EXTERNAL SCIENTIFIC REPORT

APPROVED: 23 September 2020

doi:10.2903/sp.efsa.2020.EN-1938

Establishment of an a priori protocol for the implementation and interpretation of an in-vitro testing battery for the assessment of developmental neurotoxicity

Stefan Masjosthusmann[1#], Jonathan Blum[2#], Kristina Bartmann[1], Xenia Dolde[2], Anna-Katharina Holzer[2], Lynn-Christin Stürzl[1], Eike Hagen Keßel[1], Nils Förster[3], Arif Dönmez[1], Jördis Klose[1], Melanie Pahl[1], Tanja Waldmann[2], Farina Bendt[1], Jaffar Kisitu[2], Ilinca Suciu[2], Ulrike Hübenthal[1], Axel Mosig[3], Marcel Leist[2], Ellen Fritsche[1]

[1]IUF - Leibniz Research Institute for Environmental Medicine, Auf'm Hennekamp 50, 40225 Düsseldorf, Germany

[2]University of Konstanz, UKN – University of Konstanz, Universitätsstr. 10, 78464 Konstanz, Germany

[3]RUB – Ruhr University Bochum, Faculty for Biology and Biotechnolgy, Bioinformatics Group, Bochum, Germany

#authors contributed equally

Abstract

In this project we set up a human cell-based DNT *in vitro* testing strategy that is based on test methods with high readiness and data generated therefrom. The methods underwent a fit-for-purpose evaluation that considered four key elements: 1. The test system, 2. the exposure scheme, 3. the assay and analytical endpoint(s) and 4. the classification model. This testing battery was challenged with 119 chemicals for which rich toxicological information was available (for some of them also on their DNT hazard). Testing was performed in 5 test systems measuring 10 DNT-specific endpoints and additional 9 viability/ cytotoxicity-related parameters. For approximately half of the compounds, additional and complementary data from DNT *in vitro* tests was added by the US-EPA. This extended battery was also evaluated. Testing results revealed that the test methods of this current DNT *in vitro* battery are reliable and reproducible. The endpoints had to a large extent low redundancy. Battery performance, as assessed with compounds well-characterized for DNT hazard had a sensitivity of 82.7 % and a specificity of 88.2 %. Gap analyses suggested that radial, astro- and microglia as well as myelination endpoints may be added to the battery. Two case studies, one for screening and prioritization of 14 flame retardants, and one on hazard characterization of 2 pesticides, were presented. Hypothetical AOPs were developed based on the latter case study. In conclusion, the DNT testing strategy explored here is a very promising first approach for DNT hazard identification and characterization. The performance is encouraging and may be improved by inclusion of further tests. Some uncertainties in DNT *in vitro* battery testing outcomes could be reduced by incorporating test data and modelling approaches related to *in vitro* and *in vivo* toxicokinetics of test compounds.

Key words: developmental neurotoxicity, DNT, testing battery, in vitro

Question number: EFSA-Q-2018-00308

Correspondence: pesticides.peerreview@efsa.europa.eu

Suggested citation: Masjosthusmann S, Blum J, Bartmann K, Dolde X, Holzer A-K, Stürzl L-C, Hagen Keßel E, Förster N, Dönmez A, Klose J, Pahl M, Waldmann T, Bendt F, Kisitu J, Suciu I, Hübenthal 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 as-sessment of developmental neurotoxicity. EFSA supporting publication 2020:EN-1938. 152 pp. doi:10.2903/sp.efsa.2020.EN-1938

ISSN: 2397-8325

Table of contents

1. Introduction

The intelligence of a population is its largest socio-economic potential. Therefore, it is of paramount importance to ensure individual evolvement of maximum intellectual potential. Poisoning disasters with e.g. polychlorinated biphenyls, mercury or lead demonstrate that chemicals can interfere with developmental processes of the human brain (Grandjean & Landrigan, 2006). Not only poisonings with high chemical exposure raise concerns for children's brain developments, but also low dose exposure especially to a mixture of compounds is thought to contribute to the increasing incidences of neurodevelopmental disorders currently observed (Bennett et al, 2016; Grandjean & Landrigan, 2014) Thereby, a reduction in mean intelligence quotient (IQ) of a population by only 5 points leads to a decrease in the number of gifted and simultaneously an increase of less gifted by 57% each (Figure 1; Schmidt, 2013). These data point to the individual as well as societal impact of neurodevelopmental toxicity not only at high, but also at low exposure levels.

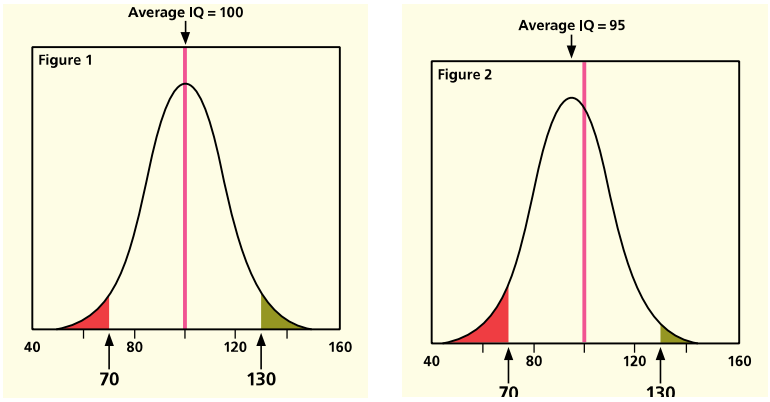


Figure 1: Socio-economic impact of a reduction in the average population IQ by 5 points. A reduction in the average population IQ by only 5 points means that the proportion of mentally retarded in a population increases by 57%, while the proportion of gifted decreases by 57% (source: Schmidt, 2013) .

Animal experiments with rats are currently the gold standard in developmental neurotoxicity (DNT) testing. These are specified in the OECD test guideline 4260F[1]. Such animal tests are currently required as so-called Tier II studies in Europe for pesticides, biocides and chemicals, and in USA for pesticides upon triggers. Such triggers are indications of toxicity from neurotoxicological or reproductive toxicity studies, a known neurotoxic mode of action (MoA), a structure-activity alert or a positive indication for endocrine disruption. European regulators are currently considering DNT studies as Tier II regulatory requirements not only for pesticides, but also for other chemicals and drugs.

To date, DNT studies have been conducted on only 35 of the 479 pesticide active substances currently approved in EU. Testing for more pesticides, and especially other chemicals of the chemical universe, is not feasible with the current guideline study for several reasons: (i) the guideline study is very time- and cost-intensive (1 year/compound that costs up to € 1.000.000), (ii) it is ethically questionable, since for testing one substance about 140 dams and 1,000 juveniles are required, (iii) there are uncertainties in its methodology, evaluation, and regulation, (iv) their predictivity for protection of the human brain is

[1] OECD/OCDE Guideline 426 for the testing of chemicals-Developmental Neurotoxicity study, October 2007.

questionable due to differences in exposure, developmental timing and pharmaco-/toxicodynamics between humans and rodents (Dorman et al, 2001; Kaufmann, 2003; Tsuji & Crofton, 2012).

To protect children's brains, the current consensus amongst academia, industry and regulators is that an alternative to the current DNT guideline is needed that allows DNT hazard assessment faster, cheaper, accurate and with human relevance (Bal-Price et al, 2015a; Crofton et al, 2011; Fritsche et al, 2017). Such an alternative is currently thought to be an *in vitro* testing strategy that contains established assays for neurodevelopmental endpoint detection covering neurodevelopmental processes as a function of developmental timing (Bal-Price *et al*, 2012; Fritsche, 2016; Fritsche *et al*, 2017). A test battery and not a single DNT assay is needed because an available single *in vitro* test system cannot fully mimic the immense complexity of human brain development, maturation and senescence. Such tests need to be medium to high throughput assays generating relevant and reliable data that (i) help with the prioritization of chemicals for future testing and (ii) can be used to guide risk management decisions (Bal-Price et al, 2012, 2015a; Crofton et al, 2011; Fritsche et al, 2017). Therefore, the establishment of DNT test and DNT test battery performance standards and testing strategy guidance is the next step towards implementation of alternative DNT testing (Fritsche et al, 2017).

Based on this background, there are two major points relevant for EFSA concerning alternative DNT testing:

1) EFSA wants to see the proposed draft *in vitro* testing battery used for the following regulatory purposes:

- immediately, for screening of chemicals and prioritization,
- followed by further harmonization in an international acceptance process through OECD, with an acknowledgement that improvements will continue to be made as science advances and regulatory acceptance increases and that changes in regulation i.e. data requirement, are needed.

2) An agreement was achieved on the need for a draft framework for regulatory use of DNT data through an integrated approach to testing and assessment (IATA). The framework should be driven by problem formulation as defined by decision-making needs. It should make efficient use of resources. As the potential impact of the regulatory decision increases, data needs and resources use will increase to reduce the scientific uncertainty in estimates of risk and impact.

1.1. Background and Terms of Reference as provided by the requestor

This contract was awarded by EFSA to:

Contractor: IUF – Leibniz Research Institute for Environmental Medicine and the University of Konstanz.

Contract title: Implementation and interpretation of an in-vitro testing battery for the assessment

of developmental neurotoxicity

Contract number: OC/EFSA/PRAS/2017/01

2. Goal and Objectives

The overall goal of this procurement is to accelerate the development and use of *in vitro* test methods, conducted in human-relevant cell systems, capable of cost- and time-efficient testing of chemicals for the potential to disrupt the development of the nervous system i.e. prediction of neurodevelopmental hazards to human health. To reach this goal, the overall objective of the contract is to provide an adequate scientific background for a DNT testing battery and facilitate the preparation of a guidance for alternative developmental neurotoxicity (DNT) testing by producing wet lab data and delivering data interpretation and user guidance. Here, the focus is on the application of a DNT testing battery covering neurodevelopmental processes essential for brain development in a temporal context.

The contract was divided into two phases (Figure 2) containing the following work packages:

Phase 1: Development of the test system

Objective 1 (WP1): Based on the background documents given in the EFSA tender specification and the personal know-how of the applicants, the partners defined a DNT *in vitro* testing battery. This battery is based on key neurodevelopmental processes essential for brain development and maturation. It covers most relevant cell types (e.g. neurons and glial cells) and reflects the fetal and perinatal periods of human nervous system development. An important cell basis for individual *in vitro* tests used in the battery are human induced pluripotent stem cells (hiPSC). Human neural progenitor cells (hNPC) add an additional high value. Rodent cells were used in cases, where there was no human cell-based assay available for a necessary endpoint at project start.

Objective 2 (WP2): Based on the information collected in WP1, the partners performed an analysis on redundant assays. For this purpose, multiple statistical approaches were used, like principal component and correlation analyses, to identify the smallest sets of test methods that give full information on DNT hazard. During this procedure, mainly redundancy-reducing approaches were used, to suggest a minimum test battery. However, suggestions were also made for incorporation of some assay overlap to increase the robustness of predictions.

Objective 3 (WP3): After identification of available *in vitro* methods that can contribute to a DNT testing battery, the applicability domain for such a battery was described on three levels: (i) which type of chemicals can be tested and identified (physicochemical properties; industry sectors; chemical classes), (ii) which general biological features are reflected (cell-cell/cell-matrix interactions; xenobiotic metabolism; differentiation stages), (iii) which general regulation pathways were reflected (i.e. which AOP key events and signaling cascades can be assessed). The partners performed a gap analysis of the battery by comparing the applicability domains to the regulatory and scientific needs. As an important basis for this work**,** human DNT compounds' modes-of-action (MoA) for selected compounds were described according to the current state-of-knowledge. These data were cross-checked against the activity pattern of the compounds in the *in vitro* assays and against knowledge of pathways, e.g. from tool compound testing, in the *in vitro* methods.

Objective 4 (WP4): A roadmap for the 'fit for purpose' validation of the individual tests and the whole battery was developed based on knowledge from WP1-WP3. The fit-for-purpose validation was based on the establishment of a quantifiable set of readiness criteria and performance standards, combined with the definition of use scenarios. Readiness requirements were defined for various use scenarios and compared to the readiness state of test methods and the entire test battery. One important readiness criterion is assay sensitivity and specificity (with respect to test chemicals, and with respect to AOP key events). It is envisaged that the available literature data for these criteria are limited and need to be supplemented by new data. Clear identification of this data need drove the compound selection and test plan in WP5.

Objective 5 (WP5a/b); WP5a: Development of a chemical library (n = 120) for challenging the *in vitro* test battery. This chemical library was carefully assembled to contain: (i) toxicologically data-rich compounds; (ii) negative controls to determine test specificity (n=17); (iii) positive controls (with documentation of the evidence that they trigger DNT in humans or solid information for other mammalian species; n=29).

WP5b: Testing of a chemical library (as assembled in WP5) to challenge the test battery. Testing was done in a tiered fashion according to a defined screening plan (e.g. range finding followed by concentration-response testing and hit confirmation testing). Primary data was handled, annotated and stored according to the FAIR principles. A biostatistics pipeline was developed using internationally accepted biostatistical methods for deriving secondary data, i.e. benchmark concentrations (BMC) and benchmark responses (BMR). The derived data was used to originate (i) an estimate of various performance parameters, such as sensitivity, specificity and balanced accuracy of individual tests and of the entire battery; (ii) a better specification of the applicability domain of the tests.

Phase 2: Reporting

Objective 6 (WP6): Based on WP4 and WP5 the (i) biological and (ii) toxicological application domains of the individual assays and the whole battery were defined. A main focus was on presence/absence as well as functionality of physiological pathways in individual assays and the testing battery. A comprehensive coverage of essential developmental regulations is necessary for the test battery, and definition of gaps determine the overall applicability domain. The analysis also points to a lack of important neurodevelopmental processes that are not yet available in *in vitro* assays. These processes are subject to recommendations for further test development.

Objective 7 (WP7): A draft guidance for *in vitro* DNT testing (using the battery of tests), for processing of the data, and for interpretation of the results are provided. Test strategies (parallel/tiered) are suggested for different scenarios. Here, also recommendations are given on how to combine the hazard data from the test battery, via calculations of free compound concentrations in cells, and *in vitro*-to-*in vivo* extrapolations (IVIVE) to relevant *in vivo* exposure levels. The guidance also includes possible necessities for species comparisons by *in vitro* testing in rodent cell systems.

Objective 8 (WP8): The partners developed two types of case studies to demonstrate how the battery test data can be used within an IATA for regulatory decisions. The case study compounds were part of the library of WP5a. For one type of case study (A), we selected a small set of data-poor compounds i.e. flame retardants. We investigated how the data produced by the DNT test battery can be used in a weight-of-evidence approach to prioritize the compounds for further testing. For another type of case study (B), EFSA chose the well-studied pesticide deltamethrin, for which rich *in vivo* data are available. It will be investigated, how the data from the test battery, and the thereby improved mechanistic understanding, can improve various aspects of risk assessment if they are used within an IATA together with other available data.

In the following sections, the outcomes of these WP are described in detail.

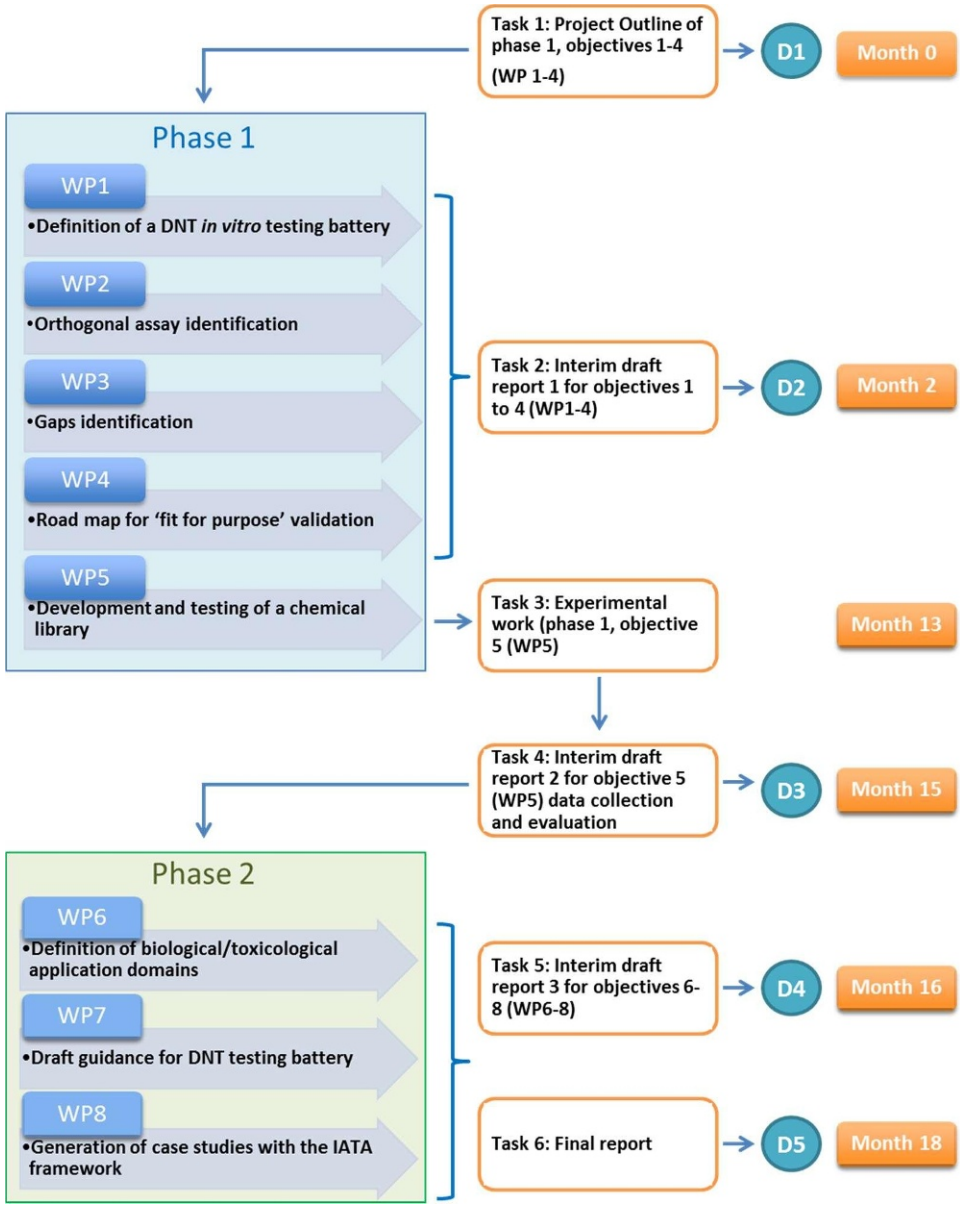


Figure 2: Project outline with tasks within work packages (WP), deliverables (D) and timing.

3. WP1: Definition of an *in vitro* Testing Battery covering Key Neurodevelopmental Processes

The objective of WP1 is to define a DNT *in vitro* testing battery based on the background documents (Fritsche, 2016a; Bal-Price *et al*, 2015a, Fritsche et al, 2015) given in the EFSA tender specification and the personal know-how of the applicants. This battery is based on key neurodevelopmental processes (Figure 3) essential for brain development and cover fetal periods. The cell basis for individual *in vitro* tests used in the battery is based on human induced pluripotent stem cells (hiPSC) and commercially available primary human foetal neural progenitor cells (hNPC). These assays were previously identified as relevant because they study compounds' effects on fundamental and indispensable neurodevelopmental key events (Bal-Price et al. 2018).

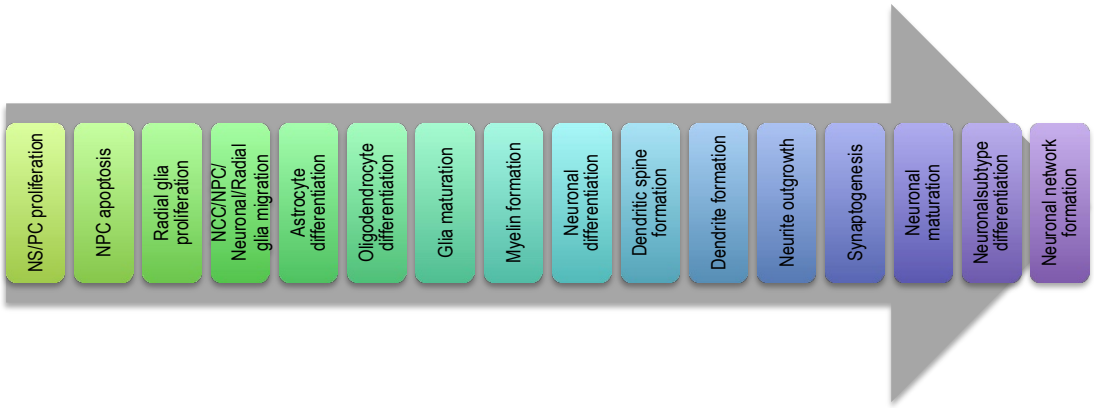


Figure 3: Neurodevelopmental processes necessary for brain development. From: Fritsche, 2016.

The idea of a DNT *in vitro* testing battery is to assemble as many relevant and testable neurodevelopmental endpoints as possible, here relevant means processes biologically plausible and essential. In addition, multiplexing endpoints in assays is warranted as this will reduce the number of assays/substance and thus limit resources like time and money.

3.1. DNT Test Methods assembled into an *in vitro* DNT Testing Battery

An endpoint measurement, e.g. used in academic research, differs from a set-up test method (Crofton et al, 2011). While an endpoint measurement can be developed into a test method, a test method is needed for the assembly of a testing battery that has the purpose of being applied for regulatory purposes. The total DNT endpoint testing battery as envisioned for the future contains measurable endpoints that have already been set up for test methods as depicted in Figure 4. The endpoints underlying the test methods were chosen according to basic neurodevelopmental biology/physiology and have been discussed intensively and agreed on amongst DNT experts during the last years (Lein *et al*, 2005; Bal-Price *et al*, 2012; Fritsche, 2016; Bal-Price *et al*, 2015a). The battery proposed under this procurement covers neurodevelopmental processes from the fetal and postnatal phase through different stages of cell development (Figure 4; Bal-Price et al. 2018). These test methods are human cell-based and encompass four different cell systems, i.e. primary human neural progenitor cells (hNPC), neural crest cells (NCC),

human immature dorsal root ganglia (iDRG) and LUHMES (Lund human mesencephalic) cells. Each cell system allows studying one or more neurodevelopmental process(es) that take(s) place in the culture dish. Theses cell systems were characterized on the molecular and cellular levels to ensure that they are the appropriate matrices for neurodevelopmental endpoint evaluation. Briefly, hNPC were purchased from Lonza as primary fetal cortical NPC. FACS analyses revealed that these cells are > 98% nestin+ and > 96% SOX2+ (Hofrichter et al, 2017) confirming their NPC nature. These NPC proliferate in culture (Moors *et al*, 2009; Baumann *et al*, 2014, 2015; Nimtz *et al*, 2019). During the hNPC differentiation process microarray gene expression analyses indicate a drastic transcriptome change reflecting a multitude of neurodevelopmental processes including cell differentiation (Masjosthusmann et al, 2018) that are underpinned by their phenotypic descriptions: GFAP+ and/or nestin+ radial glia differentiation (and their migration; (Moors et al, 2007, 2009, 2012; Baumann et al, 2015; Schmuck et al, 2017), neuronal ((Moors et al, 2009, 2012; Baumann et al, 2015; Schmuck et al, 2017) and oligodendrocyte differentiation ((Fritsche et al, 2005; Moors et al, 2009; Baumann et al, 2015; Dach et al, 2017). Here, specific gene expression products for radial glia, neurons, astrocytes and oligodendrocytes increase in expression during the early differentiation process (Masjosthusmann et al, 2018). NCC are differentiated from hiPSC line IMR90_clone#4. FACS analysis revealed that of cryo-preserved cells after thawing ~95% express either of the two NCC markers HNK-1 or p75, with 70-80% of the cells being double-positive (Nyffeler *et al*, 2017). Immunostaining for nestin and F-actin indicates a NCC typical cell morphology (Nyffeler *et al*, 2017). Human iDRG neurons are differentiated from hiPSC line SBAD2. During differentiation process immunostaining indicates that a large subpopulation of cells is SOX10 positive, which is a neural crest marker and precursor of the peripheral nervous system (Hoelting *et al*, 2016). After thawing on DoD8 cells start to grow out neurites which are positive for the neuronal marker TUJ1 and have lost the precursor marker SOX10 (Hoelting *et al*, 2016). The LUHMES cell line can upon tetracycline induction be directed from a proliferating into a neuronal differentiation status. With increasing differentiation duration, the mRNA levels of Fox-3/NeuN (marker of post-mitotic neurons) and βIII-tubulin (neuronal cytoskeletal protein) are increasing. In contrast, mRNA levels of CDK1 (cell cycle regulator) are decreased (Scholz *et al*, 2011). These results underline the post mitotic character of the differentiated LUHMES cells. Increasing gene expression of typical dopaminergic neurodevelopmental markers like DAT, DRD2 and TH demonstrate the dopaminergic phenotype of the LUHMES cells (Scholz *et al*, 2011). Protein and lipid compositions of media used for NCC, iDRG and LUHMES test systems are given in (Krebs *et al*, 2020b). More details on the test systems relevant for the test methods can be found below. Gene expression data for these test systems with regards to electrophysiologically-relevant molecules are collected in Annex A6. For properly representing the overall state-of-the-art concerning DNT *in vitro* testing for regulatory purposes, we added results from test methods established and performed by Tim Shafer, US-Environmental Protection Agency (US-EPA; Figure 4; (Frank et al, 2017; Shafer et al, 2019; Harrill et al, 2011) as well as from a recently developed and yet unpublished test method (IUF) sponsored by the Danish EPA (DK-EPA, grant number DK-EPA, grant number MST-667-00205). These test methods use test systems suitable for assessing neuronal network formation and activity. The US-EPA rat neuronal network formation (rNNF) assay is based on rat mixed primary cortical cultures containing neurons and glia cells (Frank et al, 2017). The IUF human NNF (hNNF) assay employs a cell kit (SynFire®, NeuCyte, USA) that contains hiPSC-derived excitatory and inhibitory neurons as well as human primary astrocytes (https://www.neucyte.com/technology) in a fixed ratio. Both, the rNNF and hNNF assay measure network formation over time using Axion Maestro MEA systems with a 768-channel amplifier (Axion Biosystems, USA) and an R-based tool for data analyses. Additionally, the manufacturer's standalone tools, Neural Metric Tool and AxIS Metric Plotting Tool (Axion Biosystems, USA) were used for the data evaluation of the hNNF assay. Gene expression data also for these test systems with regards to electrophysiologically-relevant molecules are collected in Annex A6. Although these two test methods were not sponsored by this project, they contribute to the overall knowledge base of the effects of these 119 compounds. In this report only the data previously published

by the US-EPA will be recognized, the US-EPA aims at contributing data for all compounds. Yet COVID-19 has interrupted their experimental schedule and completion of the testing and analysis of the data is therefore delayed.

Knowing that this battery is not complete, i.e. endpoints still needing test method development and endpoints not being represented at all, gaps in the testing battery are pointed out in Chapter 6 of this Report 'WP3: Identification of Gaps in the Testing Battery'.

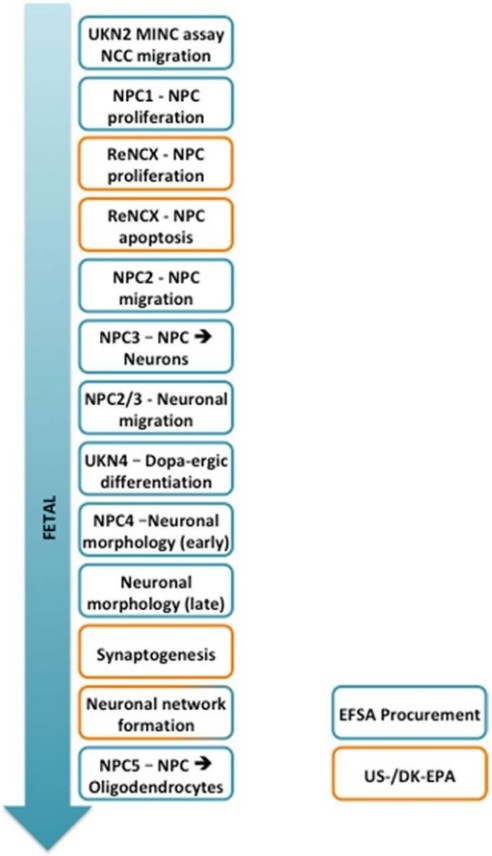


Figure 4: Assembly of a DNT *in vitro* testing strategy covering neurodevelopmental endpoints during fetal stages of brain development.

Within the project chemical testing was performed in the test methods NPC1-5 and UKN2, 4 and 5. In addition, data from the US-EPA test methods hNP1-proliferation and –apoptosis, rat neurite initiation, neurite maturation, synaptogenesis and neuronal network formation (NNF) are added for compounds were available. Data from the DK-EPA-sponsored test method using human NNF will be shown for two case studies. In the following paragraphs the readiness of the 'core' methods, i.e. NPC1-5 and UKN2,4 and 5, the biological background and the experimental procedures for compound testing are described. A 'matrix' summarizing chemicals studied by test methods is given in Annex A1.

3.1.1. Primary hNPC Proliferation Assay (NPC1 Test)

NPC proliferation is a fundamental neurodevelopmental KE that, when disturbed, like in Zika virus infected primary NPC, leads to microcephaly in children (Tang *et al*, 2016; Devakumar *et al*, 2018). Proliferation of primary hNPC of fetal origin (Lonza, Verviers, Belgium) grown as neurospheres in 3D is studied by measuring increase in sphere size over time using phase contrast microscopy (Baumann *et al*, 2015, 2014; Gassmann *et al*, 2010, 2012; Moors *et al*, 2009; Schreiber *et al*, 2010; Tofighi *et al*, 2011; Nimtz et al. 2019) and/or by measuring BrdU incorporation after 3 days *in vitro* (DIV) using a luminescence-based BrdU Assay (Roche) and a luminometer (Baumann *et al*, 2014, 2015; Nimtz *et al*, 2019). This assay is part of a 'high content DNT test', the 'Neurosphere Assay' (NPC1-5) and uses withdrawal of growth factors as an endpoint-specific control (Baumann *et al*, 2016; Nimtz *et al*, 2019).

3.1.2. The cMINC Neural Crest Cell Migration Assay (UKN2 Test)

Neural crest cells (NCC) differentiate during neurulation from the cells of the neural roof plate. These cells eventually give rise to over 100 different cell types in the human body, including the peripheral nervous system, melanocytes, cardiomyocytes or facial connective tissue (Huang & Saint-Jeannet, 2004). One major feature of NCC is their migration to the different parts of the developing embryo. Once they arrive at their final destination, they differentiate into the according cell type. A large percentage of developmental disorders (e.g. congenital heart defects, orofacial clefts, Hirschsprung's disease) are caused by inhibition of neural crest cell migration. These kinds of alterations can be induced by genetic factors (Lee et al, 2009) or exposure to pharmaceuticals (e.g. valproic acid, Fuller et al, 2002) and pesticides (e.g. triadimefon, Menegola et al, 2005). For the migration inhibition of NCC (cMINC assay, UKN2), hiPSC are differentiated into NCC, which are subsequently used for the migration assay (Nyffeler et al, 2017). As endpoint-specific controls, Cytochalasin D, a known inhibitor of actin polymerisation and taxol, which inhibits the breakdown of microtubules are used. As cell migration requires a dynamic variability of the cytoskeleton these two compounds lead to migration inhibition.

3.1.3. Primary hNPC Migration Assay (NPC2 Test)

Cortex development takes place during the fetal phase of development. It involves radial glia migration leading to the development of a scaffold that is subsequently used by neurons to migrate and reach their final cortical destination. In contrast to rodents, human brain is gyrencephalic and radial glia composition of gyrencephalic species differs from non-folded brain surface species and determines gyrencephaly (Borrell & Götz, 2014). Thus, NPC migration is a fundamental neurodevelopmental KE that, when disturbed, e.g. in methylmercury exposed children, leads to alterations in cortex development (Choi, 1989).

Primary hNPCs of fetal origin (Lonza) grow as neurospheres in 3D (see NPC1). Cells that migrate out of the sphere display radial glia morphology and are NESTIN, SOX-2 and PAX-6 positive (Moors et al, 2007, 2012; Schmuck et al, 2017; Edoff et al, 2017; Tofighi et al, 2011; Barenys et al, 2016; Moors et al, 2009; Baumann et al, 2015, 2016). NPC migration of these radial glia cells has been extensively characterized (Edoff et al, 2017; Moors et al, 2007, 2009; Gassmann et al, 2010, 2012; Baumann et al, 2016). It can be assessed by brightfield microscopy in unstained cells because radial glia build the scaffold for the other cell types' migration and form the leading migration edge. Moreover, high content image analyses (HCA) and the Omnisphero software (Schmuck et al, 2017) can be used to automatically acquire information on migration of Hoechst-stained spheres (Baumann et al, 2016; Schmuck et al, 2017). As an endpoint-specific control, epidermal growth factor (EGF) is used to accelerate and increase migration (unpublished data), whereas

the src kinase inhibitor PP2 is used to reduces migration distance (Baumann et al, 2015; Nimtz et al, 2019). Secondarily, after radial glia differentiation, neurons and oligodendrocytes arise (Schmuck et al, 2017). In this multicellular secondary 3D model (Alépée et al, 2014) with the help of convolutional neural network (CNN) immunostained neurons (TUBB3) and oligodendrocytes (O4) are identified and their position in the NPC migration area detected using the Omnisphero software (Schmuck et al, 2017). With this the a) radial glia cell, b) early neuronal and c) oligodendrocyte migration can be assessed (Schmuck et al, 2017). The NPC2 assay can be multiplexed with early fetal neuronal and oligodendrocyte differentiation (NPC3 and 5) as well as neuronal morphology (NPC4), yet these are described as separate assays as they can also be studied without migration measures (Baumann et al, 2015; Nimtz et al, 2019).

3.1.4. Primary hNPC Neuronal Differentiation Assay (NPC3)

During NPC migration out of the plated neurosphere (NPC2) cells differentiate into GFAP+/NESTIN+ radial glia, TUBB3+ neurons and O4+ oligodendrocytes (Edoff et al, 2017; Moors et al, 2009,2012; Baumann et al, 2015) over a period of one to five days (Schmuck et al, 2017). Neuronal cells are identified by TUBB3 positive staining within the migration area of each neurosphere five days after plating by using a CNN and the Omnisphero software (Schmuck et al, 2017). The number of identified neurons is normalized to the number of Hoechst+ nuclei to calculate the percent of neurons in the mixed culture migration are of each sphere. NPC3 can be multiplexed with NPC4 (neuronal morphology, see below) or NPC2 (radial glia and neuronal migration); in the latter, information on neuronal (TUBB3+ cell) positioning is further processed to values of neuronal migration (Schmuck et al, 2017). In addition, multiplexing of NPC3 with NPC2 and NPC5 (oligodendrocyte differentiation and positioning, see below) after five days *in vitro* reveals information on neuronal and oligodendrocyte differentiation and migration within one assay (Schmuck, unpublished data). EGF inhibits neuronal differentiation and is used as an endpoint-specific control (Nimtz et al, 2019; Baumann et al, 2015).

3.1.5. Neuronal Morphology (neurite length and area) of young neurons differentiated from hNPC (NPC4 Test)

The outgrowth of neurites is a major process during brain development. It is needed for the formation of dendrites and axons and is therefore a prerequisite for cell connectivity of neurons. A disturbed or impaired neurite outgrowth during human brain development is thought to be one reason for the development of autism spectrum disorders. Therefore, this test method was developed in order to more rapidly assess chemical toxicity on the growth of neurites.

The NPC4 assay is an extension of the NPC3 assay. All neurons that were identified in NPC3 are analyzed for their morphology with a skeletonization algorithm using the Omnisphero software. Thereby the morphological characteristic of young TUBB3+ neurons is determined as neurite length and neurite area (Schmuck et al, 2017).

3.1.6. The NeuriTox Neurite Outgrowth of CNS Neurons Test (UKN4 Test)

For the establishment of this test method, immortalized primary (LUHMES) cells derived from an 8-week old mesencephalon were used (Scholz et al, 2011). These cells are kept in a progenitor status by overexpression of v-myc under the control of a TET-off promotor. Upon silencing of the v-myc expression

the neuronal progenitors differentiate into mature post-mitotic neurons in 6 days. In order to assess effects of chemicals on neurite outgrowth the differentiating cells are plated after two days of differentiation into 96-well plates and are treated for 24 hours (Krug *et al*, 2013). Then the cells are stained with Hoechst and calcein and imaged for total neurite area with an automated microscope. The viable cells and the neurite area are determined by double positivity and measurement of calcein-positive pixels by the software of the microscope. As an endpoint specific control for neurite growth, narciclasine is used that activates Rho which induces stress fibre contraction and thereby reduces neurite outgrowth. It was also shown that neurite enhancement is possible via inhibition of stress fibre contraction using HA-1077 a Rho-associated kinase inhibitor or blebbistatine, which inhibits myosin II (Krug *et al*, 2013).

3.1.7. The PeriTox Neurite Outgrowth of PNS Neurons Test (UKN5 Test)

Besides the neurite outgrowth of CNS neurons, also the neurites of PNS neurons are sensitive targets of chemicals. A prominent example is the development of neuropathies during chemotherapy after treatment with platinum compounds (Quasthoff and Hartung, 2002). In addition, acrylamide is a known toxicant that induces neuropathies in humans.

In order to differentiate immature human dorsal root ganglia cells, hiPSC are used, differentiated for 8 days resulting in neural crest cell like neuronal precursors. These progenitor cells can then be frozen in liquid nitrogen. After thawing, the cells immediately start to grow neurites. Therefore, one hour after thawing the cells are treated for 24 hours with different chemicals and stained with Hoechst and calcein. For imaging and quantification of viable cells and neurite area the principle is the same as in the UKN4 test method (Hoelting et al, 2016). As an endpoint specific control for neurite inhibition, vincristine (microtubule toxicant), colchicine (microtubule polymerisation inhibitor) and cytochalasin D (actin polymerisation inhibitor) are used, as they all interfere with the cytoskeleton and inhibit the neurite growth. Narciclasine is also used as an endpoint specific control for neurite inhibition as it activates Rho.

3.1.8. Oligodendrocyte Differentiation (NPC5)

Oligodendrocytes are necessary for brain function as they are the myelinating cells of the central nervous system. Loss in oligodendrocyte number or function produces reduced myelin in the brain that is followed by severe neurodevelopmental deficits as in the Alan-Hernon-Dudley syndrome due to a pre- and postnatal brain hypythyroidism. They are generated from NPC via oligodendrocyte progenitor cells, undergoing different maturation stages until becoming myelinating oligodendrocytes. The O4 epitope used for immunocytochemical oligodendrocyte identification in this study hence identifies cells that are committed to the oligodendrocyte lineage starting to be expressed in pre-oligodendrocytes (Figure 5).

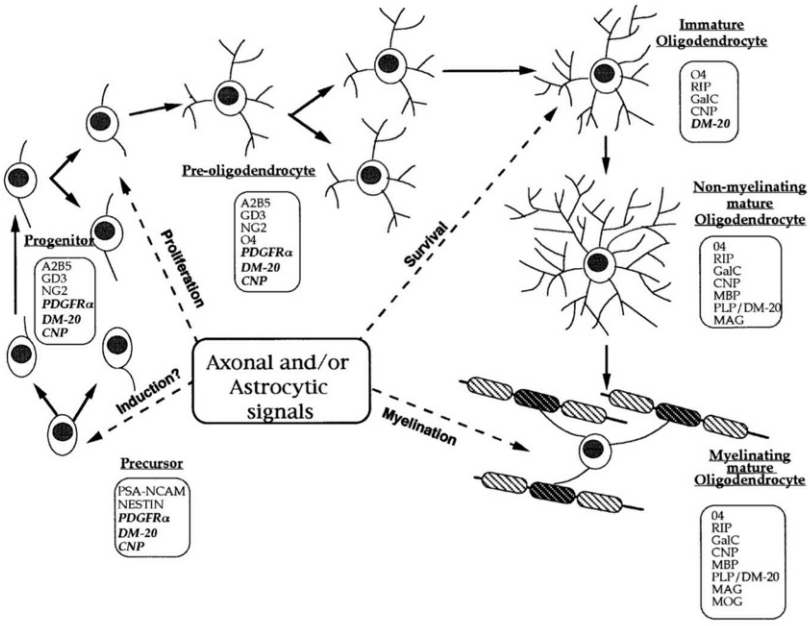


Figure 5: Schematic representation. Here, we use the term 'oligodendrocyte' for addressing cells that are committed to the oligodendroglia lineage of the developmental stages of cells of the oligodendrocyte lineage. Schematic drawing of the morphological and antigenic progression from precursor cells to myelinating mature oligodendrocytes, through progenitors, preoligodendrocytes, and immature nonmyelinating oligodendrocytes. Timing of neuronal and astrocytic signalling is indicated. Stage-specific markers are boxed. RNAs are in italics. From: Baumann & Pham-Dinh, 2001.

Proper oligodendrocyte differentiation and maturation is thus a prerequisite for successful neurite myelination. With the NPC5 assay, oligodendrocyte differentiation and maturation are studied. Oligodendrocytes differentiate in co-culture with astrocytes and neurons in secondary 3D structures from hNPC within 5 DIV (Dach et al, 2017; Alépée et al, 2014; Moors et al, 2009; Schreiber et al, 2010; Schmuck et al, 2017) into a mixed culture of pre-oligodendrocytes, immature, and non-myelinating, mature oligodendrocytes (Figure 5). The hNPC oligodendrocyte differentiation and maturation was recently put into an assay format (NPC5; (Dach et al, 2017; Nimtz et al, 2019). DAPI-stained nuclei that co-localize for the epitope O4 after 5 DIV are identified. The number of identified O4+ oligodendrocytes divided by the number of total nuclei in the migration area reveals % of differentiated oligodendrocytes (NPC5; Baumann et al, 2016; Barenys et al, 2017; Dach et al, 2017; Schmuck et al, 2017). Because oligodendrocyte precursor/progenitor cells do not express O4, the NPC5 test method assesses oligodendrocyte differentiation from the precursor/progenitor state into the oligodendrocyte lineage. Here, all subsequent oligodendrocyte maturation stages are assessed as they all express the O4 epitope. The process of direct neurite myelination is not part of the NPC5 assay.

3.2. Readiness Evaluation of DNT Test Methods

A systematic approach to building a test battery should first determine the readiness of individual alternative DNT methods. A general set of readiness criteria has been proposed by OECD (2014a), and these have been clustered in four categories (Table 1). Such guidance has been considered here in compiling specific readiness criteria for DNT test methods, and in devising a preliminary scoring system to obtain indications on the readiness status of various published tests.

Table 1 Example for ranking parameters for in vitro methods for detection of chemicals. Ranking parameters were established by the OECD for thyroid-disrupting chemicals to determine readiness of tests for validation (OECD, 2014).

CATEGORY 1 **Initial High Priority Considerations**	**CATEGORY 2** **Method performance considerations**
Biological plausibility Extrapolation to humans or broadly applicable across vertebrates/phyla Availability of resources Reference chemicals	Within-laboratory reproducibility Between-laboratory reproducibility Assay variability Accuracy Assay specificity/assay sensitivity
CATEGORY 3 **Technical capability**	**CATEGORY 4** **Other practical considerations**
Dynamic range/concentration test range Detection/adjustment of confounding factor and/or Incorrect/inconclusive measurements and/or other bias Response characterization	Technological transferability/proprietary elements Transparency of the method Documentation of development and utility of the method

Ranking parameters were established by the OECD for thyroid-disrupting chemicals to determine readiness of tests for validation (OECD, 2014). The criteria in **Category 1** are considered of highest priority. Each criterion within this category is considered to have equal weight, and all are essential to demonstrate the readiness of the assay. For instance, the assessment of the biological plausibility is considered very important in defining readiness of the method for validation. However, criteria in this category are hard to quantify. Moreover, many DNT tests cover multiple mechanisms and processes with varying levels of plausibility and data on their *in vivo* relationship. Thus, the practical value of such criteria for DNT methods needs to be considered case-by-case. The criteria for **Category 2** are better defined and quantifiable. They relate to the evaluation of reliability and efficacy of the method. Sufficient positive and negative compounds should be included to assess specificity and sensitivity, and focus should be given to the robustness of the assay. Regarding **Category 3**, the criteria are also relevant to assay performance evaluation. However, the particular performance issues described under this category are considered to be of less significance during initial phases of test development and evaluation. **Category 4** contains criteria for the methods that are considered as good to meet, in order to gain broad acceptance.

