, 1703–1722.
- Klose, J., Li, L., Pahl, M., Bendt, F., Hubenthal, U., Jungst, C., Petzsch, P., Schauss, A., Kohrer, K., Leung, P.C., Wang, C.C., Koch, K., Tigges, J., Fan, X., Fritsche, E., 2022. Application of the adverse outcome pathway concept for investigating developmental neurotoxicity potential of Chinese herbal medicines by using human neural progenitor cells in vitro. *Cell Biol. Toxicol.*
- Klose, J., Pahl, M., Bartmann, K., Bendt, F., Blum, J., Dolde, X., Forster, N., Holzer, A.K., Hubenthal, U., Kessel, H.E., Koch, K., Masjosthusmann, S., Schneider, S., Sturz, L.C., Woeste, S., Rossi, A., Covaci, A., Behl, M., Leist, M., Tigges, J., Fritsche, E., 2021a. Neurodevelopmental toxicity assessment of flame retardants using a human DNT in vitro testing battery. *Cell Biol. Toxicol.*
- Klose, J., Tigges, J., Masjosthusmann, S., Schmuck, K., Bendt, F., Hubenthal, U., Petzsch, P., Kohrer, K., Koch, K., Fritsche, E., 2021b. TBBPA targets converging key events of human oligodendrocyte development resulting in two novel AOPs. *ALTEX* 38, 215–234.
- Koch, K., Bartmann, K., Hartmann, J., Kapr, J., Klose, J., Kuchovska, E., Pahl, M., Schluppmann, K., Zuhr, E., Fritsche, E., 2022. Scientific validation of human neurosphere assays for developmental neurotoxicity evaluation. *Front Toxicol* 4, 816370.
- Krebs, A., Nyffeler, J., Karreman, C., Schmidt, B.Z., Kappenberg, F., Mellert, J., Pallocca, G., Pastor, M., Rahnenfuhrer, J., Leist, M., 2020a. Determination of benchmark concentrations and their statistical uncertainty for cytotoxicity test data and functional in vitro assays. *ALTEX* 37, 155–163.
- Krebs, A., Nyffeler, J., Rahnenfuhrer, J., Leist, M., 2018. Normalization of data for viability and relative cell function curves. *ALTEX* 35, 268–271.
- Krebs, A., van Vugt-Lussenburg, B.M.A., Waldmann, T., Albrecht, W., Boei, J., Ter Braak, B., Brajnik, M., Braunbeck, T., Brecklinghaus, T., Busquet, F., Dinnyes, A., Dokler, J., Dolde, X., Exner, T.E., Fisher, C., Fluri, D., Forsby, A., Hengstler, J.G., Holzer, A.K., Janstova, Z., Jennings, P., Kisitu, J., Kobolák, J., Kumar, M., Limonciel, A., Lundqvist, J., Mihalik, B., Moritz, W., Pallocca, G., Ulloa, A.P.C., Pastor, M., Rovida, C., Sarkans, U., Schimming, J.P., Schmidt, B.Z., Stober, R., Strassfeld, T., van de Water, B., Wilmes, A., van der Burg, B., Verfaillie, C.M., von Hellfeld, R., Vrieling, H., Vrijenhoek, N.G., Leist, M., 2020b. The EU-ToxRisk method documentation, data processing and chemical testing pipeline for the regulatory use of new approach methods. *Arch. Toxicol.* 94, 2435–2461.
- Krebs, A., Waldmann, T., Wilks, M.F., Van Vugt-Lussenburg, B.M.A., Van der Burg, B., Terron, A., Steger-Hartmann, T., Ruegg, J., Rovida, C., Pedersen, E., Pallocca, G., Luijten, M., Leite, S.B., Kustermann, S., Kamp, H., Hoeng, J., Hewitt, P., Herzler, M., Hengstler, J.G., Heinonen, T., Hartung, T., Hardy, B., Gantner, F., Fritsche, E., Fant, K., Ezenad, J., Exner, T., Dunken, T., Dietrich, D.R., Coecke, S., Busquet, F., Braeuning, A., Bondarenko, O., Bennekou, S.H., Beilmann, M., Leist, M., 2019. Template for the description of cell-based toxicological test methods to allow evaluation and regulatory use of the data. *ALTEX* 36, 682–699.
- Krug, A.K., Balmer, N.V., Matt, F., Schonenberger, F., Merhof, D., Leist, M., 2013a. Evaluation of a human neurite growth assay as specific screen for developmental neurotoxicants. *Arch. Toxicol.* 87, 2215–2231.
