Concept Article

Mapping Out Strategies to Further Develop Human-Relevant New Approach Methodology (NAM)-Based Developmental Neurotoxicity (DNT) Testing

Eike Cöllen[1,2], Kristina Bartmann[3,4], Jonathan Blum[1], Kelly Carstens[5], Ivana Celardo[1,6], Nivedita Chatterjee[7], Marco Corvaro[8], Nadine Dreser[1], Ellen Fritsche[3,9,10], Thomas Hartung[6,11], Helena T. Hogberg[12], Thomas Knudsen[13], Katharina Koch[3,4], Anna Kreutz[14], Malene Lislien[15], Viktoria Magel[1,2], M. Sue Marty[16], Giorgia Pallocca[1,6], Anna Bal-Price[17], Costanza Rovida[6], Magdalini Sachana[18], Timothy J. Shafer[5], Lena Smirnova[11], Ilinca Suciu[1], Yaroslav Tanaskov[1,2], Silvia Tangianu[1,6], Chiara Wolfbeisz[1,2] and Marcel Leist[1,6]

[1]In vitro toxicology and biomedicine, inaugurated by the Doerenkamp-Zbinden foundation, University of Konstanz, Konstanz, Germany; [2]INVITE-2, Graduate school of the LBW for in vitro technologies, Konstanz-Sigmaringen, Germany; [3]DNTOX GmbH, Düsseldorf, Germany; [4]IUF – Leibniz Research Institute for Environmental Medicine, Düsseldorf, Germany; [5]Center for Computational Toxicology and Exposure, Office of Research and Development, United States Environmental Protection Agency, Research Triangle Park, NC, USA; [6]CAAT-Europe, University of Konstanz, Konstanz, Germany; [7]NanoSafety Group, International Iberian Nanotechnology Laboratory, Braga, Portugal; [8]Corteva Agriscience Italia, S.r.l., Rome, Italy; [9]Department of Pharmaceutical Sciences, University of Basel, Basel, Switzerland; [10]SCAHT – Swiss Centre for Applied Human Toxicology, Basel, Switzerland; [11]Center for Alternatives to Animal Testing, Department of Environmental Health and Engineering, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA; [12]National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, Division of Translational Toxicology, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA; [13]US EPA Center for Computational Toxicology and Exposure, Research Triangle Park, NC, USA (retired); [14]Inotiv, Morrisville, NC, USA; [15]Department of Chemical Toxicology, Norwegian Institute of Public Health, Oslo, Norway; [16]Dow Chemical Company, Midland, MI, USA; [17]European Commission Joint Research Centre, Ispra, Italy; [18]Environment Health and Safety Division, Environment Directorate, Organisation for Economic Cooperation and Development (OECD), Paris, France

Received January 9, 2025;
Epub January 14, 2025;

Correspondence:
Marcel Leist, PhD,
In vitro Toxicology and Biomedicine,
Dept inaugurated by the Doerenkamp-Zbinden foundation,
University of Konstanz,
Universitaetsstrasse 10,
78464 Konstanz, Germany
(marcel.leist@uni-konstanz.de)

ALTEX 42(2), 308-322.
doi:10.14573/altex.2501091

Abstract

On occasion of the DNT5 meeting in Konstanz, Germany (April 2024), participants brainstormed on future challenges concerning a regulatory implementation of the developmental neurotoxicity (DNT) *in vitro* test battery (DNT-IVB). The five discussion topics below outline some of the key issues, opportunities, and research directions for the next several years: (1) How to contextualize DNT hazard with information on potential maternal toxicity or other toxicity domains (non-DNT)? Several approaches on how to use cytotoxicity data from new approach methodologies (NAMs) were discussed. (2) What opportunities exist for an immediate or near-future application of the DNT-IVB, e.g., as a prioritization step or add-on to other information? Initial examples are already emerging; the data can be used even if the battery is not converted to a defined approach. (3) How to establish data interpretation procedures for multi-dimensional endpoints that reduce dimensionality and are suitable for classification? A decision framework is required on how to use the DNT-IVB in a regulatory context. Machine learning may provide novel classification models. (4) How can a battery of molecular initiating events (MIEs) be smartly linked to the DNT-IVB? At what tier of an overall strategy would MIEs be evaluated, and how would one optimally balance cost vs information yield. (5) What is the way forward to scientific validation of DNT NAMs and the DNT-IVB? A large set of animal data would be required for conventional approaches, while mechanistic information may establish relevance in other ways.

Plain language summary

A meeting on developmental neurotoxicity (DNT) testing was held in Konstanz, Germany in April 2024 (DNT5 meeting). A major topic of discussion was the DNT *in vitro* test battery (DNT-IVB) and how this set of cell-based animal-free test methods may be used in a regulatory context. Opportunities for future developments were addressed in discussion groups: How specific DNT readouts can be combined with more general toxicity data; opportunities for applying the DNT-IVB as soon as possible; and what decision framework will be needed on how to use the DNT-IVB for regulatory purposes. Moreover, the use of signaling assays (evaluating the interaction of test compounds with receptors, enzymes and transporters) in combination with the DNT-IVB was discussed. Finally, ideas and concepts to achieve scientific validation of the DNT-IVB were collected.

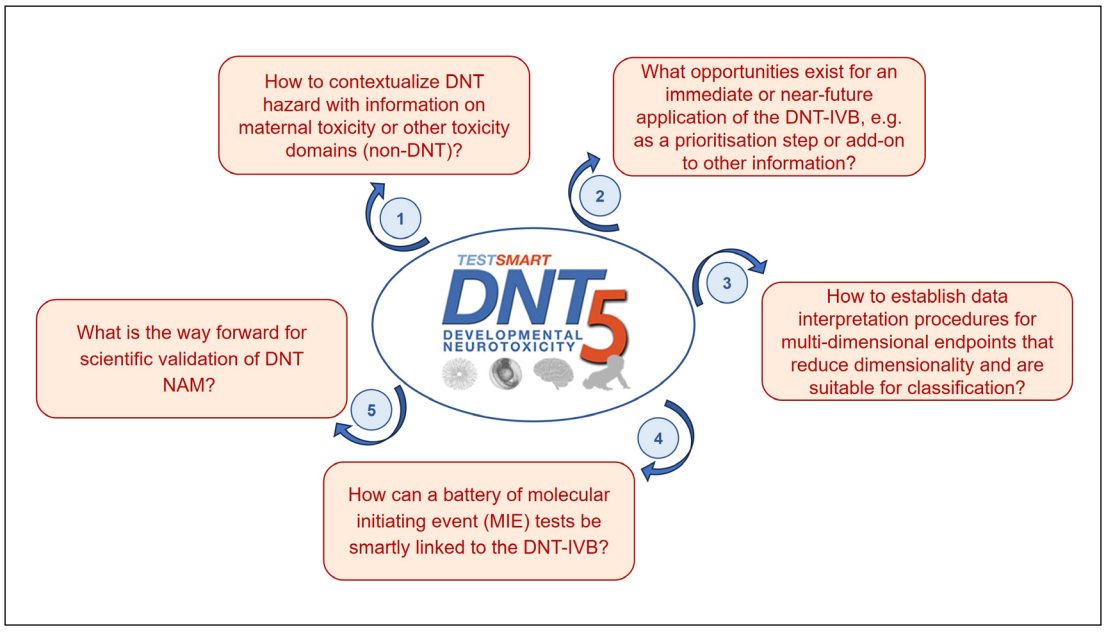

Fig. 1: Key topics discussed at breakout sessions of the DNT5 meeting in Konstanz

1 Setting the stage

The 5th International Conference on Developmental Neurotoxicity Testing (DNT5) took place in Konstanz, Germany on April 7-10, 2024 (Celardo et al., 2025). It was organized by the Center for Alternatives to Animal Testing in Europe (CAAT-Europe) and brought together stakeholders from around the globe, including research scientists, regulators, industry representatives, non-governmental organizations, and service providers. They discussed[1] the actions to take towards the implementation of predictive, time-efficient and human-relevant *in vitro* DNT methods in risk assessment and regulatory decision-making (detailed program[2]). The scientific organizing committee included experts from CAAT at both the University of Konstanz and Johns Hopkins University, the Joint Research Centre (JRC) of the European Commission (represented by EURL-ECVAM), the Swiss Centre for Applied Human Toxicology (SCAHT), the US Environmental Protection Agency (US EPA), and the US National Institute of Environmental Health Sciences (US-NIEHS) (details on webpage[3]).

Many conference talks focused on the implementation of the DNT *in vitro* test battery (DNT-IVB), co-developed with the European Food Safety Authority (EFSA) and the US EPA. The OECD Initial Recommendations on Evaluation of Data from the DNT-IVB (OECD, 2023a) refers to the evaluation of data from the DNT-IVB and represents an important step towards the full acceptability of an *in vitro* approach for the assessment of DNT (see overarching summary[4]). However, the way towards a full replacement of animal testing in all regulatory contexts requires further work. To organize this efficiently, it is important to clearly define the challenges and the potential solutions. The conference included a session on *Hot Topics for Novel Testing Strategies in DNT*, in which participants worked on the key challenges in small groups (Fig. 1). The outcome of their discussions is summarized below.

2 Challenge 1: How to contextualize DNT hazard with information on maternal toxicity or other toxicity domains (non-DNT)?

2.1 Outline of the challenge

The traditional animal-based test strategy assesses not only DNT endpoints but also quantifies maternal toxicity. Moreover, other

[1] The discussion summaries compiled in this article should not be understood as views of the institution of employment of the participants. This article is neither a consensus statement nor a complete roadmap for future developments. The balance of some topics (according to their importance or weight) might not be perfect. Moreover, not every concept is fully explained with all origins, implications and applications.