For assay scoring, the DNT consortium (Bal-Price et al, 2015a; Fritsche et al, 2017) created an assay scoring system during a workshop (Bal-Price et al. 2018). This easy-to-use scoring system assigns a maximum score to each criterion (see fourth column in Table B5, Annex B), and establishes a simple tool for clustering of scores (Table B3, Annex B). The scores were assigned on the basis of publicly available information

extracted from publications. More information on the readiness criteria of the specific assays used in this project is found under '7. WP4: Fit-for-Purpose validation'.

4. WP2: Analyses of *in vitro* Assays for Redundant Assays

For analysing DNT assays for redundancy, two fundamentally different approaches were applied:

1. The first approach used statistical methods like principal component and correlation analyses to identify the smallest sets of test methods that give full information on DNT hazard. For this, experience was used from other cross systems comparisons in international research consortia (e.g. the testing of the NTP 80 compound library; or the cross-systems testing of EU-ToxRisk). For instance, a principal component analysis (according to test hits in all DNT assays) reveals clustering of tests that have a high level of redundancy, and a wide spread over tests that are not redundant.
2. The second approach is based on biological knowledge and expert judgement which important neurodevelopmental processes are represented by the tests. This approach was used to make sure that as many key neurodevelopmental processes as possible are covered by the test battery.

Analyses of the first approach are shown in WP7. The outcome of the second approach is depicted in Figure 6. Here, possibly redundant assays dealing with corresponding key neurodevelopmental processes (orange, on the left), are grouped. However, although dealing with similar subjects, the assays differ in cell types, developmental timing and models, like primary NPC versus hiPSC-derived NPC. Apart from this procurement, the US-EPA added possibly redundant assays for proliferation, neuronal network formation and synaptogenesis to the battery. Due to the COVID-related delay in testing, however, the data for the US-EPA assays is not complete for the 120 chemicals. In addition, the human NNF assay only has data for two case study compounds. Therefore, the assessment of redundancy is somewhat restricted, yet still very helpful.

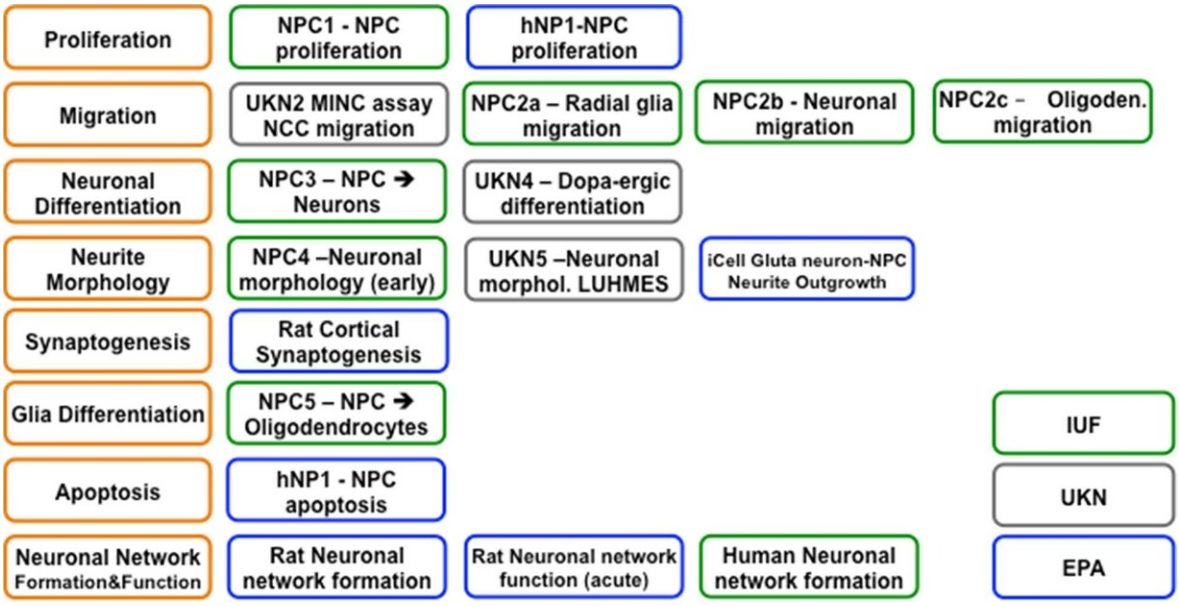


Figure 6: Redundancy assay evaluation of a DNT *in vitro* testing strategy covering neurodevelopmental endpoints during fetal stages of brain development.

For helping to understand potential differences in redundant assay performance described in WP7, we assembled basic test method characteristics and present them in a reduced information table format for easy use (Table 2 - Table 5). From these tables it is obvious that although e.g. measuring migration, cell types used for migration analyses are completely different. Because timing of occurrence (e.g. neural crest cells migrating very early during embryogenesis for organogenesis and e.g. oligodendrocytes in accordance to their formation migrate rather late in the developmental process for improving neuronal function) and also stimuli for migration are very different across cell types, we do not expect a large overlap of migration effects between the assays UKN2, NPC2a,b and c. In the case of neuronal network formation, for example, the situation is less clear. Here, the test methods do not differ in basic cell types (neurons and astrocytes), yet in species (primary rat versus hiPSC-derived human) and assay duration (12 days versus 35 days). Therefore, the performance of assays, even the ones covering the same endpoints, might be very specific for cell type, species, duration or even method of endpoint assessment.

Table 2 Assessment of test method characteristics for possibly redundant assays: NPC proliferation

Test Method	Test System	Dimensionality	Assay duration	Method of evaluation
NPC1	primary human NPC (Lonza)	3D	72 h	Sphere Area/BrdU incorporation
hNP1	neuroepithelial derived stem cells (cell lineage WA09; ArunA)	2D	26 h	BrdU incorporation

Table 3 Assessment of test method characteristics for possibly redundant assays: NPC migration

Test Method	Test System	Dimensionality	Assay duration	Method of evaluation
UKN2	hiPSC-derived neural crest cells	2D	24 h	Measurement of gap filling
NPC2a	primary human NPC (Lonza) differentiated into radial glia in a mixed culture	secondary 3D	72 h& 120 h	Measurement of spontaneous radial glia travelling distance out of the sphere
NPC2b	primary human NPC (Lonza) differentiated into neurons in a mixed culture	secondary 3D	120 h	Measurement of spontaneous neuronal travelling distance out of the sphere
NPC2c	primary human NPC (Lonza) differentiated into oligodendrocytes in a mixed culture	secondary 3D	120 h	Measurement of spontaneous oligodendrocyte travelling distance out of the sphere

Table 4 Assessment of test method characteristics for possibly redundant assays: neuronal morphology

Test Method	Test System	Dimensionality	Assay duration	Method of evaluation
NPC3	primary human NPC (Lonza) differentiated into neurons in a mixed culture	secondary 3D	120 h	Neurite length (µm) and neurite area (pixel) based on TUBB3 positive neurons
UKN4	Lund human mesencephalic (LUHMES) differentiating to dopaminergic neurons	2D	24 h	Neurite area (pixel) based on calcein AM positive cellular structures
UKN5	Sensory neurons differentiating from the hiPSC line SBAD2	2D	24 h	Neurite area (pixel) based on calcein AM positive cellular structures

Table 5 Assessment of test method characteristics for possibly redundant assays: neuronal network formation

Test Method	Test System	Dimensionality	Assay duration	Method of evaluation
rNNF	Primary rat cortical neurons	2D	12 days	Electrical activity on MEAs
hNNF	hiPSC-based GABAergic/glutamatergic neurons and primary human astrocytes (NeuCyte, USA)	2D	35 days	Electrical activity on MEAs

MEA multi electrode array

5. WP3: Identification of Gaps in the Testing Battery

5.1. General Procedure

After identification of available *in vitro* methods that can contribute to a DNT testing battery (WP1/2), a gap analysis was performed. For one, this uses information on the readiness of assays, i.e. endpoints measurable versus test methods available, second, it uses redundancy information from WP2 and third, it used expert knowledge on basic neurobiology. However, here the focus is not on potential overlaps (as in WP2), but on the gaps not yet filled. While the readiness analyses were performed and test methods identified, the strategy for the expert knowledge on basic neurobiology is as follows:

Step1: the applicability domain for the DNT test battery is described ('what can it detect?')

Step2: the regulatory and scientific needs are described ('what compounds should be detected?')

Step3: by comparison of step1 vs step2, gaps are identified.

Details to step 1: the applicability domain for the DNT battery is described on three levels in order to appropriately consider physiology, toxicology and chemical compound properties:
(i) which general biological features are reflected (cell-cell/cell-matrix interactions; xenobiotic metabolism; differentiation stages),
(ii) which type of chemicals can be tested and identified (physicochemical properties; industry sectors; chemical classes),
(iii) which general regulation pathways are reflected (i.e. which AOP key events and signaling cascades can be assessed).

The expert knowledge on basic neurobiology thereby points out gaps that are not covered at all by the battery.

5.2. Specific DNT Battery Evaluation

5.2.1. Description of the applicability Domain of the DNT testing battery:

The proposed DNT testing battery is the sum of its single assays (Figure 6; Table B1, Annex B). Therefore, the single assays describe the applicability domain of the battery on the cell biological level. Figure 7 summarizes necessary neurodevelopmental endpoints and indispensable cell types that orchestrate brain development.

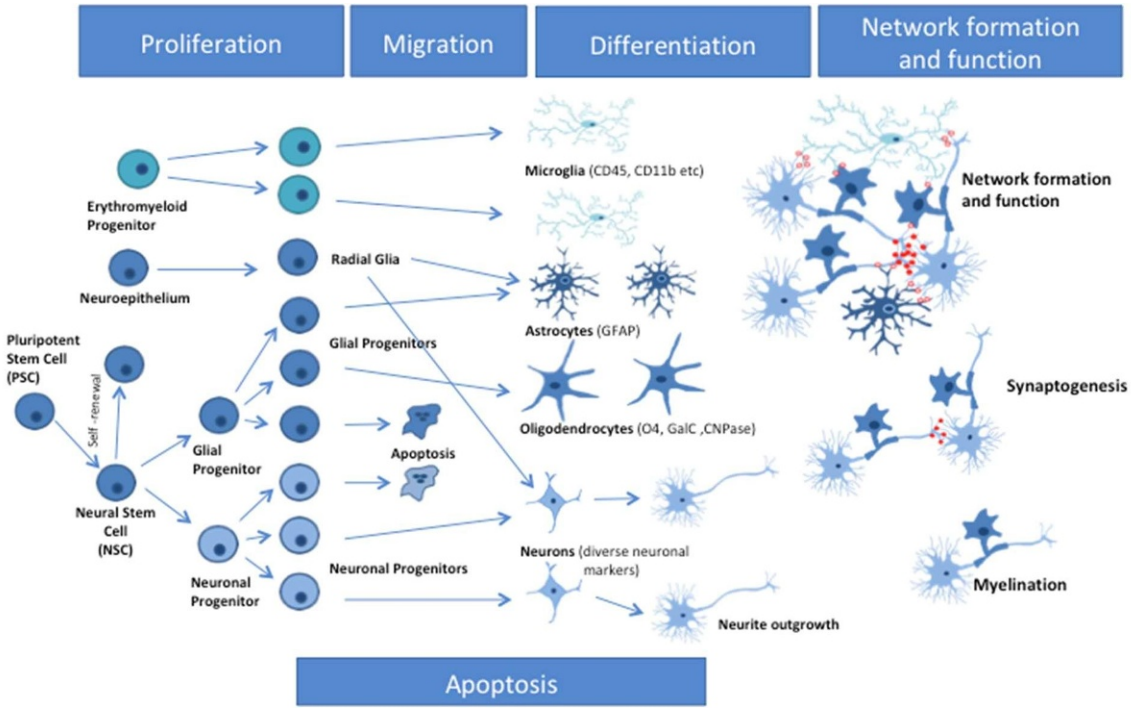


Figure 7: Basic neurodevelopmental endpoints necessary for brain development (Bal-Price et al. 2018).

A comparison of these cell types and processes with the test methods present in the current DNT battery demonstrates that certain gaps already exist in the battery on the level of the presence of test methods

for analyzing DNT endpoints (Table B1, Annex B). However, the majority of endpoints can be evaluated with the established test methods (Bal-Price et al. 2018):

- Neural Crest Cell (NCC) Migration
- NPC proliferation
- NPC apoptosis
- Radial glia migration
- NPC differentiation into neurons
- Neuronal migration
- Neuronal morphology
- Synaptogenesis
- Oligodendrocyte differentiation
- Oligodendrocyte migration
- Neuronal network formation with glutamatergic and GABA-ergic functioning

These test methods were previously set up with mainly human test systems, i.e. cells, which represent the different cell types of the developing brain, like hiPSC-derived NCC, or primary or hiPSC-derived NPC, which have the ability to differentiate into neurons, astrocytes and oligodendrocytes. Some of the cell systems are multicellular thus containing cell-cell as well as cell-matrix interactions (e.g. NPC2-5).

Development is an on-going process over time involving organogenesis during the embryonic phase (until gestation week 8) and further organ growth during the fetal period. Due to these different tasks at distinct developmental phases and time points, sensitivity to chemicals might be dissimilar. This shapes the need for implementing different stages of development into the battery. Therefore, the current DNT *in vitro* battery contains test systems originating from hiPSC and fetal cells, thus representing fetal neurodevelopmental stages.

Concerning metabolism, it is known that the fetal brain has a limited phase 1 metabolism. However, detoxifying phase 2 metabolism might protect against threats like oxidative stress. Primary hNPC, for example, have the capacity to up-regulate protective strategies upon reactive oxygen species (ROS) exposure. However, their capacity only reaches about a third of the glutathione-dependent protective level than primary rat NPC. Here, expression levels of genes involved in the antioxidative defense were comparable between the *in vitro* systems and developing human and rat brains *in vivo* (Masjosthusmann *et al*, 2019) pointing to the physiological relevance of the test system.

In addition to the already well established test methods of the current DNT *in vitro* testing battery, a variety of additional and crucial endpoints for brain development have been studied on an academic level. These endpoints have to be taken to the next level, i.e. to test method development (Crofton et al. 2011). A high priority should be given to set-up methods, where the basic science is already established and the endpoints measured follow a scientific reason. These include:

- Neural rosette formation
- hiPSC-derived NPC proliferation
- hiPSC-NPC neuronal differentiation
- Neuronal subtype differentiation
- Human synaptogenesis
- Astrocyte differentiation and maturation
- Astrocyte reactivity
- Microglia activity

Concerning astro- and microglia assays, there are several test systems on their way that might be added to the battery in the future. Rat 3D spheres contain astro- and microglia and have been used for toxicity testing for many years. However, recent developments in hiPSC-derived glia cell models will most likely be the test systems of choice in the future.

With the above-mentioned test systems a large variety of compounds and compound classes can be evaluated. These include industrial chemicals, cosmetics ingredients and pharmaceuticals. Compounds need to be soluble in a solvent at a solubility where the solvent does not produce effects by itself in the *in vitro* systems. Compounds that have to be tested with more sophisticated exposure methods like 'passive dosing' include highly volatile or lipophilic compounds. Such methods are currently not established for routine applications yet would be very helpful to include in the future. Also, nanomaterial testing has to be performed with caution as the particle behaviour changes when moving from an air into a fluid environment.

One tool currently well accepted for novel hazard and risk assessment strategies is the AOP (adverse outcome pathway) concept (Carusi et al, 2018). The AOP consists of multiple key events (KE) across different layers of organization from the molecule to the individual or population, which are causally linked by KE relationships (KER). The AOP provides certainty that the effect measured *in vitro* is linked to an adverse outcome (AO) relevant for chemical regulation. Comparing AOP-based KE to the neurodevelopmental processes that we propose to include in the DNT *in vitro* testing battery identified additional gaps. The DNT-AOPs submitted to the AOP-Wiki are rather limited (Table B2, Annex B), yet helpful. MIEs (molecular initiating events) addressed in the AOPs currently available on the AOP Wiki are leading to KE altering intracellular calcium levels, TH homeostasis or modulating NMDA receptors. In addition, four putative DNT AOPs have been published that add data. Targets for MIEs within these AOPs are N-Methyl- d-aspartic acid (NMDAR), SH-groups of proteins, ryanodine receptors (RyR) or the extracellular matrix (ECM) protein laminin. As MoA of DNT compounds reach far beyond interference of compounds with these receptors/pathways, at this point the development of an only on AOP-based testing strategy is not feasible. Development of AOPs for DNT is highly warranted, however, considering the multiple neurodevelopmental targets that might possibly serve as sites of molecular initiation and the time it takes for developing AOPs that are in accordance with the OECD guidelines for AOP development, such an AOP-informed testing strategy will have to be envisioned for the future. As there is a clear gap in AOP development, it is highly likely that the data generated within this project using these standardized assays from the *in vitro* battery will contribute to further AOP development. From the presently available AOPs, however, we already identify gaps that cannot be assessed with the currently proposed testing strategy:

- Altered function of Glu transporters in astrocytes (early KE)
- Direct impaired function of synaptic proteins related to Glu or indirect through disruption of their NO regulation (early KE)
- RyR sensitization (early KE)
- Reduced release of BDNF (intermediate KE)
- Altered dendritic arborization (intermediate KE)
- Altered synaptogenesis (late KE)
- Reduced neuronal survival, differentiation and synaptogenesis (late KE)

This analysis shows that lack of methods or methodologies cover levels of organization from early to late KE. Also, the intermediate KE neurite morphology for dendritic arborisation evaluation as well as neuronal subtype differentiation were amongst the methods already identified above as high priority.

Another aspect of gap analysis is the evaluation of neurodevelopmental processes/assays that we have not covered in the proposed DNT testing battery yet. Expert judgement suggests several aspects that need covering in the future. These include but are not restricted to:

- Astrocyte differentiation

- Developing brain region-specific astrocytes
- Developing astrocyte-endothelial cell interaction (blood-brain-barrier)
- Placental barrier
- Microglia-interaction during development
- Interference of compounds with growth factors, neurotropins, trophic neurotransmitters or ion channels/receptors (see WP6)
- Endocrine aspects of brain development
- Learning of neuronal networks
- Myelination of axons

From the priority side, first the 'low hanging fruit' should be gone after, i.e. set-up of test methods that already have science ready for further development. This includes for example influence of compounds on expression of BDNF and synaptogenesis (Pistollato et al, 2020). Also the differentiation of mainly GABAergic neurons was currently established (Nimtz *et al*, 2019) allowing assessment of GABAergic function. However, also new developments like endocrine-related DNT (current development in the Horizon 2020 project EnDpoiNTs) should be adopted into the battery as soon as respective assays are available.

While this gap analysis is important for future interpretation and application of the DNT testing battery, it was beyond the scope of the procurement to establish such new methods.

5.2.2. Regulatory and scientific Needs of the DNT *in vitro* Testing Battery

From the regulatory point of view, it is of outmost importance that most of the compounds' MoA for DNT are covered in the detection ability of the DNT testing battery. Especially for understanding 'negatives' in test results, such information is of high relevance. This is not easy to achieve, because the necessary knowledge on human neurobiology as well as on human DNT MoA is rather sparse. In addition, there is a plethora of unknown DNT MoA, with new compounds constantly adding to the list, that hampers high certainty. Nevertheless, from the current knowledge base we strive for best MoA description of the DNT *in vitro* battery with the highest level on scientific understanding.

With regards to basic neurobiology, a variety of pathways contribute to neurodevelopment. Due to human mutations, some of them are known to be pathognomonic for humans. In addition, compounds that are developmentally neurotoxic to humans or animals *in vitro* or *in vivo* possess information on their MoA. Such information on a) DNT compounds and b) pathways is summarized in Tables A3 and A4 of Annex B, respectively (Fritsche, 2016). These data link pathways to neurodevelopmental processes. During this procurement we therefore tackled the issue from these two angles. The approach we took was to select a number of chemicals that are known to trigger DNT in vivo (Aschner *et al*, 2017; Mundy *et al*, 2015). We examined in detail, how the tests of the battery responded to these compounds (hit calls or non-hit calls). From this we derived an initial estimate of sensitivity and we got an impression which types of compounds are well-detected or fail to be detected. This approach was expanded by selecting a number of chemicals that do not trigger DNT in vivo (according to all available evidence and reasoning). Examination of the data on these compounds gave information on potential false negatives (FN) and on types of compounds (and their modes of action) with liability to trigger such effects. Altogether, such comparisons were used to estimate the battery specificity (for results see 8.4.4). This is a pragmatic approach because for most of the compounds their MoA concerning DNT is not known. In addition, we analyzed the results for 24 compounds with known effects on signalling pathways for parts of the battery (for results see WP6). Information on the ability of the battery concerning MoA in general, even if for specific compounds their molecular and cellular mechanisms are unknown, reduces uncertainty. Moreover, the application domains of the individual assays are less important than the application domain of the whole battery. In the end,

the whole battery is the entity that will lead to further decision-making, the individual assays are only used in the second step, e.g. for further targeted *in vivo* testing.

In addition to these biological and toxicological issues, which will be specifically addressed in WP5a and WP6, standard performance characteristics are crucial. This concerns assay performance including endpoint stability and reproducibility. As this battery contains assays that mimic basic processes of neurodevelopment, one important control measure is the change of a measured endpoint in the test system over time. This includes e.g. growth of neurospheres, migration of radial glia or neural crest cells, differentiation of LUHMES cells or NPC into neurons, radial glia or oligodendrocytes, or growth of neurites. The ability of the DNT battery test systems to reproducibly undergo developmental changes over measuring time is thus the first prerequisite for assay performance. In the second step, reproducible modulations of measured endpoints indicate assay reproducibility. Here, ESC are a valuable tools (Crofton et al. 2011). Data for assay reproducibility for negative and positive controls is presented in WP5b (8.4.1).

6. WP4: Fit-for-Purpose Validation

We developed a roadmap for the 'fit-for-purpose' validation of the individual tests and the whole battery. This fit-for-purpose validation expands on previous work performed in the field, particularly for DNT (Crofton et al, 2011; Basketter et al, 2012; Bal-Price et al, 2015b; Leist et al. 2012, 2014; Bal-Price et al. 2015b). With the resources and time constraints given, a classical regulatory validation was neither feasible nor helpful. However, some elements of the modular validation approach, as suggested by (Hartung et al, 2004; Hartung, 2007, 2010), were used. For instance, an evaluation of the technical robustness of a test and of its classical performance criteria (Crofton et al, 2011; Leist et al. 2010), such as signal to noise ratio, linear range, detection limit, performance of negative and positive controls etc. are part of a fit-for-purpose validation.

Regulatory applications of the battery include 'screening and prioritization' as well as 'hazard characterization as a step in the process of risk assessment with targeted *in vivo* testing'. For the battery performance for screening and prioritization, it is important that a cellular disturbance is detected at a correct potency level. Here, it is less important what the specific endpoint hit really is within the battery, because for the prioritization exercise it does not matter if a compound affects e.g. proliferation or synaptogenesis as long as the substance is prioritized correctly according to its potency for DNT. We expect some signaling pathways to exert some redundancies across different neurodevelopmental endpoints. For example, the Epidermal Growth Factor Receptor (EGFR) antagonist PD153035 increases neurogenesis, while it reduces oligodendrogenesis and radial glia migration (Masjosthusmann *et al*, 2018). However, pathways do not necessarily have to be redundant across neurodevelopmental processes, for example src kinase solely contributes to radial glia migration, while neurogenesis and gliogenesis are unaffected (Moors *et al*, 2007). Therefore, the screening and prioritization exercise has to cope with some uncertainty. When the battery is used for hazard characterization as a step in the process of risk assessment including prioritization for targeted *in vivo* testing, the situation is different. Here, we need to not only determine a compound's potency on any neurodevelopmental endpoint, but in addition gain a deep understanding of the mechanistic basis of the neurodevelopmental disturbance(s). Here, any uncertainty in the battery might be an issue, especially for the following targeted *in vivo* evaluations. For both scenarios, screening and prioritization as well as hazard characterization, assay performance in itself is of high relevance for scientific validation as already pointed out in WP3, 5.2.2.

Here, we use knowledge from WP1-WP3 as well as historical and new data on compound testing and assay performance for scientific validation. The first step of the fit-for-purpose validation is based on the establishment of a quantifiable set of readiness criteria (Table B5, Annex B). These criteria are exemplified

for the assay UKN5 in detail (Table B6, Annex B) and summarized in Table B7, Annex B. In a second step, all assays were scored for these criteria (Table B8, Annex B). In the third step, use scenarios were defined. For each use scenario (e.g. prioritization and screening), the most relevant readiness criteria were combined to an overall score. The score compared to the maximum reachable score indicated the readiness of the respective test method for the given use scenario. This clustering of scoring criteria is an important concept, as it allows individual scores for phases of test development. Phase I concerns all criteria that can be fulfilled during initial test method development. Phase II criteria refer to the test method performance based on e.g. evaluation of replicates to conclude on robustness and reproducibility. Phase III is optional as a proper screening is not always feasible for each test method, i.e. 2nd and 3rd tier methods. This allows a distinction of readiness for e.g. academic research purposes, screening and prioritization, or regulatory risk assessment.

In addition, assay and whole test battery evaluation was performed according to the OECD GD211 (ENV/JM/MONO(2014)35). However, OECD GD211 is targeted mainly at regulators, and therefore leaves scientists less familiar with regulation uncertain as to what level of detail is required and how individual questions should be answered. Moreover, GD211 pays little attention to the description of the test system (i.e., cell culture) and the steps leading to it being established. Therefore, we modified OECD GD211 to adapt it to the common needs of test developers and regulators. We developed an annotated toxicity test method template (ToxTemp) (i) to fulfil all requirements of GD211, (ii) to guide the user concerning the types of answers and detail of information required, (iii) to include acceptance criteria for test elements, and (iv) to define the cells sufficiently and transparently (Krebs *et al*, 2019). This ToxTemp is used for describing the test methods of this DNT *in vitro* testing battery. ToxTemp for the individual test method descriptions can be found in Annex C-G.

This fit-for-purpose evaluation procedure is further worked out to make it broadly available to other fields and stakeholders. A 'validation score card' (VSC), could give quick information to regulators that need to consider data from an *in vitro* test. A scoring scheme is exemplified in Bal-Price *et al.* (2018a) and further outlined in Krebs *et al*. (2020b) . As basis for a scoring system, a template was developed (ToxTemp, Krebs *et al*, 2019). An international consensus process (including EURL-ECVAM) will be required to assign actual points to the individual pieces of information, and to make such a scheme broadly acceptable.

In line with the new roadmaps of the EPA and the FDA on the use and implementation of NAMs in risk assessment, the focus of the validation process would be shifted from classical to fit-for-purpose validation: Classical validation puts strong emphasis on correlation of test results to a gold standard. In a fit-for-purpose validation, we first make sure the assays are robust. Then it is evaluated how well the scientific rationale is described, and what exactly the test is used for. Correlative information will then be added with progressing time and use of the test.

Structure of the VSC:

I. Model characteristics/Fit for purpose
II. Test performance criteria/robustness
III. Biological rationale for test relevance
IV. Toxicological rationale for test relevance

Our roadmap plans to provide following information:

For I: Which neurodevelopmental process is modelled. Why is this process important and relevant? What kind of toxicants can be expected to be detected? A throughput evaluation will also be added here

For II: Baseline variation; statistically-based benchmark response (SNCD), as described by Sand et al. (2017), historical control values for positive controls

For III: Which AOP KE/MIE are reflected by the test method? Which developmental signaling pathways of relevance are reflected? Is the sensitivity to end-point specific controls evaluated?

For IV: Which toxicants are detected? Which ones are not detected? Which ones are expected to be detected/not detected.

With the above-explained procedures, the fit-for-purpose validation performed here dissected each test method into its four key elements (Schmidt 2017): **1. The test system; 2. The exposure scheme; 3. The assay and analytical endpoint(s); and 4. The classification model**. Each of these evaluation domains were sub-specified, where required and appropriate, and general considerations on **biological and toxicological relevance**, on the **suitability for high-throughput testing** and on experience with known DNT compounds were added. This information is specified for each test method in WP5 and in the ToxTemp (Annex C-G).

7. WP5a: Selection of Test Chemicals

The test compound set used in this study was assembled from different groups of chemicals. The core of our selection consisted of 'human DNT compounds'. These are chemicals that have been shown to be associated with DNT in the human population (Mundy et al, 2015; Aschner et al, 2017; Grandjean & Landrigan, 2006, 2014). They are therefore here also termed as "human positive". For purposes of the evaluation of the test battery performance, this group is used as gold standard. Our expectation was that these compounds should be positive in at least one assay of the DNT testing battery. By contrast, if one or more of these compounds test negative in the battery, we would conclude that there still may be gaps in the biological processes covered by the battery. It should be noted that the total number of "human positive" DNT compounds is small (20, with only 9 tested in UKN and IUF assays) and allows only a limited evaluation of the battery.

Therefore, a second group of compounds was added to our test compound collection; chemicals that have been investigated in animal models (guideline or non-guideline studies) for DNT (Mundy et al, 2015; Aschner et al, 2017). They are described as "*in vivo* positive" (n=53), "*in vivo* negative" (n=1) or "*in vivo* uncertainty" (n=2). This group is important because of the *in vivo* data anchoring, and also for allowing a comparison of human and animal effects. Some of these compounds were selected for IATA case studies by the OECD DNT Expert group. In our compound overview, such chemicals are additionally labelled as "IATA". A third important group consisted of compounds selected for their known (or expected) safety with regards to DNT (described as "negative control"; n=30). For these substances there is evidence (e.g. clinical use experience) that they do not cause DNT in humans (Aschner et al, 2017). They are important to determine the specificity of the test battery. Other compounds were selected for testing, because they are model compounds for an endpoint assessed in the battery ("model compound") or because of their close relationship with one of the chemicals of the above groups (described as "Metabolite/Related").

Additional selection/filtering criteria for the compound set were physicochemical properties. Compounds that are not soluble in cell culture media or that evaporate quickly were de-selected, as they pose a major challenge for *in vitro* test systems. Some compounds were de-selected, as they were not commercially available. Annex A1 gives an overview on all selected compounds, including their subgroup assignment (i.e. the reason why they were selected). The full set of compound consists of 183 test chemicals of which 166 were tested in at least one assay.

8. WP5b: Testing of a Chemical Library to challenge the Test Battery

8.1. Chemical Testing Strategy

For the chemical testing strategy, we conducted a two-step testing paradigm with different approaches.

In the first approach (applied for UKN2, 4 and 5) the highest testable compound concentration (defined by the solubility and/or availability of the compounds from ToxCast; in most cases 20 µM) was tested for specific DNT effects as well as time-matched influences on cell viability. As a cut-off to separate effect from no-effect compounds a Benchmark Response of 20 was used for all assays. In case both endpoints were negative (no response observed at the selected BMR), no additional testing was performed, and the substance was classified as negative. Only if the response in the specific DNT endpoint alone or cytotoxicity and specific DNT endpoint were higher than the BMR, additional testing was performed. In this case, testing was performed in a serial dilution (1:3) of the compound to receive concentration-response information for the individual compound and endpoint. Thereby the start concentration can be different and depended on prior knowledge of the compounds (e.g. activity in other cell-based assays), solubility and availably.

A total of 6 concentrations were tested in UKN2 and UKN5 and 10 concentrations in UKN4 in three independent experiments (representing different hiPSC differentiations for UKN2 and UKN5 and different cell passage numbers in UKN4) with technical replicates (3 in UKN4 and UKN5; 4 in UKN2) per condition over all endpoints. For each endpoint BMRs were calculated and the respective BMC values were used for the classification model calculations. If the response in the endpoint was lower than the respective BMR (defined for each assay; see 8.3.8) the compound was classified as negative.

In the second approach (applied to NPC1-5) we tested the complete concentration response of each compound in a serial dilution (1:3) consisting of 7 concentration in two independent experiments (different human individuals) with five replicates per condition over all endpoints. The start concentration thereby depended on prior knowledge of the compound, solubility and availability of the compounds from ToxCast (in most cases 20 µM to not exceed a DMSO concentration of 0.1%). For each endpoint a BMR was defined based on the variability between the experiments. If the response in any endpoint (specific DNT endpoint or general cell viability) was higher than the respective BMR, the compound was classified as "hit". In this case, the testing in the assay in which the endpoint was affected was repeated with at least one more experiment. If the hazard classification was not clear based on two (for potentially negative classification) or three (for potentially positive classification) independent experiments (e.g. due to inconclusive concentration response relationships or high experiment to experiment variability), further experiments were performed. If an endpoint was not affected based on at least two independent experiments, the compound was classified as negative for this endpoint.

A defined set of compounds was tested in two screening projects at the University of Konstanz before the compound distribution through ToxCast. Information on tested concentration is given in column "pre-project" in Annex A1.

In detail, 55 compounds were pre-tested for this project in a concentration-response screen with the highest tested concentration being 100 µM with individual exceptions based on solubility or *in vivo* information. Moreover, 18 compounds (provided by the US National Toxicology Program (NTP)) were tested in a 20 µM pre-screen with a subsequent follow-up concentration-response approach for hit candidates. Data is publicly available online https://sandbox.ntp.niehs.nih.gov/neurotox/ and published (Delp *et al*, 2018). BMCs are given in Annex A1 and concentration-response curves of the 55 compound set are included in Annex O.

8.2. Test Battery and Experimental Procedure

This chapter gives an overview over the test systems and test methods used in this test battery. For a description of the full experimental details see the respective DB-ALM SOPs attached as Annex I-M. For a full test method description according to the "Guidance Document for Describing Non-Guideline *in vitro* Test Methods" (GD211; OECD 2014) see Annex C-G. Experimental procedures and data handling were performed according to the Guidance Document on Good *in vitro* Method Practices (GIVIMP; Pamies et al, 2016; OECD, 2018).

8.2.1. Primary hNPC Proliferation Assay (NPC1 Test)

Test system

Human neural progenitor cells (hNPC) are generated from primary human brain at gestational week (GW) 16 -19. For this project four male individuals (Lot no.: 0000391398 (GW19), 0000549062 (GW16), 0000516385 (GW16), 0000553745 (GW16)) were purchased from Lonza Verviers SPRL (Verviers, Belgium). Human NPC are cultured as floating neurospheres in proliferation medium (B27), consisting of DMEM (Life Technologies, Darmstadt, Germany) and Hams F12 (Life Technologies) (3:1) supplemented with 2% B27 (Life Technologies), 1% penicillin and streptomycin (Pan-Biotech, Aidenbach, Germany), 20 ng/ml epidermal growth factor (EGF, Life Technologies) and 20 ng/ml recombinant human fibroblast growth factor (FGF, R&D systems, Wiesbaden, Germany). The cultures are maintained at 37°C with 5% CO2. In the first 3 to 4 weeks after thawing the spheres are allowed to grow. After this period the cells are passaged every week by mechanical chopping of the spheres with a tissue chopper (Mcllwain Tissue Chopper, Vibratome). Over the whole cultivation time the cells are fed every 2 to 3 days by replacing half the medium.

Test method

see 3.1.1

NPC1 Assay

Each compound is tested in serial dilutions (1:3) with 7 concentrations and a solvent control (SC) plated in five replicate wells per condition in a 96-well U-bottom plate. Additionally, each plate includes a positive control (PC; B27 media without growth factors), a BrdU background control (BGBrdU; B27 media without addition of BrdU labeling reagent), a lysis control (LC; cells that were lysed 30 minutes before the assay is ended, serving as a positive control for the cytotoxicity assay and viability assay) and a background control (BG; B27 media without cells, serving as a background for the cytotoxicity and viability assays; Figure 8). The BrdU background control and positive control are used to assess the acceptability of single values in the experiment (see 8.3.5) All outer wells are filled with 200 µL sterile deionized water each to avoid edge effects. Plating of liquids and preparation of compound dilutions is performed automatically using a liquid handling system (MICROLAB STAR® M; Hamilton).

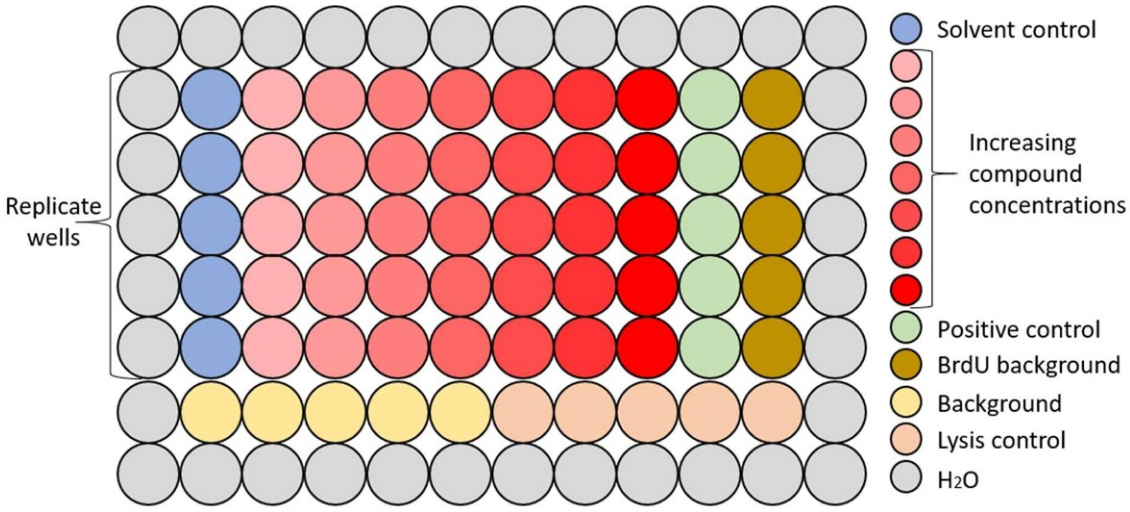


Figure 8: Plating scheme for NPC1.
Solvent control depends on the solvent of the compound that was tested. 7 compound concentrations are plated in a 1:3 dilution from lowest (left) to highest (right) concentration. Positive control for cell proliferation is proliferation media without growth factors. BrdU background is used for the proliferation assay by BrdU (NPC1b). Background and lysis control are used for cell viability and cytotoxicit assay.

1 or 3 days after passaging (depending on the day the experiment is started), spheres with a diameter of 0.3 mm are selected manually from different cell culture dishes. These spheres are plated manually in a 96-well U-bottom plate (one sphere per well) in 100 µL of the respective condition. The plates are maintained for 72 h at 37°C with 5% CO2 at a pH between 7.2-7.6. Each day the size of each sphere is determined by brightfield microscopy using a high content imaging device (ArrayScan VTI; Thermo Fisher Scientific). The increase (as slope) in pixel area over 72 h is calculated as an indirect measure of cell proliferation. After 56 h the BrdU labeling solution is added to each well except for the BrdU background control. After 72 h a lactate dehydrogenase (LDH) assay (CytoTox-ONE membrane integrity Assay; Promega) and an alamar blue (CTB) assay (CellTiter-Blue Assay; Promega) is performed to assess effects on general cytotoxicity (membrane integrity) and cell viability (mitochondrial activity), respectively. All pipetting steps for the LDH and CTB assays are performed using a liquid handling system (MICROLAB STAR® M; Hamilton). Afterwards, DNA synthesis (as BrdU incorporation) is measured as an indicator for cell proliferation using a BrdU assay (Cell Proliferation ELISA; Roche; Figure 9).

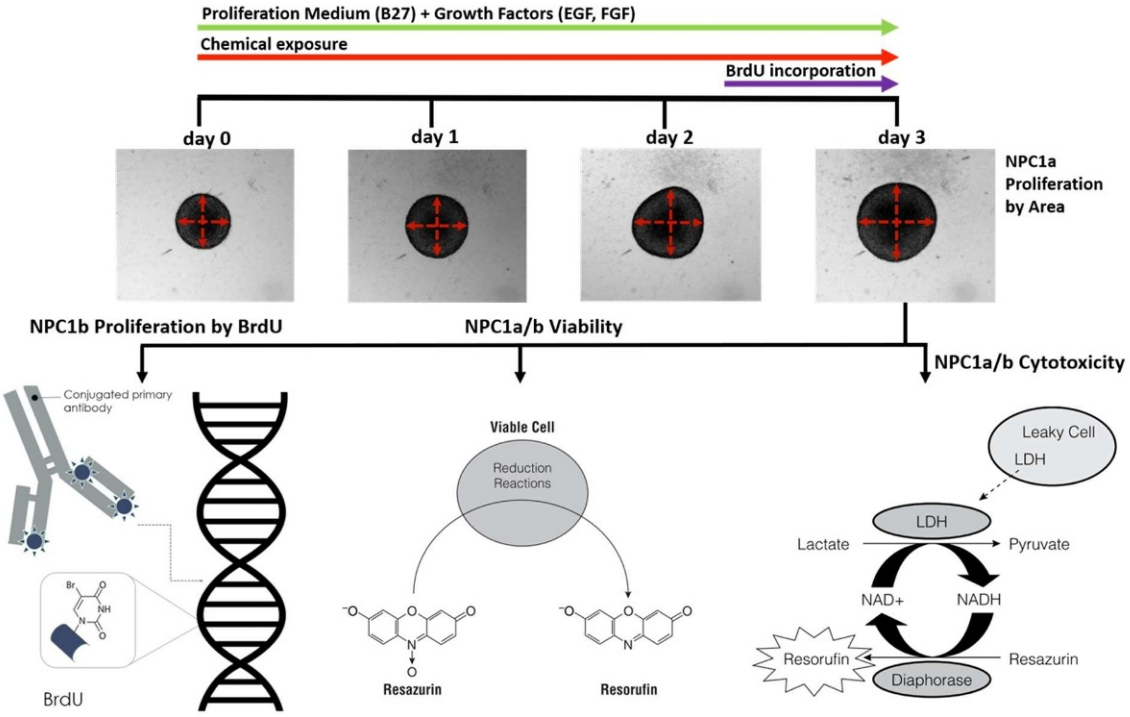


Figure 9: Assay layout NPC1.
Spheres are plated in 96-well U-bottom plates and exposed to increasing compound concentration in proliferation media over a cultivation time of 72 h. Sphere size is determined every day. The assay is terminated by the assessment of cell viabiltiy, cytotoxicity and proliferation by BrdU.

Endpoint assessment:

Proliferation by area [72h] (NPC1a) is assessed as the slope of the sphere size increase (amount of pixel) over 72 h measured by brightfield microscopy using high content imaging at 0 h, 24 h, 48 h and 72 h.

Proliferation by BrdU [72h] (NPC1b) is assessed as BrdU incorporation (a direct measure of DNA synthesis) over the last 16 h of compound exposure. It is measured as a chemiluminescence signal (relative luminescence unit) in a multi plate reader after 72 h.

Viability (proliferation) [72h] is assessed as mitochondrial activity by measuring the amount of resazurin reduced to resorufin as fluorescence signal (relative fluorescence unit) in a multi plate reader in the last two hours of the 72 h differentiation and compound treatment period.

Cytotoxicity (proliferation) [72h] is assessed as membrane integrity by measuring the LDH dependent reduction of resazurin to resorufin in the supernatant (which is removed before the viability assay is started)

as fluorescence signal (relative fluorescence unit) in a multi plate reader after 72 h of differentiation and compound treatment.

8.2.2. Primary hNPC Migration Assay (NPC2 Test)

Test system

See 8.2.1 "Test system" for description of the basic culture. Test system description for NPC2 is identical to NPC3-5 and will thus only be described here.

To initiate migration and differentiation, NPCs kept at least for 3 to 4 weeks in culture before use are plated on poly-D-lysine/laminin (Sigma Aldrich) coated 96-well flat bottom plates in differentiation medium (N2). The differentiation medium consists of DMEM (Life Technologies) and Ham's F12 (Life Technologies) at a ratio of 3 to 1 supplemented with 1% N2 (Life Technologies) and 1% penicillin and streptomycin (Pan Biotech). NPCs migrate for 5 days and thereby differentiate into radial glia cells, neurons, oligodendrocytes and astrocytes (Baumann *et al*, 2015; Nimtz *et al*, 2019; Bal-Price *et al*, 2018). For letting spheres migrate and differentiate, they are incubated at 37°C and 5% CO2 at a pH between 7.2 and 7.6. Cells are fed on day 3 by replacing half the media with fresh media. The basic neurosphere culture is tested for mycoplasma contamination every three month. Control for fungal or bacterial contamination is performed by visual inspection at each feeding day.

Test method

See 3.1.3

For this project the assays NPC2-5 are all performed within one experimental run with different endpoints measured at different timepoints. The methodological assay description is therefore placed below the introduction of all assays (under the subheading "NPC2-5 Assay").