- Krug, A.K., Gutbier, S., Zhao, L., Poldt, D., Kullmann, C., Ivanova, V., Forster, S., Jagtap, S., Meiser, J., Lepar, G., Schildknecht, S., Adam, M., Hiller, K., Farhan, H., Brunner, T., Hartung, T., Sachinidis, A., Leist, M., 2014. Transcriptional and metabolic adaptation of human neurons to the mitochondrial toxicant MPP(+). *Cell Death Dis.* 5, e1222.
- Krug, A.K., Kolde, R., Gaspar, J.A., Rempel, E., Balmer, N.V., Meganathan, K., Vojnits, K., Baquie, M., Waldmann, T., Ensenat-Waser, R., Jagtap, S., Evans, R.M., Julien, S., Peterson, H., Zagoura, D., Kadereit, S., Gerhard, D., Sotiriadou, I., Heke, M., Natarajan, K., Henry, M., Winkler, J., Marchan, R., Stoppini, L., Bosgra, S.,

- Westerhout, J., Verwei, M., Vilo, J., Kortenkamp, A., Hescheler, J., Hothorn, L., Bremer, S., van Thriel, C., Krause, K.H., Hengstler, J.G., Rahnenfuhrer, J., Leist, M., Sachinidis, A., 2013b. Human embryonic stem cell-derived test systems for developmental neurotoxicity: a transcriptomics approach. *Arch. Toxicol.* 87, 123–143.
- Lee, J., Braun, G., Henneberger, L., Konig, M., Schlichting, R., Scholz, S., Escher, B.I., 2021. Critical membrane concentration and mass-balance model to identify baseline cytotoxicity of hydrophobic and ionizable organic chemicals in mammalian cell lines. *Chem. Res. Toxicol.* 34, 2100–2109.
- Lee, J., Escher, B.I., Scholz, S., Schlichting, R., 2022. Inhibition of neurite outgrowth and enhanced effects compared to baseline toxicity in SH-SY5Y cells. *Arch. Toxicol.* 96, 1039–1053.
- Lein, P., Locke, P., Goldberg, A., 2007. Meeting report: alternatives for developmental neurotoxicity testing. *Environ. Health Perspect.* 115, 764–768.
- Leist, M., Efremova, L., Karreman, C., 2010. Food for thought ... considerations and guidelines for basic test method descriptions in toxicology. *ALTEX* 27, 309–317.
- Leist, M., Hartung, T., Nicotera, P., 2008. The dawning of a new age of toxicology. *ALTEX* 25, 103–114.
- Leist, M., Hasiwa, N., Daneshian, M., Hartung, T., 2012. Validation and quality control of replacement alternatives – current status and future challenges. *Toxicology Research* 1, 8–22.
- Leist, M., Hasiwa, N., Rovida, C., Daneshian, M., Basketter, D., Kimber, I., Clewell, H., Gocht, T., Goldberg, A., Busquet, F., Rossi, A.M., Schwarz, M., Stephens, M., Taalman, R., Knudsen, T.B., McKim, J., Harris, G., Pamies, D., Hartung, T., 2014. Consensus report on the future of animal-free systemic toxicity testing. *ALTEX* 31, 341–356.
- Leontaridou, M., Urbisch, D., Kolle, S.N., Ott, K., Mulliner, D.S., Gabbert, S., Landsiedel, R., 2017. The borderline range of toxicological methods: quantification and implications for evaluating precision. *ALTEX* 34, 525–538.
- Li, S., Huang, R., Solomon, S., Liu, Y., Zhao, B., Santillo, M.F., Xia, M., 2017. Identification of acetylcholinesterase inhibitors using homogenous cell-based assays in quantitative high-throughput screening platforms. *Biotechnol. J.* 12.
- Loser, D., Hinojosa, M.G., Blum, J., Schaefer, J., Brull, M., Johansson, Y., Suci, I., Grillberger, K., Danker, T., Moller, C., Gardner, I., Ecker, G.F., Bennekou, S.H., Forsby, A., Kraushaar, U., Leist, M., 2021a. Functional alterations by a subgroup of neonicotinoid pesticides in human dopaminergic neurons. *Arch. Toxicol.* 95, 2081–2107.
- Loser, D., Schaefer, J., Danker, T., Moller, C., Brull, M., Suci, I., Uckert, A.K., Klima, S., Leist, M., Kraushaar, U., 2021b. Human neuronal signaling and communication assays to assess functional neurotoxicity. *Arch. Toxicol.* 95, 229–252.
- Lotharius, J., Falsig, J., van Beek, J., Payne, S., Dringen, R., Brundin, P., Leist, M., 2005. Progressive degeneration of human mesencephalic neuron-derived cells triggered by dopamine-dependent oxidative stress is dependent on the mixed-lineage kinase pathway. *J. Neurosci.* 25, 6329–6342.