[2] https://tinyurl.com/4tz7ks4a

[3] https://www.uni-konstanz.de/dnt5/about-the-conference

[4] https://tinyurl.com/5h7yf7rx

Abbreviations: AOP, adverse outcome pathway; BMR, benchmark response; C&L, classification and labelling; DA, defined approach; DART, developmental and reproductive toxicity; DIP, data interpretation procedure; DNT-IVB, developmental neurotoxicity *in vitro* battery; EOGRTS, extended one-generation reproductive toxicity study; IATA, integrated approaches to testing and assessment; ITS, integrated testing strategies; IVIVE, *in-vitro*-to-*in-vivo* extrapolation; KE, key event; KNDP, key neurodevelopmental process; MEA, multi-electrode array; MIE, molecular initiating event; ML, machine learning; NAM, new approach methodology; NGRA, next generation risk assessment; OECD, Organisation for Economic Cooperation and Development; POD, point of departure; WoE, weight-of-evidence

forms of developmental and reproductive toxicity (DART) may occur. This may lead to an indirect (non-specific) DNT readout, and thus a misclassification. When DNT is only observed under conditions of pronounced maternal toxicity, or when DNT only occurs at doses much higher than other toxicity, for example, developmental bone malformations, compounds are usually not classified as DNT toxicants. The concept of relative toxicity, i.e., comparing different test endpoints, has also been applied to (some) NAMs. For instance, (i) functional impairments (e.g., inhibited migration of neural precursors or inhibited neurite outgrowth) are contrasted with general cytotoxicity, or (ii) in the ECVAM-validated embryonic stem cell test (EST) the impaired differentiation of stem cells to cardiac cells is compared to effects on non-neural cells, e.g., fibroblasts (Seiler and Spielmann, 2011).

A further issue to be considered is that some legislations require that experimental data (usually obtained in maximum tolerated dose studies on the same animal species) need to be obtained to set the doses used for DNT studies. For instance, the REACH regulation (EC, 2006) stipulates that the highest DNT study dose level should be in the range of the maximum tolerated dose for general toxicity. This type of top-concentration anchoring may be taken into consideration for the *in vitro* testing strategy.

2.2 Guiding questions for discussing the challenge:

Q1: Is it feasible to suggest (and evaluate) a standard "maternal toxicity" predictor within a DNT *in vitro* test battery?

Q2: For assays that measure general cell function, such as cytotoxicity or inhibited proliferation: How can they be qualified to yield DNT-specific information?

Q3: Is there a general and reliable guidance to ensure that effects in functional endpoints relevant to NAM are not caused by non-specific cytotoxicity, or by effects on other organs (e.g., in model organisms)?

2.3 Points to consider for further progress

The current *in vivo* strategy for dose-setting is related neither to exposure nor to bioactivities. The anchoring to a maximally tolerated dose is very strict, and studies performed with lower doses, e.g., in the context of extended one-generation reproductive toxicity studies (EOGRTS), are likely to be rejected.

The DNT-IVB was not designed to reveal maternal toxicity (Blum et al., 2023; Carstens et al., 2022). This is not considered necessary under the framework of next generation risk assessment (NGRA), where the focus is on testing the safety of relevant exposure levels. However, the possibility to add an *in vitro* range finding study to determine a point of departure (POD) for general toxicity on a diverse set of cell types could be considered. Alternatively, ToxCast data or the scientific literature could be mined for a general cytotoxic threshold concentration. There was also discussion on whether *in vitro* and *in silico* information could be used to facilitate dose setting for OECD TG 426 (OECD, 2018a) or for EOGRTS studies: (i) modelling data could be used to derive relevant internal exposures instead of going for maximal tolerated exposures; (ii) NAM-derived cytotoxicity data may guide dose-range finding studies by setting an upper threshold for a realistic internal exposure.

It was generally agreed that cytotoxicity, trivial as it may sometimes appear, can be a complex issue, and its interpretation often requires specific knowledge on the assay platform and on the overall context. Three examples illustrate the various aspects of cytotoxicity:

1. *Setting the POD for cytotoxicity in a NAM:* Different approaches are found in the literature, and there is no consensus yet on the best compromise to ensure sufficient sensitivity without losing necessary specificity. Different endpoint measures for viability and cytotoxicity (e.g., ATP content vs lactate dehydrogenase release) may yield different results; moreover, some viability and cytotoxicity measures provide data on cell populations (e.g., resazurin reduction) while others yield information on individual cells (e.g., nuclear dye uptake). Usually, the POD is defined by a benchmark response (BMR), i.e., a deviation from the baseline. This may be 5%, 10% or 20%, or it may be a dynamic value determined by the baseline variability. Thus, the highest non-cytotoxic concentration of a compound in a given test system is not just determined by the test item characteristics but also depends on the BMR, the analytical method, and the selected modelling approach (curve fits, etc.) (Keßel et al., 2023).
2. *Cytotoxicity may serve as reference point for other (functional or more specific) endpoints of a NAM:* For instance, in the UKN2 / NPC2 (migration assays) or UKN4 / NPC4 (neurite assays) tests of the DNT-IVB, cytotoxicity is measured in addition to a functional endpoint (migration/neurite growth). The overall assay outcome uses a ratio of functional endpoint metrics and cytotoxic potency. Thus, data on cytotoxicity are an essential part of the data interpretation procedure (DIP). In this context, cytotoxicity measurements affect the classification of a compound as hit or non-hit. Altering the method to assess cytotoxicity or neglecting the cytotoxicity endpoint would result in different assay outcomes and, possibly, drawing incorrect conclusions.
3. In some cases, cytotoxicity may be the main functional endpoint of a NAM. This is in line with the classification of cell death (apoptosis) as one of the key neurodevelopmental processes (KNDP). In some NAMs, cytotoxicity may manifest as inhibited proliferation (i.e., reduced cell numbers at the end of the assay exposure period). Proliferation of, e.g., neuroprogenitor cells or neural crest cells is also a KNDP. This is evidenced by specific neural tube defects after exposure to anti-proliferative agents (Guan et al., 2015; Wang et al., 2018). Thus, there are cases where cytotoxicity indicates disturbance of a KNDP (e.g., NPC1 assay). The interpretation of such data can be a challenge. A frequently used solution is to compare the cytotoxic potency of a compound on different cell types. If the viability or proliferation of a developing cell sub-population is disturbed more potently than that of an adult and/or non-neuronal cell

model, then this may indicate a specific developmental toxicity. Several adult cell types may be required for such a comparison. In some (*in vivo*) cases, the effects of anti-proliferative agents are not determined by relative potencies but by the timing of exposure. The toxicant methylazoxymethanol is known for this effect (Penschuck et al., 2006). Such compounds are not a problem in practice if the DNT-IVB is used in parallel to other NAM-based evaluations. Proliferation inhibitors are likely to be identified in other screens (Jaklin et al., 2022).

The above considerations are important because the discrimination between a "general toxicant" and a compound causing DNT determines classification and labelling (C&L) (EC, 2008). While classification as a general toxicant, e.g., a hepatotoxicant (specific target organ toxicity, STOT-RE), has few regulatory consequences, classification as a developmental and reproductive toxicant (DART) has major consequences: the compound then falls into the carcinogenic, mutagenic, and reproductive toxicant (CMR) category. DNT is part of the DART category. This highlights the need for reliable guidance that ensures that NAM-based "DNT endpoints" are not caused by non-specific cytotoxicity (i.e., cytotoxicity that occurs in a NAM of the DNT-IVB at concentrations that affect a functional (DNT-specific) endpoint). When using model organisms, care needs to be taken to ensure that DNT-specific endpoints are not altered because of cytotoxicity or other adverse effects on other organs (outside the nervous system). The discussion on how NAMs may be used for C&L is still at an early stage.

Another discussion topic was the aspect of time; it was noted that exposure duration and time of sampling (assessment of endpoints) can massively affect NAM outcomes.

1. For instance, different cell subpopulations may be killed at different times. This can make detection difficult and complicates long-term evaluation.
2. Another neglected aspect is the acute short-term neurotoxicity of anesthetics (or narcotics). This may trigger DNT when used during pregnancy or when infants are exposed. Such effects are not well captured by standard cytotoxicity measures in NAMs. Only a combination of functional NAMs may be informative.
3. Another complication is how non-lethal manifestations of cytotoxicity affect various cell functions. For instance, there are hundreds of examples where a certain experimental condition leads to death of > 50% of all cells in a population within 24 h, but where no cell death is measurable under the same conditions after 3 h. It is very hard to determine whether or not a functional property of the cells (e.g., certain neuronal signaling functions), assessed after 3 h, is affected by the cytotoxicity that has already been initiated but only becomes measurable much later (Waldmann et al., 2017).

The above examples indicate that knowledge on overt cytotoxicity or even non-lethal cell stress responses can be important for the interpretation of assay results and for assessment of the toxicological implications of assay hits (Jaklin et al., 2022; Judson et al., 2016; Krug et al., 2013; Leist et al., 2010; Shah et al., 2016). This information should contribute to weight-of-evidence (WoE) approaches in the context of integrated approaches to testing and assessment (IATAs).

3 Challenge 2: What opportunities exist for an immediate or near-future application of the DNT-IVB, e.g., as a prioritization step or add-on to other information?

3.1 Outline of the challenge

The assembled DNT-IVB, as described in the Initial Recommendations document (OECD, 2023a), derived from work published earlier by Blum et al. (2023) and Carstens et al. (2022), already provides a large set of ready-to-use test methods for certain applications. Several additional tests have been proposed and/or developed for complementation of the battery (Tal et al., 2024) in a tiered testing strategy. These assays are intended to indicate adverse effects on the differentiation and signaling of various neural cell types, or they add new biological components (e.g., neuroinflammation) or behavioral readouts (battery of tests available in zebrafish).

While this extended platform may be used for many questions concerning human safety, practical use examples are the most important driving force for further developments. Case studies inform about the opportunities and challenges related to expanding, condensing or structuring the DNT-IVB (e.g., in tiers), support the transfer to contract research organizations, and show how to best interpret the outcome data. It has become clear that the DNT-IVB is already applicable for prioritization and screening, while more time is likely needed before its application in full risk assessment of, e.g., pesticides and drugs. However, there may be other immediate applications for specific questions besides compound screening. It is important to involve all relevant stakeholders to actively identify opportunities to use the DNT-IVB.

3.2 Guiding questions for discussing the challenge:

Q4: Are there areas of safety evaluation in industry where the DNT-IVB could be added to existing methods to reduce uncertainties or to provide additional information or to support investigative toxicology?

Q5: Are there regulatory questions (e.g., on contaminants and metabolites) where data from the DNT-IVB may be used to fill information gaps?

Q6: What are the most important data/information needs for using the DNT-IVB from now on in filtering and pre-screening functions?

3.3 Points to consider for further progress

In the scope of regulatory applications, the DNT-IVB could be used as a screening tool in addition to existing methods. Data from the battery could help prioritize chemicals within large groups before performing safety assessments using animals. Examples, where such use has already occurred, include organophosphorus flame retardants and PFAS (Carstens et al., 2023; Klose et al., 2022; OECD, 2022b). This use may be extended to food contact materials, food colorants and dyes, or various metals. One way

to structure and coordinate such an approach is the DNT Health Effects Innovation (HEI) program within the Division of Translational Toxicology (DTT) at US NIEHS, which has prepared lists of chemicals to be tested using DNT-IVB assays.