8.2.3. Primary hNPC neuronal differentiation assay (NPC3)

Test system

see 8.2.2

Test method

see 3.1.4

8.2.4. Neuronal Morphology (neurite length and neurite area) of young neurons differentiated from primary fetal hNPC (NPC4)

Test system

see 8.2.2

Test method

see 3.1.5

8.2.5. Oligodendrocyte differentiation (NPC5)

Test system

see 8.2.2

Test method

see 3.1.8

NPC2-5 Assay

Each compound is tested in serial dilutions (1:3) with 7 concentrations and a solvent control (SC) plated in five replicate wells per condition in PDL/laminin coated 96-well F-bottom plates. Additionally, each plate includes two lysis controls (LC; one for 72 h and one for 120 h; cells are lysed 45 minutes before the assay is ended, serving as positive controls for the cytotoxicity and viability assays) and a background control (BG; N2 media without cells; as background for the cytotoxicity and viability assays; Figure 10). All outer wells are filled with 200 µL sterile deionized water to avoid edge effects. Coating of the plates, plating of liquids and preparation of compound dilutions is performed automatically using a liquid handling system (MICROLAB STAR® M; Hamilton).

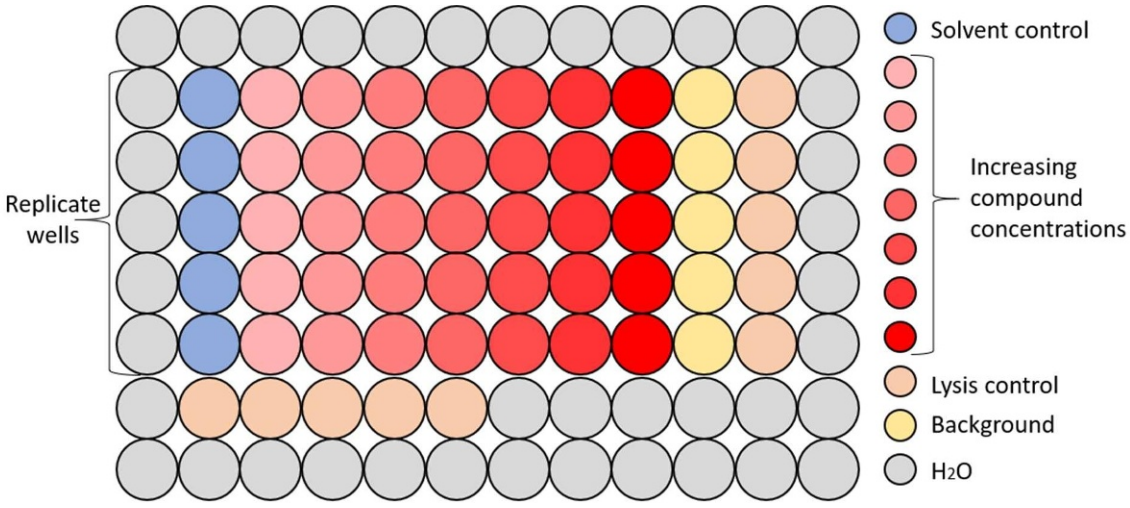


Figure 10: Plating Scheme for NPC2-5.
Solvent control depends on the solvent of the compound that was tested. 7 compound concentration are plated in a 1:3 dilution from lowest (left) to highest (right) concentration. Background and lysis control (5 replicates for each timepoint; 72 and 120 h) are used for cell viability and cytotoxicity assay.

In addition to the plates for compound testing, one control plate that includes the endpoint specific positive controls (ESC) for NPC2, 3 and 5 is prepared with every experimental run (one experimental run can include up to 15 experiments that are plated on one day). This plate includes five replicate wells with 0.1 µg/mL BMP7 as ESC for oligodendrocyte differentiation, 20 ng/mL EGF as ESC for migration and neuronal differentiation and 10 nM PP-2 as ESC for migration. ESC are run for each individual and passage number used for compound testing on the same day. The control plate also contains two LC (for 72 h and 120 h) and one BG control (Figure 11).

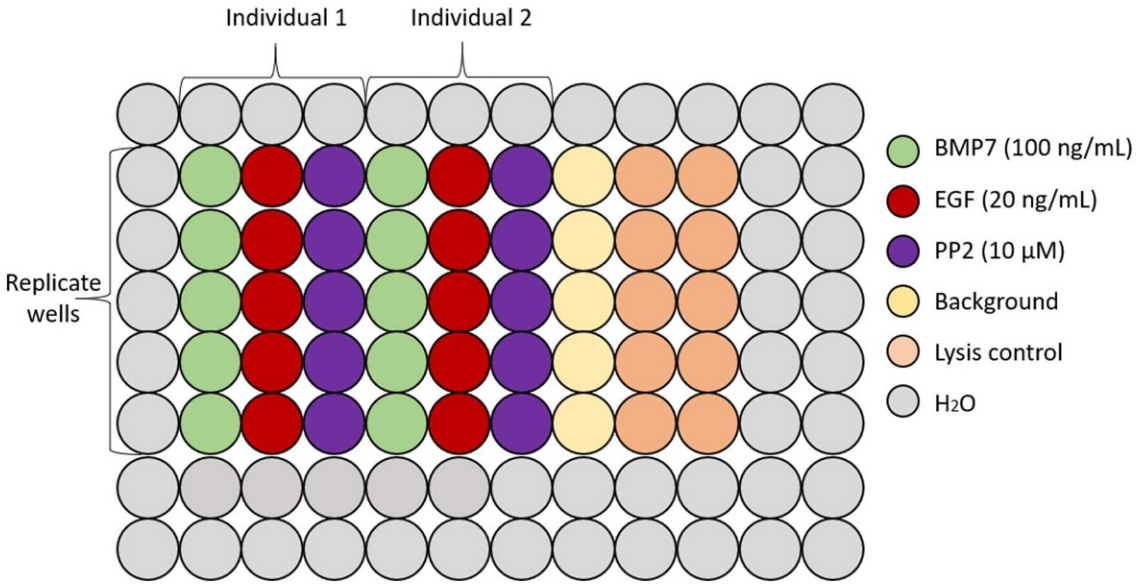


Figure 11: Plating Scheme for control plate of NPC2-5.
BMP7 is used as ESC for oligodendrocyte differentiation. EGF is used as ESC for migration and neuronal differentaion. PP2 is used as ESC for migration. Each control is plated for each individual and passage used within one experimental run. Background and lysis controls (5 replicsates for each timepoint; 72 h and 120 h) are used for cell viability and cytotoxicit assays.

Spheres with a diameter of 0.3 mm are randomly selected 2 days after passaging from different cell culture dishes. The spheres are manually plated in the middle of each well of a PDL/laminin-coated 96-well F-bottom plate in 100 µL of the respective test solution. Thereby the sphere selection for each well is randomized. The plates are maintained for 72 h at 37°C with 5% CO2. Radial glia cell migration (NPC2a) is determined by brightfield microscopy using a high content imaging device (ArrayScan VTI; Thermo Fisher Scientific) 72 h after plating. Also after 72 h, cells are fed by replacing half the test solution with a freshly prepared test solution. The 50 µL of media removed for feeding are used to assess cytotoxicity with the LDH assay (CytoTox-ONE membrane integrity Assay; Promega). The plates are maintained another 48 h at 37°C and 5% CO2. The assay is terminated 120 h after plating with a cytotoxicity assay (see above) and an alamar blue (CTB) assay (CellTiter-Blue Assay; Promega) to assess the effect on general cell viability. The spheres are fixed with 4% paraformaldehyde for immunocytochemical (ICC) stainings (Figure 12). All pipetting steps for feeding, LDH and CTB assay are performed using a liquid handling system (MICROLAB STAR® M; Hamilton).

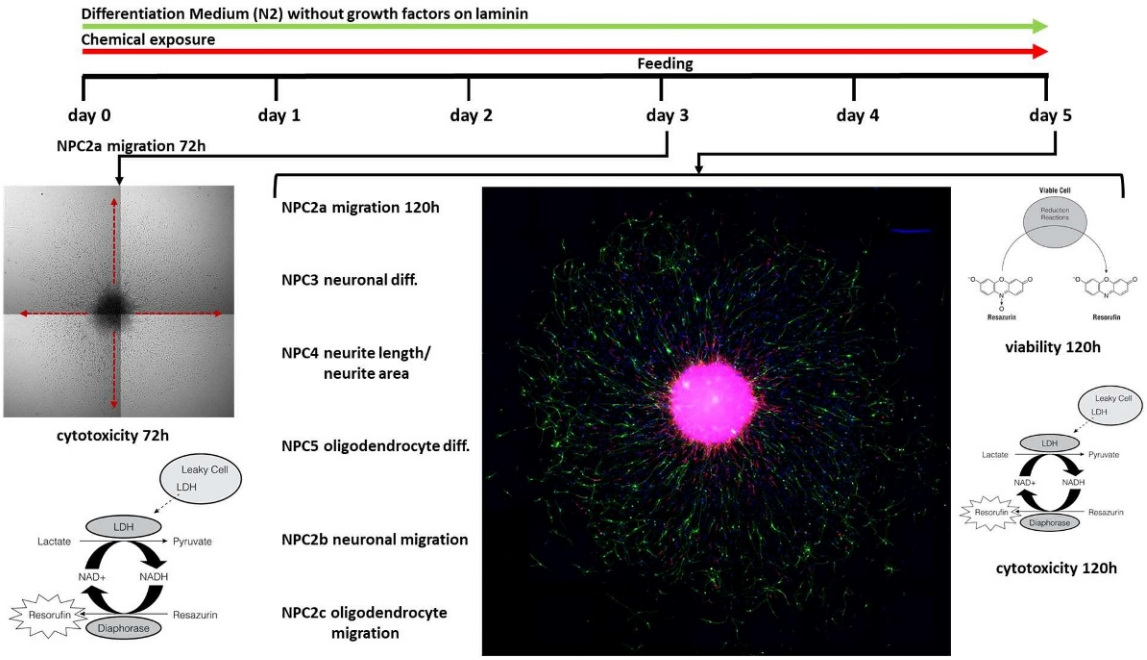


Figure 12: Assay layout NPC2-5.
Spheres are plated in 96-well F-bottom plates and exposed to increasing compound concentrations in differentiation media over a cultivation time of 120 h. Migration distance and cytotoxicity is determined after 72 h. The assay is terminated by the assessment of cell viabiltiy and cytotoxicity as well as cell fixation after 120 h. Immunocytochemistry is performed for Hoechst positive nuclei, TUBB3 positive neurons and O4 positive oligodendrocytes. Using the ICC images, migration after 120h, neuronal and oligodendocyte differentiation, neuronal morphology and neuron/oligodendrocyte specific migration is assessed.

After fixation, nuclei are stained with Hoechst, neurons and oligodendrocytes by ICC with antibodies against the neuronal marker TUBB3 and the oligodendrocyte marker O4, respectively. Fluorescence microscopy is performed using a high content imaging device (ArrayScan VTI; Thermo Fisher Scientific). Nuclei, as Hoechst positive objects, are detected using the HCS Studio: Cellomics Scan software (version 6.6.0; Thermo Scientific). The 100 single images for each well are puzzled into one image using Omnisphero (Schmuck et al, 2017). For the annotation of neurons and oligodendrocytes a convolutional neural network (CNN) running on Keras implemented in Python 3 was trained to identify both cell types. Information on the position of neurons and oligodendrocytes is again imported into the Omnisphero software which then automatically analyses migration distance of radial glia, neuron and oligodendrocytes as well as neuronal morphology (see 8.3 for more information on the software and algorithms used).

Endpoint assessment:

Migration distance radial glia [72 h] (NPC2a) is assessed by manually measuring the radial migration from the sphere core on brightfield images. Thereby, the migration distance is measured as number of pixel which is converted to µm. Migration analysis after 72 h is performed manually because the algorithm

used in Omniphero for automated migration analysis is not able to identify the migration area on brightfield images of living cells.

Migration distance radial glia [120 h] (NPC2a) is assessed by automatically identifying the migration area of each sphere of Hoechst stained nuclei on fluorescence images. The mean distance between the edge of the sphere core and the outer ring of the migration area in µm gives the migration distance.

Migration distance neurons [120 h] (NPC2b) is the mean distance of all neurons from the edge of the sphere core to the position of each neuron (see "neuronal differentiation") and is given as ratio of the "migration distance radial glia [120 h]".

Migration distance oligodendrocytes [120h] (NPC2c) is the mean distance of all oligodendrocytes from the edge of the sphere core to the position of each oligodendrocyte (see "oligodendrocyte differentiation") and is given as ratio of the "migration distance radial glia [120 h]".

Cell number [120 h] is measured as the number of Hoechst positives objects in the migration area detected on the fluorescence images from each sphere after 120 h. Nuclei detection is done using the HCS Studio: Cellomics Scan software (version 6.6.0; Thermo Scientific).

Neuronal differentiation [120 h] (NPC3) is determined as number of all TUBB3 positive cells in percent of the amount of Hoechst positive nuclei in the migration area after 120 h of differentiation. The annotation of neurons is done by a CNN running on Keras implemented in Python 3.

Neurite length [120h] (NPC4) is assessed as neurite length in µm based on the skeletonization of each annotated neuron (see "neuronal differentiation") that reaches a pre-defined intensity threshold.

Neurite area [120h] (NPC4) is assessed as the area in amount of pixel for each skeletonized neuron (see neurite length).

Oligodendrocyte differentiation [120h] (NPC5) is determined as the number of all O4 positive cells in percent of the amount of all Hoechst positive nuclei in the migration area after 120 h of differentiation. The annotation of oligodendrocytes is done by a CNN running on Keras implemented in Python 3.

Viability (migration/differentiation) [120h] is assessed as mitochondrial activity by measuring the amount of resazurin reduced to resorufin as fluorescence signal (relative fluorescence unit) in a multi plate reader in the last two hours of a 120 h differentiation and compound treatment period. As measures for migration and viability are causally linked (Nimtz *et al*, 2019), specificity of effect is only determined by comparison with cell viability when migration distance is NOT affected. In case a compound affects migration, cytotoxicity (see below) is the reference.

Cytotoxicity (migration) [72h] is assessed as membrane integrity by measuring the LDH-dependent reduction of resazurin to resorufin in the supernatant (removed from each well after 72h incubation) as fluorescence signal (relative fluorescence unit) in a multi plate reader after 72 h/120 h of cell differentiation and compound treatment.

Cytotoxicity (migration/differentiation) [120h] is assessed as membrane integrity by measuring the LDH-dependent reduction of resazurin to resorufin in the supernatant (which is removed before the viability assay is started) as fluorescence signal (relative fluorescence unit) in a multi plate reader after 120 h of cell differentiation and compound treatment.

8.2.6. The cMINC Neural Crest Cell Migration Assay (UKN2 Test)

See 3.1.2

Test system

Neural crest cells (NCCs) are differentiated from the human induced pluripotent stem cell (hiPSC) line IMR90_clone #4 (WiCell, Wisconsin). Culturing of hiPSCs is performed according to standard protocols (Thomson et al., 1998). Differentiation from hiPSC line IMR90 into NCCs is performed according to the modified protocol of Mica et al., (2013).

Generation of NCCs is started on day minus 3 of differentiation (DoD-3), by replating hiPSCs on Matrigel-coated 6-well plates (Falcon) at a density of 20 000 cells/cm² in conditioned KCM containing 10 µM Rock inhibitor and 10 ng/ml FGF2. From DoD-3 until DoD11' cells are cultured at 37°C with 5% CO2 with a daily medium change after 24 h (Figure 13). In detail, on DoD-2 medium is changed to KCM containing 10 µM Rock inhibitor and 10 ng/ml FGF2. On DoD-1 medium is changed to KCM containing only 10 ng/ml FGF2. From DoD0' till DoD11' cells receive KSR medium (Knock out DMEM, 15% knock out serum replacement, 1% GlutaMax, 1% MEM NEAA solution, 50 µM 2-mercaptoethanol) which is gradually replaced by 25% increments of N2-S medium (DMEM/F12, 1.55 mg/ml glucose, 1% GlutaMax, 0.1 mg/ml apotransferrin, 25 µg/ml insulin, 20 nM progesterone, 100 µM putrescine, 30 nM selenium). From DoD0' to DoD2' medium is supplied with 20 ng/ml Noggin. From DoD0' to DoD3' with 10 µM SB431542 and from DoD2' until DoD11' with 3 µM CHIR 99021. At DoD11' cells are detached and resuspended in N2-S medium supplemented with 20 ng/ml EGF and 20 ng/ml FGF2 and seeded as droplets (10 µl) on poly-L-ornithine (PLO)/Laminin/Fibronectin coated 10 cm dishes. From DoD11' until DoD39' cells are cultured at 37°C with 5% CO2 and expanded by weekly splitting. Seeding as droplets is not necessary anymore and medium is changed every second day. On DoD39' cells are detached, washed and resuspended in freeze medium (FBS with 10% DMSO) and frozen at a concentration of 4 Mio cells per milliliter at -80°C overnight. After 24 h cells are stored in liquid nitrogen until further use.

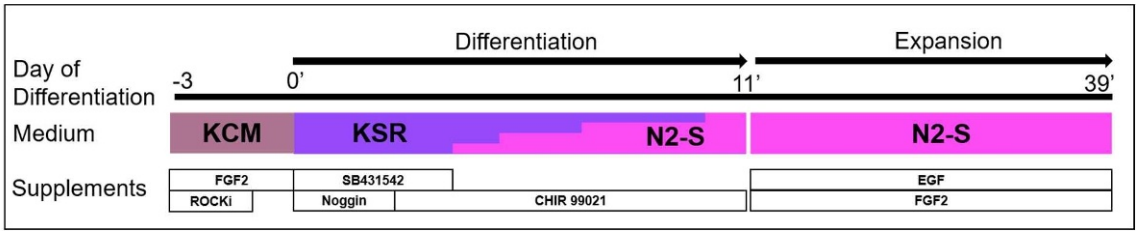


Figure 13: Differentiation scheme of IMR90 cells to neural crest cells.
After replating cells into 6-well plates a daily medium change is performed from DoD-3 until DoD11'. First three days cells receive KCM supplied with FGF2. For first two days, ROCK inhibitor Y-27632 is added additionally. From DoD0' until DoD11' cells receive KSR medium which is gradually replaced by N2-S. During the first two days medium is supplied with Noggin and during the first three days with SB431542. From DoD1' until DoD11' medium is supplied with CHIR99021. From DoD11' until DoD39' cells are expanded in N2-S medium supplied with EGF and FGF2.

UKN2 Assay

To perform the cMINC assay, the NCCs are thawed and seeded on day of migration -1 (DoM-1) into 96-well plates around stoppers (Figure 14A). After 24h stoppers are removed to allow cells to start migration into the cell-free area in the centre. On DoM1' the cells are exposed for 24h to the tested compounds. On DoM2' cells are stained with H-33342 and calcein-AM and imaged to assess viability and migration using a high content imaging microscope (Cellomics ArrayScanVTI; Thermo Fisher Scientific) as described before in detail (Nyffeler et al, 2017). A software specially designed for this assay (RingAssay software; http://invitrotox.uni-konstanz.de/) is used to calculate the number of calcein and H-33342 double-positive cells which migrate into the centre, the region of interest (ROI) (Figure 14B). To assess viability, cells outside the ROI are imaged. An automated algorithm is used to calculate all calcein and H-33342 double-positive cells resulting in a value of viable cells as described before ((Nyffeler et al, 2017; Stiegler et al, 2011; Figure 14B). Both parameters, viable cells and cells migrated into the ROI are normalized to the mean of the untreated solvent control.

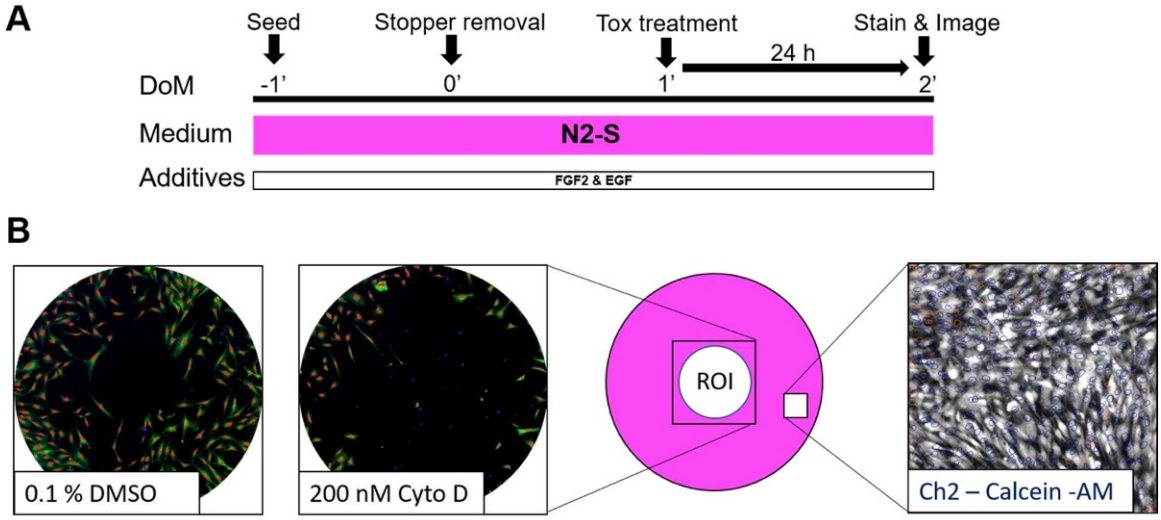


Figure 14: Performance of the UKN2 (cMINC) assay

(A) A new batch of neural crest cells (NCCs) is thawed and seeded into 96-well plates in N2-S medium containing FGF2 and EGF. Cells are seeded around stoppers to create a circular cell-free area. Stoppers are removed after 24 h to allow cells to start migration on day of migration 0' (DoM0'). On DoM1' cells are exposed for 24h to tested compounds. Thereafter, cells are stained with calcein-AM and H-33342 and high-content imaging is performed. (B) Left side represents images of migration from a solvent control (0.1% DMSO) and a positive control (200 nM CytoD). Four images for migration were taken to cover the region of interest (ROI) and calcein and H-33342 double-positive cell numbers were determined by an automated algorithm (marked with red dots). On the right, a representative image of viability estimation. Therefore, all double-positive cells are determined (blue). Images for viability assessment are taken outside the ROI.

In order to apply the UKN2 assay in a high-throughput screening approach, all plates are run in a defined plate layout including a 0.1% DMSO solvent control and 200 nM CytoD as a positive control (Figure 15). Both controls are used for the acceptance criteria of the assay. In detail, a run is only valid if: (i) viability is > 90% in the 0.1% DMSO solvent control (ii) viability is > 90% in 200 nM CytoD treated cells (iii) 200 nM CytoD treated cells migrated < 75% compared to the 0.1% DMSO solvent control. Figure 14B displays example images of a run with accurate solvent and positive controls.

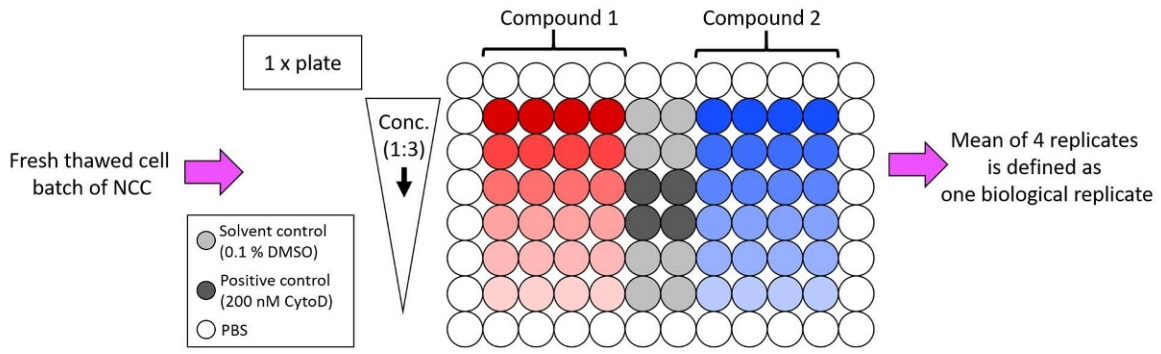


Figure 15: UKN2 plate layout.
Standard plate layout in the UKN2 assay used for concentration-response testing. Each compound concentration was tested in four wells per plate (technical replicates). Biological replicates represent an independent experiment on another day with a different batch of NCCs thawed.

Endpoint summary:

Migration inhibition: The number of migrated cells into the cell free zone is quantified 48h after stopper removal and 24h after start of compound treatment. Cells are stained with calcein-AM and H-33342 and high content imaging is performed. For migration quantification a software tool is used (http://invitrotox.uni-konstanz.de/).

Cell viability: Cell viability is measured outside the migration zone in the same well. Viability is defined as the number of H-33342 and calcein double-positive cells, viable cells are determined by an automated algorithm.

Migration and viability are normalized to untreated controls.

8.2.7. The NeuriTox Neurite Outgrowth of CNS Neurons Test (UKN4 Test)

See 3.1.6

Test system

Lund human mesencephalic (LUHMES) cells are handled in culture as described before (Lotharius et al, 2005; Scholz et al, 2011).

All flasks are pre-coated with 50 µg/ml PLO and 1 µg/ml fibronectin. For maintenance, LUHMES cells are seeded in T175 flasks (Sarstedt) in proliferation medium (PMed: AdvDMEM/F12 supplemented with 2 mM glutamine, 1 x N2 supplement and 40 ng/ml FGF) at 37°C with 5% CO2. Cells are passaged every second or third day if they reach about 80% confluency. For this, cells are seeded with 5 Mio (28 500 cells/cm²) in 20 ml or 3 Mio (17 000 cells/cm2) in 30 ml for two or three days, respectively. For differentiation, 8 Mio (45 000 cells/cm2) cells are seeded on day of differentiation -1 (DoD-1) in PMed for pre-differentiation (Pre-Diff; Figure 16). On DoD0' medium is changed to differentiation medium (DMed: AdvDMEM/F12 supplemented with 2 mM glutamine, 1 x N2 supplement, 2.25 µM tetracycline, 1 mM dibutyryl cAMP and 2 ng/ml GDNF).

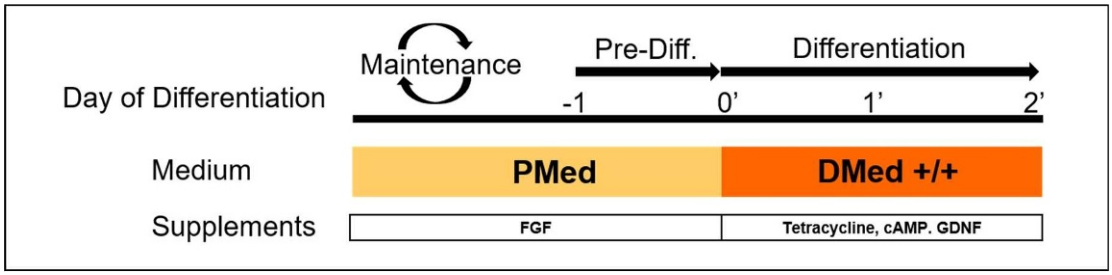


Figure 16: Maintenance and differentiation of LUHMES cells.
LUHMES cells are maintained in proliferation medium (PMed) supplemented with FGF. Differentiation is started on DoD0' by switching to differentiation medium (DMed) supplemented with tetracycline, cAMP and GDNF.

UKN4 Assay

To perform the NeuriTox assay, cells are cultivated in their proliferation status and on day of differentiation 0 (DoD0) the medium is changed into medium containing tetracycline to activate the differentiation process of the LUHMES cells (Figure 17A). On DoD2' cells are reseeded into 96-well plates. After 1h of attachment, the cells are exposed for 24h to tested compounds. Afterwards, cells are stained with H-33342 and calcein-AM and imaged via a high-content imaging microscope (Cellomics ArrayScanVTI; Thermo Fisher Scientific) to assess viability and neurite area (NA) i.e. neurite outgrowth of the cells as described before in detail (Krug *et al*, 2013; Stiegler *et al*, 2011). An imaging algorithm is used to distinguish between live and dead cells. All stained nuclei (H-33342 positive) are used to calculate the number of cells in total in channel 1 (Figure 17B). All H-33342 and calcein double-positive cells are then defined as viable objects in channel 2. Viability in each well was estimated by dividing the amount of viable cells to total cell number. To assess the NA, the algorithm calculates a likely area of the cell soma and subtracts this area from all calcein-

positive pixels imaged, resulting in the NA (Figure 17C). As the seeded precursor cells show no such area, this calculated NA represents the neurite outgrowth of the cells within 24h. In a final step for data display, the single well data was always normalized to the mean of untreated controls.

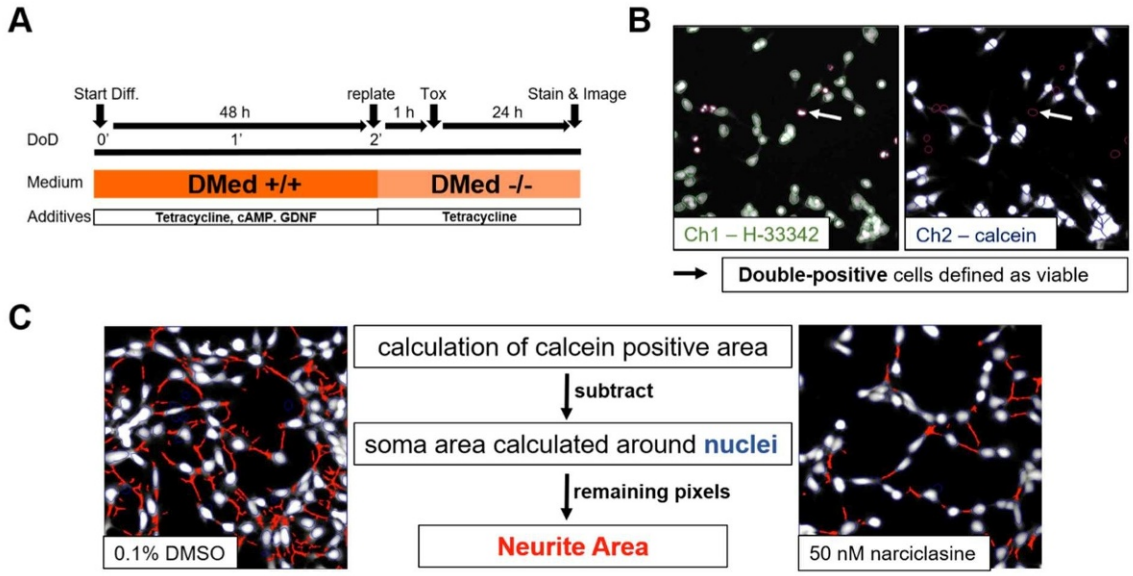


Figure 17: Performance of the UKN4 (NeuriTox) assay
(A) LUHMES cells are differentiated for two days in differentiation medium (DMed) containing tetracycline, cAMP and GDNF. On day of differentiation 2' (DoD2') cells are replated into 96-well plates in DMed containing only tetracycline. Cells are allowed to attach for 1 h before treatment. After 24h exposure to compounds, the cells are stained with calcein-AM and H-33342 and high-content imaging is performed. (B) Viability calculation is performed by an automated algorithm using two channels (Ch). All calcein and H-33342 double-positive cells (blue) are determined as viable cells. This value was normalized to the total number of cells, i.e. to all H-33342 positive cells (white arrow indicates a H-33342 single-positive cell). (C) Representative images of a solvent control (0.1% DMSO) and a positive control (50 nM narciclasine). An automated algorithm calculates the neurite area (NA, red) via subtraction of a calculated soma area (blue) from all calcein positive pixels.

In order to apply the UKN4 assay in a high-throughput screening approach, all plates were run in a defined plate layout including a 0.1% DMSO solvent control and 50 nM narciclasine as a positive control (Figure 18). Both controls were used for the acceptance criteria of the assay. In detail, a run was only valid if: (i) viability was > 90% in the 0.1% DMSO solvent control (ii) viability was > 90% in 50 nM narciclasine treated cells (iii) 50 nM narciclasine treated cells had a NA< 75% compared to the 0.1% DMSO solvent control. Figure 17C displays example images of a run with accurate solvent and positive controls.

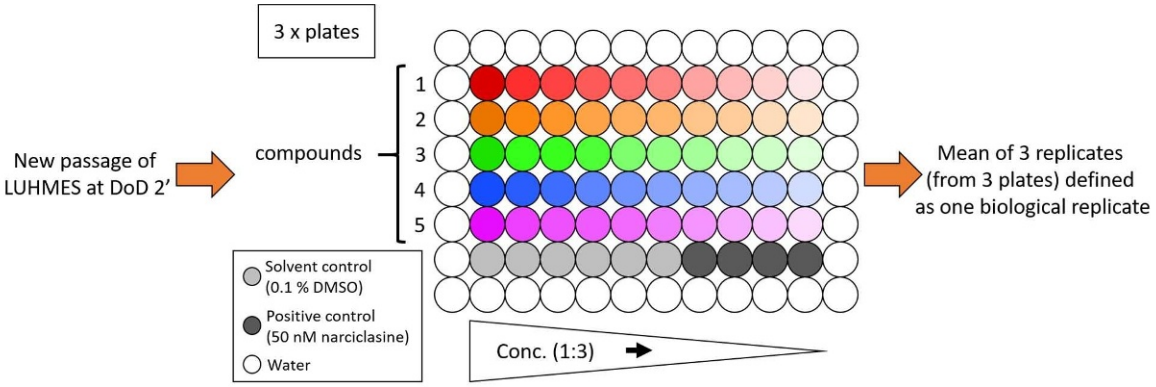


Figure 18: UKN4 plate layout

Plate layout in the UKN4 assay used for concentration-response testing. Three technical replicates for every compound concentration were distributed over three plates. Biological replicates represent an independent experiment on another day with a different LUHMES cell culture passage.

Endpoint summary:

Neurite outgrowth: Is assessed 24 h after start of compound treatment. Cells are stained with calcein-AM and H-33342 and high content imaging is performed. An automated algorithm is used to identify neurite area.

Cell viability: Cell viability is measured in the same well via an automated algorithm. Viability is defined as the number of H-33342 and calcein double-positive cells (viable cells) in relation to single H-33342 positive cells (all cells).

Neurite outgrowth and viability are normalized to untreated controls.

8.2.8. The PeriTox Neurite Outgrowth of Neural Crest Cell Test (UKN5 Test)

See 3.1.7

Test system

Sensory neurons are differentiated from the hiPSC line SBAD2. SBAD2 cells were derived and characterized at the University of Newcastle from Lonza fibroblasts CC-2511, Lot 293971 with the tissue acquisition number 24245 (Baud et al., 2017). Culturing of hiPSCs is performed according to standard protocols (Thomson et al., 1998). Differentiation from the hiPSC line SBAD2 into a sensory neuron phenotype is performed via combined small molecule inhibition according to slightly modified standard protocols (Chambers et al., 2012; Hoelting et al., 2016)

Generation of sensory neurons is started on DoD-3 by replating hiPSCs on Matrigel-coated plates. For this, cells are detached and resuspended in KCM containing 10 µM Rock inhibitor and 10 ng/ml FGF2. KCM was conditioned by lasting 24 h on mitomycin C-inactivated MEFs. Cells are seeded on Matrigel-coated 6-well plates (Falcon) at a density of 30 000 cells/cm² in conditioned KCM containing 10 µM Rock inhibitor

and 10 ng/ml FGF2. Cells are cultured at 37°C with 5% CO2 with a daily medium change after 24 h (Figure 19). In detail, on DoD-2 medium is changed to KCM containing 10 µM Rock inhibitor and 10 ng/ml FGF2. On DoD-1 medium is changed to KCM containing only 10 ng/ml FGF2. On DoD0' neural differentiation is initiated. Therefore, from DoD0' until DoD10' cells receive KSR medium (Knock out DMEM, 15% knock out serum replacement, 1% GlutaMax, 1% MEM NEAA solution, 50µM 2-mercaptoethanol) which is, from DoD4' onward, gradually replaced by 25% increments of N2-S medium (DMEM/F12, 1.55 mg/ml glucose, 1% GlutaMax, 0.1 mg/ml apotransferrin, 25 µg/ml insulin, 20 nM progesterone, 100 µM putrescine, 30 nM selenium). From DoD0' to DoD4' medium is supplied with 35 ng/ml Noggin, 600 nM Dorsomorphin and 10 µM SB431542 to initiate neuralization via dual-SMAD inhibition. From DoD2' to DoD10', three further pathway inhibitors are added, 1.5 µM CHIR99021, 1.5 µM SU5402 and 5 µM DAPT. On DoD10' cells are detached and resuspended in freeze medium (FBS with 10% DMSO) and frozen at a concentration of 8 Mio cells per milliliter at -80°C overnight. After 24 h cells are stored in liquid nitrogen until further use.

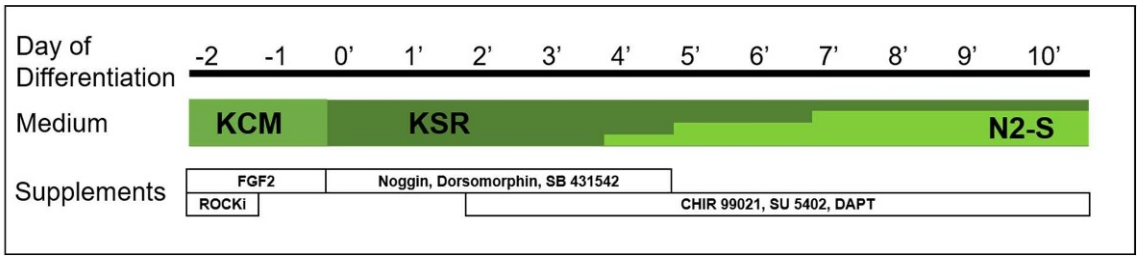


Figure 19: Differentiation scheme of SBAD2 cells to sensory neurons.
After replating cells on 6-well plates a daily medium change is performed from DoD-2 until DoD10'. First two days, KCM is added supplied with FGF2. For the first 24 h ROCK inhibitor Y-27632 is added additionally. From DoD0' until DoD10' cells receive KSR medium which is gradually replaced by N2-S medium. During this period, the medium is supplied with six small molecules (Noggin, Dorsomorphin, SB 431542, CHIR99021, SU5402, DAPT) in the indicated combination and time frame.

UKN5 Assay

To perform the PeriTox assay, the cells are thawed and seeded into 96-well plates for 1 h to allow attachment. Cells are exposed for 24 h to tested compounds (Figure 20A). Afterwards, cells are stained with H-33342 and calcein-AM and imaged via a high-content imaging microscope (Cellomics ArrayScanVTI, Thermo Fisher Scientific) to asses viability and neurite area i.e. neurite outgrowth of the cells. An imaging algorithm is used as described above in detail for the UKN4 assay to quantify viability (Figure 20B) and neurite area (Figure 20C) of the cells.

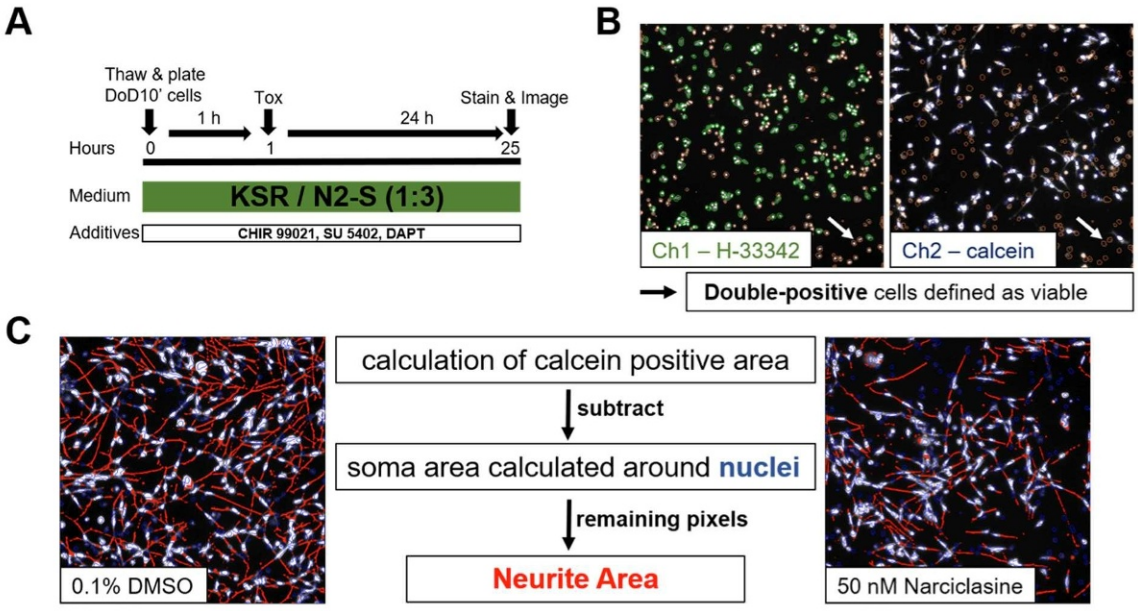


Figure 20: Performance of the UKN5 (PeriTox) assay
(A) For 10 days differentiated neuronal precursor cells are thawed and seeded into 96-well plates in medium consisting of 25% KSR and 75% N2-S supplemented with CHIR99021, SU5402 and DAPT. Cells are allowed to attach for 1h before treatment. After 24 h exposure to compounds, the cells are stained with calcein-AM and H-33342 and high-content imaging is performed. (B) Viability calculation is performed by an automated algorithm using two channels (Ch). All calcein and H-33342 double-positive cells (blue) are determined as viable cells. This value is normalized to the total number of cells, i.e. to all H-33342 positive cells (white arrow indicates a H-33342 single-positive cell).
(C) Representative images of a 0.1% DMSO solvent control and a positive control (50 nM narciclasine). An automated algorithm calculated the neurite area (NA, red) via subtraction of a calculated soma area (blue) from all calcein positive pixels.

In order to apply the UKN5 assay in a high-throughput screening approach, all plates were run in a defined plate layout including a 0.1% DMSO solvent control and 50 nM narciclasine as a positive control (Figure 21). Acceptance criteria were the same as for the UKN4 assay (Viability > 90% and NA < 75%). Figure 20C displays example images of a run with accurate solvent and positive controls.

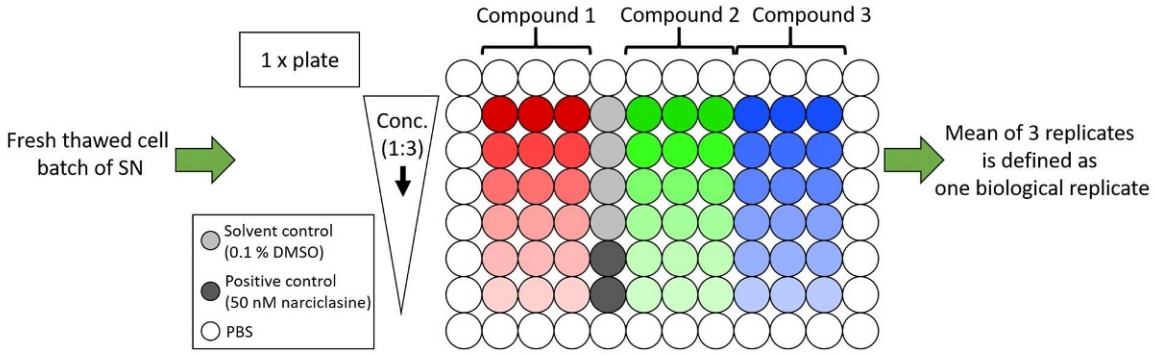


Figure 21: UKN5 plate layout
Plate layout in the UKN5 assay used for concentration-response testing. Each compound concentration was tested in three wells per plate (technical replicates). Biological replicates represent an independent experiment on another day with a different differentiation batch of DoD10' cells thawed.

Endpoint summary:

Neurite outgrowth: Is assessed 24 h after start of compound treatment. Cells are stained with calcein-AM and H-33342 and high content imaging is performed. An automated algorithm is used to identify neurite area.

Cell viability: Cell viability is measured in the same well via an automated algorithm. Viability is defined as the number of H-33342 and calcein double-positive cells (viable cells) in relation to single H-33342 positive cells (all cells).

Neurite outgrowth and viability are normalized to untreated controls.

8.3. Data Generation and Analysis

8.3.1. Plate reader-based endpoints

NPC1

Data on cell viability, cytotoxicity and proliferation by BrdU is generated using a microplate reader that measures fluorescence (viability/cytotoxicity) or chemiluminescence intensity (proliferation by BrdU) in each well.

8.3.2. Image-based endpoints

NPC2-5

Image acquisition

Images of immunochemical stainings of three channels (386 nm for Hoechst stained nuclei, 647 nm for TUBB3 stained neurons, 488 nm for O4 stained oligodendrocytes) are acquired in a 200-fold magnification and a resolution of 552x552 pixel using the ArrayScan VTI (Thermo Fisher Scientific). Thereby 64-100 images are acquired for each well of a 96-well plate. Exposure for each channel is manually determined for each plate.