- Lupu, D., Andersson, P., Bornehag, C.G., Demeneix, B., Fritsche, E., Gennings, C., Lichtensteiger, W., Leist, M., Leonards, P.E.G., Ponsoby, A.L., Scholze, M., Testa, G., Tresguerres, J.A.F., Westerink, R.H.S., Zalc, B., Ruegg, J., 2020. The ENDPoINT's project: novel testing strategies for endocrine disruptors linked to developmental neurotoxicity. *Int. J. Mol. Sci.* 21.
- Makris, S.L., Raffaele, K., Allen, S., Bowers, W.J., Hass, U., Alleva, E., Calamandrei, G., Sheets, L., Amcoff, P., Delrue, N., Crofton, K.M., 2009. A retrospective performance assessment of the developmental neurotoxicity study in support of OECD test guideline 426. *Environ. Health Perspect.* 117, 17–25.
- Masjosthusmann, S., Becker, D., Petzsch, B., Klose, J., Siebert, C., Deenen, R., Barenys, M., Baumann, J., Dach, K., Tigges, J., Hubenthal, U., Kohrer, K., Fritsche, E., 2018. A transcriptome comparison of time-matched developing human, mouse and rat neural progenitor cells reveals human uniqueness. *Toxicol. Appl. Pharmacol.* 354, 40–55.
- Masjosthusmann, S., Blum, J., Bartmann, K., Dolde, X., Holzer, A.-K., Stürzl, L.-C., Keßel, E.H., Förster, N., Dönmez, A., Klose, J., Pahl, M., Waldmann, T., Bendt, F., Kisitu, J., Suci, I., Hubenthal, U., Mosig, A., Leist, M., Fritsche, E., 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 Support. Pub.* 17, 1938E.
- Meisig, J., Dresner, N., Kapitza, M., Henry, M., Rotshteyn, T., Rahnenfuhrer, J., Hengstler, J.G., Sachinidis, A., Waldmann, T., Leist, M., Bluthgen, N., 2020. Kinetic modeling of stem cell transcriptome dynamics to identify regulatory modules of normal and disturbed neuroectodermal differentiation. *Nucleic Acids Res.* 48, 12577–12592.
- Modafferi, S., Zhong, X., Kleinsang, A., Murata, Y., Fagiani, F., Pamies, D., Hogberg, H. T., Calabrese, V., Lachman, H., Hartung, T., Smirnova, L., 2021. Gene-environment interactions in developmental neurotoxicity: a case study of synergy between chlorpyrifos and CHD8 knockout in human BrainSpheres. *Environ. Health Perspect.* 129, 77001.
- 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., Shafer, T.J., Cowden, J., 2015. Expanding the test set: chemicals with potential to disrupt mammalian brain development. *Neurotoxicol. Teratol.* 52, 25–35.
- Nimtz, L., Hartmann, J., Tigges, J., Masjosthusmann, S., Schmuck, M., Kessel, E., Theiss, S., Kohrer, K., Petzsch, P., Adjaye, J., Wigmann, C., Wiczorek, D., Hildebrandt, B., Bendt, F., Hubenthal, U., Brockerhoff, G., Fritsche, E., 2020. Characterization and application of electrically active neuronal networks established from human induced pluripotent stem cell-derived neural progenitor cells for neurotoxicity evaluation. *Stem Cell Res.* 45, 101761.
- Nimtz, L., Klose, J., Masjosthusmann, S., Barenys, M., Fritsche, E., 2019. The neurosphere assay as an in vitro method for developmental neurotoxicity (DNT) evaluation. In: Aschner, M., Costa, L. (Eds.), *Cell Culture Techniques*. Springer New York, New York, NY, pp. 141–168.
- Nunes, C., Singh, P., Mazidi, Z., Murphy, C., Bourguignon, A., Wellens, S., Chandrasekaran, V., Ghosh, S., Zana, M., Pamies, D., Thomas, A., Verfaillie, C., Culot, M., Dinnyes, A., Hardy, B., Wilmes, A., Jennings, P., Grillari, R., Grillari, J., Zurich, M.G., Exner, T., 2022. An in vitro strategy using multiple human induced pluripotent stem cell-derived models to assess the toxicity of chemicals: a case study on paraquat. *Toxicol. Vitro* 81, 105333.
- Nyffeler, J., Karreman, C., Leisner, H., Kim, Y.J., Lee, G., Waldmann, T., Leist, M., 2017. Design of a high-throughput human neural crest cell migration assay to indicate potential developmental toxicants. *ALTEX* 34, 75–94.