Regarding regulatory applications, it is important to note that DNT-IVB data may already now be used in a WoE context in combination with animal and epidemiological data. Examples for the support of positive (Dobreniecki et al., 2022) or negative (Carstens et al., 2022) classifications are already available. WoE methodologies are a key element of IATAs. This framework, which has been promoted by the OECD, allows the combined use of various model systems and data categories and permits the incorporation of DNT-IVB data, as exemplified by a case study on deltamethrin (OECD, 2022f). However, this requires the conversion of *in vitro* concentration data into corresponding human doses by *in-vitro*-to-*in-vivo* extrapolation (IVIVE). This procedure requires physiologically based kinetic (PBK) models that are parameterized concerning, e.g., blood-brain barrier and placental barrier permeability. The challenge of high uncertainties with such data and procedures is currently being addressed by activities of the OECD towards a guidance document on IVIVE. In parallel, more basic research is required to understand barrier transport of compounds that are transporter-dependent.

Another research priority to further evolve the DNT-IVB is the conversion of assay hits to an overarching statement on DNT hazard, ideally combined with an uncertainty assessment (Paparella et al., 2020; Smirnova et al., 2024). It has been suggested that additional data and experimental methods may be used, possibly in a tiered testing strategy, to follow up on hits (chemicals identified as active in the DNT-IVB) (Magel et al., 2024). In parallel, it was suggested that some negative results may also need to be discussed in light of additional data. In this context, consideration would need to be given, e.g., to the role of metabolites for potential *in vivo* (human) toxicity (Blum et al., 2023). One potential development of the DNT-IVB is that it may be broadened, e.g., by addition of MIE and KNDP assays, and an increased use of model organisms from many further developers, e.g., in the Partnership for the Assessment of Risks of Chemicals (PARC) project (Tal et al., 2024) or supported by the US NIEHS (Kreutz et al., 2024). In a subsequent development phase, it may be determined which assays are redundant and how a resource-efficient tiered strategy may be constructed.

As DNT-IVB data has already been used successfully for IATA development and to support the construction of adverse outcome pathways (AOPs) (Crofton et al., 2024; Grillberger et al., 2023; Koch et al., 2022; OECD, 2022a,c-f), it is advisable to develop guidance for data users on how to retrieve DNT-IVB data, e.g., from ToxCast (using the CompTox Chemicals Dashboard), and how to apply this information in an IATA. The regulatory uptake of the DNT-IVB requires training of regulators and data producers (industry) in evaluating such data. It is also important to make it more broadly known that extensive assay descriptions (in ToxTemp format (Krebs et al., 2019)) are available for all DNT-IVB assays. These descriptions may be found in Appendix B of the OECD Initial Recommendations document (OECD, 2023a). This extensive background material represents a useful starting point to review the readiness of assays in the DNT-IVB and to understand data generation (Blum et al., 2025), relevance, and uncertainties.

4 Challenge 3: How to establish data interpretation procedures (DIPs) for multi-dimensional endpoints that reduce dimensionality and are suitable for classification?

4.1 Outline of the challenge

The setup and performance evaluation of test methods has been covered by regulations and guidance. They generally refer to single methods, usually measuring one endpoint. Only very few defined approaches (DA), i.e., the algorithmically combined use of more than one NAM (and data thereof), have been established, i.e., for acute local ocular irritation (TG 467) (OECD, 2024) and skin sensitization (TG 497) (OECD, 2023b). Test batteries have also been assembled for estrogen receptor agonist activity (Miller et al., 2017) or for androgen receptor interaction (Judson et al., 2020). Evaluation schemes have been suggested for such integrated testing strategies (ITS), but these are not formalized at a regulatory level. The topic of ITS has been discussed for decades, and the old concept has been largely replaced by the IATA concept (see Challenge 5). The solutions that made it into regulatory toxicology are very limited. The issue of how to stage an optimal tiering scheme and how to combine test endpoints in a universal DIP has not yet been solved. This challenge is also pertinent to NAM-based DNT testing.

For instance, multi-electrode array (MEA) assessments may have 10-30 endpoints, and gene expression datasets may have up to 20,000 endpoints. Moreover, several such endpoints may be combined in even more combinatorial endpoints (e.g., the expression changes of several genes may be combined in many ways to biological "signatures", such as gene ontologies, pathways or modules of a weighted gene correlated network analysis (WGCNA) map). Two major problems arise from this: (i) the false discovery rate can increase under such conditions; (ii) quantification of a test outcome, i.e., the setup and performance evaluation of a DIP (sometimes also termed prediction model, PM), is challenging. The DIP converts test outcome data (numbers with or without a unit) into a toxicological statement (considered "toxic" or considered "non-toxic") (Blum et al., 2025). A typical set of questions in this context is how many tests or endpoints a compound should activate (hit status) to be considered a toxicant. For transcriptome assays the question may be, how to establish the number of genes (and/or their fold change) that needs to change to consider a compound a DNT-toxicant. This requires rules on which relevance is assigned to each specific gene set and, for instance, whether up-regulations are assigned the same weight in a WoE approach as down-regulations (Cherianidou et al., 2022; Pallocca et al., 2016; Shinde et al., 2017; Waldmann et al., 2014).

Combining transcriptome analyses with functional readouts can help understand hit patterns, i.e., changes in gene expression that are followed by a relevant cell biological response (Dreser et al., 2020; Klose et al., 2021, 2022, 2023; Meisig et al., 2020).

4.2 Guiding questions for discussing the challenge:

Q7: Is there a biological and statistical rationale for considering each and any endpoint as independent; should all NAM data be given equal weight/importance?

Q8: How to deal with endpoint combinations – are they independent endpoints?

Q9: Can one use a test system without a defined DIP?

4.3 Points to consider for further progress

In dealing with multiple endpoints/NAMs, some fundamental decisions need to be made. First, it is important to distinguish whether one is dealing with a "test battery" (in the sense of an ITS) or a "battery of tests". The former often has tiers, decision points, and complex but defined DIPs. The latter is a compilation of independent assays (or endpoints), usually performed in parallel. The ITS outcome would be used on a case-by-case basis to give additional information to a group of experts, e.g., in the context of an IATA. A second question, not entirely unrelated to the first, is how performance measures would be obtained. There is a broad range of options. On one end of the spectrum are, e.g., formal validation and the statistics-based design of a DIP (e.g., by machine learning (ML)). Measures of test sensitivity/specificity may be obtained from this process. On the other end of the spectrum, data would be used on a compound-by-compound basis in a WoE process that also draws on expert knowledge and many other types of information. Classical performance measures would not be obtained initially. The confidence-building process would rather be based on a series of case studies.

An important note in this context is that there is often an overrepresentation of positive controls and toxicants when test (battery) performance is evaluated. However, negative controls are essential for calculations of test method specificity. As balanced accuracy (the standard performance metric for all tests) is calculated from sensitivity and specificity, overall test performance cannot be calculated in the absence of a sufficient number of controls. It is therefore important that data on a large number (20-100, dependent on the scope) of controls is collected over time. Guidance on the choice of such control items is already available (Aschner et al., 2017; Kadereit et al., 2012; Martin et al., 2022). Several ideas on multi-endpoint data integration have been developed for exemplification purposes (Jochum et al., 2024; Kreir et al., 2024).

It should be noted that the latter approach is used for data from traditional animal experimentation: There is no overall DIP that integrates the many endpoints (histopathological, functional, behavioral, and clinical). Although experienced evaluators may look for consistent patterns, decisions can be taken based on individual endpoints. Moreover, false discovery rates are considered negligible (no statistical corrections are performed). In other words, the significance of a single endpoint (out of dozens to hundreds) is considered relevant for determining the study outcome. A big question in the field is whether this approach may be transferred to NAM batteries. At present, the DNT-IVB is mostly such a battery of tests.

Part of the discussion was concerned with ML approaches as a potential solution to interpret results from the DNT-IVB. It was suggested that experience may be gained from other fields, where ML is applied to a set of equally weighted endpoints. Another suggestion of the group was that all stakeholders (e.g., regulators, industry and researchers) need to tightly collaborate to determine best practices for data interpretation: statistical prediction models for multi-dimensional endpoints, established by researchers, are sometimes not aligned with regulatory needs. Research scientists may build complex ML models to refine small statistical details, while regulators and the regulated community need transparent models that allow clear decisions.

The "training", i.e., the setup of a prediction model, e.g., by ML, also presents some issues. Training a model from the DNT-IVB on *in vivo* animal study data includes many unknowns and uncertainties, as the animal data used as input may have inherent shortcomings, but also as the transferability of animal data to humans is not known for most compounds. Even with clear input data, the establishment of classifiers is more challenging than it may appear. For instance, a binary output of "DNT hazard" and "no DNT hazard" is appealing for regulatory decision making. However, such classifications are problematic for data in the borderline range (i.e., close to the classification threshold). Solutions to this are the addition of a borderline range (Gabbert et al., 2022) or a specific uncertainty analysis for compounds close to the borderline (Delp et al., 2018). Alternatively, one may choose a non-binary, probabilistic output (Leist et al., 2014; Maertens et al., 2022). Major challenges of the latter approach are (i) the biological interpretation of the model results and (ii) the availability of training chemicals. The value of purely statistical models may be enhanced by coupling them to biological concepts like AOPs. To further expand this concept, models may be developed that indicate the likelihood of progression within an AOP.

When building a classifier, a well-known challenge is the "curse of dimensionality"[5]. This often leads to "overfitting", when a model has many more features (i.e., endpoints, e.g., genes) than samples (i.e., number of experiments that measure all the endpoints in the presence of different test chemicals). A practical consequence of this is that the classifier performs well with the training data but may perform poorly with a test dataset. The reason is usually that the model is too closely aligned with a limited set of data points (often due to lack of more available data points/samples) and therefore potentially fails to predict additional samples or future observations. To avoid this, a rule of thumb suggests that five times more reference chemicals should be used than model features (or endpoints)[5]. Transcriptome analyses rarely follow this rule. Even with an IVB with relatively few endpoints (e.g., 20-40 endpoints for the DNT-IVB), ≥ 100-200 calibration chemicals are required to avoid overfitting. A workaround is to combine endpoints into a new measure that is considered a single endpoint. This procedure is known as dimensionality reduction, like the weighted gene correlated network analysis (WGCNA) in toxicogenomics (Kadarmideen and Watson-Haigh, 2012; Lang-

[5] https://en.wikipedia.org/wiki/Curse_of_dimensionality

felder and Horvath, 2008). This allows building toxicogenomics mappers (Callegaro et al., 2021; Schüttler et al., 2019) with a limited number of modules (e.g., 50-200) compared to the total number of genes (about 20,000). Notably, not all endpoints account for a full additional dimension. Sometimes endpoints are highly correlated with one another (e.g., various cell death measures or various oxidative stress-responsive genes). This may also apply to various DNT-IVB endpoints and needs to be considered in classifier-building.