Nuclei identification

For nuclei identification the spot detector BioApplication of the Cellomics scan software (Version 6.6.0; Thermo Scientific) is used. The Spot Detector BioApplication analyzes individual Hoechst-stained objects or spots (= cell nuclei) in the nuclei channel (386 nm) that fulfill pre-defined criteria on object intensity and object size that were established based on untreated NPC. The output of the nuclei detection are the nuclei coordinates (as csv file) that provide the exact position of each nucleus in the puzzled images (see below).

Puzzling of raw images

All single images are exported and converted (16 to 8 bit) from the Cellomics scan application (V 6.6.0). Single images are then puzzled to one high resolution png-image of the entire well for each channel using the Omnisphero software. During this process the csv file with nuclei coordinates is used to identify the

position of each nucleus in the puzzled image. For each predefined acquisition setting (magnification and resolution) in the original image, a corresponding custom puzzling protocol is created that is imported into Omnisphero. This is required because the position and orientation of each single image differs depending on the acquisition setting.

Annotation of neurons and oligodendrocytes

The identification of neurons (TUBB3 channel) and oligodendrocytes (O4 channel) is done automatically using a CNN based on the Keras architecture implemented in Python 3. Training of the CNN was done with manually annotated experiments. For the training set a range of experiments that represent many possible morphological features of neurons and oligodendrocytes were used. From each of those experiments, some wells were handpicked to be used as the validation set during the training phase. To test the performance of the networks for the whole migration area, two complete wells of a control experiment were annotated manually. In this control experiment a performance of >95% for neuronal as well as oligodendrocyte annotation was achieved.

Detection of migration area and cell type positioning

With the acquisition of nuclei positions the migration distance can be calculated after 120 h. Therefore, a density distribution mask is calculated automatically by Omnisphero (Schmuck et al, 2017) using the Hoechst-stained nuclei. By scanning the images of the nuclei channel, the algorithm determines more and less Hoechst-dense image areas. By identifying the densest area in the image, the sphere core is detected. For identifying the radial glia migration distance (NPC2a), it is assumed, that the nuclei density decreases with increasing distance to the sphere core. Once the density hits a predefined threshold, the outer boundaries are determined, the sphere is mapped out and the migration distances identified using the circular outer boundary for each migrated sphere subsequently presented as one mean migration value/sphere. For assessing neuronal (NPC2b) and oligodendrocyte (NPC2c) migration, the distance of each annotated neuron and oligodendrocyte (see above) from the sphere core is determined and the mean migration distance calculated. This value is related to the total radial glia migration distance. Hence, NPC2b and NPC2c are values given as percent of NPC2a.

Morphological analysis

After the neurons are annotated by the CNN, the neurite endpoints are assessed by Omnisphero. First, the neurite channel image is pre-processed to a binary image to eliminate image- and staining-artifacts using the image processing algorithms in Omnisphero. The remaining neurite image is then projected onto the nuclei channel image and subsequently all nuclei that cannot be linked to at least one neurite and are not annotated as neuron (see above) are removed. In the next step, each neuron and its corresponding neurite(s) are isolated into "tiles" for determining the ratio between neurons and nuclei. This number gives the number of neurites/nuclei. A skeletonization algorithm is run on each isolated neuron tile. This determines the individual length and area of each neurite. The final selection of neurite tiles for the calculation of the mean neurite area and mean neurite length is based on plausibility criteria including shape and the maximum length/area of a neurite, that were defined by manual inspection of artefacts in control experiments.

UKN

Image acquisition

For image acquisition, an automated microplate reading microscope (Array-Scan VTI, Cellomics, Thermo Fisher Scientific) equipped with a Hamamatsu ORCA-ER camera (resolution 1024 × 1024) is used. Images were recorded in 2 channels at excitation/emission wavelengths of 365 ± 50/535 ± 45 to detect H-33342 in channel 1 and 474 ± 40/535 ± 45 to detect calcein in channel 2. In both channels, a fixed exposure time and an intensity histogram-derived threshold is used for object identification (Stiegler et al, 2011).

UKN2
Four fields outside the migration area are imaged using a 10x objective (2 x 2 pixel binning) to obtain images for viability analysis. Additional four fields in the center of the well are imaged using a 5x objective (2 x 2 pixel binning). Images of the center are joined to obtain one micrograph which covers a total area of 2590 x 2590 µm (Nyffeler et al., 2017).

UKN4
Ten fields per well are imaged using a 20x objective (2 x 2 pixel binning; Krug et al, 2013)

UKN5
Nine fields per well are imaged using a 10x objective (2 x 2 pixel binning; Hoelting et al, 2016)

Viability image algorithm

For a quantitative assessment of viable cells, the following image analysis algorithm is used: nuclei were identified as objects in channel 1 according to their size, area, shape, and intensity which were predefined on untreated cells using a machine-based learning algorithm, and manual selection of nuclei to be classified as intact. Nuclei of apoptotic cells with increased fluorescence are excluded. A virtual cell soma area (VCSA) is defined around each nucleus by expanding it by 0.3 µm into each direction. Calcein-AM staining, labeling live cells, is detected in channel 2. The algorithm quantified the calcein intensity in the VCSA areas. Cells having an average calcein signal intensity in the VCSAs below a predefined threshold are classified by the program as "not viable." Valid nuclei with a positive calcein signal in their cognate VCSA are counted as viable cells. A positive calcein signal is based on measurements of the average intensity (normal cells: 1300 ± 115, threshold: < 50) and the total integrated intensity (normal cells: 186,000 ± 23,600, threshold < 1000) of cells (Stiegler *et al.*, 2011).

Migration assessment

A software tool (http://invitrotox.unikonstanz.de/) was developed for migration assessment in UKN2. The software estimates the most likely position of the previously cell-free area to set thresholds for color intensity for both dyes, and to count the number of H-33342 and calcein double-positive cells in the region of interest (ROI). The radius of the ROI is set to 0.7-0.9 mm. The radius of the reduced ROI is chosen in a way that at least 150 but not more than 300 cells are in the ROI in the untreated condition. A JPEG image of the final picture was generated to visually control the obtained result. The chosen settings were then transferred to all wells. Finally, the software counts the number of viable cells in the ROI for all wells of the plate and a data table is generated with the results (Nyffeler et al, 2017).

Neurite area assessment

Neurite area is assessed using the following image analysis algorithm: similar to viability measurement the nuclei are identified as objects in channel 1 according to their size, area and shape. All calcein-positive pixels of the field (beyond a given intensity threshold) are defined as viable cellular structures (VCSs). The threshold is dynamically determined for each field after flat field and background correction and intensity normalization to 512 gray values and is set to 12% of the maximal brightness (channel 63 of 512). The VCS defines the sum of all somata and neurites without their assignment to individual cells. In an automatic calculation, the VCSAs, defined in the H-33342 channel, are used as filter in the calcein channel and subtracted from the VCS. The remaining pixels VCS – VCSA in the calcein channel are defined as neurite area (Krug et al, 2013)

8.3.3. Data management

For storage and management of all the testing data generated within this project, a database (DB) was set up. This database is maintained and run on a server at the Ruhr University Bochum. It operates on PostgreSQL and is designed to store data from different projects, assays and institutions due to a generalisation of (meta-)data. The database is designed to solely store and manage data for later data evaluation.

In order to import experimental data into the database, a user interface (UI) was programmed and is running in Java 1.9. The DB client is able to read in high throughput specially designed Microsoft Excel sheets that were used for initial storage of meta- and experimental data. Thereby one sheet is created for every experiment. The algorithm implemented in the client is able to read and interpret these sheets to store matching compounds, wells and controls coherent in the database.

In addition, a basic error check is implemented for the raw data, that checks the sheets for correct or missing information, such as compound name, CAS No., etc.

From the database the data is exported for further data evaluation in a user-defined way. Therefore, the user can select data via the export client. This is done via so-called "limiters", that can be set via the client UI. This function can limit exported data to e.g. ignore outliers, focus on a single endpoint, compound or assay. Since limiters work on *any* meta data, exports can be narrowed down to very specific database entries, such as single days, wells or even individual response values. The data export we used here for further data evaluation is a compound-based export. In this form, every response is grouped by their corresponding compound. One sheet is created for every compound present in the database, containing the response of every replicate and endpoint.

8.3.4. Data Evaluation

To enable assay specific data analysis, a data evaluation tool was developed in R. It ensures flexible data analysis and visualization by user defined settings. The compound sheets exported from the database represents the data input format for the evaluation tool and links the database to the evaluation workflow. Multiple export sheets can be stored in a batch folder, which are than read automatically by the evaluation tool, enabling high throughput data analysis.

The evaluation tool was written in R and uses the data.table package for data management, as well as the drc and bmd package for curve fitting and plotting.

8.3.5. Acceptability criteria

Before the data is imported into the data base (DB) and subsequently into the evaluation pipeline, the experiments are manually checked against a set of acceptability criteria. Thereby, single wells, single concentrations (conditions), the whole endpoint or even whole experiments might be excluded from the analysis due to one of the following reasons:

NPC1-5

Exclusion of single wells:

- methodological/technical issues during the experiment (e.g. problems with liquid handling system, problem in staining, scanning) in single wells.
- for NPC1b (BrdU) wells are excluded in case the raw BrdU chemiluminescence values is too low:

  For the solvent control the raw value is considered to be low, if the raw intensity is between the Background BrdU and the positive control.

For the treatment conditions the raw value is considered to be low if one to two of five replicate values have a raw value intensity between background BrdU and positive control and clearly differ from the other three to four values.

Exclusion of one or more concentration (condition) within one experiment:

- methodological/technical issues during the experiment (e.g. problems with liquid handling system, problem in staining, scanning) in one or more condition.
- less than two replicate values are present.

Exclusion of one endpoint:

for NPC3-5 the respective endpoint is excluded if the percentage of neurons or oligodendrocytes in the SC is below 1.5% (if percentage of neurons and oligodendrocytes is below 1.5% the whole experiment is excluded from evaluation).

Exclusion of whole experiments:

- methodological/technical issues during the experiment (e.g. problems with liquid handling system, problem in staining, scanning) that apply to the whole experiment.
- irregular concentration-response behavior.

UKN

Single wells can be excluded based on noted methodological issues from the operator during the experiment e.g. contamination of single wells.

Following acceptance criteria have to be fulfilled, otherwise whole experiments are excluded:

- viability and specific endpoint > 90% in solvent control wells

- reduction in specific endpoint in cells treated with positive control compound by at least 25% relative to solvent control
- viability > 90% in in cells treated with positive control compound

8.3.6. Data pre-processing

Data pre-processing describes all processing steps of raw data that are necessary to obtain the final response values used for normalization and curve fitting. The pre-processing steps thereby depend on the test method and endpoint and are described below:

NPC1

Proliferation by area (NPC1a): slope of the development in sphere size over 3 days of proliferation (d0, d1, d2, d3). The calculated slope is used as raw data input for the DB and is thus not calculated in the R-based evaluation tool.

Proliferation by BrdU (NPC1b): subtraction of mean BrdU background from each raw response value.

$$Background\ corrected\ response\ [RLU] = raw\ response\ [RLU] - Background\ BrdU\ [RLU]$$

Viability (NPC1): subtraction of mean background from each response value.

$$Background\ corrected\ response\ [RFU] = raw\ response\ [RFU] - Background\ [RFU]$$

Cytotoxicity (NPC1): no pre-processing. For this endpoint the background can be neglected due to a different normalization (see 8.3.7)

NPC2-5

Migration distance radial glia 72 h (NPC2a): mean of four replicate measures of each sphere. The mean of four measures/well is used as raw data input for the DB and is thus not calculated in the R based evaluation tool.

Migration distance radial glia 120 h (NPC2a): no pre-processing.

Cell number 120 h (NPC2-5): no pre-processing.

Neuronal differentiation (NPC3): the number of all neurons is divided by the number of all cells in the migration area.

$$neuronal\ differentiation\ [\%] = \frac{\#\ neurons}{\#\ cells} * 100\ \%$$

Migration distance neurons 120 h (NPC2b): mean migration distance of all neurons in the migration area divided by migration distance radial glia 120 h.

$$migration\ distance\ neurons\ [\%] = \frac{mean\ migration\ distance\ all\ neurons\ [\mu m]}{migration\ distance\ radial\ glia\ 120\ h\ [\mu m]} * 100\ \%$$

Oligodendrocyte differentiation (NPC5): number of all oligodendrocytes is divided by the number of all cells in the migration area.

$$oligodendrocyte\ differentiation\ [\%] = \frac{\#\ oligodendrocytes}{\#\ cells} * 100\ \%$$

Migration distance oligodendrocytes 120 h (NPC2c): mean migration distance of all oligodendrocytes in the migration area divided by the migration distance radial glia 120 h.

$$migration\ distance\ oligo.[\%] = \frac{mean\ migration\ distance\ all\ oligo.\ [\mu m]}{migration\ distance\ radial\ glia\ 120\ h\ [\mu m]} * 100\ \%$$

Viability (NPC2-5): see NPC1.

Cytotoxicity (NPC2-5): no pre-processing. no pre-processing. For this endpoint the background can be neglected due to a different normalization (see 8.3.7)

Neurite Area (NPC4): no pre-processing. The raw values are directly used for normalization.

Neurite length (NPC4): no pre-processing. The raw values are directly used for normalization.

UKN4 and UKN5

For UKN4 and UKN5 pre-processing for the viability endpoint is performed. Calcein-Hoechst double positive cells are classified as selected objects. Hoechst single positive cells are defined as valid objects.

$$Viable\ cells = \frac{selected\ objects}{valid\ objects}$$

8.3.7. Data normalization and curve fitting

The data evaluation tool provides three methods for normalization: 1) Normalization to the solvent control, 2) normalization to the mean of the first three concentrations, 3) normalization to the mean of the solvent control and the first two concentrations. Each of these can then be combined with a re-normalization to the starting point of the curve or a re-normalization to the upper asymptote of the curve.

For the data presented in this report the data was normalized to the median solvent control and re-normalized to the starting point of the curve. For the cytotoxicity assays a different normalization was used:

$$\text{normalized response} = \frac{\text{lysis control (median)} - \text{response}}{\text{lysis control (median)} - \text{solvent contrl (median)}}$$

The R package drc is used to calculate the optimal fit for each experiment.

For calculations of curve fits and BMCs, the data from independent experiments is pooled (median of all replicate values for one concentration for NPC assays, mean of all replicate values for one concentration for UKN assays). Several non-linear models are run with the concentration response data of each endpoint and the Akaike´s information criteria is used to determine the best fit.

For re-normalization of the data, for each independent experiment a new curve fit is generated. The experiments data is then renormalized towards the response value of the curves starting point. After every

experiment is renormalized, the data from the re-normalized experiments is pooled again, a new plot is created based on the pooled data and a final re-normalization towards the curve starting point response is performed.

8.3.8. BMC calculation

For deriving a reference point (RP) or point of departure (Pod) the Benchmark Dose (BMD) approach as recommended by the EFSA Scientific Committee (Hardy *et al*, 2017b) was applied. The BMD or for *in vitro* toxicity testing Benchmark Concentration (BMC) approach makes use of all data points that define the fitted concentration response curve. Thereby the BMC is defined as the concentration that is associated with a specific change in response, the Benchmark Response (BMR). The BMR is a value of effect size and should be defined as an effect size that is higher than the general variability of the measured endpoint. The BMR is therefore determined based on the variability of the respective endpoint.

For NPC1-5 the BMR value is defined based on the deviation between experiments. Therefore, we normalized the mean response value of the lowest compound concentration (assuming that the lowest concentration does not affect any of the endpoints) to the solvent controls and calculated the SD between the lowest concentration over all experiments (**between experiment variation**).

For NPC1 the between experiment variations are as follows:

proliferation by BrdU (NPC1b):	14.8 %
proliferation by Area (NPC1a):	12.6 %
viability (proliferation, NPC1):	5.23 %
cytotoxicity (proliferation, NPC1):	1.4 %

For NPC1 the BMR is defined as at least 1.5 x the SD (between experiment variation) as BMR30.

For NPC2-5 the between experiment variations are as follows:

migration distance radial glia 72 h (NPC2a):	4.6 %
migration distance radial glia 120 h (NPC2a):	5.0 %
migration distance neurons 120 h (NPC2b):	8.8 %
migration distance oligodendrocytes 120 h (NPC2c):	6.5 %
cell number 120 h (NPC2-5):	10.5 %
neuronal differentiation 120 h (NPC3):	21.8 %
oligodendrocyte differentiation 120 h (NPC5):	27.2 %
neurite length 120 h (NPC4):	11.9 %
neurite area 120 h (NPC4):	12.7 %
Viability (migration/differentiation) 120 h (NPC2-5):	6.9 %,
Cytotoxicity (migration/differentiation) 72 h (NPC2-5):	2.3 %
Cytotoxicity (migration/differentiation) 120 h (NPC2-5):	2.7 %

For NPC2-5 the BMR is defined as at least 1.5 x the SD (between experiment variation) as BMR10 for migration distance radial glia 72 h and 120 h, cytotoxicity 72 h and 120 h and BMR30 for all other endpoints.

For neuronal and oligodendrocyte differentiation the BMR is defined as BMR30 although 1.5x the variation is above the BMR. Both endpoints show a high biological variation and therefore would lead to a higher BMR according to the procedure described above. However, for better comparison between related endpoints (e.g. neuronal differentiation and neurite length) and to assure no false negative classification because of a high and variable (can be different for different experiments) biological variation, a lower BMR was selected. Here it is important to consider that the BMR has to be interpreted together with a measure of uncertainty. Therefore, confidence intervals are calculated for each BMC and are a measure of the actual variation between the experiments used to define the respective BMC.

For UKN2, 4 and 5:

If none of both parameters is affected by 20% i.e. if no BMR20 is reached, the compound is classified as a "no hit" compound in the pre-screening (see Annex H). Standard BMR in all assays for hit consideration is the BMR25. However, BMR20 was chosen as a more conservative value in the pre-screen to make sure not to miss possible hit compounds in the pre-screening process.

If BMR20 is reached the compound was followed-up for concentration-response testing. For all compounds where concentration response testing was performed we determined the BMC10 and BMC25 values for all assays.

Based on the BMR and the concentration response curve generated as described in 8.3.7, the evaluation tool calculates the BMC, as well as upper and lower confidence limits (BMCU and BMCL respectively) by parametric bootstrapping in combination with median resampling. The confidence intervals are used to access the uncertainty of the BMC. In each bootstrap resampling, for each condition, the same amount of responses as in the original dataset are artificially created, thus providing an artificial dataset. From the artificial dataset, the median of each individual experiment is calculated for each condition, resulting in one resampled median value per experiment for each condition. Based on this resampled median dataset, the optimal curve fit is calculated for the resampled median dataset. Finally, one parametric bootstrapping sample is generated from the resampled median data. This means that the deviation between the resampled median of independent experiments is given as the input for the bootstrap function and is considered for generating new values. The parametric resampled data is then used to generate a curve fit based on the assumptions from the optimal fit calculated before the bootstrap resampling, resulting in one bootstrap BMC. This procedure is repeated 200 times, providing a list of 200 BMC values (one for each median resampling). This list is used for the calculation of the final BMC, BMCU and BMCL of the according dataset. For the BMC calculation, each value below or above the test range is replaced with the highest test concentration as values below the test range are biologically not plausible and values above the test range have a high degree of uncertainty. The BMC is calculated as the mean of all adjusted BMC values. For the confidence intervals, BMCs below the test range and above 1.5x the highest concentration are replaced by 1.5x the highest concentration. The BMCU and BMCL are then calculated as the 0.95 and 0.05 quantile of these adjusted BMC values.

8.3.9. Classification models for compound classification

This chapter describes individual decision trees as classification model (CM) for compound classifications as a first step in data interpretation. Here, the distinction between 'specific hits' and 'unspecific hits' for each compound has to be made. Therefore, DNT *in vitro* data effects on the DNT-specific endpoints are compared to effects on cytotoxicity/viability in the same assay. If the classification model separates cyto-toxicity/viability BMC from the respective BMC of the specific endpoint, the compound is identified as a

'specific hit'. If BMCs of cytotoxicity/viability and the specific endpoint cannot clearly be separated or overlap, it is defined as 'unspecific hit'. In case no BMC for either cytotoxicity/viability or specific endpoint can be determined, it is a negative compound, or 'no hit'. In case the separation between cytotoxicity/viability and the specific endpoint is not clear, the compound is defined as a 'borderline hit'. This last category is important to implement for not generating too many false-negatives due to statistical reasons (Leontaridou et al, 2017). Examples for each category are given in Figure 22.

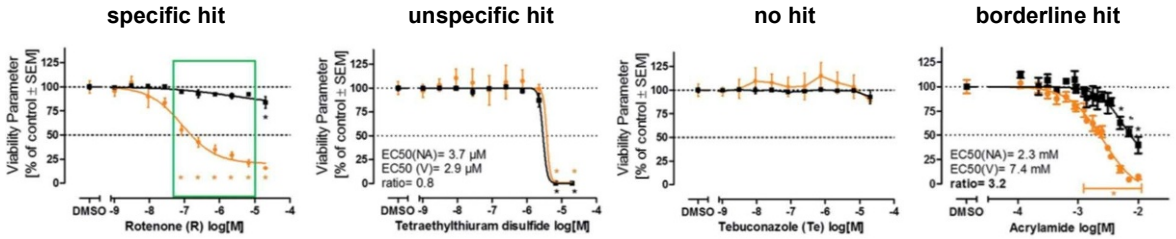


Figure 22: Examples for compound classification according to classification models. Here, compounds with a specifically determined EC50 ratio (Viability/Neurite Outgrowth) were classified (see description above; from Delp et al, 2018).

Despite the common goal of assay hit definition, different approaches are applied for the different assays. Here, especially the procedures between the IUF and the UKN labs differ and are subsequently described in detail.

The CM applied for NPC1-5 is based on a comparison between the confidence intervals (CI) for the BMC of the DNT-specific endpoints and the respective cytotoxicity/viability. Thereby a compound is classified as a "specific hit" if the CI do not overlap, meaning that the upper confidence limit of the specific endpoint (BMCUs) is lower than the lower confidence limit of the cytotoxicity/viability (= unspecific endpoint; BMCLus). A compound is classified as "borderline" if the lower confidence limit of cytotoxicity/viability overlaps by less than or equal to 10% with the CI of the specific endpoint. The compound is classified as "unspecific hit" if the overlap between the BMCUs and the BMCLus is greater than 10%. In case no confidence interval is available for cytotoxicity/viability because the BMR is not reached, the BMCLus is set as the highest tested concentration.

Examples for (i) "no hit", (ii) "specific hit", (iii) "unspecific hit" and (iv) "borderline" in the endpoints migration distance radial glia (NPC2a), oligodendrocyte differentiation (NPC5), neurite length (NPC4) for the compounds acrylamide, deltamethrin, hexachlorophene and carbaryl, are shown in Figure 23. The decision process for the CM is described in the flow chart in Figure 25.

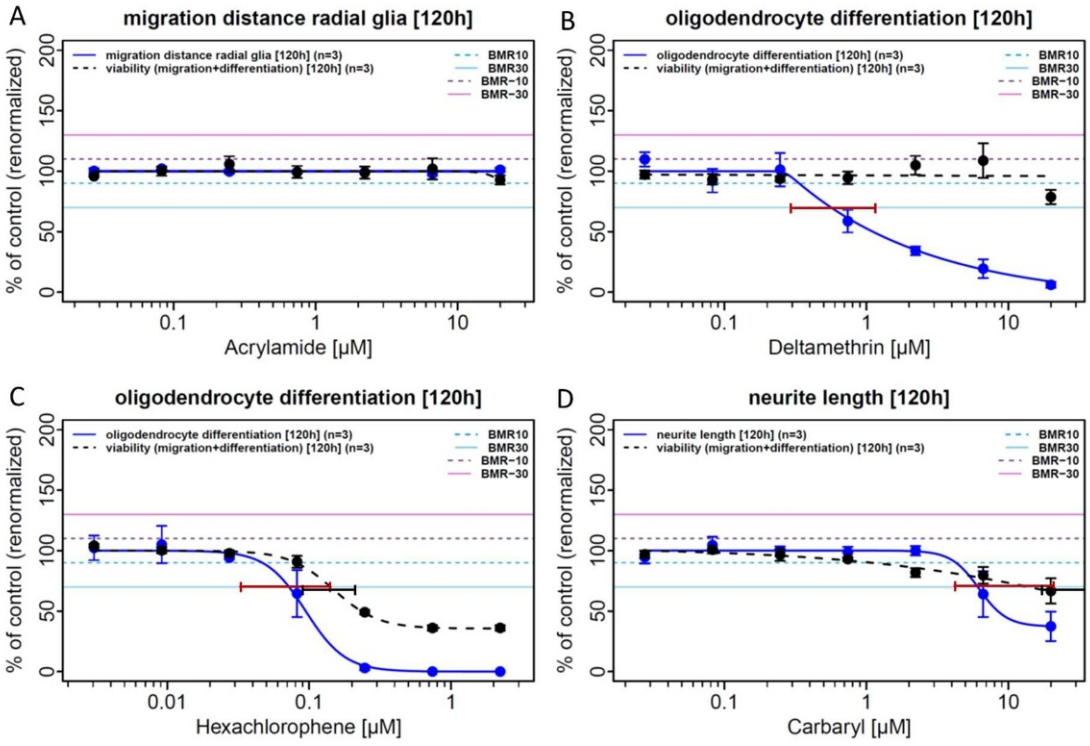


Figure 23: Concentration response curves demonstrating the four types of classifications according to the CM for NPC1-5. Shown is a "no hit" (A), "specific hit" (B), "unspecific hit" (C) and "borderline" (D) classification. The red and black bars represent the confidence intervals for specific and unspecific endpoints, respectively.

For the Classification model of NPC1-5 there are three special cases that need to be considered during the classification process:

1. If the cell number or radial glia migration after 120 h are affected by the compound, only the cytotoxicity and not the viability readout is used as a reference for all -NT specific NPC2-5 endpoints (measured at 120 h) in the CM. The reason is that the viability measures metabolic activity and therefor strongly depends on the number of cells in the migration area. A reduction in cell migration and cell number is therefore accompanied by a reduction in the cell viability assay (Nimtz *et al*, 2019). The Cytotoxicity assay on the other hand measures the release of LDH by dead cells into the media and is therefore not directly cell number-dependent (see example in Figure 24A, B)
2. If the upper confidence limit is above 1.5x the highest test concentration and can thus not be calculated, it is set to 1.5x the highest test concentration. This is often the case for flat curves or if the effect is only seen at the highest concentration. In both cases the data does not give enough information for an accurate calculation of the confidence intervals of the BMCs. Because of a high degree of uncertainty these cases are considered separately and the classification is done by expert judgement. Based on expert judgment compounds were classified as "specific hit" if the deviation between the response of the specific endpoint and the unspecific endpoint in the highest test concentration is at least 1x the BMR of the respective specific endpoint, and if the standard errors are not overlapping (see example in Figure 24C, D).
3. For endpoints that allow the generation of a BMC10us (all cytotoxicity endpoints in NPC1-5), the BMC10us is compared to the BMC30s in case no BMC30us is available for the classification model.

If the classification of this comparison is "unspecific hit", the classification needs to be reviewed by an expert, as the BMC10us is lower than the BMC30us leading to a higher probability for a false classification. To avoid such false classification, expert judgment based on the criteria described in point 2 is applied (see example in Figure 24E, F).

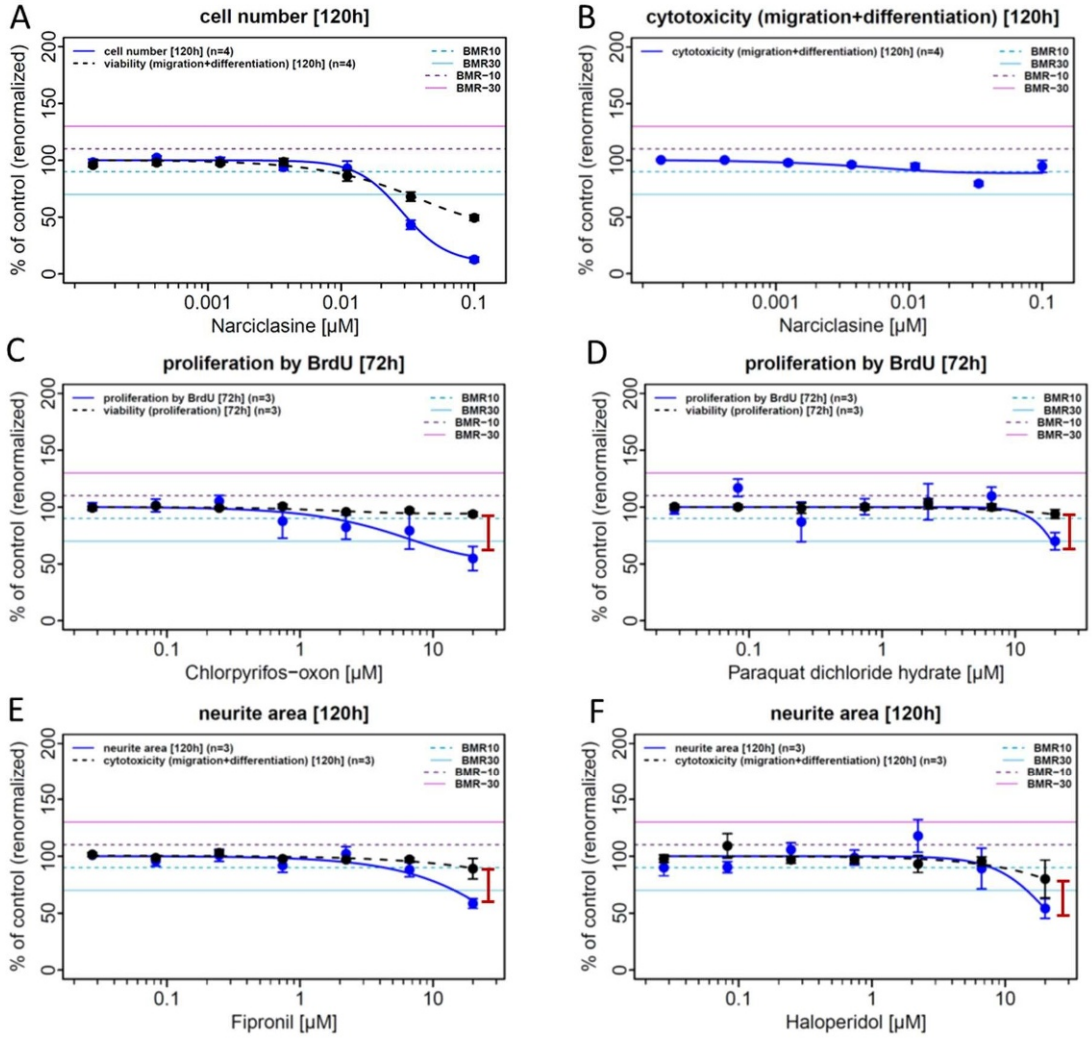


Figure 24: Concentration response (CR) curves showing examples for the three special cases that are considered in the CM used for NPC1-5. (A, B) Narciclasine reduces cell viability due to a decrease in cell number. Therefore, cytotoxicity is used for the classification of all NPC2-5 endpoints measured after 120 h narciclasine exposure. (C, D) The fitting models for both CR curves do not allow the calculation of an upper confidence limit. According to the criteria described in point 2 (difference between the response of the DNT-specific and unspecific endpoint at the highest concentration are above the BMR, in this case 30%; red bar), the effect of chlorpyrifos-oxon on NPC proliferation (NPC1) is classified as specific (C) while the effect of paraquat dichloride hydrate on NPC proliferation (NPC1) is classified as unspecific (D). (E, F) In both cases the effects on the specific endpoint (neurite area; NPC4) are accompanied by an effect on cytotoxicity (measured as BMC10), however, for fipronil the effect on the specific and unspecific endpoints at the highest test concentration can be distinguished due to the >30% difference between them (red bar; see criteria mentioned above; → "specific hit"). In contrast, the effect of haloperidol cannot be distinguished from cytotoxicity due to the <30% difference between the endpoints (→ "unspecific hit").

Besides the above-mentioned classification each hit in the NPC1-5 assay is classified as "uncertainty" if the upper confidence limit of the specific endpoint could not be calculated or if the lower and upper BMC limit are 4x lower or higher than the BMC. Here it is important to mention that large CI or BMCs without an upper confidence limit in the data set do not necessarily reflect large deviations in the raw data, but might be due to flat curves slope (Figure 24B) or effects at only the highest concentration with not sufficient concentration response information (Figure 24C, D). Over all, the approach using CI of BMC is a rather conservative approach and used for NPC assays to statistically account for their relatively high variation due to the primary nature of the cells.

The uncertainty classification does not influence the initial hit identification of the compound in a screening approach, as a hit that is classified as 'uncertainty' is still considered as a hit. However, it should be considered during compounds' prioritization. If the battery is used for hazard characterization, the hit should be confirmed by additional experiments in an optimized concentration range. Alternatively, information on the free *in vitro* concentration can be consulted to determine if the effective concentration is of *in vivo* relevance when compared to tissue concentrations. This is especially reasonable if the BMC is in the high micromolar range (e.g. above 10 µM).

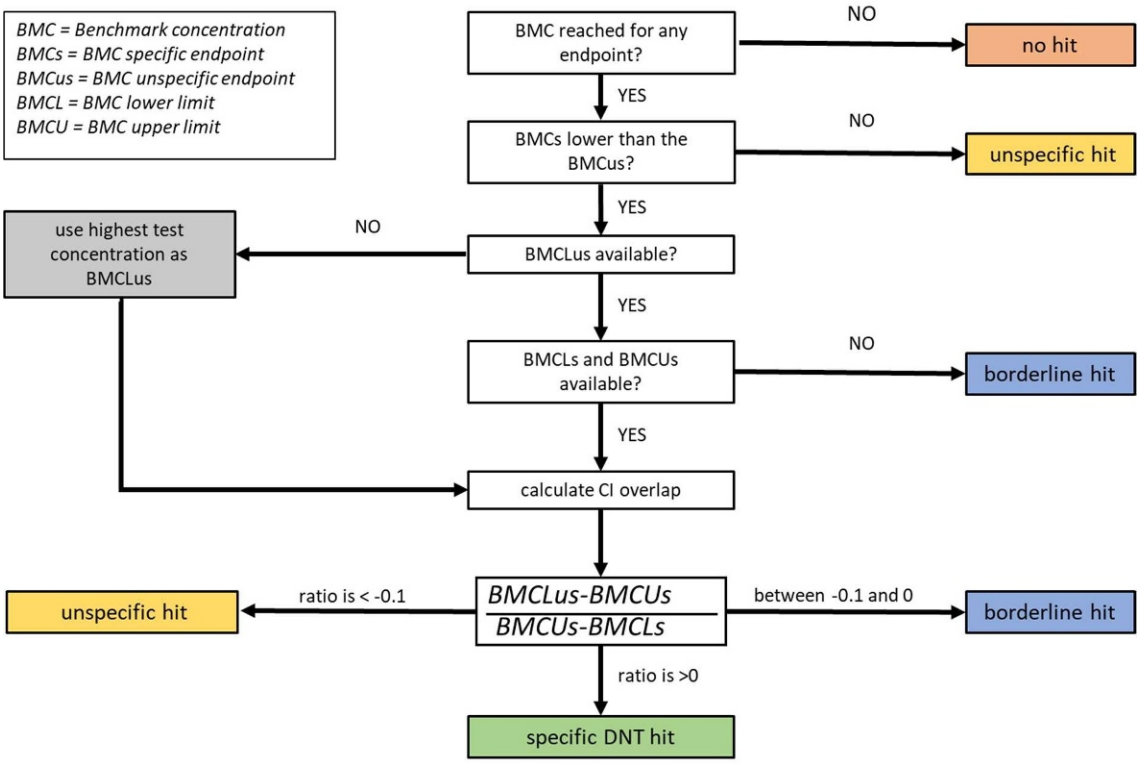


Figure 25: Decision tree for the CM applied to NPC1-5. Overview over the decisions leading to the classification of a compound in one of the four categories: "no hit", "unspecific hit", "specific hit" and "borderline".

The CM applied for UKN2-5 is based on a ratio cut-off for the ratio between the BMC for cell viability and the specific endpoints. This is in line with the respective classification models suggested in previous publications (Hoelting *et al*, 2016; Krug *et al*, 2013; Nyffeler *et al*, 2017).

In more detail, to assess the category of each compound the first question is, if any parameter shows a response above the BMR20. If this is not the case, the compound is classified as negative compound and not followed up with any further testing. If the compound leads to a response above the BMR20, the CM

is applied. To be classified as a specific DNT hit compound, the following criteria have to be fulfilled: Ratio BMC10 viability/BMC25 migration ≥1.3 in UKN2 assay; Ratio BMC25 viability/BMC25 neurite area ≥4 or ≥3 in UKN4 and UKN5 assay respectively (Figure 26). The grey dotted box (Figure 26) describes the handling of compounds that did not reduce viability to the required BMRs as described above.

If this is the case in the UKN4 or UKN5 assay, 2x the highest tested concentration (HTC) is used if viability is not affected or 1x HTC is used if the viability is affected. In the context of such cases, the viability is defined more cautiously as affected if a BMR10 is reached. Moreover, if the calculations do not reach the ratio required to define a compound as a specific hit, the BMCL25 instead of the BMC25 is used for the same calculation. Is the ratio cut-off is still not reached, the compound is defined as cytotoxic, otherwise the compound is categorized as a borderline compound.

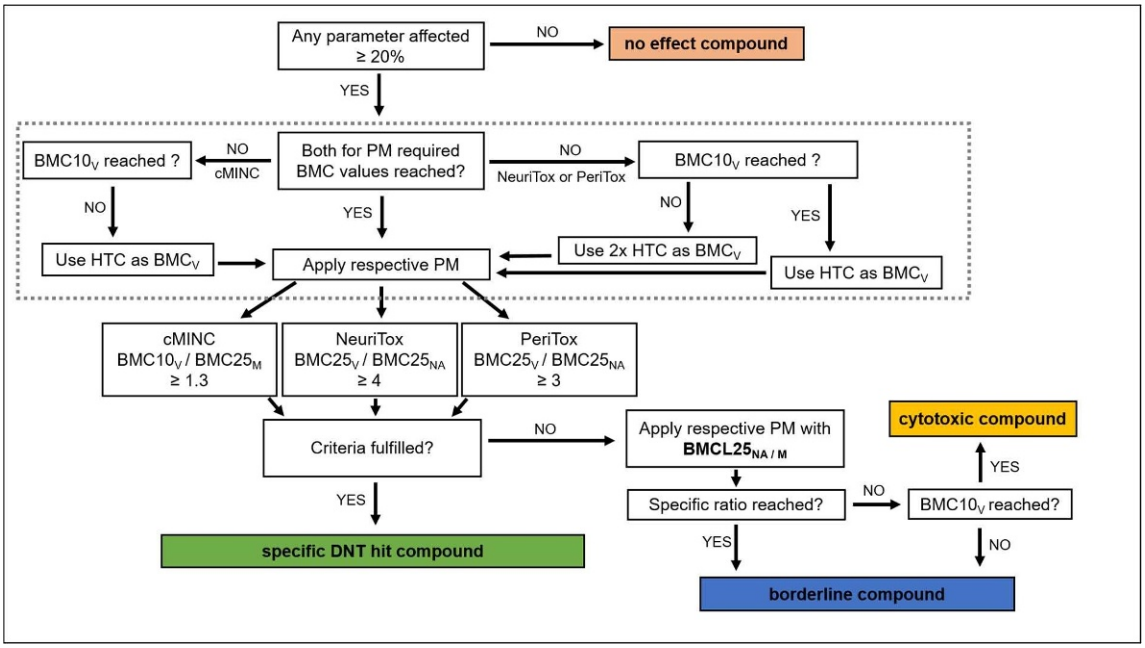


Figure 26: Decision tree for compound categorization including borderline category. Overview of the decisions leading to the classification of a compound in one of the four categories no effect, cytotoxic (unspecific hit), borderline or specific for UKN2-5.

These compound classifications for individual endpoints are subsequently needed for the CM of the whole battery. This is described in WP7 (10.3 How are the results interpreted?).

8.4. Testing Results

8.4.1. Assay reproducibility

The ability of the DNT battery test systems to reproducibly undergo developmental changes over measuring time is the first prerequisite for assay performance. Table 6 summarizes the developmental measures as ranges of raw values for each endpoint and gives an idea of inter-experimental differences that represent biological as well as technical variability.

Table 6 Ranges of raw neurodevelopmental endpoint measures over time for all endpoints assessed in NPC1-5 and UKN2, 4, 5.

Assay	Neurodevelopmental Endpoint	Raw values (5th percentile – 95th percentile)
NPC1	proliferation area (NPC1a)	958 – 2889 slope
	proliferation BrdU (NPC1b)	186320 – 6726919 RLU
	viability 72 h (NPC1)	2947 – 4378 RFU
	cytotoxicity 72 h (NPC1)	1143 – 3408 RFU
NPC2-5	mig. dist. radial glia 72 h (NPC2a)	796 – 1128 µm
	mig. dist. radial glia 120 h (NPC2a)	884 – 1225 µm
	mig. dist. neurons (NPC2b)	2873 – 5826 cells
	mig. dist. oligodendrocytes (NPC2c)	22.73 – 43.5 %
	cell number (NPC2-5)	29.35 – 48.95 %
	neuronal diff. (NPC3)	3.1 – 16.3 %
	neurite length (NPC4)	22.4 – 61.6 µm
	neurite area (NPC4)	151.3 – 411.5 pixel
	oligodendrocyte diff. (NPC5)	1.9 – 11.0 %
	cytotoxicity [72 h] (NPC2-5)	1363 – 2374 RFU
	cytotoxicity [120 h] (NPC2-5)	1346 – 2320 RFU
	viability [120 h] (NPC2-5)	7409 – 12342 RFU
UKN2	migration [24 h]	163.4 – 471.0 # cells
	viability [24 h]	1099 – 2251 # cells
UKN4	neurite Area [24 h]	46300 – 106108 pixel
	viability [24 h]	0.8772 – 0.9770 %
UKN5	neurite Area [24 h]	168521 – 384161 pixel
	viability [24 h]	0.5241 – 0.9157 %

RFU: relative fluorescence unit (background corrected)
RLU: relative luminescence unit (background corrected)

Reproducibility of the assays' raw data is visualized in a dot plots based on the variability of the pre-processed (8.3.6 Data pre-processing) solvent control data between experiments within this procurement. For a better comparison of variability across the different endpoints' raw data (median for NPC/mean for UKN for each experiment), data is normalized to the overall mean across all experiments for each endpoint (Figure 27). While some endpoints such as "migration distance of radial glia" or viability UKN4 are very stable across multiple experiments (COV of 9.6 % and 4.5 % respectively), other endpoints such as proliferation by BrdU, neuronal or oligodendrocyte differentiation or NCC migration show higher variability with inter-experiment COV of 65.38%, 48.4 %, 54.4 % and 35 % respectively.

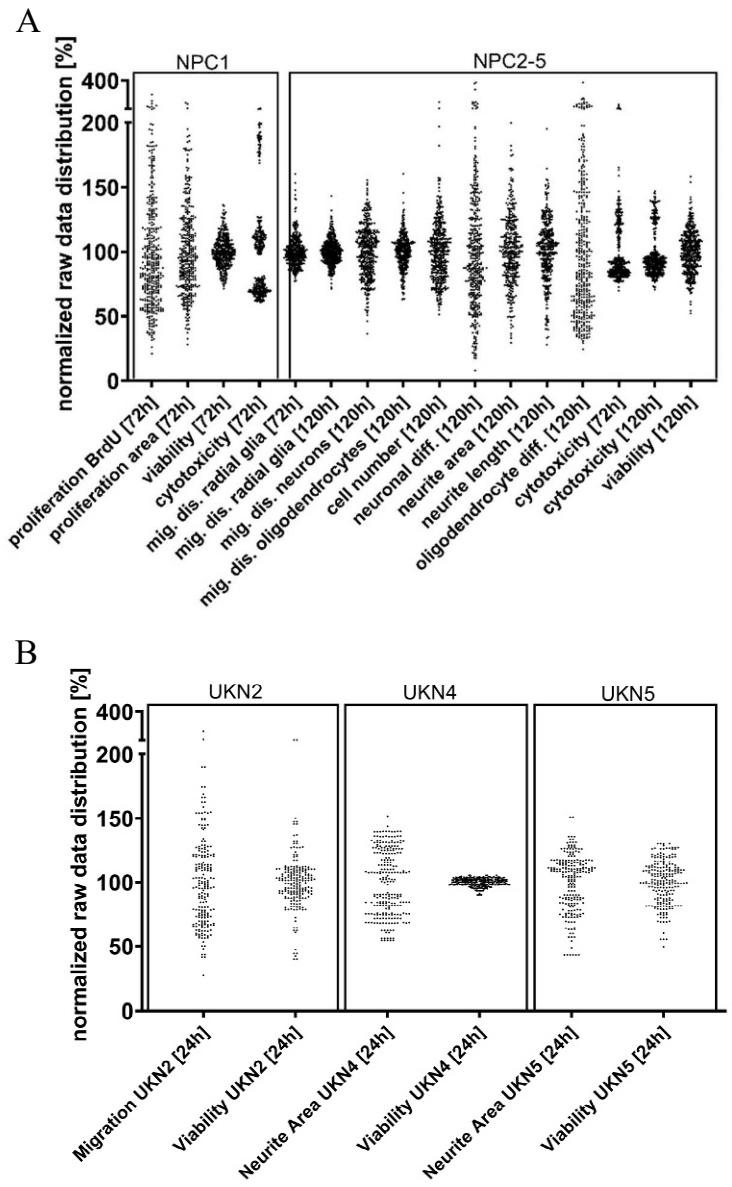


Figure 27: Inter-experimental variability of raw data for solvent controls of each endpoint assessed in NPC1-5 (A; n=377-436) and UKN2, 4, 5 (B; n=204-241) within this procurement.