- OECD, 2007. Test No. 426. *Developmental Neurotoxicity Study*.
- OECD, 2021. Guideline No. 497: *Defined Approaches on Skin Sensitisation*.
- Pallocca, G., Grinberg, M., Henry, M., Frickey, T., Hengstler, J.G., Waldmann, T., Sachinidis, A., Rahnenfuhrer, J., Leist, M., 2016. Identification of transcriptome signatures and biomarkers specific for potential developmental toxicants inhibiting human neural crest cell migration. *Arch. Toxicol.* 90, 159–180.
- Pallocca, G., Leist, M., 2022. On the usefulness of animals as a model system (part II): considering benefits within distinct use domains. *ALTEX* 39, 531–539.
- Pallocca, G., Mone, M.J., Kamp, H., Luijten, M., Van de Water, B., Leist, M., 2022a. Next-generation Risk Assessment of Chemicals - Rolling Out a Human-Centric Testing Strategy to Drive 3R Implementation: the RISK-Hunt3r Project Perspective. *ALTEX*.
- Pallocca, G., Nyffeler, J., Dolde, X., Grinberg, M., Gstraunthaler, G., Waldmann, T., Rahnenfuhrer, J., Sachinidis, A., Leist, M., 2017. Impairment of human neural crest cell migration by prolonged exposure to interferon-beta. *Arch. Toxicol.* 91, 3385–3402.
- Pallocca, G., Rovida, C., Leist, M., 2022b. On the usefulness of animals as a model system (part I): overview of criteria and focus on robustness. *ALTEX* 39, 347–353.
- Paparella, M., Bennekou, S.H., Bal-Price, A., 2020. An analysis of the limitations and uncertainties of in vivo developmental neurotoxicity testing and assessment to identify the potential for alternative approaches. *Reprod. Toxicol.* 96, 327–336.
- Patterson, E.A., Whelan, M.P., Worth, A.P., 2021. The role of validation in establishing the scientific credibility of predictive toxicology approaches intended for regulatory application. *Comput Toxicol* 17, 100144.
- Piersma, A.H., van Benthem, J., Ezendam, J., Kienhuis, A.S., 2018. Validation redefined. *Toxicol. Vitro* 46, 163–165.
- Pistolato, F., Carpi, D., Mendoza-de Gyves, E., Paini, A., Bopp, S.K., Worth, A., Bal-Price, A., 2021. Combining in vitro assays and mathematical modelling to study developmental neurotoxicity induced by chemical mixtures. *Reprod. Toxicol.* 105, 101–119.
- Products, E.Panell o.P.P., Residues, t., Hernández-Jerez, A., Adriaanse, P., Aldrich, A., Berny, P., Coja, T., Duquesne, S., Focks, A., Marinovich, M., Millet, M., Pelkonen, O., Pieper, S., Tiktak, A., Topping, C., Widenfalk, A., Wilks, M., Wolterink, G., Crofton, K., Hougaard Bennekou, S., Paparella, M., Tzoulaki, I., 2021. Development of Integrated Approaches to Testing and Assessment (IATA) case studies on developmental neurotoxicity (DNT) risk assessment. *EFSA J.* 19, e06599.
- Ritz, C., Baty, F., Streibig, J.C., Gerhard, D., 2015. Dose-response analysis using R. *PLoS One* 10, e0146021.
- Ryan, K.R., Sirenko, O., Parham, F., Hsieh, J.H., Cromwell, E.F., Tice, R.R., 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., Terron, A., 2019. International regulatory and scientific effort for improved developmental neurotoxicity testing. *Toxicol. Sci.* 167, 45–57.
- Sachana, M., Willett, C., Pistolato, F., Bal-Price, A., 2021. The potential of mechanistic information organised within the AOP framework to increase regulatory uptake of the developmental neurotoxicity (DNT) in vitro battery of assays. *Reprod. Toxicol.* 103, 159–170.
- Schmidt, B.Z., Lehmann, M., Gutbier, S., Nembo, E., Noel, S., Smirnova, L., Forsby, A., Hescheler, J., Avci, H.X., Hartung, T., Leist, M., Kobolak, J., Dinnyes, A., 2017. In vitro acute and developmental neurotoxicity screening: an overview of cellular platforms and high-throughput technical possibilities. *Arch. Toxicol.* 91, 1–33.
- Schmuck, M.R., Temme, T., Dach, K., de Boer, D., Barenys, M., Bendt, F., Mosig, A., Fritsche, E., 2017. Omnisphero: a high-content image analysis (HCA) approach for phenotypic developmental neurotoxicity (DNT) screenings of organoid neurosphere cultures in vitro. *Arch. Toxicol.* 91, 2017–2028.