5 Challenge 4: How can a battery of molecular initiating event (MIE) tests be smartly linked to the DNT-IVB?

5.1 Outline of the challenge

Ideally, some DNT-related hazard information would be available for all chemicals, but this is not the case. Instead, such information is available for only a small fraction of chemicals on the market: Until 2024, DNT guideline studies had been performed for only 230 chemicals. Data and regulatory conclusions from these studies have now been compiled and are scheduled for release/publication in early 2025[6]. Amongst the tested chemicals, there are many pesticides (n = 104). The reason for this is that a data requirement for DNT testing is triggered for plant protection products by certain "alerts". These are, e.g., chemical structural features, a known neurotoxic mode of action or certain observations (clinical signs of neurotoxicity and routine pathological endpoints that indicate neurotoxicity). At present, alerts revealed during the performance of acute, short- and long-term studies in young adult rats play a dominant role. The handling of such alerts and the waiving of testing requirements may differ between the European Union and other countries, and it is affected by the type of chemical under investigation.

For industrial chemicals, the production tonnage and alerts determine whether DNT data are required as part of an extended one-generation study (EOGRTS, OECD TG 443) in the European Union. Alerts may also trigger a guideline DNT study (according to OECD TG 426). EOGRTS has been in use only for a few years, and DNT data from this procedure are only available for 45 compounds. Biocides have special regulations. Until a few years ago, DNT assessment was rare in this domain, but data production appears to have accelerated. It is likely that some more data will become available in the near future. This overview shows that classical regulations limit DNT assessment to the most relevant subset of chemicals (highest likelihood of a potential hazard). As only about 0.1% of industrial chemicals have been assessed for DNT (Fritsche et al., 2017, 2018), there is limited information in this toxicological area. An efficient and broad use of the DNT-IVB may fill this knowledge gap and allow testing of more chemicals in commercial use.

For reasons of cost and resource limitations, comprehensive *in vitro* DNT studies may not be feasible for every chemical. It is thus important to develop decision trees to guide when and how DNT testing is initiated. One trigger could be a screen of the most relevant biochemical DNT targets (e.g., receptors and ion channels) in the course of a general/overall *in vitro*/*in silico*-based hazard assessment. Various cell-based phenotypic assays may be added to such a battery to extend biological coverage. Such a basic NAM-based screen could become a standard requirement also for traditional test programs. Pre-clinical assessment of substances within the pharmaceutical industry sector uses such approaches already. As screen throughput and speed are increasing, this may be extended to most other sectors.

Within an NGRA strategy (purely NAM-based), the outcome of an initial screen (as described above) may determine further steps. If no bioactivity is detected, the compound would be classified as low priority (low hazard potency). Further steps would then depend on the problem formulation of the investigation. For instance, a second tier of testing may be used for compounds with expected human exposure or a high likelihood of having some bioactivity. If a test compound interacts with certain targets or is a hit in certain cell-based assays (e.g., a MEA-based assay for neurofunctional or cardiac effects), decisions need to be taken on how this information is interpreted and whether more resource-intensive assays should be run. One of the follow-up steps may be a full characterization in the DNT-IVB.

In order to use MIE assays as either alerts, extensions or follow-ups of the DNT-IVB, it is important to consider the concentrations of test compounds to be tested. The two obvious options are: (i) testing at "relevant" concentrations or (ii) testing at a very high concentration. There are advantages and disadvantages to both approaches. Testing at concentrations corresponding to realistic exposures leads to relevant and biologically (and toxicologically) more meaningful data. The downside is that there is a higher likelihood of obtaining false negative or inconclusive data. Testing at high concentrations, e.g., corresponding to a dose of 1000 mg/kg bodyweight, generates fewer false negatives but is likely to be associated with false positives. This may require a highly resource-demanding follow-up of the hits (Smirnova et al., 2024).

5.2 Guiding questions for discussing the challenge:

Q10: What could be the criteria, within an NGRA strategy for a data-poor compound, to initiate a DNT evaluation (e.g., by the DNT-IVB)?

Q11: Are there parameters/criteria that would suggest not to test for DNT or to stop testing?

Q12: How can one assemble a commonly accepted (ideally also by regulators) set of trigger (*in vitro*) tests for DNT testing?

5.3 Points to consider for further progress

When talking about MIE assays, it is important to note that there is very little experience with this in the DNT field. A set of tests would need to be assembled, and pilot studies/case studies would be needed to understand their performance. In a first step, a comprehensive list of relevant molecular targets needs to be estab-

[6] Crofton, K. M. (2025). Deliverable for EFSA contract #EOI/EFSA/2022/01, in press.

lished (Burbank et al., 2024). There are several approaches to compiling a list of potential MIEs. First, one may include typical MIEs and key events (KEs) of established AOPs relevant to DNT. Second, one may leverage knowledge on the many signaling pathways known to be relevant to DNT. Picking the molecular triggers and modifiers from these would complement the list of relevant targets. Third, one may scrutinize the list of chemicals known (or highly suspected) to trigger DNT and extract their known targets from the literature. The collective list would include mitochondrial toxicants, cytoskeletal toxins, compounds interfering with important neurotransmitter receptors or transporters, inhibitors of neurotransmitter degradation, and several cell signaling modifiers that modify tyrosine kinase activity, the Notch pathway, Wnt signalling, and some kinase cascades or nuclear receptor systems.

Once a comprehensive list of relevant targets has been assembled (and possibly sorted according to priorities), the questions of screening technology and capacity need to be addressed. It is likely that quantitative structure-activity relationship (QSAR) models and other *in silico* approaches (e.g., virtual docking) will have a high future performance (accuracy, capacity to make statements on potency, etc.). They may be a core component of an initial screen of such a target panel. Besides the building of computational models, read-across would be an important method (Rovida et al., 2021). A variation of classical, structure-based read-across that is important to consider here is functional read-across, e.g., read-across based on similar outcome patterns in a set of *in vitro* assays (Vrijenhoek et al., 2022; Rovida et al., 2021). Addition of several biochemical assays (e.g., assessing binding to or modulation of target functions) would at present be a necessary component of an initial screen.

The use of data from MIE assays may depend on the problem formulation and on additional available information. Some information from MIE assays clearly represents a DNT alert and would already suggest a serious hazard potential that would need to be clarified. An example may be compounds that potently activate glutamate or nicotinic receptors.

MIE testing (testing for interactions between a test chemical and a target) is not considered suitable as a general and only criterion to initiate the full DNT-IVB. The number of AOPs and thus also the number of MIEs clearly associated with DNT is still too limited, and more basic work is required (see above). While a comprehensive set of alerts cannot be defined on the basis of current DNT knowledge, a list of potential candidates may contain typical targets of neurotoxicity (Masjosthusmann et al., 2018). The use of few, many or all such targets for testing would depend on available resources and on the problem formulation. At present, such a testing strategy may be used as a high-throughput approach to generate at least some information on the large number of lower production volume chemicals that are relatively data-poor. They may then be prioritized for further testing, e.g., in the DNT-IVB.

For very data-rich chemicals (e.g., drugs, pesticides), screening in the DNT-IVB may be a default approach. In this setting, a battery of MIE assays may take two other roles: (A) increasing the sensitivity of the battery (further reducing the false negative rate) and (B) hit follow-up for mechanistic characterization and toxicological risk assessment.

Regarding A, one of the potential outcomes of the testing of a compound in the DNT-IVB is that there are no hits within a relevant concentration range. One regulatory consequence could be that the compound is considered "negative", i.e., not associated with a DNT hazard (comparable with an *in vivo* testing outcome where the no-observed effect level (NOEL) is at the highest tested dose). For regulators to accept such a conclusion, they would like to be confident that the test battery is highly sensitive and that the likelihood for a false negative is very low (equal or lower than in the OECD TG 426 study). Adding some MIE assays to the DNT-IVB, or performing them in a second tier, may be an approach to increase such confidence. For instance, opioid, cannabinoid and nicotinic receptors may be added. Possibly also components of the thyroid signaling system may be tested in addition, if such information is not available from other test batteries.

Regarding B, the KNDP assays of the current DNT-IVB are "phenotypic assays" that capture a large group of targets and pathways. Once a compound is an assay hit, a sequence of follow-up activities is required to support toxicity predictions. One of those is providing a mechanistic understanding of the compound and thus identifying its target. Such knowledge allows more robust predictions on whether findings from the test method (model) are relevant for humans. To some degree, mechanistic studies can be performed within the DNT-IVB assays. This can also lead to target identification. In some cases, the phenotypic assays can be converted to target assays by focusing the response to few or single targets (Barenys et al., 2017; Bartmann et al., 2023; Klose et al., 2021; Nyffeler et al., 2018). A straight-forward example is an MEA assay or a Ca^{2+} signaling assay that reacts to many modulators of receptors, transporters, and ion channels, or the neurosphere differentiation assay, capturing a wide variety of pathways relevant for migration and differentiation. If a hit is obtained, one may follow up on this by using a panel of specific inhibitors and activators and observing how they change the test compound response (Loser et al., 2021b,c). An expansion of this approach is to use the test system that forms the basis of the relevant DNT-IVB NAM and to add other endpoints to it. In other cases, additional assays, including an MIE battery may be required. Especially in drug discovery, many strategies have been developed to identify targets from phenotypic screens, e.g., using combined omics technologies and/or advanced bioinformatics approaches (Chaput et al., 2020; Meier et al., 2024; Suciu et al., 2023; Vincent et al., 2020, 2022). An important experience from drug screening is that hit follow-up not only has the function of defining mechanisms and targets but also is essential to eliminate false positives from initial screens. This is an important consideration also for the acceptability and practical implication of a decision strategy based on the DNT-IVB.