The final results (concentration response curves and BMC values) were obtained from the combination of independent experiments. To reduce the inter-experimental bias and to make the data more comparable, each response value was normalized to the respective mean or median of the solvent control. To quantify the inter-experimental variability after normalization, the SC cannot be used as this value is set to 100 % for each experiment. Instead, we determined the variability between all experiments based on the response of the lowest test concentration. For this approach it is assumed that the lowest test concentration does not affect any of the endpoints measured. Figure 28 demonstrates the influence of the normalization on the overall variability in each endpoint. Here, the normalization leads to a clear reduction in inter-experimental variability making the endpoints' assessment more reproducible. However, there are still

differences in variability between different endpoints again indicating that not all endpoints are able to identify the same effect size.

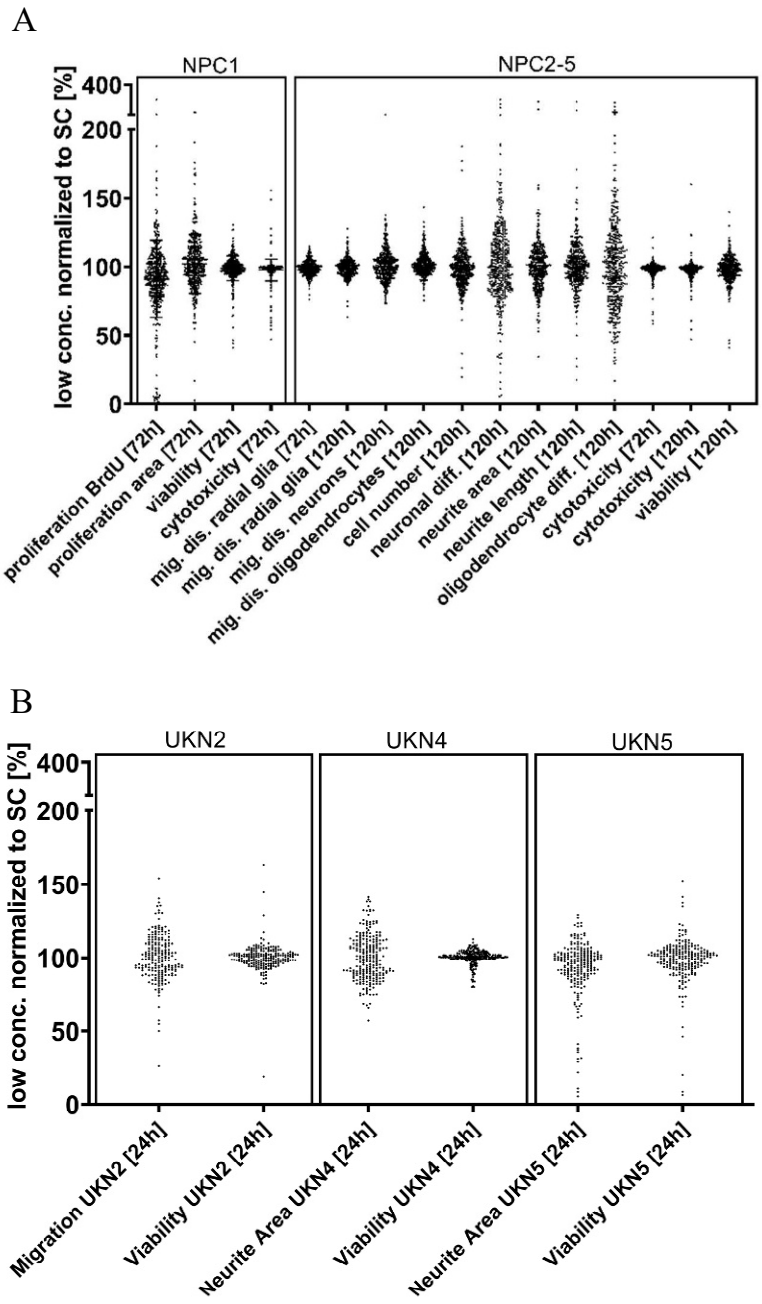


Figure 28: Inter-experimental variability of lowest concentrations normalized to solvent control for each end-point assessed in NPC1-5 (A; n=379-436) and UKN2, 4, 5 (B; n=204-241) within this procurement.

In the second step, we studied the modulation of the measured DNT-specific endpoints by ESC compounds. Therefore, the variation of the response value normalized to solvent control for each ESC compound was assessed across all experiments (Figure 29). These data demonstrate a reproducible modulation of end-points by the respective compounds in the analyzed endpoints, which indicate assay reproducibility.

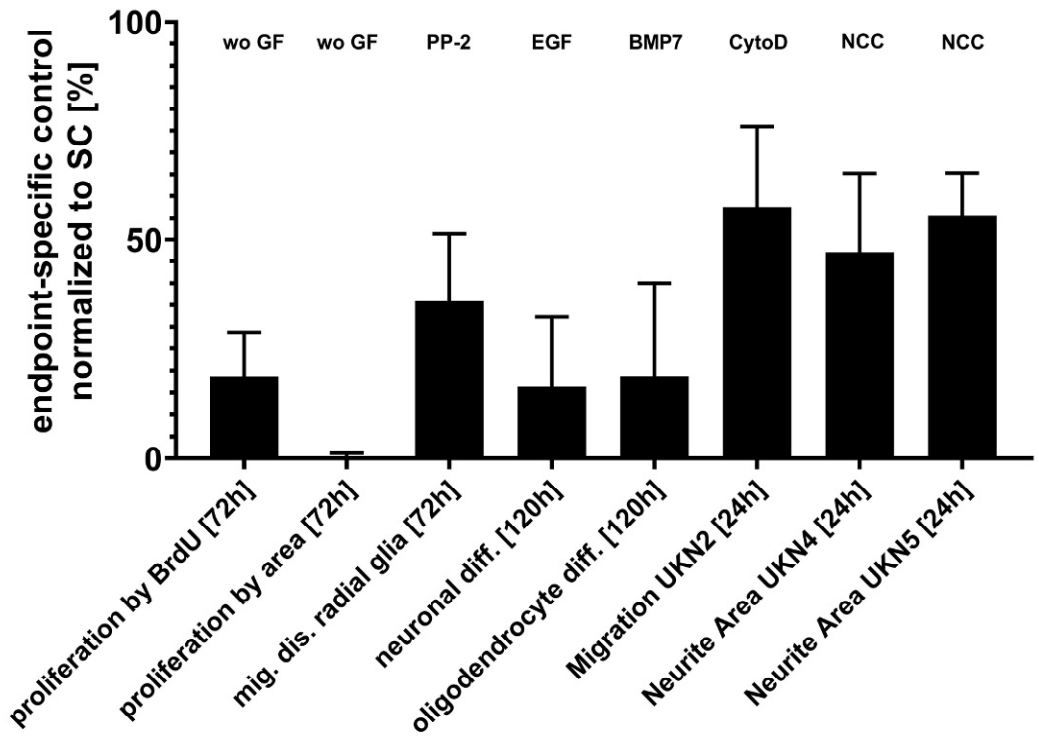


Figure 29: Inter-experimental variability of endpoints specific positive controls (ESC) normalized to the solvent control. Proliferation media without growth factors (wo GF) serves as PC for proliferation by BrdU and by area, PP-2, epidermal growth factor (EGF), bone morphogenetic protein 7 (BMP7), Cytochalasin D (CytoD) and Narciclasine (NCC) serve as ESC for migration distance of radial glia, neuronal differentiation, oligodendrocyte differentiation, migration in UKN2 and neurite area in NPC1-5 and UKN2, 4, 5, respectively. The data is based on n= 40 – 400 experiments and shows the mean normalized response including the SD.

Table 7 summarizes variability and reproducibility of the assays described above:

- Intra-experimental variability (SC) is the mean COV±SD of the COV for the replicates within the SC of each experiment across n>200 experiments.
- Inter-experimental variability (raw) is the variation as COV between the raw SC of all independent experiments (n>200; see Figure 27).
- Inter-experimental variability (low conc.) is the variability as COV between the response of the lowest test concentration of all experiments (n>200) after normalization to SC. It is assumed that the lowest test concentration does not affect any of the endpoints measured (see Figure 28).
- Inter-experimental variability (ESC) is the variability between experiments (n>40) after normalization based on the response of the respective ESC (see Figure 29).

Table 7 Summary of assay variability and reproducibility

Assay	Neurodevelopmental Endpoint	Intra-experimental variability (SC)	inter-experimental variability (raw)	Inter-experimental variability (low con.)	Inter-experimental variability (ESC)
NPC1	proliferation area (NPC1a)	19.4 ±9.4	34.3 %	21.4 %	1.1 %
	proliferation BrdU (NPC1b)	13.2 ±5.9	65.4 %	28.1 %	10.3 %
	viability 72 h (NPC1)	5.8 ±3.5	11.6 %	9.1 %	
	cytotoxicity 72 h (NPC1)	1.3 ±1.6	37.7 %	7.9 %	
NPC2-5	mig. dist. r. glia 72 h (NPC2a)	5.1 ±1.9	11.3 %	5.2 %	15.5 %
	mig. dist. r. glia 120 h (NPC2b)	5.6 ±2.3	9.6 %	6.1 %	
	mig. dist. neurons (NPC2c)	10.7 ±4.3%	19.8 %	11.5 %	
	mig. dist. oligoden. (NPC2d)	9.4 ±4.9%	13.1 %	10.9 %	
	cell number (NPC2c)	12.4 ±4.8%	23.3 %	14.4 %	
	neuronal diff. (NPC3)	23.0 ±8.9 %	48.4 %	30.5 %	16.0 %
	neurite length (NPC4)		25.9 %	23.3 %	
	neurite area (NPC4)		28.6 %	23.1 %	
	oligodendrocyte diff. (NPC5)	35.1 ±9%	54.3 %	34.2 %	21.5 %
	cytotoxicity [72 h]	1.4 ± 1.8%	24.4 %	8.6 %	
	cytotoxicity [120 h]	1.5 ±1.8%	37.6 %	6.6 %	
	viability [120 h]	8.4 ±3.3%	15.7 %	9.2 %	
UKN2	migration [24 h]	14.8±4.3%	35.4 %	15.6 %	32.2 %
	viability [24 h]	6.0±3.0%	21.1 %	9.7 %	
UKN 4	neurite Area [24 h]	11.7±4.5%	25.4 %	15.3 %	38.5 %
	viability [24 h]	3.8±3.5%	3.2 %	4.3 %	
UKN 5	neurite Area [24 h]	7.8±4.3%	22.3 %	21.0 %	17.5 %
	viability [24 h]	5.5±3.4%	16.8 %	15.7%	

8.4.2. Evaluation by endpoint

Within this procurement 119 compounds were tested in the assays NPC1-5 and UKN2,4-5 (Bal-Price *et al*, 2018). 163 compounds were tested in UKN2,4, 5 at the University of Konstanz (UKN) and 123 compounds were tested in NPC2-5 at the IUF – Leibniz Research Institute of Environmental Medicine in Düsseldorf (IUF). The US Environmental Protection Agency (EPA; not an official partner of this procurement) tested 79 compounds in the NNF and 39 in other EPA assays. Because of the Covid-19 pandemic, testing at the US-EPA has been delayed. However, as soon as the data is generated, it will add to the overall DNT *in vitro* database.

For evaluation of the results of the battery, we focus on the 119 compounds that were tested in the UKN and NPC assays. However, for an assessment of the effect that additional assays add to the battery as well as some redundant assay analyses, additional assessments are added that include the US-EPA assays with a smaller sets of chemicals.

DNT *in vitro* testing within this procurement led to a generation of testing data for 119 compounds across 22 endpoints (6 endpoints with 3 UKN assays and 16 endpoints with 5 NPC assays with 4 of the 5 assays multiplexed) including their respective viability and cytotoxicity information. These data produced more than 2000 concentration-response curves (Annex N, O). From the concentration-response curves, BMCs were calculated for each endpoint by using respective BMR. A summary of the results of all BMC values of the whole study is presented in Annex A1 and summarized in the heatmap in Annex P.

As a first data check, we cherry-picked some compounds with well-described MoA from the heatmap to understand the mechanistic rationale of the generated results. In the following paragraphs these few compounds are spot-lighted.

Domoic Acid: The neurotoxic effects of domoic acid have been attributed to its analogy to glutamate and its high affinity binding to excitatory glutamate receptors (Wikipedia) causing excitatory neurotoxicity. Within the battery, domoic acid only produced a 'hit' in the rNNF assay. This makes sense considering its MoA, because the other neuronal models do at the investigated stages of development not express functional glutamatergic synapses (despite some present gene products for glutamate receptors and transporters in UKN2, 4 and 5; Annex A6). Therefore, this result supports the mechanistic specificity of the DNT *in vitro* battery with test systems representing different neurodevelopmental maturation stages.

Narciclasine: Narciclasine is a plant growth inhibitor isolated from Narcissus bulbs. It inhibits cell proliferation and tumor cell invasion by modulating the Rho/ROCK/LIMK/cofilin signaling pathway leading to actin stress fibre stabilization (Lefranc et al, 2009). In mice *in vivo,* rho kinase modulation primarily affects neurons in comparison to NPC or glia cells (Bye et al, 2016). Also, in the DNT *in vitro* battery, narciclasine exerts cell type-specific effects with highest sensitivity of neuronal differentiation compared to radial glia and oligodendrocytes. However, the Rho kinase pathway seems to contribute to many different neurodevelopmental processes because almost all endpoints with the exception of oligodendrocyte migration were effected by low concentrations of narciclasine.

Methylmercury chloride: According to AOP17, interference of compounds with SH-groups of proteins leads to neurodegeneration. Methylmercury chloride has this known MoA also leading to oxidative stress, which is quite general (Bal-Price et al, 2015b) as potentially all cells exposed to a sufficient amount can be affected. Depending on cell-specific uptake and defence as well as proteins affected, cell-type specific susceptibility is observed. This is reflected well in the battery. Methylmercury chloride affects almost all tested endpoints of the DNT *in vitro* battery, yet with different potencies. Here neurons seem to be the most sensitive target population.

In the next step, BMCs were used to classify compounds as 'specific hit', 'unspecific hit', 'no hit' and 'borderline hit' (see 8.3.9 for detailed in formation). This classification of compounds uses the respective most sensitive endpoint (MSE) defined by the lowest BMC of the respective compound identified during the testing (Annex A1). This MSE varies across compounds as depicted by the pie diagrams in Figure 31a-c. Here, the MSE are presented across all tested compounds for three scenarios, i.e. (a) only IUF and UKN assays (119 compounds), (b) IUF, UKN and US-EPA rat neuronal network formation (rNNF) assay (60 compounds) and (c) IUF, UKN and all US-EPA assays (27 compounds). Of the 22 endpoints tested in scenario (a), 10 endpoints are represented as MSE at least once. Within the 119 compounds tested by IUF and UKN, 36% were classified as negatives. The residual 64% of compounds were represented by MSE in NPC5 (35%) > NPC1a (7%) > UKN2 (5%) = UKN4 (5%) > NPC2a (3%) = NPC4 (3%) = UKN5 (3%) = NPC3 (3%) > NPC1b (1%) (Figure 30a). Endpoints not affected as the most sensitive ones were, NPC2b,c. Here only DNT specific endpoints were considered. This changed when the 60 compounds tested by the IUF, UKN and the US-EPA (rNNF assay) in scenario (b) were analyzed. Now 20% of the compounds were negatives and the overrepresented MSE was rNNF (50%) followed by NPC5 (22%) > UKN4 (3%) > NPC1a (2%) = UKN5 (2%) = NPC4 (2%; Figure 30b). In scenario (c) only 27 compounds were analyzed shifting the ranking to rNNF (70%) > negatives (7%) > NPC5 (11%) > synaptogenesis assay (4%;Figure 30c). This evaluation is preliminary due to the restricted data currently present from the US-EPA. Therefore, an analysis of the minimal requirement of the battery is not possible yet. In addition, even with the full compound set of 119 evaluated it has to be noted that depending on a compound's MoA, different assays will give the highest alert. Thus, the selection of compounds strongly determines the battery's minimal requirement. Therefore, surely more compounds than the +100 covering more compound classes and thus MoA are needed to determine the DNT *in vitro* battery's composition in the end. As a conclusion from the 119 compounds testing data (scenario (a)) all assays seem to be relevant for the battery because the two specific endpoints that never became MSE, i.e. NPC2b, c, are multiplexed with NPC2-5, respectively, are therefore no stand-alone-assays and thus cannot be omitted by themselves. More testing in the future will allow for more specific analyses regarding the minimum battery requirement.

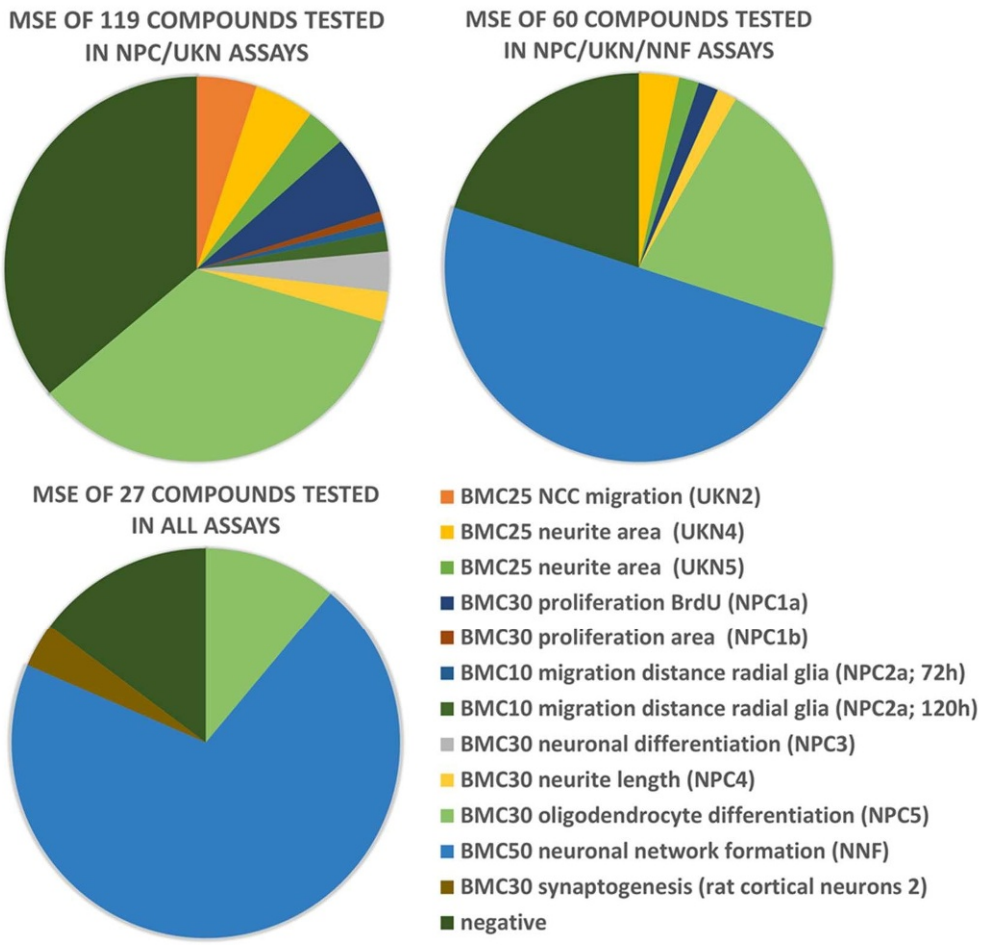


Figure 30: MSE across the Compound set.
The MSE is represented as the endpoint with the lowest BMC across the different assays and the compound sets. (a) 119 compounds tested in UKN and NPC assays, (b) 60 compounds tested in UKN, NPC and rNNF assay, (c) 27 compounds tested in all assays of the battery.

By applying the MSE and related confidence intervals, the 119 tested compounds (scenario (a)) were classified. Here, based on the MSE 58 compounds were classified as 'specific hit', 16 compounds were classified as 'unspecific hit' due to primarily a change in cell viability and 43 chemicals did not exert an effect on any endpoint and were thus classified as 'no hit'. The residual 2 compounds were classified as 'borderline' compounds (Figure 31). The confidence intervals shown in the graph define the uncertainty around each BMC. A CI that is spanning over several order of magnitude or is above the 1.5 x the highest test concentration is typically an indication of an effect only in the highest test concentration. In this case the concentration response curve modelling contains a high degree of uncertainty which is reflected in the wide CI. Test data with wide CIs as well as CIs above the test range always need to be assessed separately by expert judgment in order to control the classification (see 8.3.9 for more information on how these compounds are classified). In addition to the CIs, the graph shows a 3-fold BMC range which gives a concentration range of a BMC of the MSE. This range is used to estimate if BMC from different endpoints are similar or different and is applied instead of the calculated CIs because it allows a consistent assessment of differences between assays that is independent of differences in the assay variation. Another reason for the use of the 3-fold BMC range is that calculated BMC are not available for all assays. Figure 31 and Figure

32 demonstrate that this 3-fold range is a good representation of the CI. However, while the majority of CI are lower than the 3-fold range, some are also higher.

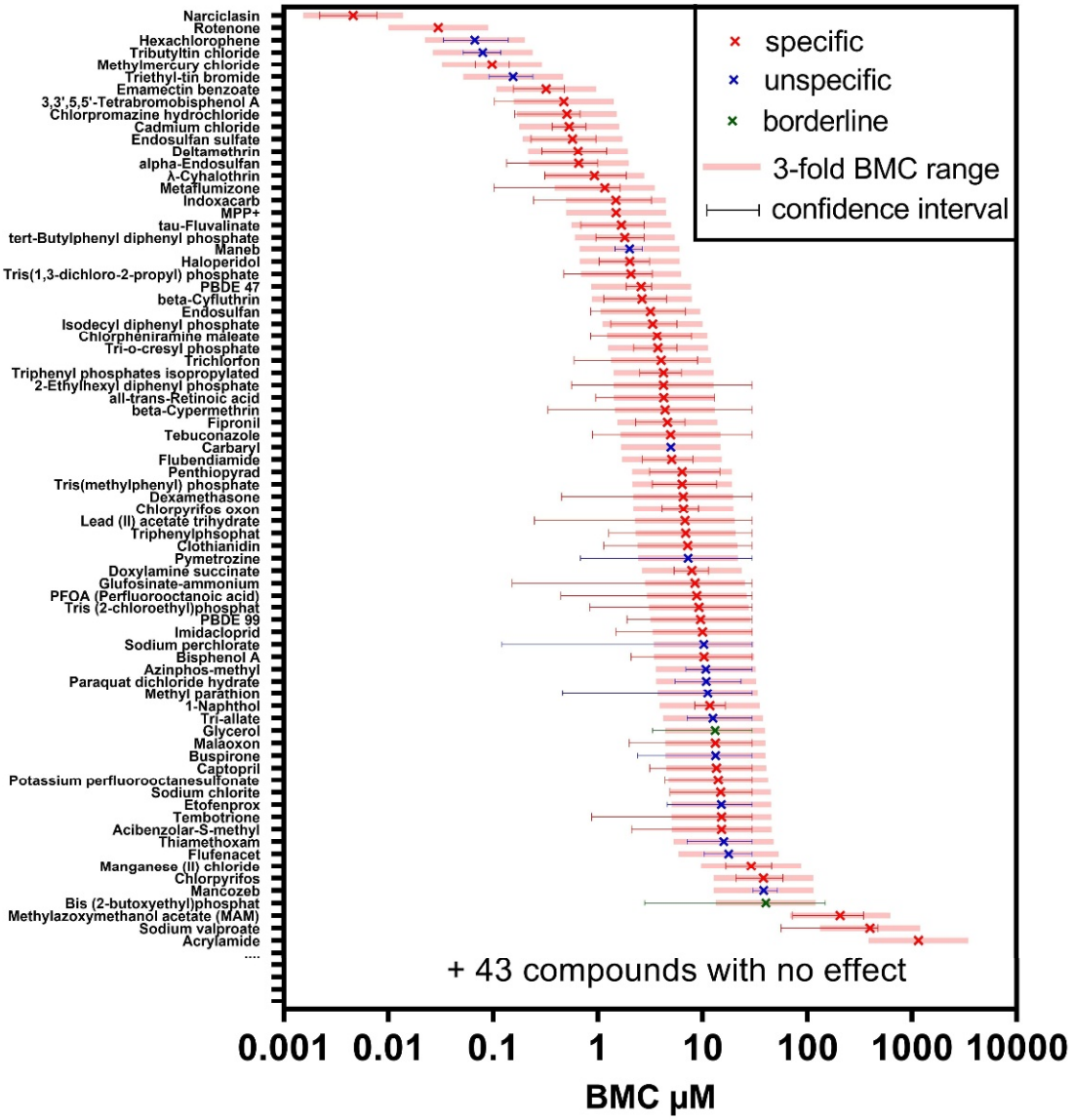


Figure 31: Hazard identification of 119 compounds tested in UKN and NPC assays. BMCs of the MSE affected by the compounds in the UKN or NPC assay are ranked from highest to lowest including the calculated confidence intervals as well as the 3-fold BMC range. The effects are classified in "specific hit", "unspecific hit", "borderline" and "no hit". Upper confidence limits that are above 1.5x the highest test concentration are given as 1.5x the highest test concentration as CIs above the test range cannot be calculated with certainty. Wide CIs (with a manually set upper confidence limit) are typically an indication of an effect only in the highest test concentration that is accompanied by a high uncertainty due to limited information on the curves shape).

Moving to scenario (b), i.e. including the rNNF assay into the analysis, the MSE identification changes. What happens when the rNNF assay is added to the NPC and UKN assay results is depicted in Figure 32. Of the 60 compounds commonly tested, the rNNF assay gives the MSE in 63 % (red crosses Figure 32). This graph

visualizes that the rNNF assay either has a much lower BMC_{50} (at least one order of magnitude) than the NPC and UKN assays (17 compounds), is in a similar concentration range than the other battery assays (8 compounds), has an BMC above the other battery assays (4 compounds), produces an effect where the residual battery classified a 'negative' (7 compounds) or was negative where the residual battery classified 'positive' (9 compounds). Hence, the rNNF assay adds important information to the overall battery results. It has to be of note, that the rNNF assay runs for 12 days with a 12 day exposure period. The residual battery assays take 1 to 5 days with respective shorter exposure duration. Hence, the longer exposure might in some cases affect magnitude of effects resulting in lower BMC values. Yet still the rNNF assay adds specific information to the overall battery as there are compounds only positive in the rNNF assay or solely positive in the NPC/UKN assays.

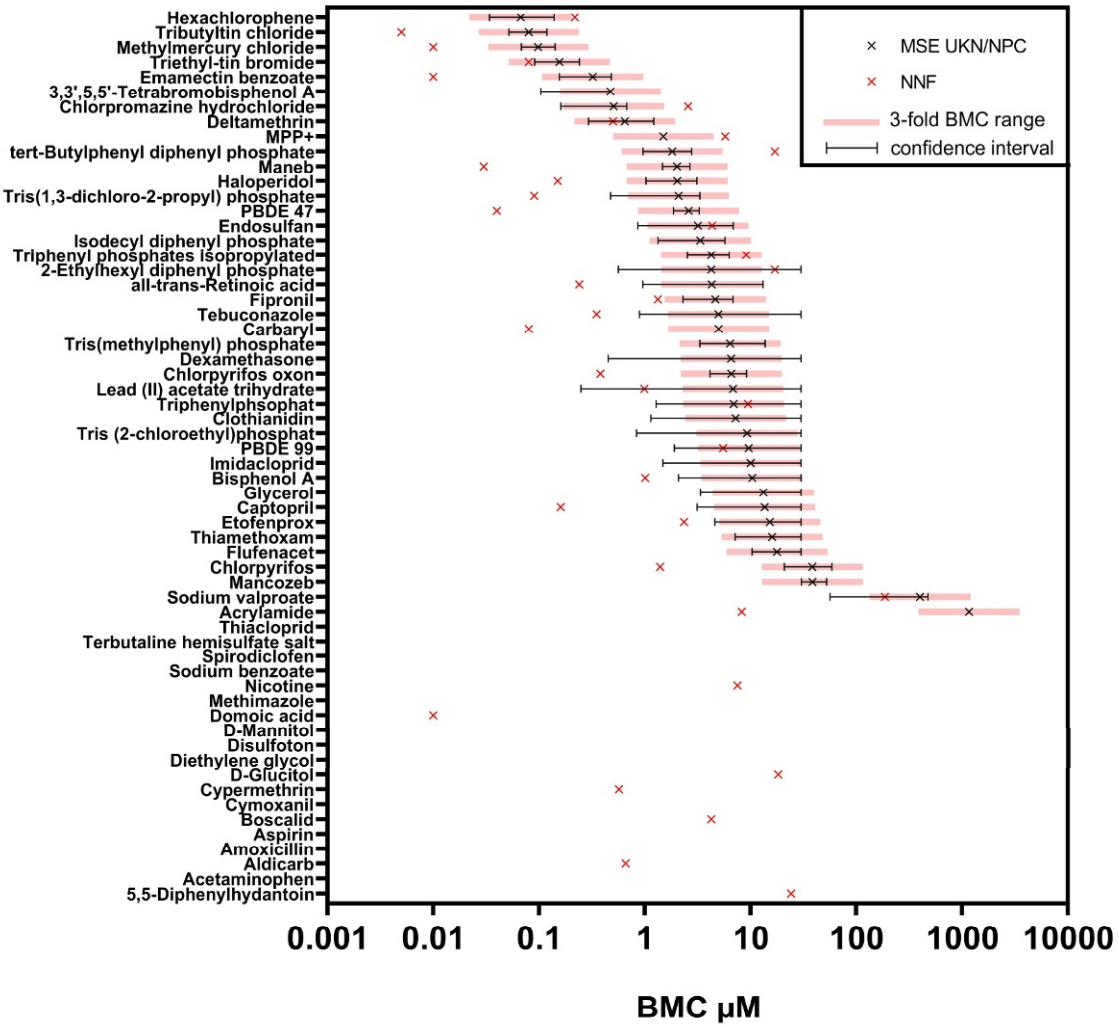


Figure 32: Hazard identification of 60 compounds tested in UKN, NPC and rNNF assays. BMCs of the MSE affected by the compounds in the UKN or NPC assay are ranked from highest to lowest including the calculated confidence intervals as well as the 3-fold BMC range. Additionally, the BMC of the rNNF is plotted in comparison. Upper confidence limits that are above 1.5x the highest test concentration is given as 1.5x the highest test concentration as CIs above the test range cannot be calculated with certainty. Wide CIs (with a manually set upper confidence limit) are typically

an indication of an effect only in the highest test concentration that is accompanied by a high uncertainty due to limited information on the curves shape).

In many cases, the MSE has the lowest BMC but is accompanied by other endpoints with the BMC in a similar range. These endpoints we interpret as equally important. To also take these endpoints into account, we applied the 3-fold BMC range of the MSE to identify if BMCs are different or similar. The scientific rationale for this MSE range is exemplified by the model compound narciclasine. The MSE for narciclasine is neuronal differentiation (NPC3) with a BMC_{30} of 0.005 µM exerting no cytotoxicity for this test method. However, within the 3-fold range of this BMC also other endpoints were changed, i.e. proliferation by area (NPC1b; BMC_{30} 0.008 µM), radial glia migration (NPC2a; BMC_{10} 0.010 µM), neurite length as well as neurite area (NPC4; BMC_{30} 0.013 µM), neurite area (UKN4; BMC_{25} 0.014 µM) and oligodendrocyte migration (NPC2c; BMC_{10} 0.047 µM). Narciclasine is an alkaloid belonging to the chemical class of phenanthridines, which is found in various Amaryllidaceae species. Different MoA for narciclasine were identified over the last years as this compound served as a lead compound for anti-tumor therapies (Fürst 2016). Narciclasine binds to the 60S subunit of the ribosome, thereby inhibiting the step of peptide bond formation during protein biosynthesis. This ribosome inhibition is made responsible for narciclasine's cell growth inhibition properties found in tumor cells. In addition, narciclasine interacts with the translation elongation factor eEF1A and increases Rho activity. These signaling pathways lead to alteration of migration, neurite growth or neuronal differentiation (Fürst 2016; Park et al. 2018; Threadgill et al. 1997). Exactly these effects were found in the phenotypic screening of the DNT *in vitro* battery showing its high predictive capacity. As narciclasine is a pharmacological compound, its MoA have been very well studied and therefore it serves as an excellent model compound. This is not the case for most environmental chemicals. Here, also multiple MoA are present, yet for many compounds unknown. Therefore, we conclude that looking at only the one MSE is not sufficient, yet multiple effects might contribute to DNT potential of compounds. Thus, when talking about the MSE, we include all endpoints that lie within the 3-fold BMC of the MSE. For the 'screening and prioritization' application, this is not as crucial as for 'risk assessment' purposes of the DNT *in vitro* battery results as the latter might lead to targeted *in vivo* testing.

Using the approach of the 3-fold MSE range, we identified the most relevant endpoint hits for each compound that was identified as a 'hit', irrespective of its specificity using the NPC and the UKN assays (Figure 33). BMC values for the different compounds range from <0.01 µM to >1000 µM with the majority of compounds showing effects between 1 and 10 µM. On the first glimpse, two different types of hits can be distinguished: one that does not contain a second endpoint hit within the 3-fold BMC range of the MSE and a second one that has less specificity for one endpoint and thus collects several endpoint BMC within the 3-fold range of the MSE. Assays belonging to the first group are NPC5>>NPC1a>UKN5>UKN4>UKN2, here the majority of single MSE are attributed to NPC5. The rest of the assays gather in different combinations within the 3-fold range of the MSE. It is to note, that in this case endpoints were not filtered for specificity, meaning that all endpoints independent of each other were plotted (Figure 33).

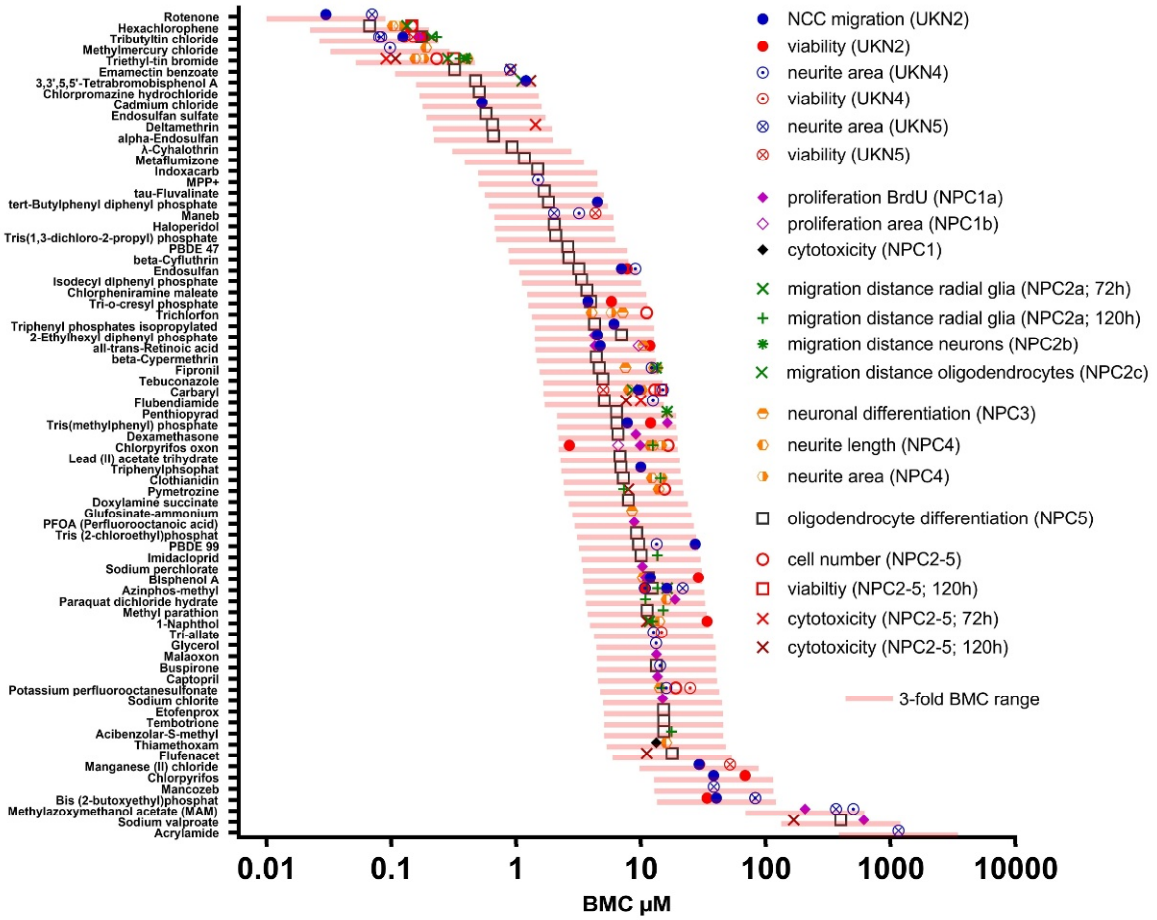


Figure 33: BMC of MSE of 'specific' and 'unspecific' hits across the NPC and UKN assays for the 119 compounds tested including the BMC of endpoints that fall into the 3-fold range of the MSE. 43 compounds tested negative.

Adding the NNF assay to the evaluation changes the picture: as expected from the first MSE analyses (Figure 32) the NNF assay becomes the most frequent MSE (Figure 30). In this assay combination, the rNNF assay produces specific effects at fairly low concentrations (Figure 34). This also holds true for oligodendrocyte differentiation (NPC5), yet not that frequent anymore. Interestingly, there is only one compound producing effects on rNNF and NPC5 at the same concentration, which is deltamethrin. For most of the chemicals, MSE for rNNF and NPC5 do not overlap. In addition to the rNNF assay and NPC5, in this smaller data set of only 60 compounds, the sole MSE is only found by UKN5 (once) and UKN4 (twice). In all other cases, multiple endpoints gather within the 3-fold range of the respective MSE.

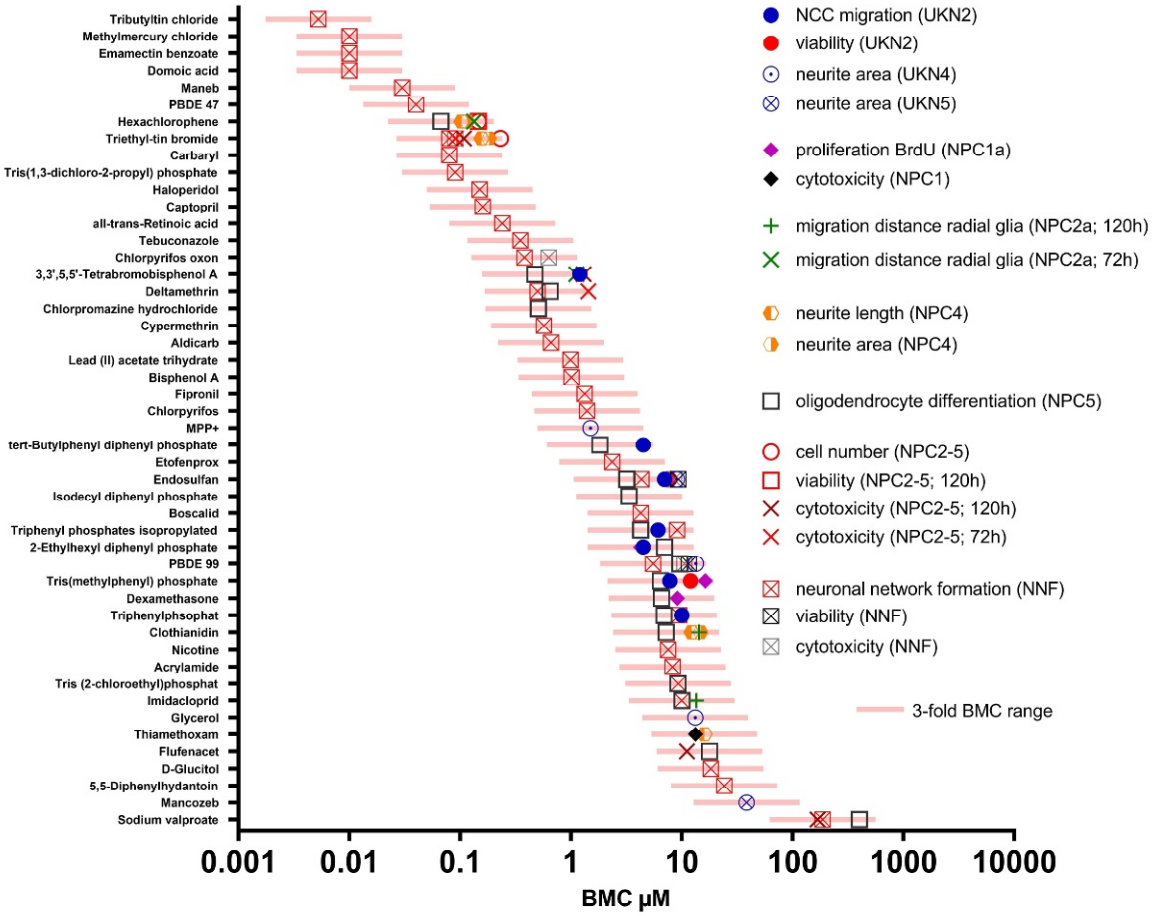


Figure 34: BMC of MSE of 'specific' and 'unspecific' hits across the NPC, UKN and rNNF assays for the 60 compounds tested including the BMC of endpoints that fall into the 3-fold range of the MSE.

An alternative way of data analysis is the use of only specific endpoint hits according to the individual assay classification models (see 8.3.9. for specific information) by omitting all BMCs from the analysis that did not distinguish between the effect of a compound on a DNT endpoint and a reduction in viability or induction of cytotoxicity. Similar to the readout in Figure 33, NPC5 is the dominant specific assay in this data set (Figure 35). Using only BMCs that are classified as "specific hit", reduces the number of compounds with a DNT hit from 76 to 64. The number of endpoints (only DNT specific endpoints) that is affected by these 64 compounds is reduced from 14 endpoints with 117 BMCs (for the 76 compounds) to 12 endpoints with 208 BMCs.