- Scholz, D., Polt, D., Genewsky, A., Weng, M., Waldmann, T., Schildknecht, S., Leist, M., 2011. Rapid, complete and large-scale generation of post-mitotic neurons from the human LUHMES cell line. *J. Neurochem.* 119, 957–971.
- Sheets, L.P., Li, A.A., Minnema, D.J., Collier, R.H., Creek, M.R., Pepper, R.C., 2016. A critical review of neonicotinoid insecticides for developmental neurotoxicity. *Crit. Rev. Toxicol.* 46, 153–190.
- Shinde, V., Hoeltgen, L., Srinivasan, S.P., Meisig, J., Meganathan, K., Jagtap, S., Grinberg, M., Liebing, J., Bluthgen, N., Rahnenfuhrer, J., Rempel, E., Stoeber, R., Schildknecht, S., Forster, S., Godoy, P., van Thriel, C., Gaspar, J.A., Hescheler, J., Waldmann, T., Hengstler, J.G., Leist, M., Sachinidis, A., 2017. Definition of transcriptome-based indices for quantitative characterization of chemically disturbed stem cell development: introduction of the STOP-Toxukn and STOP-Toxukk tests. *Arch. Toxicol.* 91, 839–864.
- Simon, J.M., Paranjape, S.R., Wolter, J.M., Salazar, G., Zylka, M.J., 2019. High-throughput screening and classification of chemicals and their effects on neuronal gene expression using RASL-seq. *Sci. Rep.* 9, 4529.
- Smirnova, L., Hogberg, H.T., Leist, M., Hartung, T., 2014. Developmental neurotoxicity - challenges in the 21st century and in vitro opportunities. *ALTEX* 31, 129–156.

- Spreng, A.S., Brull, M., Leisner, H., Suci, I., Leist, M., 2022. Distinct and dynamic transcriptome adaptations of iPSC-generated astrocytes after cytokine stimulation. In: *Cells* 11.
- Stiegler, N.V., Krug, A.K., Matt, F., Leist, M., 2011. Assessment of chemical-induced impairment of human neurite outgrowth by multiparametric live cell imaging in high-density cultures. *Toxicol. Sci.* 121, 73–87.
- Strickland, J., Truax, J., Corvaro, M., Settivari, R., Henriquez, J., McFadden, J., Gullledge, T., Johnson, V., Gehen, S., Germolec, D., Allen, D.G., Kleinstreuer, N., 2022. Application of defined approaches for skin sensitization to agrochemical products. *Front Toxicol* 4, 852856.
- Thomas, R.S., Paules, R.S., Simeonov, A., Fitzpatrick, S.C., Crofton, K.M., Casey, W.M., Mendrick, D.L., 2018. The US Federal Tox21 Program: a strategic and operational plan for continued leadership. *ALTEX* 35, 163–168.
- Tohyama, C., 2016. Developmental neurotoxicity test guidelines: problems and perspectives. *J. Toxicol. Sci.* 41, SP69–SP79.
- Tsuji, R., Crofton, K.M., 2012. Developmental neurotoxicity guideline study: issues with methodology, evaluation and regulation. *Congenital. Anom.* 52, 122–128.
- Us Epa Ord, C.f.C.T., 2022. ToxCast Database: Invitrodb Version 3.5. The United States Environmental Protection Agency's Center for Computational Toxicology and Exposure.
- USEPA, 1998. Health Effects Test Guidelines OCSPP 870.6300 Developmental Neurotoxicity Study. Washington, DC.
- van Thriel, C., Westerink, R.H., Beste, C., Bale, A.S., Lein, P.J., Leist, M., 2012. Translating neurobehavioural endpoints of developmental neurotoxicity tests into in vitro assays and readouts. *Neurotoxicology* 33, 911–924.
- Vinken, M., Benfenati, E., Busquet, F., Castell, J., Clevert, D.A., de Kok, T.M., Dirven, H., Fritsche, E., Geris, L., Gozalbes, R., Hartung, T., Jennen, D., Jover, R., Kandarova, H., Kramer, N., Krul, C., Luechtefeld, T., Masereeuw, R., Roggen, E., Schaller, S., Vanhaecke, T., Yang, C., Piersma, A.H., 2021. Safer chemicals using less animals: kick-off of the European ONTOX project. *Toxicology* 458, 152846.
- Worth, A.P., Balls, M., 2001. The importance of the prediction model in the validation of alternative tests. *Altern Lab Anim* 29, 135–144.