There are also approaches that combine advantages of phenotypic assays (which provide broad coverage of molecular targets and pathways and are high-throughput) with advantages of MIE assays (which provide detailed mechanistic information). Such as-

says comprise high-throughput transcriptomic screening and phenotypic profiling (cell painting) (Culbreth et al., 2021; Harrill et al., 2021). If they can be performed in DNT-relevant test systems (e.g., neurons, astrocytes or precursor cells) (Cherianidou et al., 2022; Colaianna et al., 2017; Jaklin et al., 2022; Klose et al., 2021, 2022; McDiarmid et al., 2024; Pallocca et al., 2016), they may be added to the DNT-IVB.

A final topic of discussion asked whether there should be not only alerts but also exclusion criteria for DNT testing. Exclusion criteria can be pivotal in decisions regarding the optimal use of resources (e.g., time, money, labor). Exposure conditions that preclude the build-up of relevant concentrations of the test compound in the nervous system (e.g., no pre- and/or postnatal brain exposure) was considered the most important criterion. This is very much in line with the NGRA approach, which offers a great opportunity to advance human health protection by providing tools for addressing new toxicological challenges like DNT (Schmeisser et al., 2023). In the absence of information on brain levels, the concentrations in maternal blood at early stages of pregnancy may be used as a trigger for inclusion (or exclusion). A more sophisticated set of rules may include information on the potency of the compound, but this may only work for test substances with rich pharmacological background information (potency in the DNT-IVB in relation to other types of toxicity is discussed in Challenge 1).

6 Challenge 5: What is the way forward for scientific validation of DNT NAMs

6.1 Outline of the challenge

Test method validation is important for regulatory acceptance of data from the DNT-IVB. Principles of assay validation are laid down in OECD GD 34 (OECD, 2005). These criteria were developed for individual assays and are currently under revision by the OECD test guideline program. The introduction of batteries of tests to be assembled in fixed interpretation procedures such as a DA poses new challenges for validation. There is currently no example for how an entire IATA, or an extensive test battery, may be validated, but such approaches are being developed by large international efforts, e.g., in the endocrine disruptor field. With this background in mind, the journey towards establishing a validation process for DNT-IVB assays requires innovative solutions. Whatever the future holds, there is no doubt that reproducibility and transparent description are the basis for any useful NAM. Concepts to evaluate and document these are manifold, but the underlying principle and the essential steps are widely agreed upon[7].

The evaluation of predictivity/relevance is currently the main hurdle and needs to advance. Classical concepts of comparing NAM-output to long lists of reference chemicals, defined by their human or animal toxicity, are not fully applicable to DNT assays and require context-specific adaptation. The underlying problems are: (i) for DNT there are no lists of good reference chemicals that would be comparable in scope (number and diversity of chemicals) and quality (certainty of human adverse effect, understanding of the effect mechanisms) to lists used earlier for the validation of NAMs in the areas of skin irritation or sensitization (Aschner et al., 2017); and (ii) DNT NAMs usually measure a direct hazard, while *in vivo* adversity depends on hazard potential plus internal exposure (Kadereit et al., 2012). Thus, some compounds may be hits in a NAM (e.g., loperamide in MEA assays) although they never reach the brain *in vivo*, and *vice versa*, some compounds may show a low potency in a NAM, but be highly active *in vivo* (e.g., accumulate in the brain or act via bioactive metabolites reaching the brain) (Holzer et al., 2022; Loser et al., 2021a). (iii) DNT NAMs typically measure perturbations in specific neurodevelopmental processes, i.e., they only cover sub-areas of human toxicity, and this is why it is difficult to consider most reference chemicals as positive controls for a given NAM. For instance, PFOS or PCB may be highly correlated to DNT in humans. However, it is not known by which mechanism they act. Thus, it is not clear for which KNDP-based assay they should be considered positive reference compounds. Each NAM would need a specific list of reference compounds known to affect the underlying processes *in vivo*, and such information is at present not always available (Aschner et al., 2017).

One way out of this deadlock has been the suggestion to use "biological assay relevance" as proxy for predictivity. This is related to the concept of "mechanistic validation" (Koch et al., 2022; Leist et al., 2012, 2014). However, the acceptability of this approach may depend on the context of use (see further discussion below). Another way out may be the definition of reference chemicals for certain (sub-)types of DNT based on their effect in well-accepted mechanistically validated NAMs.

6.2 Guiding questions for discussing the challenge:

Q13: How could relevance for a certain sub-area of DNT (e.g., oligodendrocyte development or synapse formation) be shown and converted to a measure of predictivity?

Q14: Is it possible to define "relevance controls" for certain fundamental neurodevelopmental processes, and how can the pool of such compounds be increased?

Q15: Is there something like an applicability domain for certain pathways, targets, or mechanisms that are relevant in a given NAM, and how would this be defined?

6.3 Points to consider for further progress

"Context of use" is an essential element of modern validation approaches[8] (Pamies et al., 2024; van der Zalm et al., 2022). This has resulted in concepts like "fit-for-purpose validation" (Pamies

[7] doi:10.22427/NICEATM-2

[8] https://ntp.niehs.nih.gov/sites/default/files/2024-03/VWG_Report_27Feb2024_FD_508.pdf

et al., 2024) or "determination of the readiness state" (Bal-Price et al., 2018). It may also be termed "assay qualification" as by the FDA. In all cases, there is an emphasis on problem formulation. This adaptive approach allows for varying levels of uncertainty, with increasing confidence in assays correlating with increasing regulatory acceptance. An example of this context of use is the setting of action levels (maximum allowable levels) of certain chemicals in food, especially baby food, based on DNT endpoints and exposure/risk considerations (Ockleford et al., 2018). In Europe, hazard classification systems need to be considered in addition to risk assessment. In this case, the battery would be required to provide a defined classification or prioritization outcome with pathway-specific mechanistic anchoring to bolster both hazard classification and risk assessment frameworks.

A major challenge for the validation of the DNT-IVB as a whole, but also for individual NAMs, is to quantify the performance parameters (e.g., accuracy). The work on skin sensitization (and its DA) has been suggested as an example for a path forward. However, there are some remarkable differences to DNT: First, skin sensitization is based on a single AOP (AOP:40). The DNT-IVB covers multiple AOPs with many MIEs, and the overall structure of the AOP network is only loosely defined (Bal-Price et al., 2018). In addition, it is already clear that in DNT, certain molecular and cellular KEs, like activation of the nicotinic acetylcholine receptor and myelination, are currently not covered in the battery. Therefore, in the future, it will be necessary to cover the broad spectrum of MIEs in the nervous system and KNDPs (Blum et al., 2023; Chesnut et al., 2021).

Second, there are also major differences between the DNT-IVB and the DA for skin sensitization with respect to reference chemicals. In the case of skin sensitization, a large, solid list of reference chemicals (with human data) became available (Basketter et al., 2008, 2020; OECD, 2023b) at the time when individual NAMs (and the DA) were validated. In the case of *in vitro* DNT, the scarcity of reference compounds with known effects on specific neurodevelopmental processes and their respective potency represents a challenge for the predictivity assessment of the test battery. The EFSA is retrospectively reviewing approximately 230 *in vivo* DNT studies on their outcomes[6]. This information will, hopefully, be an important basis against which DNT-IVB data can be compared. However, this still leaves the challenge of dealing with potential species differences (rats vs humans). Examples of such differences include the effects of reactive oxygen species (Klose et al., 2022; Masjosthusmann et al., 2019) or thyroid hormone actions (Dach et al., 2017; Walter et al., 2019) on neurodevelopmental processes. In addition, the *in vivo* DNT guidelines (OECD, 2018a,b)[9,10] may not be sufficiently sensitive to detect human DNT effects. It has been suggested that comparisons of NAMs and animal data may be inappropriate (Aschner et al., 2017; Fritsche et al., 2017, 2018).

General aspects that provide regulatory confidence and acceptance for a testing battery are: formal validation of each individual assay/NAM (usually dealt with by OECD GD 34 (OECD, 2005)), validation of a fixed interpretation procedure for a combination of few or all assays (DA; principles explained in OECD GD 255 (OECD, 2017a)) or verification of whether a complex construct, consisting of many NAMs and assessed in a weight of evidence assessment within an IATA, provides hazard information with sufficient confidence for regulatory decision making (usually dealt with in IATA case studies according to OECD GD 369 (OECD, 2017b)). The classical validation theory and guidance (OECD GD 34) (OECD, 2005) focusses on a single NAM and assumes that such a NAM predicts a single apical endpoint (human pathology). A corollary of this is that sets of chemicals exist that are positive or negative concerning human toxicity and are assumed to behave similarly in the NAM if it has a high predictivity (accuracy). Such background conditions are not fulfilled in the DNT field. Therefore, individual NAMs cannot be validated by classical approaches (Bal-Price et al., 2015, 2018; Leist et al., 2014; Smirnova et al., 2014, 2024). An alternative approach is mechanistic validation, strongly overlapping with concepts of fit-for-purpose validation or readiness assessment (Bal-Price et al., 2018; Pamies et al., 2024). The two central pillars of this are (i) that technical reliability and robustness/transferability of the assay is documented, and that (ii) it is shown that major physiological regulations and AOP(s) (up to the cellular level) can be observed, triggered or disturbed in a similar way as in humans (relevance). This may be demonstrated using pharmacological or molecular biological tools.

At present, the DNT-IVB is a battery of tests, but not an ITS or DA. This means that the tests are performed independently and in parallel. There is no rigid interpretation model. Instead, each positive test result (test hit) is considered a relevant alert for a potential DNT hazard. Numerous steps are required to increase confidence that a test hit has implications for toxicity classification (Smirnova et al., 2024). They include various hit follow-ups and also toxicokinetic considerations. Follow-ups may lead to discarding of hits or to a rationale that confirms their relevance and also provides some form of correlation between hit potency in a NAM and expected effects in humans (Blum et al., 2023). Further development of the DNT-IVB means that such procedures are refined, and confidence in the progress is gained with additional case studies. This can eventually lead to an increasingly standard follow-up and interpretation model.

A transition from a "battery of tests" to a tiered testing strategy may be a next step. Typical features of this approach are tiering, combined with decision points, and often a weighing of test results, e.g., in Bayesian networks (Jaworska and Hoffmann, 2010). In theory, the entire tiered testing strategy may be validated in a traditional way against human data. An historical example for such a process is the DA for skin sensitization (OECD TG 497) (OECD, 2023b). However, the complexity of the DNT-IVB is orders of magnitude larger, while there is less data on reference compounds. It is thus likely that an alternative validation approach needs to be developed.