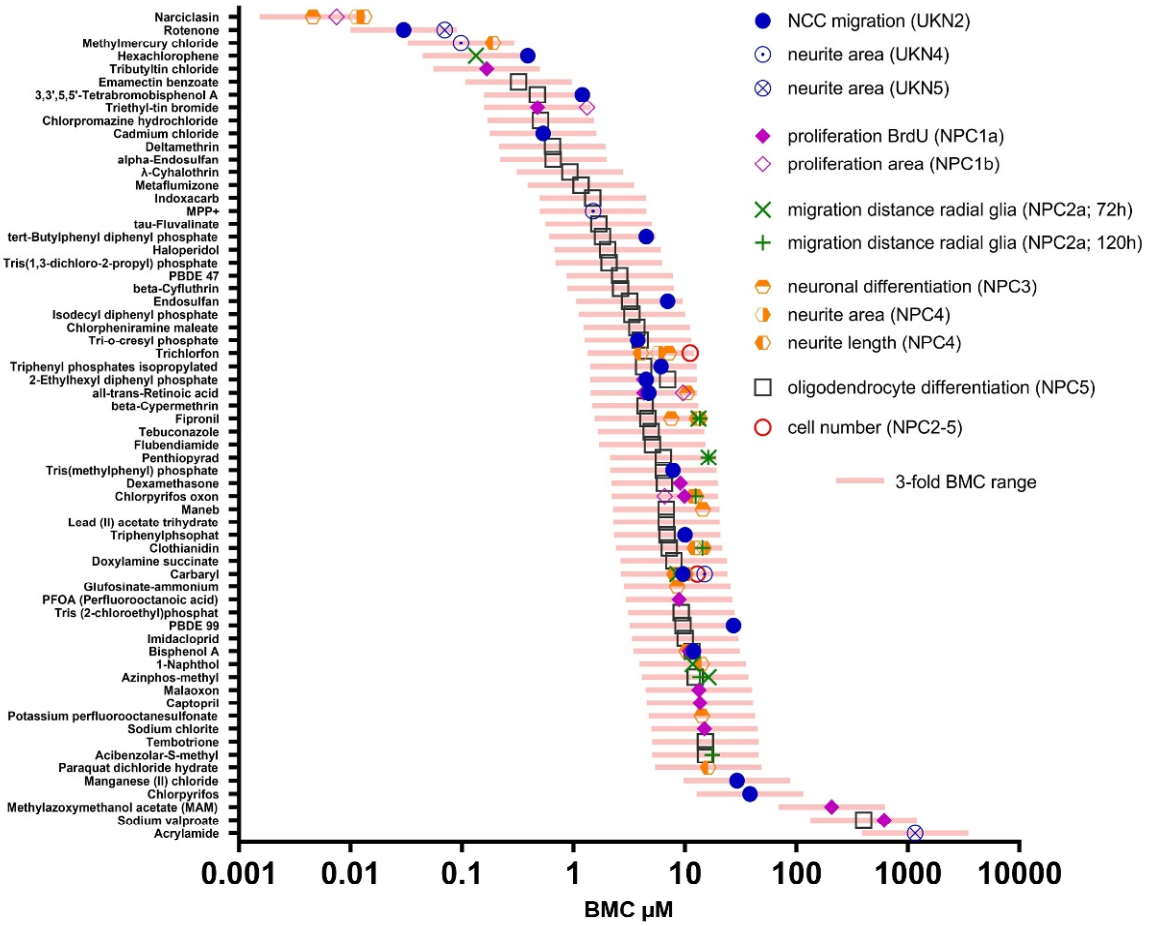


Figure 35: BMC of the MSE of only 'specific hits' across the NPC and UKN assays for the 119 compounds tested including the BMC of endpoints (specific hits) that fall into the 3-fold range of the MSE.

This analysis with only specific hits was also performed for the reduced compound set of 60 including the rNNF assay (Figure 36). Again, similar to the evaluation of the 60 compounds using 'specific' and 'non-specific' DNT hits, mainly the rNNF assay forms the MSE followed by a few specific NPC5 hits. Again, here using only BMCs that are classified as specific hits, reduces the number of compounds with a DNT hit from 48 to 42. The number of endpoints (only DNT specific endpoints) that is affected by these 42 compounds is 11 in both approaches with 67 BMCs vs. the 96 BMCs for 48 compounds.

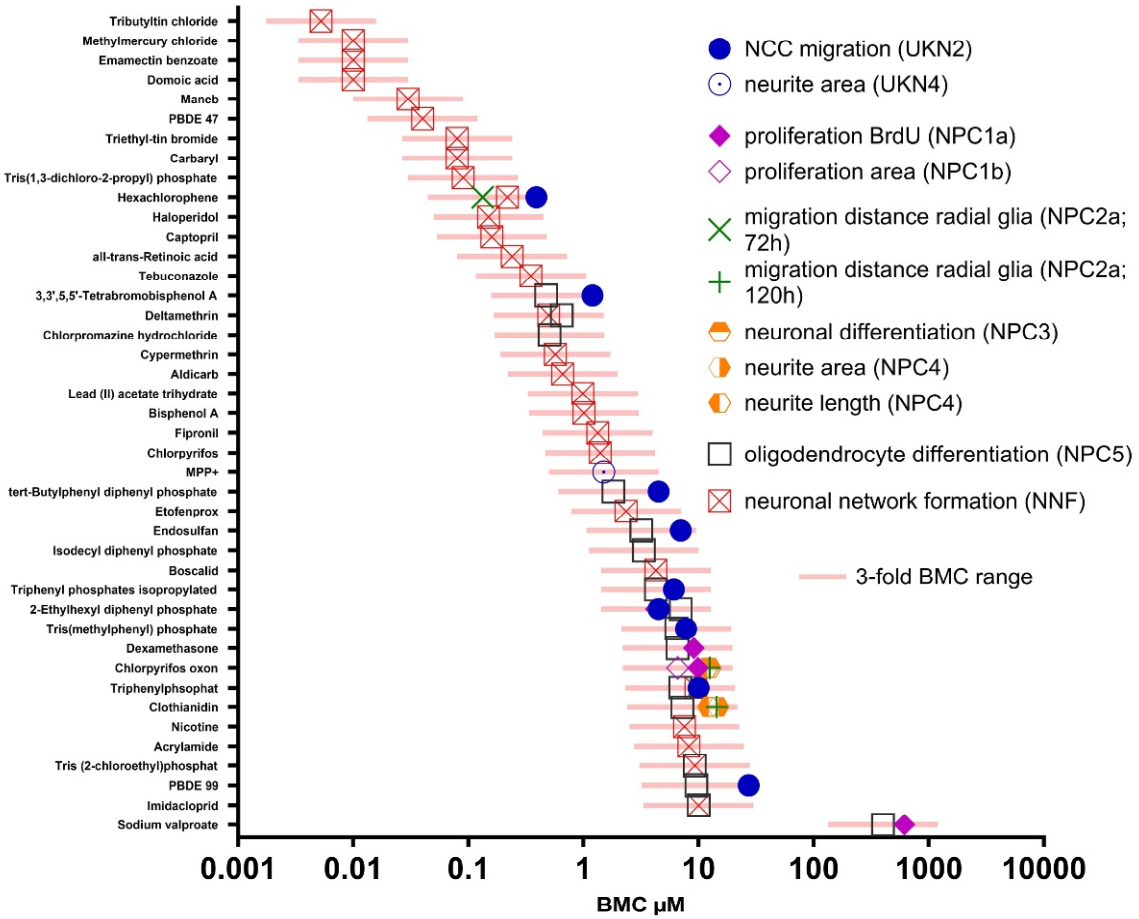


Figure 36: BMC of MSE of only 'specific hits' across the NPC, UKN and rNNF assays for the 60 compounds tested including the BMC of endpoints (specific hits) that fall into the 3-fold range of the MSE.

8.4.3. Evaluation by compound classes

Next, we analyzed how different compound classes affect endpoints of the DNT *in vitro* battery. A summary of compound classes can be found in Figure 37. Of the 119 compounds tested, 66 belong to a large variety of compound classes with less than three compounds/class. The residual 53 compounds distribute across organophosphates (24), carbamates (4), pyrethroids (7), organochlorines/fluorines (6), neonicotinoids (5), metals (6). In the next paragraphs compound classes will be discussed, which contain at least four compounds.

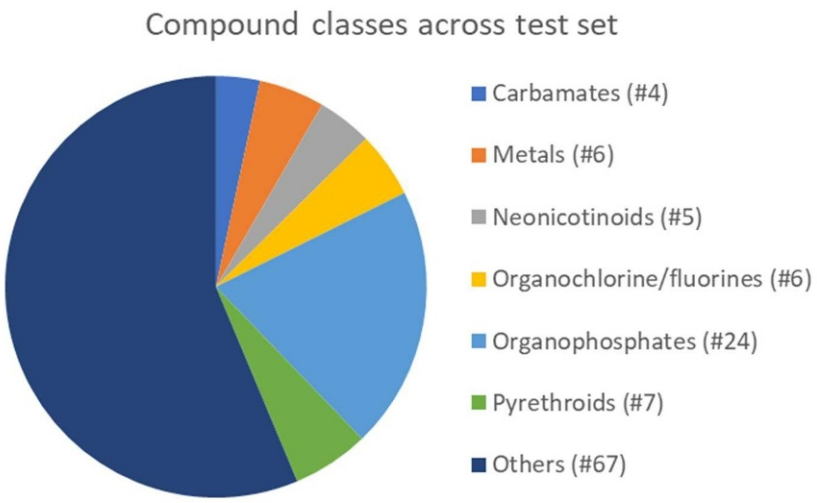


Figure 37: Distribution of compound classes across the compound set (#119).

Analyses of effects for the different compound classes are supposed to investigate if there are common effect patterns across the different endpoints. Again, these analyses were made for scenario (a) with the 119 compounds in the NPC/UKN assays and for scenario (b) with the 60 compounds in the NPC/UKN + US-EPA rNNF assay.

Organophosphates:

Scenario (a): For the compound class of organophosphates (24 compounds), DNT *in vitro* effects overall do not show a specific pattern (Figure 38). However, this compound class does not affect primary NPCs (proliferating or differentiating) viability, and also UKN assays' viability in only 7/24 (UKN2), 2/24 (UKN4) and 3/24 (UKN5) compounds. This lower effect of organophosphate compounds on primary NPC viability compared to the UKN assays might be due to the cell type (NPC in NPC1) or mixed neuron/glia cultures (NPC2-5) in comparison to the sole neural crest cell or neuronal cultures of the UKN assays as it was shown earlier that glia can be neuroprotective (Giordano et al. 2009; Pizzurro et al. 2014). The endpoints overrepresented in the organophosphate effects were oligodendrocyte differentiation (NPC5; 11/24 with 10 "specific hits") and NCC migration (UKN2; 10/24, with 8 "specific hits") followed by radial glia migration (NPC2a; 6/24, with 4 "specific hits").

Organophosphates' primary reported MoA, especially for acute neurotoxicity, is inhibition of acetylcholinesterase (AChE). However, behavioral studies in rodents, zebrafish and *C. elegans* suggest adverse effects of organophosphates on brain development after perinatal exposure already below concentrations inhibiting AChE, yet with some uncertainties (Silva 2020). Specific mechanisms for organophosphate-induced DNT in general have not been identified. One of the best studied organophosphates, chlorpyrifos, however, inhibits DNA synthesis (Whitney et al, 1995), increases apoptosis (Roy et al, 1998), inhibits neuritic projections (Qiao et al, 2004), interferes with the serotonergic system (Aldridge et al, 2003) as well as with oligodendrogenesis (Garcia et al, 2003) in rodents *in vivo* (summarized in Abreu-Villaça & Levin, 2017). For the latter, chlorpyrifos first induced protein markers for oligodendrocytes and caused a reduction later in life during adolescence. The statistical analysis of the data presented here identifies a reduction of the endpoints. However, for single compounds also inductions of

endpoints can be observed (concentrations-response curves in the Annex N, O). This is e.g. true for tris(chloroethyl)phosphate and bis-(2-butoxyethyl)phosphate. Hence, DNT *in vitro* battery results of chlorpyrifos and its metabolite, chlorpyrifos oxon, mirror the broad effect spectrum observed in *in vivo* studies. However, our non-uniform data of scattered organophosphate effects across the different endpoints suggests that there is no common, overarching MoA, besides AChE inhibition, for DNT for this compound class (Figure 38).

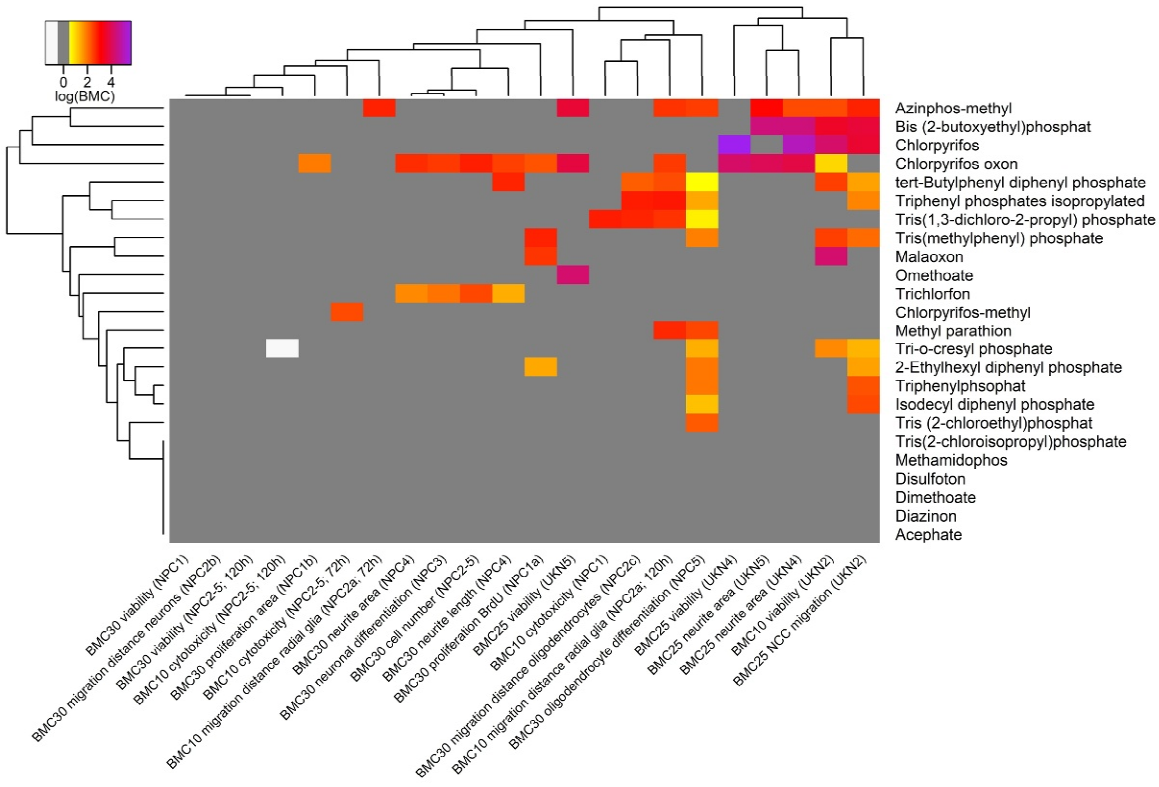


Figure 38: Organophosphate effects on NPC/UKN assays

Adding the rNNF assay to the NPC/UKN battery reduces the tested organophosphate compounds to eleven. Here, the rNNF assay as well as the NPC5 assay is altered by 8/11 compounds, followed by UKN2 (7/11; Figure 39). Interestingly, two of the organophosphates, i.e. tris(methylphenyl) phosphate and isodecyl diphenyl phosphate, altered UKN2 and NPC5 endpoints without affecting rat NNF. In contrast, chlorpyrifos oxon only affected rNNF without showing an inhibiting effect on UKN2 and NPC5. Chlorpyrifos alters rNNF and UKN2 without affecting NPC5 and tris(1,3-dichloro-2-propyl)phosphate affects rNNF and NPC5 without affecting UKN2. This supports the notion that multiple, yet unknown MoA drive neurodevelopmental toxicity of organophosphates.

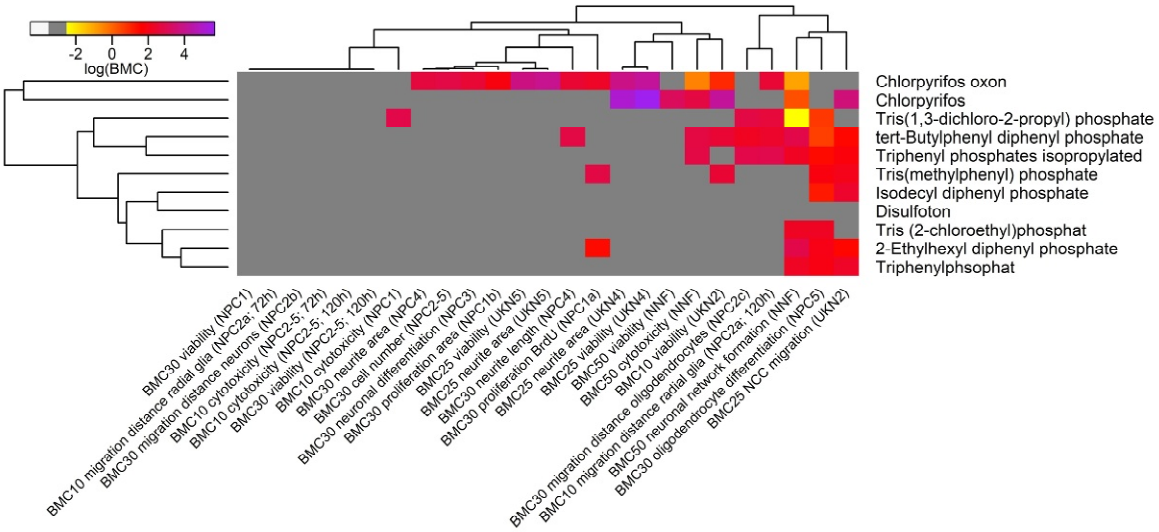


Figure 39: Organophosphate effects on NPC/UKN + NNF assays

Carbamates:

Four carbamates were tested in the battery including the NNF assay, thus only scenario (b) is discussed. What is striking in the DNT *in vitro* battery results is the great difference in assay sensitivity between the NNF assay and the NPC/UKN assays spanning two orders of magnitude for this compound class resulting in the MSE with one exception, mancozeb (negative in all assays but UKN5 with a very high BMC25 of 38 µM). Carbamates' primary neurotoxic MoA is inhibition of AChE, which, in contrast to organophosphates, is reversible. Therefore, carbamate acute neurotoxicity is generally less severe than acute organophosphate poisoning (summarized in Abreu-Villaça & Levin, 2017). This is not reflected in the DNT *in vitro* testing battery because carbamates alter NNF (MSE) at lower concentrations than organophosphates with the exceptions of chlorpyrifos-oxon and tris(1,3-dichloro-2-propyl)phosphate, where the MSE are in the same order of magnitude than carbamates (Figure 40). Hence, carbamates, similar to organophosphates, might act via alternative MoA. This notion was made earlier by a comparative *in vivo* study where the organophosphate chlorpyrifos and carbaryl induced DNT independent of AChE inhibition (Lee et al, 2015). Molecular studies using relevant *in vitro* methods or targeted *in vivo* investigations using the DNT *in vitro* testing battery as a guidance will shed more light on the underlying MoA. One carbamate, Mancozeb, did not affect NNF, but only altered UKN5 at fairly high concentrations (BMC25 38 µM).

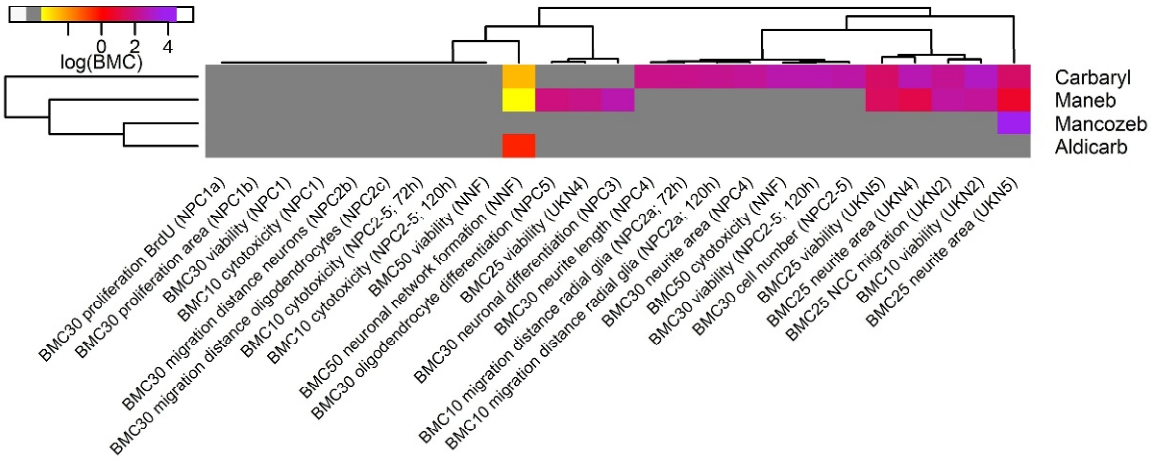


Figure 40: Carbamate effects on NPC/UKN + NNF assays

Pyrethroids:

The seven pyrethroids tested in the DNT *in vitro* battery display a more uniform pattern in the heatmap across all NPC/UKN assays than the organophosphates. Here, 6/7 compounds reduced the percentage of oligodendrocytes, yet to very different potencies (Figure 41). Deltamethrin is the only compound that alters more than one additional specific endpoint (NPC4>NPC2a>UKN2), yet at fairly high concentrations. Deltamethrin was published as a compound with effects in *in vivo* DNT studies (Mundy *et al*, 2015), supporting the battery data. MoA of deltamethrin is discussed in WP8 – Case Studies.

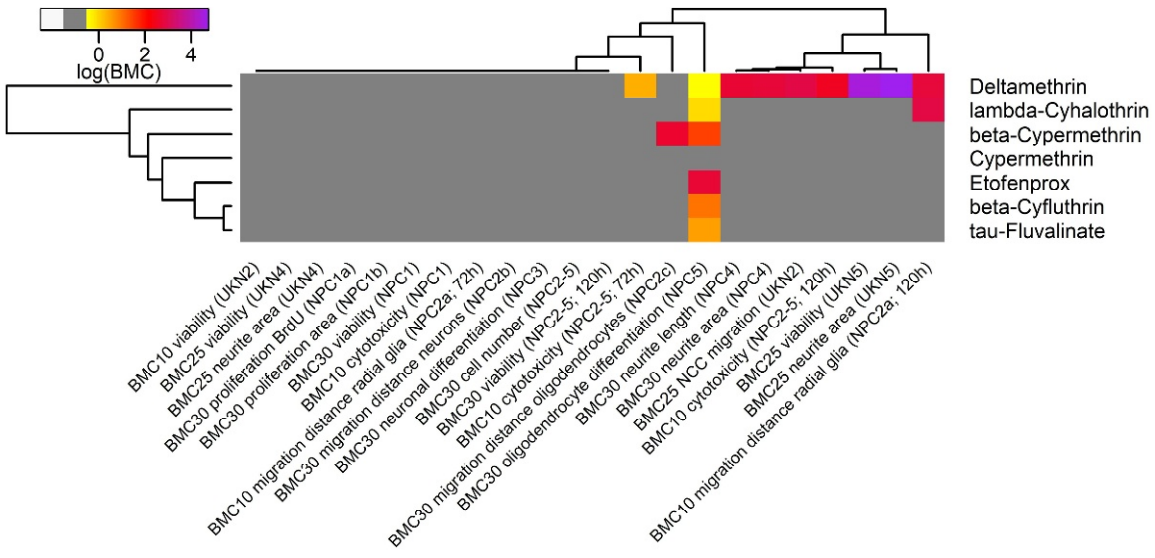


Figure 41: Pyrethroid effects on NPC/UKN assays

Adding the rNNF assay reduces the seven to only three pyrethroids tested. This sample size is rather small and thus not particularly representative. However, the observation that deltamethrin and cypermethrin reduced rNNF at fairly low concentrations and that only deltamethrin, but not cypermethrin reduced the oligodendrocyte numbers, is striking. Pyrethroids act via voltage gated sodium channels (VGSC). Such are present in membranes of neurons (Wang et al, 2017) as well as oligodendrocytes (Káradóttir *et al*, 2008; Black & Waxman, 2013; Livesey *et al*, 2016). Especially, O4+ oligodendrocyte precursor cells (OPC) are extremely sensitive to excitotoxicity (Volpe *et al*, 2011a), which they are predisposed to not only due to their NMDA and AMPA receptors (Volpe *et al*, 2011a), but also because they express active VGSC, which are down-regulated in mature oligodendrocytes (Paez *et al*, 2009). VGSC is a multi-subunit, transmembrane glycoprotein that contains an α subunit (220–260 kD) and one or more β subunits (33–36 kD). There are nine α and five β subunits. To the best of our knowledge, differential subunit expression between neurons and OPC have not been described but seem likely according to the DNT *in vitro* battery testing data (Figure 41). In addition, different pyrethroids seem to interfere with distinct subunits. In addition to – or as a result of - their action on VGSC, pyrethroids induce oxidative stress in brains of pre- or postnatally exposed rodents (summarized in Abreu-Villaça & Levin, 2017). Oligodendrocytes are specifically sensitive towards oxidative stress as they are poor in antioxidant defenses and their differentiation and maturation is inhibited by these reactive species (French *et al*, 2009; Roth & Núñez, 2016; Volpe *et al*, 2011a). Hence, two converging mechanisms explain the high sensitivity of oligodendrocyte differentiation or OPC, i.e. direct interference of pyrethroids with VGSC and production of oxidative stress. One compound, cypermethrin, strongly inhibited NNF with no effects on oligodendrocyte formation. The reason for this inconclusive result is not known and needs further investigation. For a more detailed understanding on pyrethroid effects on developing neurons and OPC, more in-depth work is needed. The outcome of exposure of the residual four pyrethroids on rat and eventually human NNF will be especially helpful.

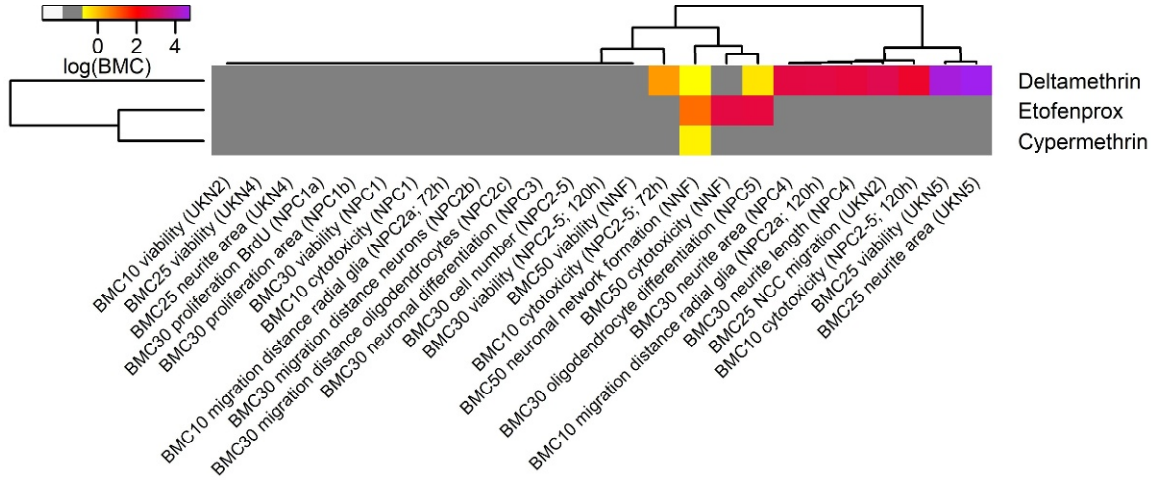


Figure 42: Pyrethroid effects on NPC/UKN + NNF assays

Organochlorines:

In the DNT *in vitro* testing battery scenario (a) six organochlorines were tested (Figure 43). Here, hexachlorophene is the most potent compound affecting almost all endpoints of the NPC/UKN assays. NPC5, oligodendrocyte differentiation, represents the MSE for hexachlorophene, however, the effect is

unspecific with a BMC of 0.07 µM. The lowest concentration where the effect can be differentiated from cell viability/cytotoxicity is 0.13 µM in the endpoints migration distance of radial glia (NPC2a).

In rodents, hexachlorophene produces a cytotoxic brain oedema and a consecutive vacuolation of white matter. This is caused by intracellular fluid influx and vacuolation, and in this respect is similar to the clinically important ischaemic brain oedema (Andreas 1993). It was suggested earlier that the spongy changes induced by hexachlorophene result from intramyelin accumulation of the fluid (Lampert et al, 1973; Towfighi, 1980) and cytotoxic insults to oligodendroglia (Kanno et al, 2012) *in vivo.*

Mechanisms associated with acute organochlorine toxicity involve blocking of sodium channels and γ-aminobutyric acid (GABA) receptors. Organochlorine-induced DNT, however, is thought to entail additional MoA including altered neurotransmitter levels and endocrine disruption (Abreu-Villaça & Levin, 2017). In the NPC/UKN assays, some sodium channels are expressed (for more details see Case Study B; Annex A6), i.e. with high probability in oligodendrocytes present in the NPC assay. These might be targets for organochlorines. In addition, proposed endocrine disruption of e.g. endosulfan involves thyroid (Schantz & Widholm, 2001) and estrogen (Briz et al, 2011) systems. However, neither thyroid hormones (besides NPC1a/b assay) nor estrogens are present in the cell culture media of the NPC/UKN assays making it unlikely that the observed effects occur due to endocrine disruption. Thus, the most likely mechanistic explanation for effects on the MSE oligodendrogenesis by organochlorine pesticides is the interference with sodium channels. However, more studies are needed to substantiate this hypothesis.

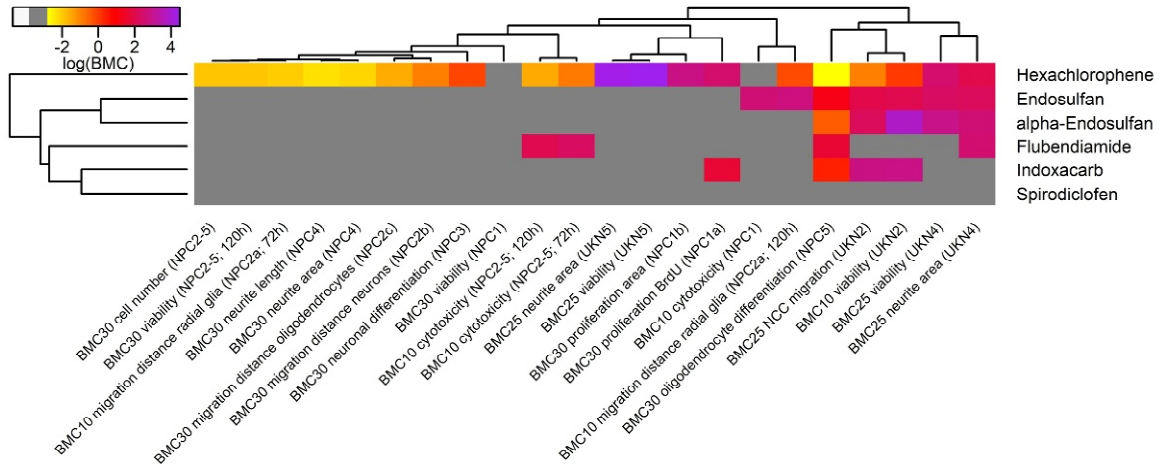


Figure 43: Organochlorine effects on NPC/UKN assays

Adding in scenario (b) the NNF assay to the organochlorine pesticide testing in NPC/UKN assays reduces the amount of tested compounds to three. This, again, is not a representative number. However, in these three commonly tested pesticides, the NNF assay does not add additional hazard information with regards to potency. Spirodiclofen is negative for the NNF assay like for the NPC/UKN assays. For endosulfan and hexachlorophene the NNF assay is less sensitive than oligodendrocyte differentiation or radial glia migration, yet lies in its 3-fold range (Figure 44**Error! Reference source not found.**).

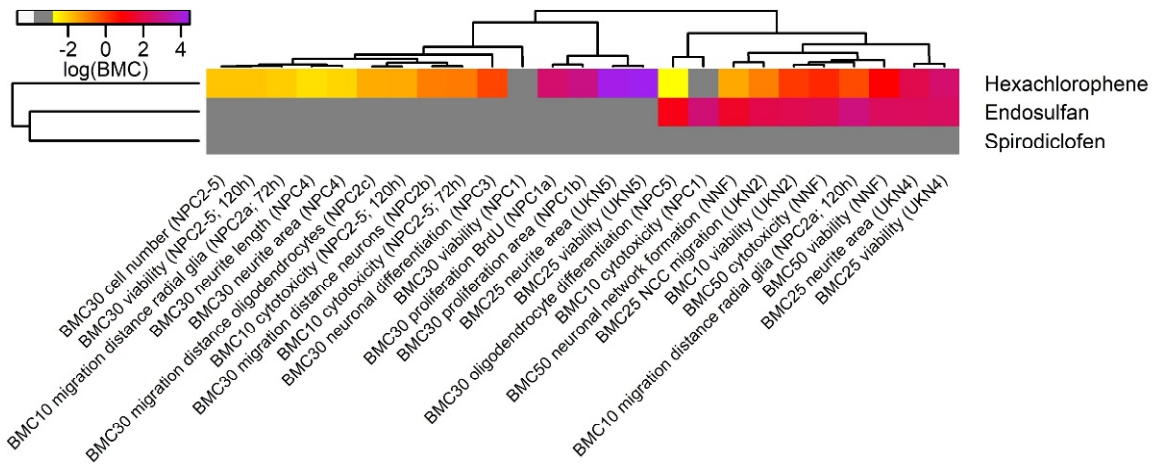


Figure 44: Organochlorine effects on NPC/UKN + NNF assays

Neonicotinoids:

Four neonicotinoids and the model compound nicotine were tested within scenario (a) in the DNT *in vitro* testing battery. Also here, effects are diverse. Dinotefuran, acetamiprid and nicotine exert no effects on any tested endpoint. Thiamethoxam has an unspecific effect on neurite length in the NPC4 assay without altering neurite area in NPC4, UKN4 and UKN5. Clothianidin reduces four endpoints, radial glia migration (NPC2a), oligodendrocyte differentiation (NPC5), neurite length and neurite area (NPC4), while imidacloprid acts only on radial glia migration (NPC2a) and oligodendrocyte differentiation (NPC5) with only NPC5 being specific (Figure 45).

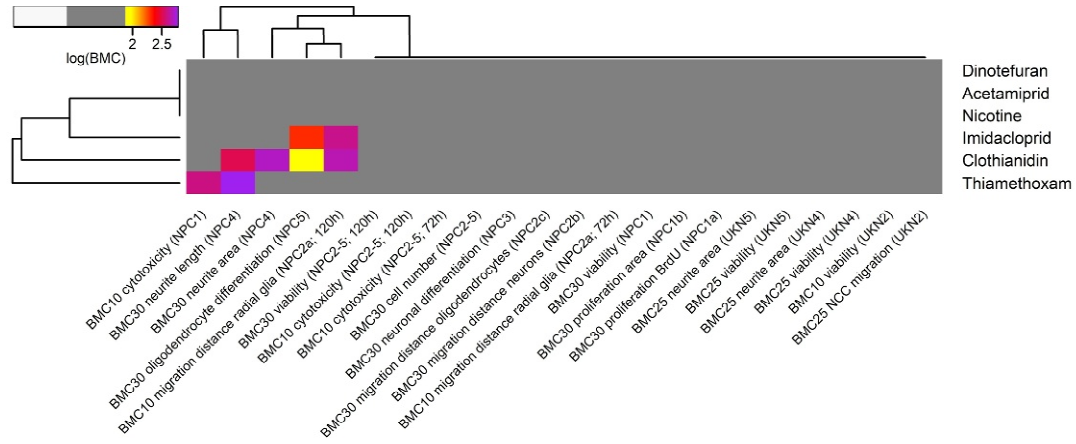


Figure 45: Neonicotinoid effects on NPC/UKN assays

Within scenario (b) the NNF assay supports the heatmap clustering of scenario (a) by adding effects of NNF for nicotine and imidacloprid to the data set (Figure 46). It is surprising that clothianidin and thiamethoxam both did not affect NNF, as the primary MoA for neonicotinoids is stimulation of nicotinic Ach receptors (nAChR) that are involved in neurotransmission. Neurodevelopmental studies with the nAChR ligand nicotine in rodents suggest that nAChR act beyond direct cholinergic neurotransmission. Their stimulation alters the proliferation/differentiation balance leading to a premature switch from replication to

differentiation (Slotkin et al, 1987). In addition, increased intracellular calcium due to nAChR stimulation provides a stop signal for neurite growth (Pugh & Berg, 1994). Nicotine stimulation of presynaptic nAChR enhances presynaptic release of Ach, dopamine, noradrenaline, serotonin, GABA and glutamate. Beyond neurons, also OPC express functional nAChR mediating intracellular calcium oscillations upon nicotine exposure *in vitro* (Rogers et al, 2001). However, the phenotypic consequence of OPC nAChR stimulation is not known. Some of these effects are reflected in the DNT *in vitro* battery, yet not in a consistent way across all compounds of this compound class. These data suggest that individual neonicotinoids might also act through off-target effects on distinct neurodevelopmental endpoints. A recent review article summarized *in vitro* and *in vivo* findings of neonicotinoids with regards to DNT (Sheets et al, 2015). More research is clearly needed to answer these and specific questions on neonicotinoids compounds' MoA.

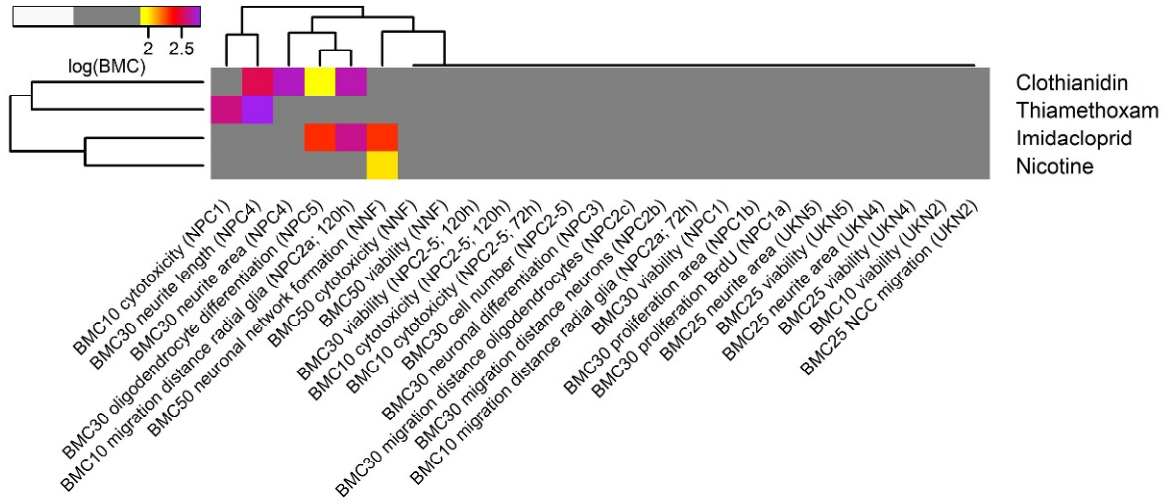


Figure 46: Neonicotinoid effects on NPC/UKN + NNF assays

Metals:

Of the nine metals shown within scenario (a) in the NPC/UKN assays, only six compounds were studied in all assays (Figure 47). The residual three are different forms of arsenicals and were included into the heatmap although not tested in all assays for completeness. With the exception of manganese(II)chloride, which tested negative in all NPC assays (and very high BMC_{25} of 29 - 50 µM in UKN2 and UKN5 assays) and lead (II) acetate trihydrate which tested positive in NPC5 (induction of oligodendrocyte formation) and also very high BMC_{25} in UKN2 and UKN5 (28 – 80 µM), most endpoints tested positive for all metals. The battery results reveal that the MSE across most compounds tested was reached by one of the UKN assays (2 x UKN2, 2 x UKN4, 1x NPC5 and 1x NPC4). Common to all the UKN assays is the single cell composition of NCC, neuronally differentiating hiPSC or LUHMES cells. Astrocytes can protect neurons from certain toxicities. This is a well-known concept that was exemplified by methylmercury *in vitro* and *in vivo* (Ni et al, 2012; Eskes et al, 2002) and possibly explains the higher sensitivity of the single cell cultures (UKN assays > NPC1 assay) compared to the mixed cultures including radial glia and astroglia (NPC2-5 assays). Thus, this compound class of metals nicely displays the different positive aspects of single as well as of mixed cell types *in vitro* cultures. Mixed cell types might converge toxicities and especially protective or toxicity-enhancing interactions of multiple cell types and thus reflect the *in vivo* situation more realistically.

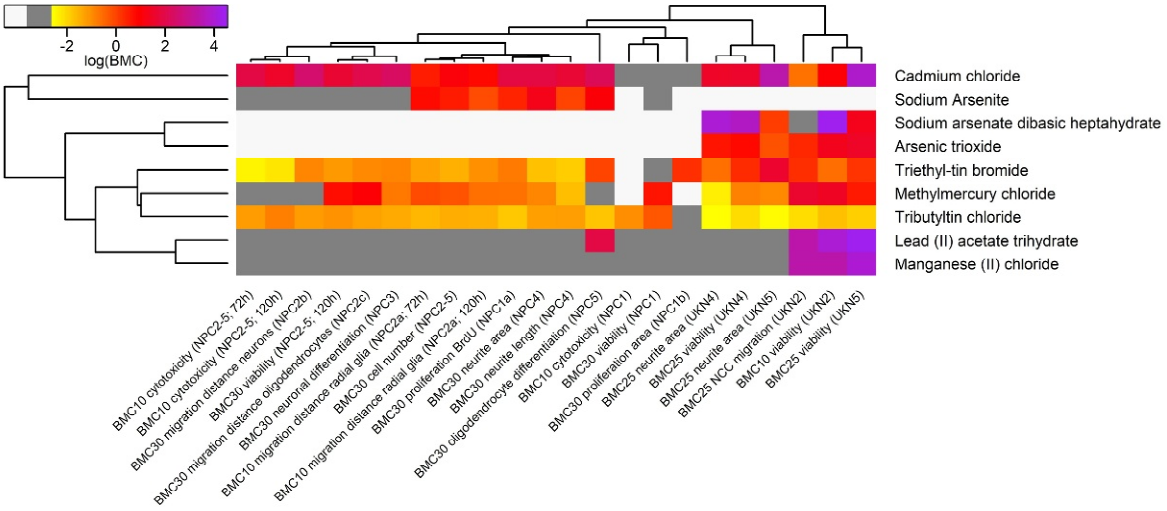


Figure 47: Metal effects on NPC/UKN assays

The addition of the NNF assay to the NPC/UKN battery adds significant hazard information data for metals to the overall outcome of the battery. In four of the five commonly tested metals the NNF assay is the MSE (with only the NPC4 assay in the 3-fold BMC range for triethyltin bromide; Figure 48). This is especially striking for lead(II)acetate. Lead specifically interferes with synapse development and function *in vitro* (Neal et al, 2010) and *in vivo* (Gąssowska et al, 2016; Nam et al, 2019; Tena et al, 2019). Exposure, effects and underlying MoA are comprehensively summarized by Neal & Guilarte (2010). Moreover, AOP13 was developed using lead as a model compound (https://aopwiki.org/aops/13; (Sachana *et al*, 2018). As lead exerts its DNT properties by interfering with signaling necessary for establishing synapses, the battery, and here especially the MSE NNF reflects well the *in vivo* situation and thus the specificity of the lead MoA. In comparison, e.g. methylmercury exerts much broader effects on different endpoints due to its ability to interfere with SH-groups of proteins (Bal-Price *et al*, 2015b) or sodium arsenite acting on developing brain cells via formation of reactive oxygen species (Masjosthusmann *et al*, 2019) reflected in the higher number of affected endpoints. In addition, this result strongly supports the necessity of the NNF assay to be part of the DNT *in vitro* testing battery.

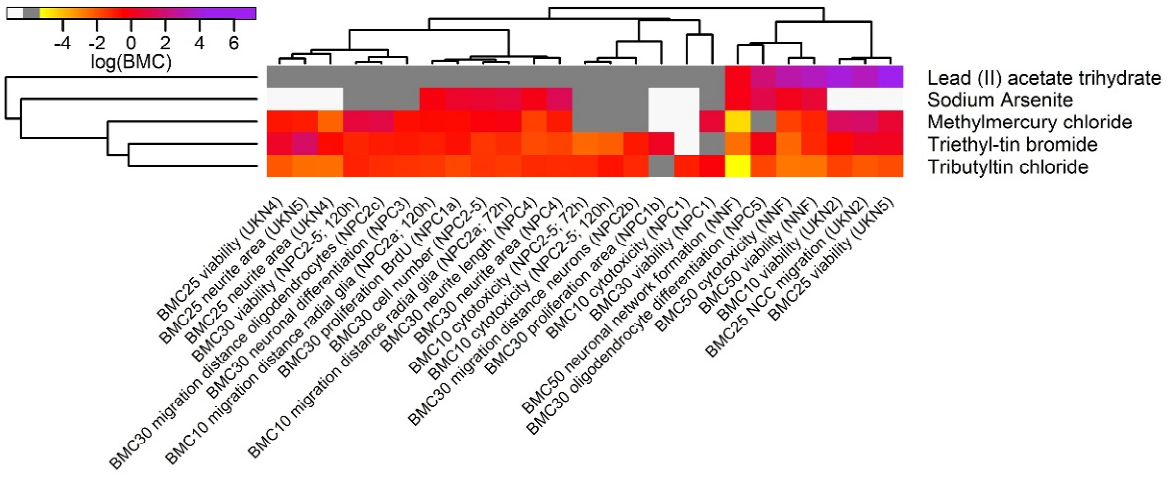


Figure 48: Metal effects on NPC/UKN + NNF assays

8.4.4. Sensitivity and specificity analyses

We first analyzed relative sensitivity of the assays in the battery (i.e. how many of all battery hits are covered by a given assay or subset of assays = "coverage"). For this, we took all data generated from UKN and IUF into consideration. Coverage assesses the sensitivity of the individual assays, i.e. how many of the overall battery hits are identified by an individual test. The identification of hits is detailed as "specific" identification (i.e. independent of cytotoxicity), as "borderline" identification (may be affected by cytotoxicity) and as "unspecific" identification (cytotoxic compound). The categories specific and borderline were also used in combination under the term "quasi-specific". These results are shown in Table 8.