9 https://nepis.epa.gov/Exe/ZyPDF.cgi/P100G6UI.PDF?Dockey=P100G6UI.pdf

10 https://nepis.epa.gov/Exe/ZyPDF.cgi/P100IRWO.PDF?Dockey=P100IRWO.pdf

Besides the confirmation of intra-laboratory reliability and the establishment of performance metrics, an important aspect of validation is assay transferability and inter-laboratory reproducibility. This will be addressed within the EFSA DNT-RAP2 project. As a test of the transferability concept, the DNT-IVB assays (at present 17) will be transferred to a contract research lab, i.e., DNTOX GmbH. The aim of this project is not only to show assay transferability, but also to make the test methods available to end-users.

7 Conclusion

There was agreement that at the current development stage, the DNT-IVB represents a good starting point for generating alerts on a potential DNT hazard of chemicals. The IVB was also considered by the discussion group to be ready for application to elucidate the mechanism of action of toxicants, in some cases in conjunction with NAMs assessing specific MIEs or cell biological pathways. There are still several challenges to be addressed before the DNT-IVB can be applied as a standalone approach for regulatory purposes, but the road is mapped. Future activities should be focused on measures to increase confidence in these methods (and their combination). Besides measures to assure reproducibility of NAM-derived data, case studies may be performed that provide a mechanistic understanding of observed toxicities. Also, more efforts need to be invested into hit confirmation and follow-up to balance sensitivity and specificity of the approach, and to provide experience and guidance on approaches to predict actual risk from DNT-IVB hits (Smirnova et al., 2024).

References

Aschner, M., Ceccatelli, S., Daneshian, M. et al. (2017). Reference compounds for alternative test methods to indicate developmental neurotoxicity (DNT) potential of chemicals: Example lists and criteria for their selection and use. *ALTEX 34*, 49-74. doi:10.14573/altex.1604201

Bal-Price, A., Crofton, K. M., Leist, M. et al. (2015). International STakeholder NETwork (ISTNET): Creating a developmental neurotoxicity (DNT) testing road map for regulatory purposes. *Arch Toxicol 89*, 269-287. doi:10.1007/s00204-015-1464-2

Bal-Price, A., Hogberg, H. T., Crofton, K. M. et al. (2018). Recommendation on test readiness criteria for new approach methods in toxicology: Exemplified for developmental neurotoxicity. *ALTEX 35*, 306-352. doi:10.14573/altex.1712081

Barenys, M., Gassmann, K., Baksmeier, C. et al. (2017). Epigallocatechin gallate (EGCG) inhibits adhesion and migration of neural progenitor cells in vitro. *Arch Toxicol 91*, 827-837. doi:10.1007/s00204-016-1709-8

Bartmann, K., Bendt, F., Dönmez, A. et al. (2023). A human iPSC-based in vitro neural network formation assay to investigate neurodevelopmental toxicity of pesticides. *ALTEX 40*, 452-470. doi:10.14573/altex.2206031

Basketter, D., Darlenski, R. and Fluhr, J. W. (2008). Skin irritation and sensitization: Mechanisms and new approaches for risk assessment. *Skin Pharmacol Physiol 21*, 191-202. doi:10.1159/000135635

Basketter, D., Beken, S., Bender, H. et al. (2020). Building confidence in skin sensitisation potency assessment using new approach methodologies: Report of the 3rd EPAA partners forum, Brussels, 28th october 2019. *Regul Toxicol Pharmacol 117*, 104767. doi:10.1016/j.yrtph.2020.104767

Blum, J., Masjosthusmann, S., Bartmann, K. et al. (2023). Establishment of a human cell-based in vitro battery to assess developmental neurotoxicity hazard of chemicals. *Chemosphere 311*, 137035. doi:10.1016/j.chemosphere.2022.137035

Blum, J., Brüll, M. and Hengstler, J. (2025). The long way from raw data to NAM based information: Overview on data layers and processing steps. *ALTEX 42*, 167-180. doi:10.14573/altex.2412171

Burbank, M., Kukic, P., Ouedraogo, G. et al. (2024). In vitro pharmacologic profiling aids systemic toxicity assessment of chemicals. *Toxicol Appl Pharmacol 492*, 117131. doi:10.1016/j.taap.2024.117131

Callegaro, G., Kunnen, S. J., Trairatphisan, P. et al. (2021). The human hepatocyte TXG-MAPr: Gene co-expression network modules to support mechanism-based risk assessment. *Arch Toxicol 95*, 3745-3775. doi:10.1007/s00204-021-03141-w

Carstens, K. E., Carpenter, A. F., Martin, M. M. et al. (2022). Integrating data from in vitro new approach methodologies for developmental neurotoxicity. *Toxicol Sci 187*, 62-79. doi:10.1093/toxsci/kfac018

Carstens, K. E., Freudenrich, T., Wallace, K. et al. (2023). Evaluation of per- and polyfluoroalkyl substances (PFAS) in vitro toxicity testing for developmental neurotoxicity. *Chem Res Toxicol 36*, 402-419. doi:10.1021/acs.chemrestox.2c00344

Celardo, I., Ashner, M., Ashton, R. S. et al. (2025). Developmental neurotoxicity (DNT): A call for implementation of new approach methodologies for regulatory purposes: Summary of the 5th international conference on DNT testing. *ALTEX 42*, 323-349. doi:10.14573/altex.2503191

Chaput, L., Guillaume, V., Singh, N. et al. (2020). FastTargetPred: A program enabling the fast prediction of putative protein targets for input chemical databases. *Bioinformatics 36*, 4225-4226. doi:10.1093/bioinformatics/btaa494

Cherianidou, A., Seidel, F., Kappenberg, F. et al. (2022). Classification of developmental toxicants in a human iPSC transcriptomics-based test. *Chem Res Toxicol 35*, 760-773. doi:10.1021/acs.chemrestox.1c00392

Chesnut, M., Paschoud, H., Repond, C. et al. (2021). Human iPSC-derived model to study myelin disruption. *Int J Mol Sci 22*, 9473. doi:10.3390/ijms22179473

Colaianna, M., Ilmjärv, S., Peterson, H. et al. (2017). Fingerprinting of neurotoxic compounds using a mouse embryonic stem cell dual luminescence reporter assay. *Arch Toxicol 91*, 365-391. doi:10.1007/s00204-016-1690-2

Crofton, K. M., Paparella, M., Price, A. et al. (2024). A developmental neurotoxicity adverse outcome pathway (DNT-AOP) with voltage gate sodium channel (VGSC) inhibi-

tion as a molecular initiating event (MiE). *EFSA J 22*, e8954. doi:10.2903/j.efsa.2024.8954

Culbreth, M., Nyffeler, J., Willis, C. et al. (2021). Optimization of human neural progenitor cells for an imaging-based high-throughput phenotypic profiling assay for developmental neurotoxicity screening. *Front Toxicol 3*, 803987. doi:10.3389/ftox.2021.803987

Dach, K., Bendt, F., Huebenthal, U. et al. (2017). BDE-99 impairs differentiation of human and mouse NPCs into the oligodendroglial lineage by species-specific modes of action. *Sci Rep 7*, 44861. doi:10.1038/srep44861

Delp, J., Gutbier, S., Klima, S. et al. (2018). A high-throughput approach to identify specific neurotoxicants/ developmental toxicants in human neuronal cell function assays. *ALTEX 35*, 235-253. doi:10.14573/altex.1712182

Dobreniecki, S., Mendez, E., Lowit, A. et al. (2022). Integration of toxicodynamic and toxicokinetic new approach methods into a weight-of-evidence analysis for pesticide developmental neurotoxicity assessment: A case-study with DL- and L-glufosinate. *Regul Toxicol Pharmacol 131*, 105167. doi:10.1016/j.yrtph.2022.105167

Dreser, N., Madjar, K., Holzer, A. K. et al. (2020). Development of a neural rosette formation assay (RoFA) to identify neurodevelopmental toxicants and to characterize their transcriptome disturbances. *Arch Toxicol 94*, 151-171. doi:10.1007/s00204-019-02612-5

EC (2006). Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. http://data.europa.eu/eli/reg/2006/1907/oj

EC (2008). Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on Classification, Labelling and Packaging of Substances and Mixtures, Amending and Repealing Directives 67/548/EEC and 1999/45/EC, and Amending Regulation (EC) No 1907/2006 (text with EEA relevance). *Off J Eur Union L 353/1*.

Fritsche, E., Crofton, K. M., Hernandez, A. F. et al. (2017). OECD/EFSA workshop on developmental neurotoxicity (DNT): The use of non-animal test methods for regulatory purposes. *ALTEX 34*, 311-315. doi:10.14573/altex.1701171

Fritsche, E., Grandjean, P., Crofton, K. M. et al. (2018). Consensus statement on the need for innovation, transition and implementation of developmental neurotoxicity (DNT) testing for regulatory purposes. *Toxicol Appl Pharmacol 354*, 3-6. doi:10.1016/j.taap.2018.02.004

Gabbert, S., Mathea, M., Kolle, S. N. et al. (2022). Accounting for precision uncertainty of toxicity testing: Methods to define borderline ranges and implications for hazard assessment of chemicals. *Risk Anal 42*, 224-238. doi:10.1111/risa.13648

Grillberger, K., Cöllen, E., Trivisani, C. I. et al. (2023). Structural insights into neonicotinoids and N-unsubstituted metabolites on human nAChRs by molecular docking, dynamics simulations, and calcium imaging. *Int J Mol Sci 24*, 13170. doi:10.3390/ijms241713170

Guan, Z., Wang, X., Dong, Y. et al. (2015). dNTP deficiency induced by HU via inhibiting ribonucleotide reductase affects neural tube development. *Toxicol 328*, 142-151. doi:10.1016/j.tox.2014.12.001

Harrill, J. A., Everett, L. J., Haggard, D. E. et al. (2021). High-throughput transcriptomics platform for screening environmental chemicals. *Toxicol Sci 181*, 68-89. doi:10.1093/toxsci/kfab009

Holzer, A. K., Suciu, I., Karreman, C. et al. (2022). Specific attenuation of purinergic signaling during bortezomib-induced peripheral neuropathy in vitro. *Int J Mol Sci 23*, 3734. doi:10.3390/ijms23073734

Jaklin, M., Zhang, J. D., Schäfer, N. et al. (2022). Optimization of the teratox assay for preclinical teratogenicity assessment. *Toxicol Sci 188*, 17-33. doi:10.1093/toxsci/kfac046

Jaworska, J. and Hoffmann, S. (2010). Integrated testing strategy (ITS) – Opportunities to better use existing data and guide future testing in toxicology. *ALTEX 27*, 231-242. doi:10.14573/altex.2010.4.231

Jochum, K., Miccoli, A., Sommersdorf, C. et al. (2024). Comparative case study on NAMs: Towards enhancing specific target organ toxicity analysis. *Arch Toxicol 98*, 3641-3658. doi:10.1007/s00204-024-03839-7

Judson, R., Houck, K., Martin, M. et al. (2016). Analysis of the effects of cell stress and cytotoxicity on in vitro assay activity across a diverse chemical and assay space. *Toxicol Sci 153*, 409. doi:10.1093/toxsci/kfw148

Judson, R., Houck, K., Paul Friedman, K. et al. (2020). Selecting a minimal set of androgen receptor assays for screening chemicals. *Regul Toxicol Pharmacol 117*, 104764. doi:10.1016/j.yrtph.2020.104764

Kadarmideen, H. N. and Watson-Haigh, N. S. (2012). Building gene co-expression networks using transcriptomics data for systems biology investigations: Comparison of methods using microarray data. *Bioinformation 8*, 855-861. doi:10.6026/97320630008855

Kadereit, S., Zimmer, B., van Thriel, C. et al. (2012). Compound selection for in vitro modeling of developmental neurotoxicity. *Front Biosci (Landmark Ed) 17*, 2442-2460. doi:10.2741/4064

Keßel, H. E., Masjosthusmann, S., Bartmann, K. et al. (2023). The impact of biostatistics on hazard characterization using in vitro developmental neurotoxicity assays. *ALTEX 40*, 619-634. doi:10.14573/altex.2210171

Klose, J., Tigges, J., Masjosthusmann, S. et al. (2021). TBBPA targets converging key events of human oligodendrocyte development resulting in two novel AOPs. *ALTEX 38*, 215-234. doi:10.14573/altex.2007201

Klose, J., Pahl, M., Bartmann, K. et al. (2022). Neurodevelopmental toxicity assessment of flame retardants using a human DNT in vitro testing battery. *Cell Biol Toxicol 38*, 781-807. doi:10.1007/s10565-021-09603-2

Klose, J., Li, L., Pahl, M. et al. (2023). 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 39*, 319-343. doi:10.1007/s10565-022-09730-4

Koch, K., Bartmann, K., Hartmann, J. et al. (2022). Scientific validation of human neurosphere assays for developmental neurotoxicity evaluation. *Front Toxicol 4*, 816370. doi:10.3389/ftox.2022.816370

Krebs, A., Waldmann, T., Wilks, M. F. et al. (2019). Template for the description of cell-based toxicological test methods to allow evaluation and regulatory use of the data. *ALTEX 36*, 682-699. doi:10.14573/altex.1909271

Kreir, M., Putri, D., Tekle, F. et al. (2024). Development of a new hazard scoring system in primary neuronal cell cultures for drug-induced acute neuronal toxicity identification in early drug discovery. *Front Pharmacol 15*, 1308547. doi:10.3389/fphar.2024.1308547

Kreutz, A., Oyetade, O. B., Chang, X. et al. (2024). Integrated approach for testing and assessment for developmental neurotoxicity (DNT) to prioritize aromatic organophosphorus flame retardants. *Toxics 12*, 437. doi:10.3390/toxics12060437

Krug, A. K., Balmer, N. V., Matt, F. et al. (2013). Evaluation of a human neurite growth assay as specific screen for developmental neurotoxicants. *Arch Toxicol 87*, 2215-2231. doi:10.1007/s00204-013-1072-y

Langfelder, P. and Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics 9*, 559. doi:10.1186/1471-2105-9-559

Leist, M., Efremova, L. and Karreman, C. (2010). Food for thought ... Considerations and guidelines for basic test method descriptions in toxicology. *ALTEX 27*, 309-317. doi:10.14573/altex.2010.4.309

Leist, M., Hasiwa, N., Daneshian, M. et al. (2012). Validation and quality control of replacement alternatives – Current status and future challenges. *Toxicol Res 1*, 8-22. doi:10.1039/c2tx20011b

Leist, M., Hasiwa, N., Rovida, C. et al. (2014). Consensus report on the future of animal-free systemic toxicity testing. *ALTEX 31*, 341-356. doi:10.14573/altex.1406091

Loser, D., Grillberger, K., Hinojosa, M. G. et al. (2021a). Acute effects of the imidacloprid metabolite desnitro-imidacloprid on human nACh receptors relevant for neuronal signaling. *Arch Toxicol 95*, 3695-3716. doi:10.1007/s00204-021-03168-z

Loser, D., Hinojosa, M. G., Blum, J. et al. (2021b). Functional alterations by a subgroup of neonicotinoid pesticides in human dopaminergic neurons. *Arch Toxicol 95*, 2081-2107. doi:10.1007/s00204-021-03031-1

Loser, D., Schaefer, J., Danker, T. et al. (2021c). Human neuronal signaling and communication assays to assess functional neurotoxicity. *Arch Toxicol 95*, 229-252. doi:10.1007/s00204-020-02956-3

Maertens, A., Golden, E., Luechtefeld, T. H. et al. (2022). Probabilistic risk assessment – The keystone for the future of toxicology. *ALTEX 39*, 3-29. doi:10.14573/altex.2201081

Magel, V., Blum, J., Dolde, X. et al. (2024). Inhibition of neural crest cell migration by strobilurin fungicides and other mitochondrial toxicants. *Cells 13*, 2057. doi:10.3390/cells13242057

Martin, M. M., Baker, N. C., Boyes, W. K. et al. (2022). An expert-driven literature review of "negative" chemicals for developmental neurotoxicity (DNT) in vitro assay evaluation. *Neurotoxicol Teratol 93*, 107117. doi:10.1016/j.ntt.2022.107117

Masjosthusmann, S., Barenys, M., El-Gamal, M. et al. (2018). Literature review and appraisal on alternative neurotoxicity testing methods. *EFSA Support Publ 15*, 1410E. doi:10.2903/sp.efsa.2018.en-1410

Masjosthusmann, S., Siebert, C., Hübenthal, U. et al. (2019). Arsenite interrupts neurodevelopmental processes of human and rat neural progenitor cells: The role of reactive oxygen species and species-specific antioxidative defense. *Chemosphere 235*, 447-456. doi:10.1016/j.chemosphere.2019.06.123

McDiarmid, A. H., Gospodinova, K. O., Elliott, R. J. R. et al. (2024). Morphological profiling in human neural progenitor cells classifies hits in a pilot drug screen for Alzheimer's disease. *Brain Commun 6*, fcae101. doi:10.1093/braincomms/fcae101

Meier, M. J., Harrill, J., Johnson, K. et al. (2024). Progress in toxicogenomics to protect human health. *Nat Rev Genet 26*, 105-122. doi:10.1038/s41576-024-00767-1

Meisig, J., Dreser, N., Kapitza, M. et al. (2020). Kinetic modeling of stem cell transcriptome dynamics to identify regulatory modules of normal and disturbed neuroectodermal differentiation. *Nucleic Acids Res 48*, 12577-12592. doi:10.1093/nar/gkaa1089

Miller, M. M., McMullen, P. D., Andersen, M. E. et al. (2017). Multiple receptors shape the estrogen response pathway and are critical considerations for the future of in vitro-based risk assessment efforts. *Crit Rev Toxicol 47*, 570-586. doi:10.1080/10408444.2017.1289150

Nyffeler, J., Chovancova, P., Dolde, X. et al. (2018). A structure-activity relationship linking non-planar PCBs to functional deficits of neural crest cells: New roles for connexins. *Arch Toxicol 92*, 1225-1247. doi:10.1007/s00204-017-2125-4

Ockleford, C., Adriaanse, P., Hougaard Bennekou, S. et al. (2018). Scientific opinion on pesticides in foods for infants and young children. *EFSA J 16*, e05286. doi:10.2903/j.efsa.2018.5286

OECD (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. *OECD Series on Testing and Assessment, No. 34*. OECD Publishing, Paris. doi:10.1787/e1f1244b-en

OECD (2017a). Report on Considerations from Case Studies on Integrated Approaches for Testing and Assessment (IATA) – First Review Cycle (2015). *OECD Series on Testing and Assessment, No. 250*. OECD Publishing, Paris. doi:10.1787/9789264274815-en

OECD (2017b). Guidance Document on the Reporting of Defined Approaches to be Used Within Integrated Approaches to Testing and Assessment. *OECD Series on Testing and Assessment, No. 255*. OECD Publishing, Paris. doi:10.1787/9789264274822-en

OECD (2018a). Developmental Neurotoxicity Study (OECD TG 426). *OECD Series on Testing and Assessment, No. 545*. OECD Publishing, Paris. doi:10.1787/9789264304741-27-en

OECD (2018b). Test No. 443: Extended One-Generation Reproductive Toxicity Study. *OECD Guidelines for the Test-*

ing of Chemicals, Section 4. OECD Publishing, Paris. doi: 10.1787/9789264185371-en

OECD (2022a). Case Study on the use of Integrated Approaches for Testing and Assessment for Developmental Neurotoxicity Hazard Characterisation of Imidacloprid and the Metabolite Desnitro-imidacloprid. *Series on Testing and Assessment, No. 366*. OECD Publishing, Paris.

OECD (2022b). Per- and Polyfluoroalkyl Substances and Alternatives in Coatings, Paints and Varnishes (CPVs): Report on the Commercial Availability and Current Uses. *OECD Series on Risk Management of Chemicals, Section 4, No. 497*. OECD Publishing, Paris. doi:10.1787/6745457d-en

OECD (2022c). Case Study on the Use of Integrated Approaches for Testing and Assessment for Developmental Neurotoxicity Hazard Characterisation of Acetamiprid. *Series on Testing and Assessment, No. 365*. OECD Publishing, Paris.

OECD (2022d). Case Study on the Use of Integrated Approaches for Testing and Assessment for DNT to Prioritize a Class of Organophosphorus Flame Retardants. *Series on Testing and Assessment, No. 364*. OECD Publishing, Paris.

OECD (2022e). Case Study for the Integration of In Vitro Data in the Developmental Neurotoxicity Hazard Identification and Characterisation Using Flufenacet. *Series on Testing and Assessment, No. 363*. OECD Publishing, Paris.

OECD (2022f). Case Study for the Integration of In Vitro Data in the Developmental Neurotoxicity Hazard Identification and Characterisation Using Deltamethrin as a Prototype Chemical. *Series on Testing and Assessment, No. 362*. OECD Publishing, Paris.