Table 8 Performance analysis and sensitivity evaluation of NPC/UKN DNT in vitro test methods across the 119 compounds tested.

Assay endpoint	All hits	Quasi specific (specific + borderline)	Specific hits (specific)
	% coverage of respective assay in brackets		
All assays combined	78 (100)	67 (100)	66 (100)
All specific endpoints combined	74 (95)	67 (100)	66 (100)
All cytotox. combined	47 (60)	-	-
UKN2 (NC migr.)	37 (47)	35 (52)	29 (44)
UKN2 CT	33 (42)	-	-
UKN4 (neurite)	34 (44)	12 (18)	6 (9)
UKN4 CT	18 (23)	-	-
UKN5 (per. neurite)	24 (31)	8 (12)	7 (11)
UKN5 CT	18 (23)	-	-
NPC1a (prol. BrdU)	24 (31)	20 (30)	20 (30)
NPC1b (prol. area)	10 (13)	9 (13)	9 (14)
NPC1 CT30	3 (4)	-	-
NPC1 CT10	8 (10)	-	-
NPC2a72h (rad. glia migr)	17 (22)	11 (16)	11 (16)
NPC2a120h (rad. glia migr)	31 (40)	9 (13)	9 (14)
NPC2b (neuron migr)	7 (9)	0 (0)	0 (0)
NPC2c (oligo. migr)	17 (22)	7 (10)	7 (11)
NPC3 (neuronal diff)	19 (24)	12 (18)	12 (18)
NPC4 (neurite length)	24 (31)	12 (18)	12 (18)
NPC4 (neurite area)	19 (24)	8 (12)	8 (12)
NPC5 (oligo. diff)	49 (63)	42 (63)	42 (64)
NPC2-5 (cell no)	18 (23)	7 (10)	7 (11)
NPC2-5 72h CT10	14 (18)	-	-
NPC2-5 120h CT10	17 (22)	-	-
NPC2-5 120h CT30	13 (17)	-	-

Of the overall 78 hits in total and 74 hits in specific endpoints (of 119 compounds) in the battery, 60% of the 78 were covered if taking only cytotoxicity measures into account. Thus, cytotoxicity assays would have identified only a bit more than half of the compounds producing hits in the battery. Relying on cytotoxicity, even using complex, DNT-relevant test systems was obviously not sufficient to identify all relevant hazard.

Vice versa, all specific/functional endpoints (e.g. migration or neurite growth) in combination identified 95% of the hits in the battery. This indicates that the specific endpoints alone are almost sufficient as assay readouts, and that they are necessary to cover all potential DNT effects.

Next, we asked, how many compounds showed specific functional toxicity, independent of cytotoxicity. For this purpose, a classification model (CM) was used within each assay. For instance, this CM specified for the UKN5 assay that a hit was considered specific (independent of cytotoxicity) when cytotoxicity was triggered at 3-fold (or more) higher concentrations than the functional endpoint. We found that 66 compounds had a DNT-specific effect in at least one assay of the battery (Annex A2). When borderline compounds were included (quasi-specific), 67 of 78 compounds were defined as specific hit candidates. Thus, nearly all of the compounds identified somehow in the battery, showed at least one DNT-specific effect (only specific hits or specific and borderline combined).

On this basis, we examined the coverage of individual assays, with respect to this pool of hits. The most sensitive assays were NPC5 and UKN2, identifying 64% and 44% of the specific hits, respectively.

As NPC5 is not a stand-alone assay, but is multiplexed with NPC 2,3 and 4, it is interesting to look at the combined coverage of NPC2-5. The coverage of these four test methods was 76% with respect to specific hits) and 82% with respect to quasi-specific hits. This is meant to give an example of subsets of assays combined. Another example would be the combination of UKN2, UKN5 and NPC5. These three tests would have a coverage of 86% with respect to specific hits) and 81% with respect to quasi-specific hits.

That multiple tests are needed because not one test covers the whole spectrum of MoA is also demonstrated in the correlation graphs (Figure 57 and Figure 58). Testing across the whole battery produces correlations with the majority <0.5 for the different test methods.

In a next step, we analyzed battery performance, i.e. sensitivity and specificity of the DNT *in vitro* testing battery by relating categorized pre-defined compounds of the 119 chemicals to the actual battery outcomes. First, we studied the battery performance with regards to the negative compounds pre-classified in the substance set. These 'negatives' did not undergo rigorous review, yet were suggested in Aschner et al. (2017) with restricted documentation. Currently, the US-EPA is performing a thorough literature review on 'DNT negatives' that is currently not finished (Tim Shafer, personal communication). This 'DNT negatives' compound set will be very valuable for specificity analyses of *in vitro* data in the future. Table 9 summarizes the result for the 17 tested, pre-defined 'negative' substances across all assays (only 7 of those were also tested in the rNNF assay). There were two false positive results in the evaluation, chlorpheniramine maleate and doxylamine succinate (BMC 3.7 µM and 7.9 µM, respectively), both affecting NPC5 (oligodendrocyte differentiation). In addition, we identified two unspecific hits where cytotoxicity was not distinguishable from the specific DNT hits, here one in NPC5 and one in the rNNF assay. The rest of the 'negative compounds' tested negative in all other assays (Annex A1). Borderline cases were not observed. For calculation of performance, we identified the 'specific hits' as false positives (FP) and the 'no hits' and 'unspecific hits' as true negatives (TN). This leads to a specificity of **88.2%** for the basic battery (all 17 'negative compounds') and a specificity of **100%** when the rNNF assay is included (with only 7 of the 17 'negative compounds' tested in the rNNF assay).

Table 9 Performance analysis for substances categorized 'negative compounds' (n=17), that were pre-selected from the total of 119 compounds. These 'negatives' did not undergo rigorous review, yet were suggested in Aschner et al. (2017) with restricted documentation.

Chemical	Source for Selection	*In vitro* Hit	MSE	rNNF assay tested
Acetaminophen	Aschner et al., 2017	-	no hit	yes
Amoxicillin	Aschner et al., 2017	-	no hit	yes
Aspirin	ML prioritization	-	no hit	yes
Buspirone	ML prioritization	-	unspecific hit (NPC5)	no
Chlorpheniramine maleate	ML prioritization	+	NPC5	no
D-Glucitol	Aschner et al., 2017	-	unspecific hit (rNNF)	yes
Diethylene glycol	Aschner et al., 2017	-	no hit	yes
D-Mannitol	Aschner et al., 2017	-	no hit	yes
Doxylamine succinate	ML prioritization	+	NPC5	no
Famotidine	ML prioritization	-	no hit	no
Ibuprofen	Aschner et al., 2017	-	no hit	no
Metformin	ML prioritization	-	no hit	no
Metoprolol	Aschner et al., 2017	-	no hit	no
Penicillin VK	ML prioritization	-	no hit	no
Saccharin	Aschner et al., 2017	-	no hit	no
Sodium benzoate	ML prioritization	-	no hit	yes
Warfarin	Aschner et al., 2017	-	no hit	no

ML = Marcel Leist; MSE = most sensitive endpoint

The next pre-defined category was the human positive compounds. These are summarized in Table 10. All 9 human DNT positives tested positive in the DNT *in vitro* battery assays with the MSE scattered across the battery assays, meaning that each compound produced at least one specific hit. Hence, there were no false negatives observed. Considering not only hit/no hit, but including potency in the evaluation, three compounds had BMC values >5 µM (BDE-99, dexamethasone, Manganese (II) chloride), one around 1 µM (chlorpyrifos) and the residual <1 µM (Annex A1). Such are nominal medium concentrations, yet final predictive capacity will have less uncertainty when potencies and internal exposure levels are included into the assessments. However, with the tools we currently have at hand, sensitivity for the 'human positive compounds' is **100%**.

Table 10 Performance analysis for substances categorized 'human positive compounds' (n=9), that were pre-selected from the total of 119 compounds.

Chemical	Source for Selection	*In vitro* Hit	MSE	rNNF assay tested
Cadmium chloride	Mundy et al., 2015 // Aschner et al., 2017	+	UKN2	yes*
Chlorpyrifos	Grandjean & Landrigan, 2014	+	rNNF	yes
Dexamethasone	Aschner et al., 2017 // Mundy et al., 2015	+	NPC5	yes
Hexachlorophene	Aschner et al., 2017	+	NPC2a	yes
Lead (II) acetate trihydrate	Grandjean & Landrigan, 2006	+	rNNF	yes
Manganese (II) chloride	Grandjean & Landrigan, 2014	+	UKN2	yes*
Methylmercury chloride*	Grandjean & Landrigan, 2006	+	rNNF	yes
BDE-47	Grandjean & Landrigan, 2014	+	rNNF	yes
BDE-99	Grandjean & Landrigan, 2014	+	NPC5	yes

*different CAS was tested; data not included in this report

Besides the 'negatives' and the 'human positives', we also have the category of pre-defined 'compounds that show findings in *in vivo* DNT studies' in our 119 compound set identified in Aschner *et al* (2017) and Mundy *et al* (2015). Results from the performance analyses of this category containing 20 compounds are depicted in Table 11. Of the compounds tested in the UKN and IUF assays, 6 were not assessed in the rNNF assay. Five of the compounds (25%; ketamine, diphenylhydantoin, domoic acid, nicotine, terbutalin hemisulfat) were negative in all tested NPC/UKN endpoints, being 'false negatives' in the truncated battery, yet 2 of those were positive in the rNNF assay and 1 of those was not tested in the rNNF assay resulting in 3 false negatives in the extended battery (Annex A5 performance analyses). The other compounds produced at least one specific hit; borderline hits were not observed. The false negatives are probably attributed to gaps in the DNT *in vitro* battery and/or the truncated battery lacking the rNNF assay. For example, ketamine hydrochloride's MoA is based on interference with NMDA and opioid receptors. NMDA receptors are present in the test system used for the rNNF (or hNNF) assay, yet not in any of the other test systems (Annex A6). Therefore, any MoA involving NMDA receptors can only be assessed in the NNF assays. In few cases, also the concentration range tested might be too low. One example is valproic acid. Human internal exposure upon epilepsy treatment is higher than the 20 µM tested as the maximum concentration for most compounds in this test set. These instances are rare, but possible. However, as these reference data are compounds that were published as DNT *in vivo* positives in rodents (Mundy *et al*, 2015; Aschner *et al*, 2017), also species differences might account for false negatives in such an analysis.

Table 11 Performance analysis for 'compounds that show findings in in vivo DNT studies' (n=20) summarized by Aschner et al. 2017 or Mundy et al. 2015, that were pre-selected from the total of 119. The fact that findings were observed alone does not imply that the compounds were considered as DNT in vivo positives by authorities.

Chemical	Source for Selection	*In vitro* Hit	MSE	rNNF assay tested
(±) Ketamine hydrochloride	Aschner et al., 2017	-	no hit	yes*
5,5-Diphenylhydantoin	Aschner et al., 2017	-	unspecific hit (rNNF)	yes
Acrylamide	Mundy et al., 2015	+	rNNF	yes
all-*trans*-Retinoic acid	Aschner et al., 2017	+	rNNF	yes
Chlorpromazine hydrochloride	Aschner et al., 2017	+	NPC5	yes
Deltamethrin	Mundy et al., 2015	+	NPC5/rNNF/hNNF	yes
Domoic acid	Aschner et al., 2017	+	rNNF	yes
Haloperidol	Aschner et al., 2017	+	rNNF	yes
Maneb	Aschner et al., 2017	+	rNNF	yes
Methylazoxymethanol acetate (MAM)	Aschner et al., 2017	+	NPC1a	no
Nicotine	Aschner et al., 2017	+	rNNF	yes
Paraquat dichloride hydrate	Aschner et al., 2017	+	NPC4	yes*
PFOA (Perfluorooctanoic acid)	Aschner et al., 2017	+	NPC1	no
Potassium perfluorooctanesulfonate	Aschner et al., 2017	+	NPC3	no
Sodium valproate	Aschner et al., 2017	+	NPC5	yes
Tebuconazole	Mundy et al., 2015	+	rNNF	yes
Terbutaline hemisulfate salt	Aschner et al., 2017	-	no hit	yes
Tributyltin chloride	Mundy et al., 2015	+	rNNF	yes
Trichlorfon	Mundy et al., 2015	+	NPC4	no
Triethyl-tin bromide	Aschner et al., 2017	+	rNNF	yes

*different CAS was tested; data not included in this report

Combining information on specific DNT hits of the 'negatives', 'human positives' and 'compounds that show findings in *in vivo* DNT studies' (n = 46) tested in NPC/UKN assays, (Table 8), one finds 26 (56.5%) positive and 20 (43.5%) negative hits. Of those, 2 (4.4%) compounds were classified as false positives and 5 (10.9%) as false negatives resulting in 84.8% correct predictions (Annex A6).

This results in an overall **sensitivity** of the NPC/UKN test method battery **(true positives/true positives+false negatives) of 82.8%** and an overall **specificity (true negatives/true negatives+false positives) of 88.2%**.

Including the rNNF assay into this equation, one finds 19 (67.9%) positive and 9 (32.1%) negative hits of the 28 compounds tested. Of those, 0 (0%) compounds were classified as false positives and 2 (7.1%) as false negatives resulting in 92.9% correct predictions (Annex A5).

This results in an overall **sensitivity** of the NPC/UKN/NNF test method battery **(true positives/true positives+false negatives) of 90.5%** and an overall **specificity (true negatives/true negatives+false positives) of 100%**.

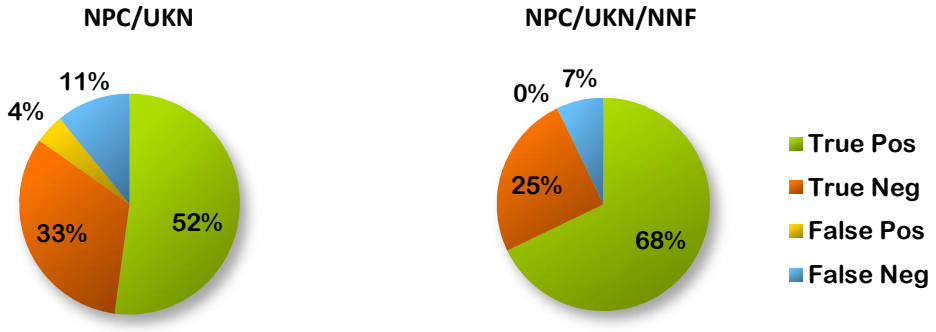


Figure 49: Performance analyses of the DNT *in vitro* testing battery for NPC/UKN and NPC/UKN/NNF assays using 46 and 28 compounds of the 119 compounds tested, respectively, that were classified as 'human positives', 'compounds that show findings in animal studies' and 'negatives'. Individual assessment of true and false positives as well as true and false negatives are depicted in Table 9 and in the Annex A5 (Performance).

Analyzing the performance of the *in vitro* test methods in relation to *in vivo* data is very difficult with the current knowledge base because there is not enough data on DNT in humans available. Another group of compounds, however, which might be helpful for performance analyses of DNT in vitro data in the future but was not considered for DNT in vitro battery performance for this procurement, consists of pesticides tested in DNT OECD or EPA guideline studies (unless used as a rationale in the Aschner et al. 2017 or Mundy et al. 2015 publications). There are mainly two reasons why these compounds were not used for battery performance analyses here, although some of those were part of the chemical test set. First, comprehensive and detailed information on substance effects in vivo is not publicly available. Here is a need for additional information gathering to fill scientific information and decision gaps of compound effects on the developing brain. For DNT in vitro battery interpretation and performance analyses, it is therefore necessary to understand any relationship to the in vivo effects observed in such guideline studies. Secondly, definition of a 'positive DNT finding' irrespective of final regulation by authorities is currently difficult. To demonstrate this issue, publicly available DNT-related effects on pesticides were collected from regulatory reviews of DNT guideline studies (Table 12, kindly provided by Kevin Crofton). Presented results illustrate that: 1) not all chemicals have been reviewed by all authorities; and, 2) for some compounds there are congruent conclusions by multiple agencies that *in vivo* DNT endpoints were affected, while for others there was disagreement on whether outcomes were effects. The lack of agreement between different authorities, coupled with the difficulty of access to regulatory reviews, makes any current comparisons between the in vivo DNT studies and in vitro battery results difficult. It is obvious that the assessment of the performance of the in vitro battery when using animal studies as a reference standard needs a detailed and uniform evaluation of the in vivo studies.

Table 12 Preliminary collection of DNT-related data on 'compounds that show findings in *in vivo* DNT studies' (+) and 'compounds that do not show findings in *in vivo* DNT studies' (-) in OECD or EPA DNT guideline studies (publicly available sources). The fact that findings were observed alone does not imply that the compounds were considered as DNT *in vivo* positives by authorities. Table kindly assembled by Kevin Crofton on expert judgment. ~ = DNT endpoints affected but agency did not conclude adversity.

Chemical Name	Australian PVMA	CalEPA	ECHA	EFSA	Food Safety Commission Japan*	Health Canada	JMPR	US EPA	Other
Acetamiprid				+	+		+	+	
Aldicarb	+							+	
Beta-cyfluthrin		-						+	
Carbaryl	-	~						+	
Chlorpyrifos	+	-						+	
Chlorpyrifos methyl				+				.	
Chlothianidin (TI 435)		-						-	
Cypermethrin zeta								-	
Deltamethrin		+						+	
Diazinon								-	
Dimethoate	+					+	-	+	
Dinotefuran		-					-	-	
Disulfoton								-	
Endosulfan	-	-						-	
Ethylenethiourea									
Etofenprox					+		+	+	
Fenamidone								-	
Fipronil	+	-						+	
Flufenacet (thiafluamide)								-	
Glufosinate ammonium							+	+	
Imidacloprid		+		+	+	+		+	
Malathion								+	
Mancozeb								-	-
Metaflumizone									
Methamidophos		-						+	
Methyl parathion		-						+	
Penthiopyrad (MTF-753)	+	~					+	+	
Perchlorate								-	
Pymetrozine		-						+	
Rotenone									
Spirodiclofen		~		-				+	

Thiacloprid (YRC 2894)								+	
Thiamethoxam		~	-		+			+	
Trichlorfon								+	
Trimethyltin								+	+

*from Toyhama, 2016

Another issue needing consideration are *in vivo* negatives. They have been used to calibrate assays and assay batteries in several toxicological domains. However, this principle cannot be easily transferred to DNT. There are several reasons, mostly linked to two areas: toxicokinetics and hazard ranking. Concerning toxicokinetics, major problems arise, when compounds do not cross the placental or fetal blood-brain-barrier or when they are quickly metabolized/inactivated by the mother. In such cases, they do not trigger DNT *in vivo,* although they may have the toxicodynamic (hazard) potential to do so. Concerning hazard ranking, compounds may be toxic to the mother or they may affect fetal development outside the CNS at lower concentrations than those required for DNT. In such cases, DNT *in vivo* would not be obvious (either not considered at all, or being neglected). Several more factors make *in vivo* anchoring problematic - for instance sensitivity of *in vivo* studies not being sufficiently high. For all these reasons, it is not scientifically helpful to determine NAM performance (specificity) by using *in vivo* negatives (only). A better approach is to carefully assemble a list of bona fide negatives. These may include some well-characterized *in vivo* negatives (as presented in Table 9). An important selection criterion for this list is toxicological evidence that there is no hazard of the compound to trigger DNT at the level of brain cells. For this reason, a hand-selected set of negatives has been used here for initial evaluation of the battery performance. At later stages, comparisons to *in vivo* negatives may be taken. However, this is not a step to be taken in a rush. As mentioned above, the US-EPA is performing a thorough literature review on 'DNT negatives' that is currently not finished (Tim Shafer, personal communication) but will eventually help increasing certainty in *in vitro* performance. For other classically validated assays, e.g. skin irritation or dermal sensitization, *in vivo* negatives have been used. However, in each case far beyond 100 compounds were available. In such cases, there is sufficient statistical power for such an approach and some classification errors may cancel out. For DNT, where only at a maximum 1-2 dozen compounds are available, this approach is statistically, and scientifically not justified, and considerable extra work needs to be invested into the classification of gold standard negatives. For these reasons paired with the uncertainty of calling a compound 'compound that does not show findings in *in vivo* DNT studies', we did not include the OECD/EPA 'negatives' (Table 12) into the performance analyses.

In conclusion, performance of the current DNT *in vitro* testing battery calculates to a **sensitivity of 82.8%** and a **specificity of 88.2% and** a **sensitivity** of **90.5%** and a **specificity of 100%** for the NPC/UKN and the NPC/UKN/NNF assays, respectively. For these analyses positive (Aschner et al., 2017; Mundy at al., 2015) and carefully selected negative compounds were employed. Increasing the number of compounds for performance analyses in the future is necessary. Here, transparent analyses of DNT *in vivo* guideline studies by different stakeholder, i.e. industry and regulatory agencies, would be very helpful. However, if gaps in the battery are filled in the future with further test methods, performance will improve, i.e. generating less false negatives. In addition, including kinetics in the analyses will make the battery more specific. Now there are hits in the battery at such high concentrations that will not have any toxicological relevance. Re-analyzing the data set with IVIVE will be very helpful in the future.

9. WP6: Information of the biological/toxicological application domain of the individual tests and the whole testing battery with gap analyses

Based on WP4 and WP5 as well as historical data of the test methods, the **(i) biological and (ii) toxicological application domains** of the individual assays and the whole battery are defined.

9.1. The biological applicability domain

For the biological applicability domain, the main question is: **which developmental processes and signalling pathways are reflected by the individual tests and thus the battery?** The neurodevelopmental processes covered by the individual assays of the DNT *in vitro* battery (**Error! Reference source not found.**) include NPC proliferation, migration of NCC, radial glia, neurons and oligodendrocytes, differentiation into neurons and oligodendrocytes, neurite outgrowth (length and area) and rNNF (the latter contributed by the US-EPA). In addition, these will be complemented by NPC proliferation, apoptosis, neurite outgrowth and synaptogenesis by testing of the same compounds by the US-EPA (data expected towards the end of 2020). Concerning the signaling pathways studied within this procurement, the compound selection was changed in the beginning of the project from the initial technical offer of testing pathway modulators and DNT compounds to only DNT compound testing (in communication with the US-EPA, to have the greatest compound overlap between all the assays in the battery). Therefore, information on the signaling pathways can only be derived from historical data of the responsible groups. In addition, the IUF is currently testing additional pathway modulators and this preliminary data will also be contributed (unpublished in Table 13). The information on signaling pathways modulating neurodevelopmental endpoints of individual test methods is summarized in Table B4 Annex B. This describes the so far tested biological applicability domain of the individual assays. The sum of pathways across all assays is then reflected by the battery and defines the biological applicability domain of the battery.

Table 13 Alteration of neurodevelopmental endpoints covered by DNT *in vitro* assays by signaling pathway modulators. All pathways are considered 'specific hits'. Pathways indicated by an asteriks were tested in all endpoints of NPC1-5.

DNT Assay	Signalling Pathway	Model Compound	Literature
NPC1	EGFR*	lack of EGF, EGFR antagonist	(Masjosthusmann *et al*, 2018)
	EP1-4*	PGE2	unpublished
	PDGFR*	CP-673451	
	CREB*	KG-501	
	mTORC*	MHY1485	
	AKT*	SC79	
	RHO*	Narciclasine	this procurement
	Electron transport chain/complex I*	Rotenone	
	Electron transport chain/complex I*	MPP+	
	PKC*	Bis-I	
NPC2	SRC	PP2	

	EGFR	AG1478, PD153035, EGF	(Moors *et al*, 2007; Masjosthusmann *et al*, 2018; Baumann *et al*, 2015)
	PKC	PMA/Bis-I	
	NO-cGMP	7-NI, ODQ, Rp-8-Br-cGMP	(Tegenge *et al*, 2011)
	ROCK	Y27632	
	WNT*	CHIR99021	unpublished
	RHO	Narciclasine	this procurement
	Electron transport chain/complex I	Rotenone	
	PKC	Bis-I	
NPC3	Notch*	DAPT	(Masjosthusmann *et al*, 2018; Baumann *et al*, 2015)
	EGFR	EGF, PD153035	
	RHO	Narciclasine	this procurement
	Electron transport chain/complex I	Rotenone	
NPC4	PKC	Bis-I	
	RHO	Narciclasine	
NPC5	BMP	BMP2, BMP7	(Dach *et al*, 2017; Masjosthusmann *et al*, 2018)
	?	Ascorbic Acid	
	Notch	DAPT	
	EGFR	PD153035	
	WNT	CHIR99021/IWP2	unpublished
	PPARγ*	Pioglitazone	
	COX-2*	Celecoxib	
	mTORC*	MHY1485	
	AKT	LY294002	
	PDGFR	CP-673451	
	PKC	Bis-I	this procurement
	RHO	Narciclasine	
UKN4	Microtubule/cytoskeleton	Colchicine	(Krug *et al*, 2013)
	RHO	Narciclasine	
	Electron transport chain/complex I	Rotenone	
	Protein homeostasis/protein synthesis	cycloheximide, brefeldin A	
UKN5	microtubule/cytoskeleton	Colchicine	(Hoelting *et al*, 2016)
	RHO	Narciclasine	
	electron transport chain/complex I	Rotenone	
	protein homeostasis/protein synthesis	Cycloheximide, brefeldin A	
Neurite outgrowth	PKC	Bis-I	(Druwe *et al*, 2016)
rNeuronal Network formation	PKC	Bis-I	(Brown *et al*, 2016)

Grouping the data now by pathways with the endpoints that each pathway alters provides an overview of specificity of pathways for neurodevelopmental processes. Most pathways are not very specific for individual processes. However, Table 13 and Figure 50 both demonstrate that multiple pathways converge on most neurodevelopmental endpoints, yet some, are specific. Also, individual pathways contribute to multiple neurodevelopmental endpoints producing some redundancy. The neurodevelopmental endpoint with the most converging pathways is oligodendrocyte differentiation, i.e. 12 different pathway modulators affect this process. Here 4 are specific modulators, not affecting any other endpoint.

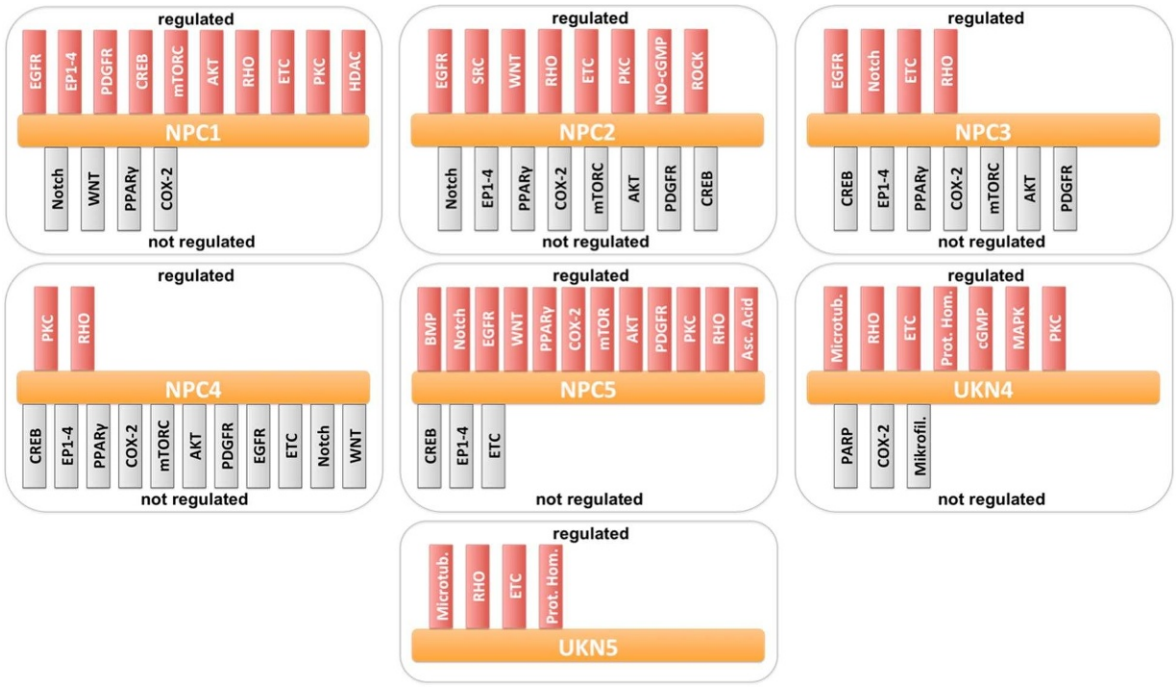


Figure 50: Outcomes for application domains of individual building blocks of the DNT *in vitro* testing battery. Orange: Test methods, Red: pathway active, grey: pathway inactive (either no response or not distinguishable from cytotoxicity).

Some pathways studied and identified as contributors to neurodevelopmental process development also play roles in neurodevelopmental disease including EGFR, FGFR, IGFR, PI3K, GSK3, AKT, mTOR, PLC, PKC, CREB, GMP-NO, RHO, NOTCH, WNT, PDGF, BDNF, TH and PGE_2 (Hevner, 2015; Bengoa-Vergniory & Kypta, 2015; Wong *et al*, 2014; Gyamfi *et al*, 2009; Dotti *et al*, 2014; Kang *et al*, 2016; Kiryushko *et al*, 2006; Lee, 2015; Lafourcade *et al*, 2013; Imayoshi *et al*, 2013; Ehrlich & Josselyn, 2016). Our data indicates that at least one, but generally more than one assay of the DNT *in vitro* testing battery is guided by each of these pathways. Hence, the DNT testing battery covers the biological application domain of these pathways well. In addition, a recent publication highlighted the role of BDNF for many DNT-related AOPs (Pistollato 2020). The described synaptogenesis assay in human iPSC-derived mixed neural cultures (e.g. established at the JRC) would hence be a valuable addition to the DNT *in vitro* battery.

Another set of molecules with implications for human neurodevelopmental disorders are ion channels. Mutations in ion channels produce channelopathies with distinct clinical outcomes. Current knowledge on channelopathies and underlying mutations was recently comprehensively summarized (D'Adamo et al. 2020). Clinical adverse outcomes (e.g. epilepsy, autism, intellectual disability) of channelopathies are caused by mutations that change ion channel/receptor activities during development. Also chemicals can interfere with such receptor functions. The DNT *in vitro* testing battery is able to assess adverse effects of

such chemicals on some channels/receptors relevant for human disease, i.e. sodium channels (organochlorines), GABAR (bicuculline), glutamate receptors (glutamate), while for other channels, i.e. potassium channels, calcium channels or hyperpolarization-activated cyclic nucleotide-gated channel 1, the applicability domain of the DNT *in vitro* battery still needs to be established.

Besides these signaling pathways, a compound class mediating cell-extrinsic signals that provide trophic functions during brain development is the class of neurotransmitters. Despite their classic role in neuronal communication by acting as synaptic chemical messengers in the mature CNS, such have gained attention as important factors to influence CNS development (Cameron et al., 1998; Nguyen et al., 2001; Ojeda & Ávila, 2019; Xing & Huttner, 2020). Neurotransmitters mediate developmental processes such as cell proliferation (Haydar et al., 2000), neuronal differentiation (Salazar et al., 2008), neuronal migration (Komuro & Rakic, 1993; Murthy et al., 2014), synaptic maturation (Fu et al., 2012), neurite growth (Anelli et al., 2013) and cell death (Ikonomidou et al., 2001). We have not studied the interference with trophic neurotransmission for neurodevelopmental processes in the here presented assays. This is a clear gap of the biological applicability domain of the DNT *in vitro* testing battery.

9.2. The toxicological applicability domain

For the toxicological applicability domain, there are three major questions: **(a) which types of chemicals can be tested? (b) which compound classes can be detected? and (c) which type of AOP key events can be assessed?**

Concerning (a), we displayed the MW and logP ranges covered by the set of chemicals (Figure 51). Most were within a MW range of 100-500, with exceptions ranging from 71 (Acrylamide) – 1008 (Emamectin benzoate). The logP values ranged from -15.2 (Sodium arsenate dibasic heptahydrate) – 8.7 (triphenyl phosphates isopropylated) with the majority of test compounds in the range of -2 – 7. The compounds were mainly pesticides, pharmaceuticals, industrial chemicals and flame retardants. The structural diversity spanned from inorganic compounds, and organometals to complex heterocycles. Types of chemicals that cannot or can only with large difficulties be tested *in vitro* are chemicals that do not dissolve in a solvent like DMSO or water as well as very volatile compounds. For the latter, there is the possibility of passive dosing in sealed wells. Yet, the test methods are not adapted for testing, using a passive dosing approach. Establishment of the assay set-ups for passive dosing will increase the chemical space that can be assessed with the DNT *in vitro* battery.

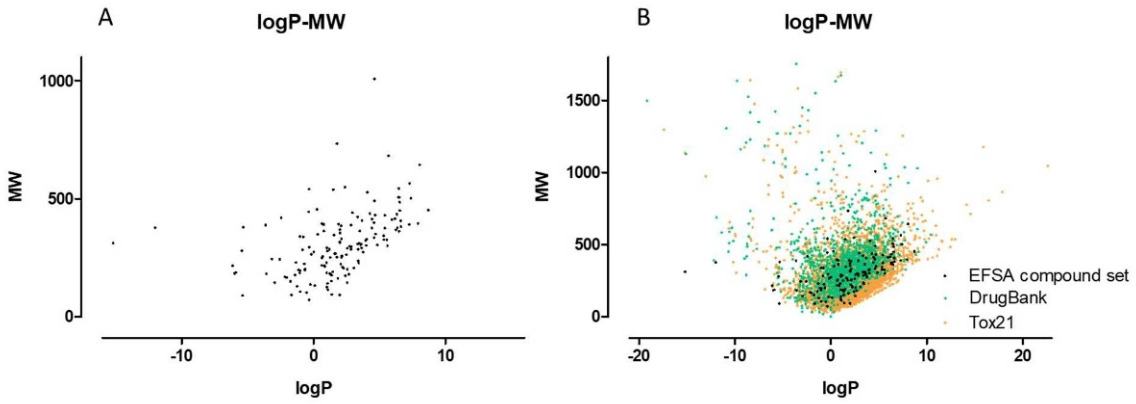


Figure 51: LogP-molecular weight scatter plots of the compound set in the EFSA project (A) compared to the Tox21 and DrugBank compound set (B).

Concerning (b), we identified compound classes that can be assessed with the DNT *in vitro* battery assays (see WP5b). From the 119 compounds tested in the battery, compound classes were chosen that are comprised of at least four individual compounds. Compound classes were identified that produced a relatively uniform or a dissociated pattern in the DNT *in vitro* battery responses. For the more uniform patterns, we distinguished carbamates, pyrethroids, organochlorines and heavy metals. Carbamates are a compound class that is identified by the battery, here the NNF assay is fundamental as it is two orders of magnitudes more sensitive than the NPC/UKN assays. Pyrethroids mainly affected oligodendrocyte formation and (in a smaller set of compounds) NNF. The scientific explanation for a possibly common MoA between these two endpoints might be the expression of VGSC in both cell types, which is discussed more in-depth within the case study on deltamethrin in WP8. Overall, the battery covers pyrethroids' MoA well. Organochlorines belong to another compound class acting via sodium channels. Again, effects on oligodendrocytes seem to be the most common endpoint across the tested organochlorines suggesting a similar MoA as for the pyrethroids. However, more data on more compounds is needed to substantiate a predictivity of the battery for this compound class. Also the kationic nature of heavy metals affects more (e.g. Pb) or less (e.g. Hg) specific effects on a variety of neurodevelopmental endpoints. Therefore, heavy metals add to the toxicological applicability domain of the battery. Compound classes that do not display a mainly uniform pattern across individual compounds include organophosphates and neonicotinoids. Here, MoA for DNT are not clear and might differ amongst the compounds of the same compound class.

In addition, Figure 31 and Figure 32 depict that some compounds' neurodevelopmental toxicities are specifically identified by certain test methods, i.e. representing cell types and/or neurodevelopmental processes. That the MSE profiles (represented by test methods within the 3-fold BMC of the MSE) of the individual compounds strongly differ across compounds shows the specific toxicological applicability domains of the assays and thus of the whole battery. Because a large variety of toxicological MoA can be assessed by the DNT *in vitro* battery as a whole, the battery will also be suited to identify the 'bad guys', e.g. for lead compound development. Excellent examples are here some metals like mercury or arsenicals, hexachlorophen, carbaryl, or chlorpyrifos oxon.

Concerning (c), we assembled the AOPs that are currently published in the AOP Wiki (Table 14). Here, AOPs were included that concern the developing and the adult nervous system. The rationale for including adult neurotoxicity AOPs is the high likelihood that neurotoxic MoA also affect the developing nervous system, i.e. interference with ion channels and neurotransmitter systems causing adult neurotoxicity will likely also affect brain development (Shah *et al*, 2018; Brummelte *et al*, 2017; Kirmse *et al*, 2018). The 13 AOPs identified in the AOP Wiki (Table 14) contain a total of 42 cellular key events (KE, Table 14). Some KE of DNT AOPs 17 and 13 are currently analyzed within the battery (Cell injury/death; Synaptogenesis, decreased; Neuronal network function, decreased). Key events of DNT AOP 8 are not applicable to the DNT battery as they take place in the liver. For DNT AOP 54 the cellular KE is currently not measured, yet could be (GABAergic interneurons, decreased), while AOP 152 cellular KE (neuronal T4 content) does not make sense to be tested outside of the *in vivo* context and KE 'hippocampal gene expression' cannot be measured with the battery. Over all, when KE of current DNT AOPs are compared to the neurodevelopmental KE of the DNT *in vitro* battery, the battery contains many more cell biologically relevant KE (e.g. NPC proliferation, migration, neuronal and glia differentiation, neurite outgrowth) than the AOP Wiki. DNT AOPs are currently not covering the biological space of neurodevelopmental processes necessary for brain development. It is thought that during this phase of DNT testing it is more likely that battery results inform new DNT AOPs than the AOPs informing the battery. Once more DNT AOPs are available, this will turn around.

Table 14 Adverse Outcome Pathways (AOP) concerning the developing or adult brain currently in the AOP Wiki (Dec 2019).

Nr	Id	Title	MIE	AO	OECD Status
1	26	Calcium-mediated neuronal ROS production and energy imbalance	unclear	energy imbalance	
2	17	Binding of electrophilic chemicals to SH(thiol)-group of proteins and /or to seleno-proteins involved in protection against oxidative stress during brain development leads to impairment of learning and memory	multiple	neuroinflammation leading to neurodegeneration	EAGMST Under Review
	8	Upregulation of Thyroid Hormone Catabolism via Activation of Hepatic Nuclear Receptors, and Subsequent Adverse Neurodevelopmental Outcomes in Mammals	multiple	neurodevelopmental outcomes	Under Development
4	205	AOP from chemical insult to cell death			
5	12	Chronic binding of antagonist to N-methyl-D-aspartate receptors (NMDARs) during brain development leads to neurodegeneration with impairment in learning and memory in aging	NMDAR	neuroinflammation and lead to neurodegeneration	TFHA/WNT Endorsed
6	48	Binding of agonists to ionotropic glutamate receptors in adult brain causes excitotoxicity that mediates neuronal cell death, contributing to learning and memory impairment.	Glutamate Receptor	learning and memory impairment.	TFHA/WNT Endorsed
7	10	Binding to the picrotoxin site of ionotropic GABA receptors leading to epileptic seizures in adult brain	GABA	leading to epileptic seizures	TFHA/WNT Endorsed
8	13	Chronic binding of antagonist to N-methyl-D-aspartate receptors (NMDARs) during brain development induces impairment of learning and memory abilities	NMDAR	impairment of learning and memory abilities	TFHA/WNT Endorsed
9	16	Acetylcholinesterase inhibition leading to acute mortality	AChE	acute mortality	Under Development
10	3	Inhibition of the mitochondrial complex I of nigro-striatal neurons leads to parkinsonian motor deficits	NADH-ubiquinone oxidoreductase (complex I), Binding of inhibitor	Motor function, impaired	TFHA/WNT Endorsed
11	54	Inhibition of Na+/I- symporter (NIS) leads to learning and memory impairment	NIS	learning and memory deficits in children	TFHA/WNT Endorsed
12	260	CYP2E1 activation and formation of protein adducts leading to neurodegeneration			
13	152	Interference with thyroid serum binding protein transthyretin and subsequent adverse human neurodevelopmental toxicity			Under Development

MIE – molecular initiating event; AO – adverse outcome

Table 15 Key Events (KE) of brain-related AOPs (only mentioned once each even if present in multiple AOPs)

AOP number	NT/DNT	KE	Measurable with DNT *in vitro* battery	Assay
26	NT	Increase, Ca++ (intracellular)	possible	
		Disruption, Mitochondrial electron transport chain	possible	
		Decrease, Mitochondrial ATP production	possible	
		Increase, ROS production	possible	
		Decreased, Nitric Oxide	possible	
17	DNT	Glutamate dyshomeostasis	possible	
		Cell injury/death	measured	NPC3, UKN4, UKN5, rNNF
		Neuroinflammation	-	
		Tissue resident cell activation	-	
		Increased pro-inflammatory mediators	possible	
		Neuronal network function, Decreased	measured	rNNF
8	DNT	n.a.	-	
205	NT	Direct mitochondrial inhibition	possible	
		Mitochondrial impairment	possible	
		Apoptosis	possible	
		Necrosis	measured	NPC3, UKN4, UKN5, rNNF
12	NT	Inhibition, NMDARs	possible	
		Decreased, Calcium influx	possible	
		BDNF, Reduced	possible	
		Neurodegeneration	measured	NPC3, UKN4, UKN5, rNNF
48	NT	Mitochondrial dysfunction 1	possible	
		Overactivation, NMDARs	possible	
		Increased, Intracellular Calcium overload	possible	
		Decreased, Neuronal network function (in adult brain)	possible	
10	NT	Reduction, Ionotropic GABA receptor chloride channel conductance	possible	
		Reduction, Neuronal synaptic inhibition	possible	
		Generation, Amplified excitatory postsynaptic potential (EPSP)		
13	DNT	Aberrant, Dendritic morphology	possible	
		Synaptogenesis, Decreased	possible	Pistollato et al. 2020
		Neuronal network function, Decreased	measured	rNNF
		Reduced, Presynaptic release of glutamate	possible	
16	NT	Accumulation, Acetylcholine in synapses	-	
3	NT	Inhibition, NADH-ubiquinone oxidoreductase (complex I)	possible	
		Impaired, Proteostasis	possible	
54	DNT	T4 in neuronal tissue, Decreased	-	
		GABAergic interneurons, Decreased	possible	
260	NT	Protein Adduct Formation	possible	
		Oxidative Stress in Brain	possible	
		Lipid Peroxidation	possible	
		Unfolded Protein Response	possible	

		General Apoptosis	possible	
152	DNT	Hippocampal gene expression, Altered	-	

NT – neurotoxicity; DNT – developmental neurotoxicity; measured – currently measured in the battery; possible – possible to measure in the battery with the current method set-up, yet this measurement is not performed; - not measurable.

AOP and KE analyses clearly indicate a few gaps of the current battery. These include quantification of synaptogenesis, neuronal subtypes (including GABAergic, glutamateric, cholinergic neurons, levels of neurotransmitters, receptor over-activation/inhibition, function of ion channels, etc.), or assessment of neuroinflammation, i.e. microglia/astrocyte activation. Moreover, BDNF expression seems to be a predictive endpoint that currently is not part of the battery, but is a relevant KE that could be incorporated into the battery.

In addition to the gaps identified amongst the KE in the AOPs, another gap in the battery is astrocyte differentiation, maturation and/or activation. Besides microglia, this cell type is clearly underrepresented within the test methods. Another battery gap is myelination. Less myelination can have different reasons including oligodendrocyte differentiation, maturation, myelin production and the complex final process of neurite wrapping with myelin sheets produced by mature oligodendrocytes. The battery contains the first cellular steps on oligodendrocyte formation, yet no direct axon myelination assay.