OECD (2023a). Initial Recommendations on Evaluation of Data from the Developmental Neurotoxicity (DNT) In-Vitro Testing Battery. *OECD Series on Testing and Assessment, No. 377*. OECD Publishing, Paris. doi:10.1787/91964ef3-en

OECD (2023b). Guideline No. 497: Defined Approaches on Skin Sensitisation. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/b92879a4-en

OECD (2024). Test No. 467: Defined Approaches for Serious Eye Damage and Eye Irritation. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/28fe2841-en

Pallocca, G., Grinberg, M., Henry, M. et al. (2016). Identification of transcriptome signatures and biomarkers specific for potential developmental toxicants inhibiting human neural crest cell migration. *Arch Toxicol 90*, 159-180. doi:10.1007/s00204-015-1658-7

Pamies, D., Ekert, J., Zurich, M. G. et al. (2024). Recommendations on fit-for-purpose criteria to establish quality management for microphysiological systems and for monitoring their reproducibility. *Stem Cell Rep 19*, 604-617. doi:10.1016/j.stemcr.2024.03.009

Paparella, M., Bennekou, S. H. and 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. doi:10.1016/j.reprotox.2020.08.002

Penschuck, S., Flagstad, P., Didriksen, M. et al. (2006). Decrease in parvalbumin-expressing neurons in the hippocampus and increased phencyclidine-induced locomotor activity in the rat methylazoxymethanol (MAM) model of schizophrenia. *Eur J Neurosci 23*, 279-284. doi:10.1111/j.1460-9568.2005.04536.x

Rovida, C., Escher, S. E., Herzler, M. et al. (2021). NAM-supported read-across: From case studies to regulatory guidance in safety assessment. *ALTEX 38*, 140-150. doi:10.14573/altex.2010062

Schmeisser, S., Miccoli, A., von Bergen, M. et al. (2023). New approach methodologies in human regulatory toxicology – Not if, but how and when! *Environ Int 178*, 108082. doi:10.1016/j.envint.2023.108082

Schüttler, A., Altenburger, R., Ammar, M. et al. (2019). Map and model-moving from observation to prediction in toxicogenomics. *Gigascience 8*, giz057. doi:10.1093/gigascience/giz057

Seiler, A. E. and Spielmann, H. (2011). The validated embryonic stem cell test to predict embryotoxicity in vitro. *Nat Protoc 6*, 961-978. doi:10.1038/nprot.2011.348

Shah, I., Setzer, R. W., Jack, J. et al. (2016). Using ToxCast™ data to reconstruct dynamic cell state trajectories and estimate toxicological points of departure. *Environ Health Perspect 124*, 910-919. doi:10.1289/ehp.1409029

Shinde, V., Hoelting, L., Srinivasan, S. P. et al. (2017). Definition of transcriptome-based indices for quantitative characterization of chemically disturbed stem cell development: Introduction of the STOP-Tox$_{ukn}$ and STOP-Tox$_{ukk}$ tests. *Arch Toxicol 91*, 839-864. doi:10.1007/s00204-016-1741-8

Smirnova, L., Hogberg, H. T., Leist, M. et al. (2014). Developmental neurotoxicity – Challenges in the 21st century and in vitro opportunities. *ALTEX 31*, 129-156. doi:10.14573/altex.1403271

Smirnova, L., Hogberg, H. T., Leist, M. et al. (2024). Revolutionizing developmental neurotoxicity testing – A journey from animal models to advanced in vitro systems. *ALTEX 41*, 152-178. doi:10.14573/altex.2403281

Suciu, I., Delp, J., Gutbier, S. et al. (2023). Definition of the neurotoxicity-associated metabolic signature triggered by berberine and other respiratory chain inhibitors. *Antioxidants (Basel) 13*, 49. doi:10.3390/antiox13010049

Tal, T., Myhre, O., Fritsche, E. et al. (2024). New approach methods to assess developmental and adult neurotoxicity for regulatory use: A parc work package 5 project. *Front Toxicol 6*, 1359507. doi:10.3389/ftox.2024.1359507

van der Zalm, A. J., Barroso, J., Browne, P. et al. (2022). A framework for establishing scientific confidence in new approach methodologies. *Arch Toxicol 96*, 2865-2879. doi:10.1007/s00204-022-03365-4

Vincent, F., Loria, P. M., Weston, A. D. et al. (2020). Hit triage and validation in phenotypic screening: Considerations and strategies. *Cell Chem Biol 27*, 1332-1346. doi:10.1016/j.chembiol.2020.08.009

Vincent, F., Nueda, A., Lee, J. et al. (2022). Phenotypic drug discovery: Recent successes, lessons learned and new directions. *Nat Rev Drug Discov 21*, 899-914. doi:10.1038/s41573-022-00472-w

Vrijenhoek, N. G., Wehr, M. M., Kunnen, S. J. et al. (2022). Application of high-throughput transcriptomics for mechanism-

based biological read-across of short-chain carboxylic acid analogues of valproic acid. *ALTEX 39*, 207-220. doi:10.14573/altex.2107261

Waldmann, T., Rempel, E., Balmer, N. V. et al. (2014). Design principles of concentration-dependent transcriptome deviations in drug-exposed differentiating stem cells. *Chem Res Toxicol 27*, 408-420. doi:10.1021/tx400402j

Waldmann, T., Grinberg, M., König, A. et al. (2017). Stem cell transcriptome responses and corresponding biomarkers that indicate the transition from adaptive responses to cytotoxicity. *Chem Res Toxicol 30*, 905-922. doi:10.1021/acs.chemrestox.6b00259

Walter, K. M., Dach, K., Hayakawa, K. et al. (2019). Ontogenetic expression of thyroid hormone signaling genes: An in vitro and in vivo species comparison. *PLoS One 14*, e0221230. doi:10.1371/journal.pone.0221230

Wang, X., Guan, Z., Dong, Y. et al. (2018). Inhibition of thymidylate synthase affects neural tube development in mice. *Reprod Toxicol 76*, 17-25. doi:10.1016/j.reprotox.2017.12.007

Discussion participants by challenge

We are grateful to conference participants taking part in the discussion of the different challenges.

Challenge 1: K. Hellsten (ECHA), T. Knudsen (US EPA), A. Kreutz (Inotiv), M. Leist, V. Magel, M. Dürr, S. Tangianu (University of Konstanz), I. Mangas (EFSA), N. Petrovič (National Institute of Public Health, Slovenia), L. Smirnova (CAAT, Johns Hopkins University), M. Whelan (European Commission JRC), C. Wray (UFZ).

Challenge 2: R. Ashton (Wisconsin University), A. Bal-Price (European Commission, JRC), O. Bandele (US FDA), K. Bartmann (IUF), J. Blum, N. Dreser (University of Konstanz), K. Crofton (R3Fellows LLC), C. Medina (Health Canada), C. Miguel Sanz (ZeClinics), R. Owen, J. Spath (UFZ), M. Sachana (OECD), T. Shafer (US EPA), E. Schirmer (BASF SE), J. Stanko (NIH), J. Victor (DNTOX), M-G. Zurich (University of Lausanne).

Challenge 3: P. Ankli (Edelweiss Connect), K. Cartsens (US EPA), S. Gutsfeld, I. Piragytė-Langa, T. Tal (UFZ), H. Heusinkveld (RIVM), J-H. Hsieh (NIEHS), Y. Kanda (NIHS Japan), U. Kraushaar (NMI), C. McPherson (NIH), A. Mie (Stockholm University), I. Mueller (Unilever), P. Müller, I. Suciu, A. K. Ueckert (University of Konstanz), J. Schaefer (NMI), M. Schenke (Johns Hopkins University), C. Thompson (Syngenta), S. Wang (University of Lausanne), M. Wojewodzic (NIPH).

Challenge 4: M. Aschner (Albert Einstein College of Medicine), A. Cediel-Ulloa (Uppsala University), N. Choksi (ToxStrategies), A. Forsby (Stockholm University), B. Hardy (Edelweiss Connect), M. Lislien (Norwegian Institute of Public Health), S. Melching-Koll (BASF SE), T. Mille (Université de Lorraine), A. Muriana (Biobide), O. Myhre (Norwegian Institute of Public Health), M. Paparella (Medical University Innsbruck), E. Pitzer (US EPA), M. T. Rigoli (University of Milan), J. Stein (SCC Consulting GmbH), C. vom Berg (Eawag), I. Celardo, C. Wolfbeisz, E. Cöllen (University of Konstanz).

Challenge 5: K. Aiello-Holden (BfR), M. Bisson (INERIS), M. Corvaro (Corteva Agriscience), S. Fitzpatrick (US FDA), M. Frangiamone (University of Lausanne), E. Fritsche (SCAHT), T. Hartung (John Hopkins University), H. T. Hogberg (NICEATM), A. Hvitved (ARDF), G. T. Knight (Wisconsin University), K. Koch, E. Kuchovska, I. Scharkin, L. M. Stark (IUF), G. Pelletier (Health Canada), C. Rovida (CAAT-Europe), V. Schiavone (ZeClinics), K. Schlüppmann (Universitätsklinikum Düsseldorf), H. Schroeder (Université de Lorraine), Y. Tanaskov (University of Konstanz).

Disclaimer

This document has been cleared for publication by the US EPA. Mention of trade names or commercial products does not constitute endorsement for use. The views expressed in this article do not necessarily represent the views or policies of the US EPA.

Conflict of interest

EF is a shareholder and scientific managing director of the company DNTOX GmbH, a contract research organization, which provides DNT IVB assay services. The other authors have no conflict of interest.

Data availability

Background information supporting conclusions of this report is available from the corresponding author [M.L.] upon reasonable request. Data have not been generated in the course of this project.

Acknowledgements

We gratefully acknowledge the funding of the meeting and its workshops by the Doerenkamp-Zbinden foundation, the German Research Foundation, and the 3R-Network of the Land Baden-Württemberg. This study was also supported by the Baden-Württemberg Ministry for Science, Research and Art (MWK Baden-Württemberg) through InViTe2. The work on the article was facilitated by grants from the BMBF (161L0243B, 016LW0146K), the European Union's Horizon 2020 research and innovation program under grant agreements No. 964537 (RISK-HUNT*3R*), No. 964518 (ToxFree) and No. 101057014 (PARC), the Horizon Europe IHI2 project Vict3R and the NIH, National Institute of Environmental Health Sciences through NIEHS Contract No. HHSN273201500010C.