Hence, from the biological and current AOP status points of view, these test methods should eventually be added to the battery for covering a larger application domain with more possible MoA. An overview over the neurodevelopmental processes covered by test methods in the current battery and the gap identification is shown in Figure 52.

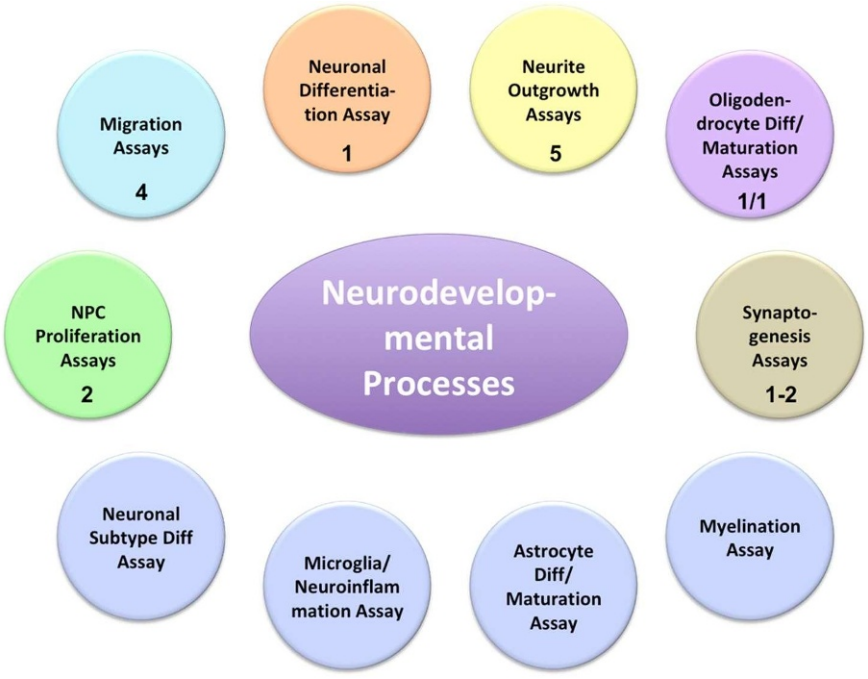


Figure 52: Gap analysis of the DNT *in vitro* testing battery. Overview over the neurodevelopmental processes covered by test methods in the current battery (number of test methods for each endpoint currently represented in the battery given in the respective circles). Gap analysis identified assays beneficial for the future battery. These developments envisioned for the future are shown in lavender. The synaptogenesis assay recently published (Pistollato et al. 2020) will be added to the battery, personal communication Anna Price, JRC.

Another important consideration, outside the AOP framework, is the role that kinetics of test compounds may play in neurodevelopmental toxicity. Here, maternal metabolism, crossing of parent compound/metabolite(s) the placental and (developing) blood-brain-barrier, and as a result finally internal fetal brain concentrations of parent compound/metabolite(s) need consideration (Figure 53). This information will transform the pure hazard information of the battery towards a risk-informed readout.

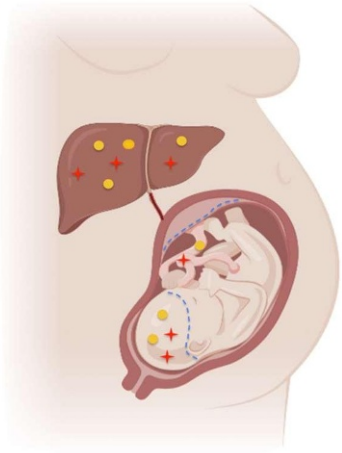


Figure 53: Schematic representation of internal exposure of fetal brain via the maternal blood stream. Parent compound is depicted as yellow circles, metabolites as red stars. Blue dashed lines indicate barriers (placental and (developing) blood-brain-barrier).

From hazard to risk?

This procurement focussed more on hazard identification than on toxicokinetic issues and risk assessment. Nevertheless, we layed down some principles to be considered for further evaluation and for follow-up work by specialists in the ADME and exposure fields. We also suggest that these principles be considered for further use of *in vitro* tests in regulatory contexts:

Principle 1: how to get from the points-of-departure (PoD) identified here to a safe daily dose or acceptable daily intake measure.

1. The free concentrations of chemicals in the assays need to be modelled. This requires information on protein and lipid content of the media, and on protein binding by the chemicals (ideally to be determined experimentally). An alternative starting point may be the predicted intracellular concentration of the chemical. This requires data on cell and culture volumes, and ideally also on cell composition. For both, the procedure has been exactly detailed in (Fisher *et al*, 2019).
2. In a next step, a generic IVIVE procedure may be used to determine the respective external dose. Multiple algorithms have been developed for this. One example is the NTP open-source IVIVE tool (https://ice.ntp.niehs.nih.gov/; (Bell *et al*, 2018). The principle of this tool is layed out in Figure 54. It needs to be noted that generic models have to make assumptions on liver metabolism rates and renal elimination rates. An important refinement is to input actual data on these processes to be determined experimentally.

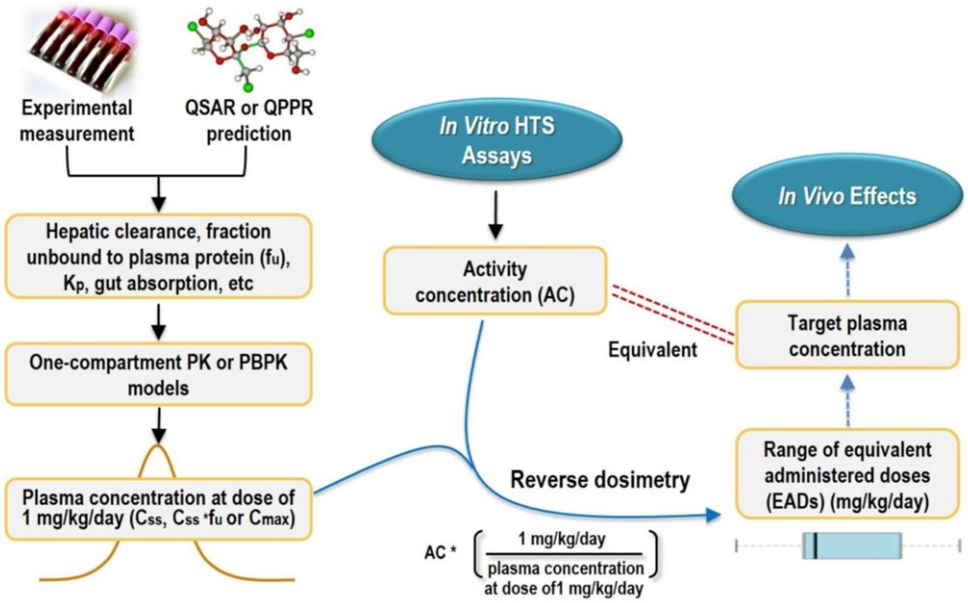


Figure 54: Predicting *in vivo* equivalent administered doses from *in vitro* activity concentration (from Chang et al, 2019).

3. Also, the generic modelling was developed for major target organs like the liver. For DNT, further modifications would be necessary. This refers mainly to the additional barriers across the placenta and the BBB. We suggest to make initial predictions of barrier crossing only considering passive diffusion across cell membranes. The next refinement would include QSAR models for pgp, the major transporter extruding chemicals from the brain. Higher levels of refinement would then consider a broader panel of transporters known to affect chemical transport across the BBB und the placenta.

In conclusion, some generic models may be used at present to obtain rough dose predictions, but they leave a high uncertainty if they are not refined by specific input parameters (transport across barriers and elimination rates in different compartments). At present, the data would therefore not be suitable to develop classification models based on hazard testing data only.

Principle 2: how to get from external doses to risk assessment

The safe threshold doses predicted from IVIVE modelling are compared here to exposure estimates. An ideal shortcut would be the availability of internal exposure data, or of biomarkers of exposure. Such data need to be obtained for various exposure scenarios and groups of the population with differing sensitivities. New large-scale exposome projects in the US and Europe may provide more and more reliable exposure data. At present there are large gaps and uncertainties. Especially for monitoring of the non-agricultural population, novel methods are required. One example may be to measure aggregated exposure over time in hair samples (Iglesias-González *et al*, 2020; Béranger *et al*, 2020).

In a generic process one would test whether there is an overlap of exposure estimates with predicted threshold dose estimates. If there is no overlap, or a separation by several log steps, then risk assessment is relatively easy, as risk is likely to be very low. If there is an overlap, then a generic solution is not possible anymore. Further work would require an exact problem formulation, which would then allow to define the relevant exposures and exposed populations for detailed modelling.

10. WP7: Development of a Draft Guidance for DNT Battery Usage

A **draft guidance** for *in vitro* DNT testing is assembled. This draft guidance focuses on five main aspects:

1. **How are the tests within a test battery and the battery itself used?**
2. **How is the data handled?**
3. **How are the results interpreted?**
4. **How do we arrive at relevant *in vivo* exposure levels?**
5. **How can be dealt with species-specific effects?**
6. **Uncertainties for regulatory use of data**

10.1. How are the tests within a test battery and the battery itself used?

Tiered Testing

There are different options for using the DNT testing battery. One scenario is a tiered testing strategy and the second one a parallel use of the test methods of the battery. A tiered testing strategy is a stepwise testing strategy where all existing information on a test substance is reviewed, in a specified order, using a weight-of-evidence process at each tier to determine if sufficient information is available for a hazard classification decision, prior to progression to the next tier. Such a procedure would imply possible sufficient information on DNT activity of compounds at each tier. In the current DNT *in vitro* testing battery, assays are used that are complementary to each other because they cover different neurodevelopmental key events (NDKE). A tiered testing strategy for using these *in vitro* assays does not have a biological rationale, as each of the NDKE is considered to be equally important for DNT; or in other words, any disturbance in each individual NDKE must be assumed to possibly lead to DNT. This logic suggests that none of the assays should be omitted from a first tier of testing and only be used at later tiers. This is supported by the testing data demonstrating that all assays appear as MSE in the test set (Figure 30). However, in real-world scenarios, non-scientific arguments also need to be considered. These may relate to cost, resources and throughput. Under such a perspective, it would make sense to run in a first tier one or two assays that are relatively cheap, or that have a high hit rate at an average cost. If a compound was tested positive, then it would be classified as DNT alert and no further testing may be required. If a compound was negative, then it would automatically be transferred to the next tier. Data from the smaller set of sixty compounds indicates that 80% of the positively tested substances in the DNT *in vitro* battery are identified as a positive hit by two of the battery assays, i.e. oligodendrocyte differentiation (NPC5) and rat neuronal network formation (NNF assay). Considering the 119 compounds without the NNF assay, almost 2/3 of the positive compounds were positively classified by NPC5 and NPC proliferation (NPC1a). Although the second scenario sounds like an economically attractive option, uncertainty will strongly increase by leaving out the residual assays of the battery.

A second situation, when a tiered strategy becomes beneficial arises as soon as *in silico*/*in vivo* approaches are added to the battery (Figure 55). Here, especially information on kinetics as a first tier could be of tremendous help. Such would include information on a compound crossing the placenta barrier, crossing the (maturing) blood-brain-barrier, or being secreted into breast milk and further taken up and distributed within the nursing child. Such a first tier would then either omit the need for further testing (if the compound does not reach the perinatal developing brain), or possibly give an internal exposure estimate (see WP6). This exposure estimate will facilitate data interpretation when compounds are tested in the *in vitro* battery

within Tier 2. This will take *in vitro* data from the testing battery beyond pure hazard identification. Here, no hazard observed in the human DNT testing battery will lead to the decision for no further testing. This decision on the rationale of an *in vitro* negative compound can only be made when the confidence in the battery is high enough that it sufficiently covers the chemical and biological space. In case an exposure-relevant hazard is observed in the battery, the next tier, Tier 3, will include another *in vitro* step. This time the compound is tested with test methods based on rat cells corresponding to the most sensitive, positive finding in the human *in vitro* assays. This step will facilitate data interpretation from *in vitro* to *in vivo* across species according to the parallelogram approach (Baumann *et al*, 2016; Lewandowski, 2003; Masjosthusmann *et al*, 2018, 2019). Moreover, it will ensure that the possible subsequent targeted *in vivo* testing in Tier 4 is likely to be reasonable with regards to toxicodynamics, which might differ between species (Gassmann *et al*, 2010; Baumann *et al*, 2016; Dach *et al*, 2017; Masjosthusmann *et al*, 2018, 2019). In case corresponding rat *in vitro* test methods do not show the effect seen in the human *in vitro* tests, there seems to be a species difference in toxicodynamics making subsequent *in vivo* testing illogical. In this case, kinetics *in vitro* and *in vivo* will help estimating if the human *in vitro* effect is relevant beyond hazard (see paragraph on IVIVE below).

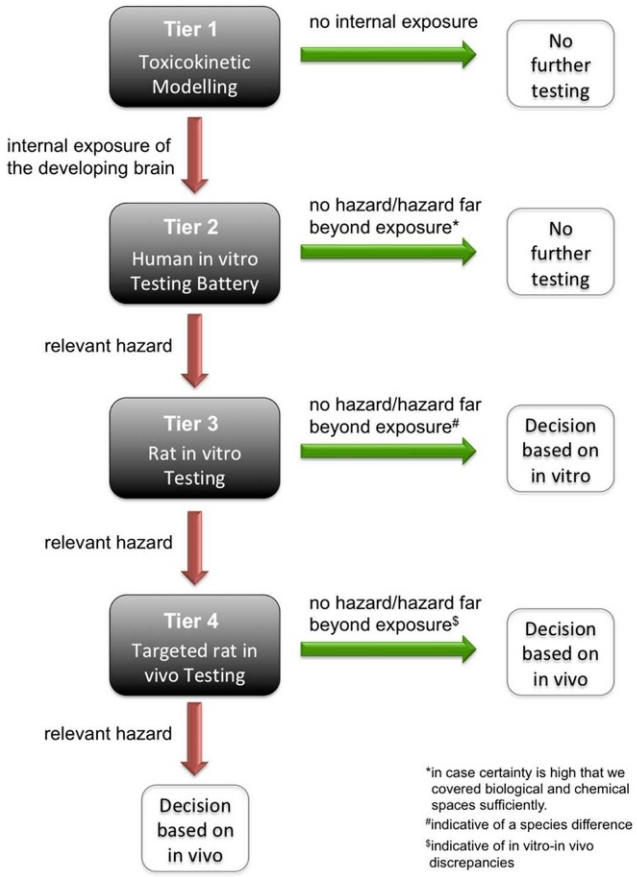


Figure 55: Proposed tiered testing strategy for DNT hazard and risk assessment.

Classification Models (CM)

The classification models include classifications for individual assays defining a compound's 'specific hit', 'unspecific hit', 'borderline hit' or 'no hit' for each test method and for the whole battery. How the classification models for the individual assays work is comprehensively described in WP5b (8.3.9). Once hits are identified, a classification model for the whole battery is applied (Figure 56). In a first step, BMC are compared across the testing results of the whole battery. Here, either only 'specific hits' can be considered, or 'specific' as well as 'unspecific' hits can be included into further considerations. The comparison of BMC including a 3-fold range of the most sensitive will lead to the identification of the MSE(s) for each compound. Behind each MSE of a test method there stands a neurodevelopmental process, which is per definition a KE of an AOP, irrespective if such an AOP currently exists or not. In case one of the MSEs can be linked to an existing AOP, it is possible to continue with an AOP-informed IATA. In addition to the identified hazard, now kinetics need to be considered and a potential risk for human exposure conditions evaluated (see 10.4 How to arrive at relevant *in vivo* exposure levels?). In the (predominant) cases where no AOP is available, *in vitro* testing results can initiate and guide AOP development. Here lies one of the great chances of such battery results for the field of regulatory science: filling the gaps of missing AOPs by inside-out AOP development. Start from the cellular KE and complete the AOP towards the MIE and towards the AO. This principle has been pursued earlier (Bal-Price *et al*, 2016).

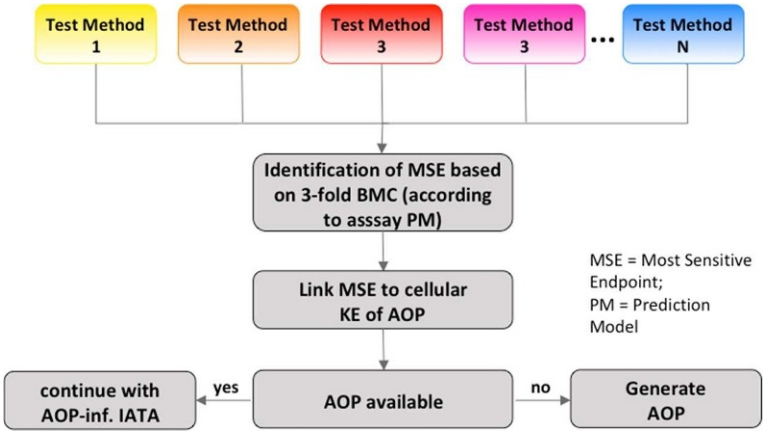


Figure 56: Proposed classification model for the battery. Test methods 1 - N stand for the assays of the DNT *in vitro* testing battery.

Similar Assays

Analyses of redundancy by BMC correlation

The testing battery contains assays covering different biological domains (e.g. proliferation and migration), but also related domains (e.g. migration of neural crest, of radial glia or of neurons). Figure 6 illustrates the different test methods covering similar neurodevelopmental endpoints. In addition, Table 2 - Table 5 describe the basics of the individual test methods for NPC proliferation, cell migration, neurite morphology and neuronal network formation in more detail with regards to test system, dimensionality, assay duration and method of endpoint assessment. Using the whole testing data set, we now asked the question if the assays of the DNT *in vitro* testing battery are redundant; i.e. whether they give results that are independent of one another. In other words, the analysis asks what the degree of redundancy is, and whether there are assays that may be fully overlapping. For this type of data exploration, we produced a matrix of all

BMC across assays and compounds tested in this study (see Annex A3). The cells in the table were filled in the following way. All compounds with a hit at below 20 µM were assigned their BMC value (hit concentration). All negative hits were arbitrarily assigned a value of 50 µM. All missing data were imputated as average of the table column. The cells were given weight factors of 1 for positive hits, of 0.1 for negative hits and of 0.01 for imputated data. Using this data basis (119 compounds x 22 endpoints = 2574 cells; 60 compounds x 25 endpoints = 1450 cells), all correlations between all assays were calculated (Pearson's method), and data were presented as correlation matrix (Figure 57 and Figure 58). In this matrix, the 22 (25) endpoints are listed both as x and y-axis. Thus, the matrix is diagonal symmetric, and each correlation value is represented twice for better visualization. For orientation, the diagonal is clearly seen as showing correlation = 1, as it stands for the correlation of each assay with itself. (yellow diagonal in Figure 57 and Figure 58). Due to covid-19, testing at the EPA was delayed and only a reduced number of compounds was tested. To account for this, a new table was established, containing only the compound that were tested in the rNNF at the EPA. Then again the correlation matrix was established (60 compounds), now containing 25 endpoints (i.e. including the NNF specific endpoint and the associated viability and cytotoxicity assays, Figure 58).

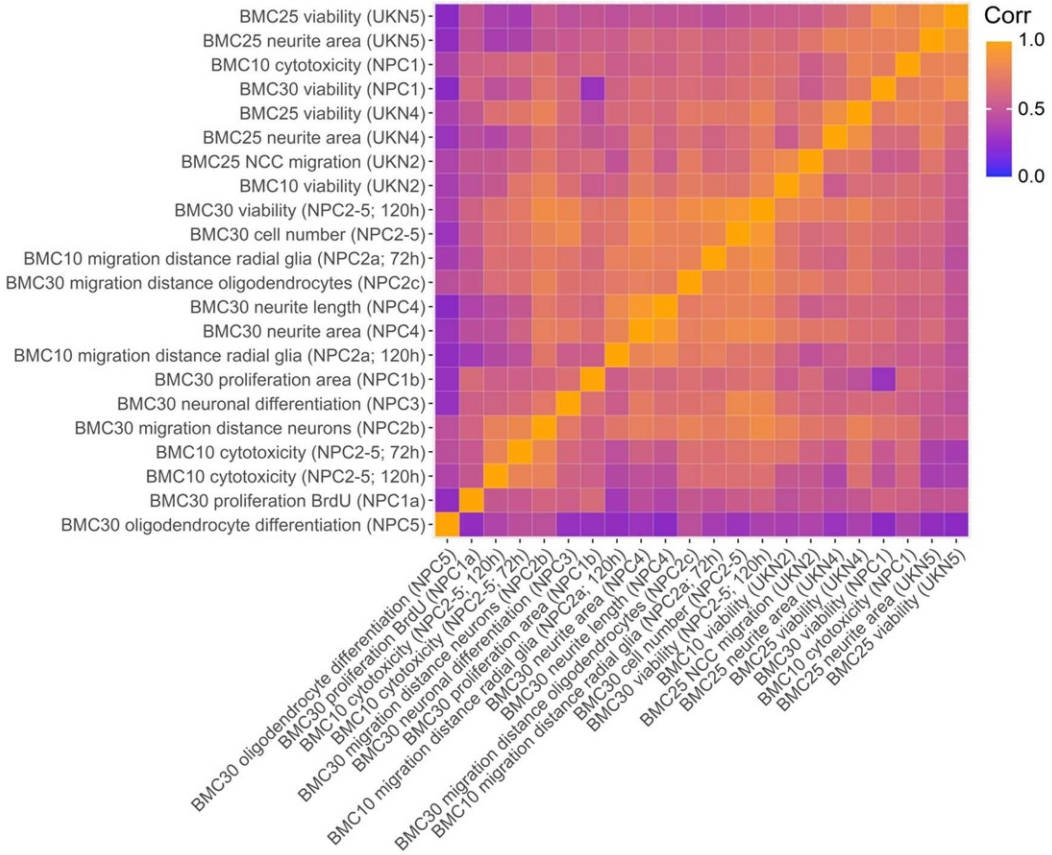


Figure 57: Correlation of test method results by comparison of assay-specific BMC across the whole chemical test set (n = 119) using NPC and UKN assays (22 endpoints). Correlation (Corr) of 1.0 (yellow) indicates 100% correlation, and 0.0 (purple) no correlation.

Discussion of correlation analysis

This correlation analysis correlates results of all test methods and viability/cytotoxicity assays for the 119 (NPC/UKN assays) or 60 (NPC/UKN/NNF assays) tested compounds with each other using a color code from purple (no correlation) to yellow (high correlation) representing correlation coefficients. A first visual inspection gives some interesting overview. First, none of the fields is as bright yellow as the diagonal indicating that no assays correlate in 100% with each other. Second, the rows/columns linked to NPC5 are most distinctly blue, i.e. this assay responded very different to the chemical set than all other assays.

A next level analysis identifies also some clear features: First, some endpoints display higher correlation indicated by brighter yellowish areas. These include UKN2,4 and 5 with their respective viability assays, NPC2-5 viability, cell number and radial glia migration (NPC2a) as well as NPC4 neurite length and neurite area. This is expected and speaks for consistency of the data. Second, a larger, more yellow block comprises several NPC2-5 endpoints. This is also consistent, as the assay endpoints are measures from the same test system, and the relatively high correlation is to be expected. Third, if one follows UKN4 (neurite area), there is a high correlation with UKN5. Both also correlate with NPC4 (neurite area). This is also expected, as all three are neurite assays. This second level analysis shows that data show some expected trends, but that assays are not simply overlapping, even if they address similar neurodevelopmental processes (but in different test systems).

A third level of analysis can look at the exact correlation numbers (n = 220 unique assay combinations). Here only examples are given. For instance, the four migration assays (Table 16) or neurite assays (Table 17) were partially, but not fully correlated. They are thus not redundant. The different cell types used e.g. in migration assays can explain these findings: while UKN2 assesses NCC migration, NPC2a-c measure radial glia, neuron and oligodendrocyte migration, respectively. In addition, cell and assay formats differ between the UKN and NPC assays: UKN2 studies migration after 24 h of compound exposure, while NPC2a-c analyze migration distances after 5 days. Moreover, single cell type culture of NCC in the UKN2 assay migrate into an empty space after a stamp removal, while NPC migration is a mixed radial glia/neuron/oligodendrocyte culture moving altogether radially out of a plated neurosphere (Table 3). That the latter three cell types do not respond similarly towards chemical insults despite their combined presence in the dish, supports the specificity of our findings.

Table 16 Correlation analyses of the migration assays

Assay 1	Assay 2	Correlation coefficient
BMC10 migration distance radial glia (NPC2a; 72h)	BMC30 migration distance oligodendrocytes (NPC2c)	0.784206395
BMC30 migration distance neurons (NPC2b)	BMC30 migration distance oligodendrocytes (NPC2c)	0.77369318
BMC10 migration distance radial glia (NPC2a; 72h)	BMC10 migration distance radial glia (NPC2a; 120h)	0.764627801
BMC10 migration distance radial glia (NPC2a; 72h)	BMC30 migration distance neurons (NPC2b)	0.735352141
BMC25 NCC migration (UKN2)	BMC30 migration distance oligodendrocytes (NPC2c)	0.725650108
BMC10 migration distance radial glia (NPC2a; 120h)	BMC30 migration distance oligodendrocytes (NPC2c)	0.724778697
BMC25 NCC migration (UKN2)	BMC30 migration distance neurons (NPC2b)	0.701020841
BMC10 migration distance radial glia (NPC2a; 120h)	BMC30 migration distance neurons (NPC2b)	0.695147246
BMC25 NCC migration (UKN2)	BMC10 migration distance radial glia (NPC2a; 72h)	0.605727072
BMC25 NCC migration (UKN2)	BMC10 migration distance radial glia (NPC2a; 120h)	0.478515457

Correlating the three neurite morphology assays, i.e. NPC 4, UKN4 and UKN5, reveals a similar picture (Table 17). Also, these data support the non-redundancy of these neurite morphology assays. This again might be explained with the different cell types and assay formats. While UKN4 and UKN5 are based on single neuronal cell types derived from hiPSC and LUHMES cells, respectively, NPC4 neurons are differentiated from primary human NPC and grow in a mixed-culture format with glia cells. In addition, compound exposure lasts for 24 h in the first and for 5 days in the latter assay. While differentiating neurons in NPC4 are not subtype-specified, UKN4 and UKN5 neurons are peripheral and dopaminergic neurons, respectively. The high correlation between NPC4 neurite length and NPC4 neurite area can be seen as a measure of consistency as data as these two neurite measures were generated from the same data of the NPC4 assay.

Table 17 Correlation analyses of the neurite morphology assays

Assay 1	Assay 2	Correlation coefficient
BMC30 neurite length (NPC4)	BMC30 neurite area (NPC4)	0.930375504
BMC25 neurite area (UKN4)	BMC25 neurite area (UKN5)	0.783632107
BMC25 neurite area (UKN4)	BMC30 neurite area (NPC4)	0.704651621
BMC25 neurite area (UKN5)	BMC30 neurite area (NPC4)	0.626592711
BMC25 neurite area (UKN5)	BMC30 neurite length (NPC4)	0.575298673
BMC25 neurite area (UKN4)	BMC30 neurite length (NPC4)	0.570412141

Besides specific correlations of similar endpoints, next, assay pairs are given that correlate well and show higher redundancy with a correlation coefficient of ~0.8 (Table 18).

Table 18 Correlation analyses of test methods: assays with higher correlation (>0.8)

Assay 1	Assay 2	Correlation coefficient
BMC30 neurite length (NPC4)	BMC30 neurite area (NPC4)	0.930375504
BMC30 cell number (NPC2-5)	BMC30 viability (NPC2-5; 120h)	0.920173209
BMC25 neurite area (UKN5)	BMC25 viability (UKN5)	0.89735506
BMC10 migration distance radial glia (NPC2a; 72h)	BMC30 viability (NPC2-5; 120h)	0.875046694
BMC25 neurite area (UKN4)	BMC25 viability (UKN4)	0.85875908
BMC25 viability (UKN5)	BMC30 viability (NPC1)	0.845931298
BMC30 migration distance oligodendrocytes (NPC2c)	BMC30 viability (NPC2-5; 120h)	0.836060859
BMC30 migration distance neurons (NPC2b)	BMC30 viability (NPC2-5; 120h)	0.835761782
BMC25 NCC migration (UKN2)	BMC10 viability (UKN2)	0.834956101
BMC30 neurite area (NPC4)	BMC30 cell number (NPC2-5)	0.827503894
BMC30 neurite area (NPC4)	BMC30 viability (NPC2-5; 120h)	0.827401208
BMC30 neuronal differentiation (NPC3)	BMC30 cell number (NPC2-5)	0.818602986
BMC10 migration distance radial glia (NPC2a; 120h)	BMC30 neurite length (NPC4)	0.816376331
BMC10 migration distance radial glia (NPC2a; 72h)	BMC30 cell number (NPC2-5)	0.812799608
BMC30 migration distance neurons (NPC2b)	BMC10 cytotoxicity (NPC2-5; 72h)	0.809334525
BMC30 neuronal differentiation (NPC3)	BMC30 viability (NPC2-5; 120h)	0.80726829

As another example, assay pairs are given that do not correlate well and are thus not redundant with a correlation coefficient of < 0.4 (Table 19).

Table 19 Correlation analyses of Test Methods: assays with low correlation (<0.4)

Assay 1	Assay 2	Correlation coefficient
BMC30 oligodendrocyte differentiation (NPC5)	BMC30 neurite length (NPC4)	0.229991699
BMC30 oligodendrocyte differentiation (NPC5)	BMC25 viability (UKN5)	0.231016332
BMC30 oligodendrocyte differentiation (NPC5)	BMC30 viability (NPC1)	0.231621923
BMC30 oligodendrocyte differentiation (NPC5)	BMC30 proliferation BrdU (NPC1a)	0.245128194
BMC30 oligodendrocyte differentiation (NPC5)	BMC25 neurite area (UKN5)	0.2454206
BMC30 oligodendrocyte differentiation (NPC5)	BMC10 migration distance radial glia (NPC2a; 120h)	0.254303807
BMC30 oligodendrocyte differentiation (NPC5)	BMC30 neuronal differentiation (NPC3)	0.267203854
BMC30 oligodendrocyte differentiation (NPC5)	BMC30 proliferation area (NPC1b)	0.27045702
BMC30 oligodendrocyte differentiation (NPC5)	BMC30 neurite area (NPC4)	0.276219184
BMC30 proliferation area (NPC1b)	BMC30 viability (NPC1)	0.277237207
BMC30 oligodendrocyte differentiation (NPC5)	BMC25 neurite area (UKN4)	0.284759648
BMC30 oligodendrocyte differentiation (NPC5)	BMC30 cell number (NPC2-5)	0.293143607
BMC30 proliferation BrdU (NPC1a)	BMC10 migration distance radial glia (NPC2a; 120h)	0.319489714
BMC25 viability (UKN5)	BMC10 cytotoxicity (NPC2-5; 72h)	0.335063122
BMC30 oligodendrocyte differentiation (NPC5)	BMC10 migration distance radial glia (NPC2a; 72h)	0.335371811
BMC30 oligodendrocyte differentiation (NPC5)	BMC10 viability (UKN2)	0.345586631
BMC25 neurite area (UKN5)	BMC10 cytotoxicity (NPC2-5; 120h)	0.346901744
BMC30 oligodendrocyte differentiation (NPC5)	BMC25 viability (UKN4)	0.355835507
BMC30 oligodendrocyte differentiation (NPC5)	BMC30 viability (NPC2-5; 120h)	0.356034343
BMC25 viability (UKN5)	BMC10 cytotoxicity (NPC2-5; 120h)	0.36022318
BMC30 oligodendrocyte differentiation (NPC5)	BMC10 cytotoxicity (NPC1)	0.372635053
BMC25 neurite area (UKN5)	BMC10 cytotoxicity (NPC2-5; 72h)	0.374200088
BMC30 oligodendrocyte differentiation (NPC5)	BMC25 NCC migration (UKN2)	0.381509829
BMC30 oligodendrocyte differentiation (NPC5)	BMC10 cytotoxicity (NPC2-5; 120h)	0.387055422
BMC30 proliferation BrdU (NPC1a)	BMC30 neurite length (NPC4)	0.38931597

The high redundancy between cell number NPC2-5 and viability NPC2-5 (Table 18) makes sense, as a reduction in viability must lead to a reduction in cell number in the migration area. Therefore, one might possibly assess neurosphere viability by quantifying nuclei number. Two more assays that have high correlation are NPC2a (radial glia migration) and NPC2-5 viability. This is not due to the fact that cells do not migrate due to cytotoxicity, but that viability is low because of a reduced migration area and thus less

cells exposed to the cell titer blue reagent. This was previously shown by adding the LDH assay to the portfolio and by studying cell viability with FACS analyses (Nimtz *et al*, 2019). Hence, the NPC2-5 viability assay cannot replace NPC2a or vice versa because a reduction in NPC2-5 viability can be due to a reduced migration area or because of reduced cell viability. Therefore, we always need information on cytotoxicity in addition, which helps distinguishing between the first and the latter.

One assay has an extremely poor correlation with all other assays that is NPC5, oligodendrocyte differentiation. This indicates that oligodendrocytes strongly differ in their chemical sensitivity from neurons and have their own toxicological applicability domain. This is supported by the assessment of the biological applicability domain of the NPC5 assays. The profile of signaling pathways contributing to oligodendrocyte differentiation strongly differs from the profiles of the other endpoints. Hence, oligodendrocyte differentiation is an asset of the DNT *in vitro* testing battery. Similarly, this holds true for the NPC proliferation assay, i.e. NPC1a (Figure 57). Over all, one can conclude from the redundancy analysis, that there is sufficient and necessary individuality across the assays and thus the present DNT *in vitro* testing battery is the minimal requirement test battery for DNT evaluation.

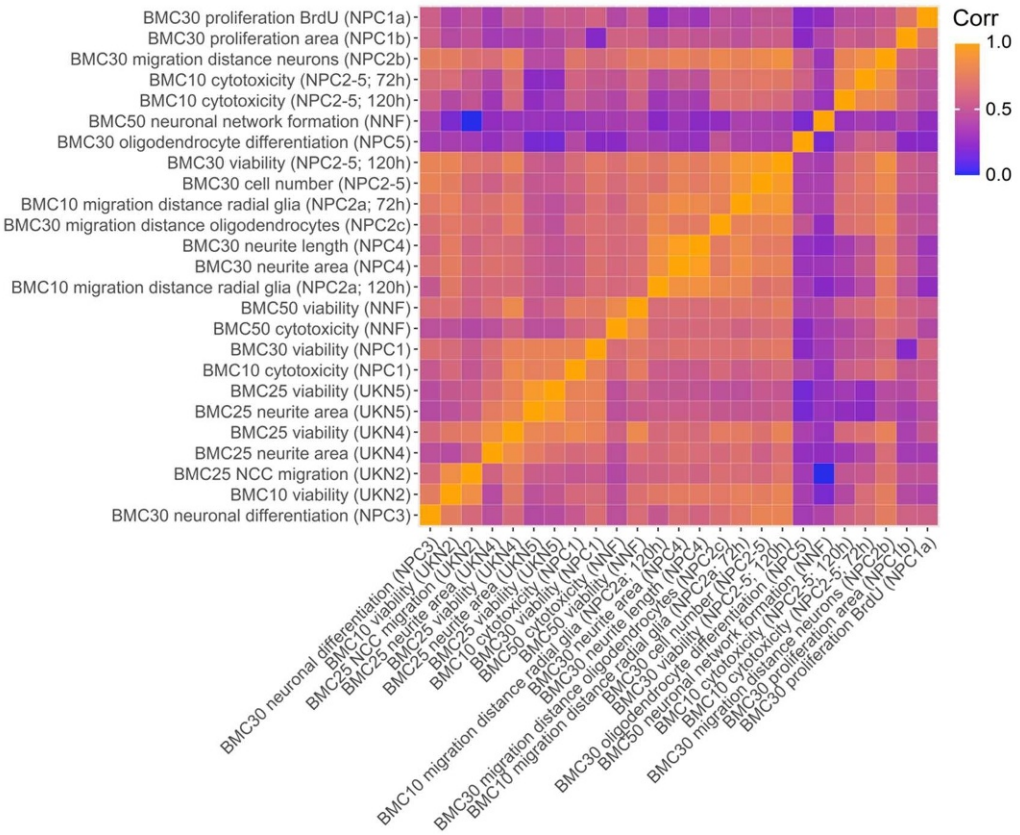


Figure 58: Correlation of test method results by comparison of assay-specific BMC across the reduced chemical test set (n = 60) using NPC, UKN and rNNF assays (25 endpoints). Correlation (Corr) of 1.0 (yellow) indicates 100% correlation, and 0.0 (purple) no correlation.

Adding the rNNF assay to the correlation table further substantiates the previous observation. The rNNF is not redundant to the NPC/UKN assays (correlation coefficients ranging from ~0.1 to ~0.3 for many of the assays). These correlation analyses are supported by the PCA analysis (Figure 59). Especially, NPC5, rNNF and the two migration assays, i.e. UKN2 and NPC2a indicate separate clusters.

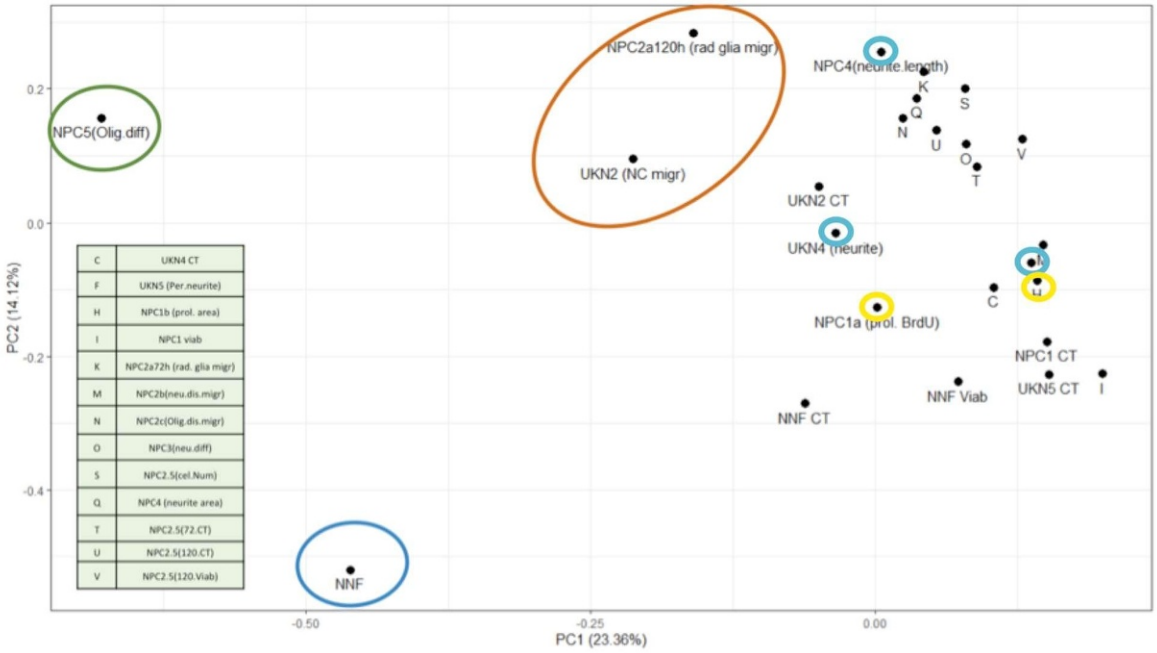


Figure 59: Principal component analysis of the 25 assay endpoints. The data matrix from Annex A3 was used for a principal component analysis of the different endpoint (each represented as a 119-dimensinal data vector). The first two principal components are displayed. This analysis shows that oligodendrocyte differentiation (green) clearly clusters away from rNNF assay (blue) and all other endpoints. These 'other endpoints' contain specific as well as cytotoxicity/viability measures and cannot be considered a cluster. Two migration assays (UKN2 and NPC2a) can also be assigned to a cluster (orange). Non-clustering neurite morphology assays are circled in turquoise. Proliferation assays are circled in yellow.

10.1.1. Analyses of redundancy by direct test method comparison

Next, we analyzed the performance of the 4 assays (UKN4, UKN5, NPC4 neurite length, NPC4 neurite area) assessing neurite morphology in more detail. Table 20 demonstrates that 43 compounds (36% of all compounds) produced 55% of all hits in the battery.

Specifically, 11 compounds were only recognized by UKN4, 3 only by NPC4 neurite length and 2 substances only by UKN5. This demonstrates the non-redundancy of the assays. Leaving out one of the assays would reduce information. However, with NPC4 we generate information on neurite length and neurite area. Here, neurite area seems to be redundant as all hits on neurite area also affect neurite length. Amongst the compounds, 13 affect all three assays, UKN4, UKN5 and NPC4. These are only 30% of the 43 compounds affecting neurite morphology and therefore further support the non-redundancy of the assays. After application of the classification model, 3 compounds are classified as "specific hit" in UKN4, UKN5 and NPC4. These are methylmercury chloride, narciclasine and rotenone. Of the 43 compounds, 15 substances affect neurite morphology with BMC <10 µM. Neurite morphology belongs only in 13 out of 74 positive hits of the 119 compounds (scenario (a)) and in 4 out of 48 positive hits of the 60 compounds (scenario (b)) to the MSE thus demonstrating its importance for the overall battery output. Again, different cell types and assay formats probably determine individual assay sensitivity (see above). Especially the importance of the UKN4 dopaminergic nature is shown by its sole sensitivity to MPP+, a model compound inducing parkinsonian symptoms also *in vivo.*

Table 20 Comparison of neurite morphology assays for the 43 out of the 119 compounds producing hits in at least one of the assays (µM).

Nr	compound name	CAS #	Selection criteria	UKN4 neurite area	UKN5 peripheral neurite area	NPC4 neurite length	NPC4 neurite area
3	1-Naphthol	90-15-3		82.6	82.7	14.1	no hit
5	3,3',5,5'-Tetrabromobisphenol A	79-94-7	in vivo data available	no hit	no hit	3.3	3.4
16	Acrylamide	79-06-1	in vivo positive	no hit	1163	no hit	no hit
19	all-*trans*-Retinoic acid	302-79-4	in vivo positive	15.4	no hit	18.1	13.7
21	alpha-Endosulfan	959-98-8		13.4	no hit	no hit	no hit
28	Azinphos-methyl	86-50-0		10.8	21.7	no hit	no hit
35	Buspirone	36505-84-7	negative control	14.3	no hit	no hit	no hit
37	Cadmium chloride	10108-64-2	human positive // in vivo positive	4.6	30.6	5.5	6.7
41	Carbaryl	63-25-2	in vivo data available	15	5	8	10.3
44	Chlorpromazine hydrochloride	69-09-0	in vivo positive	5.7	3.1	4.9	5.1
45	Chlorpyrifos	2921-88-2	human positive // in vivo data available	164.4	no hit	no hit	no hit
46	Chlorpyrifos oxon	5598-15-2	metabolite/related	47.3	55.4	11.9	14.6
48	Clothianidin	210880-92-5	in vivo data available	no hit	no hit	12.3	14.8
59	Deltamethrin	52918-63-5	in vivo positive	no hit	112.8	14.9	15.9
73	Emamectin benzoate	155569-91-8	in vivo data available	1.3	0.9	3.3	3.9
74	Endosulfan	115-29-7	in vivo data available	9.1	no hit	no hit	no hit
75	Endosulfan sulfate	1031-07-8		28.9	no hit	no hit	no hit
81	Fipronil	120068-37-3	in vivo data available	12.2	no hit	12.8	13.7
82	Flubendiamide	272451-65-7		12.5	no hit	no hit	no hit
87	Glycerol	56-81-5	negative control	13.2	no hit	no hit	no hit
88	Haloperidol	52-86-8	in vivo positive	13.7	12.6	12.2	15.3
91	Hexachlorophene	70-30-4	human positive	7.4	75.9	0.1	0.1
103	Mancozeb	8018-01-7	in vivo data available	no hit	38.5	no hit	no hit
104	Maneb	12427-38-2	in vivo positive	3.2	2	no hit	no hit
115	Methylazoxymethanol acetate (MAM)	592-62-1	in vivo positive	505	366	no hit	no hit
116	Methylmercury chloride*	115-09-3	human positive	0.1	0.4	0.2	0.4
120	MPP+	36913-39-0	metabolite/related	1.5	no hit	no hit	no hit
122	Narciclasin	29477-83-6	model compound	0.014	0.021	0.013	0.011
127	Paraquat dichloride hydrate	1910-42-5	in vivo positive	115	130	16.1	no hit
130	PBDE 47	5436-43-1	human positive	21.6	85.1	no hit	no hit
131	PBDE 99	60348-60-9	human positive	13.4	no hit	no hit	no hit
134	Penthiopyrad	183675-82-3	in vivo data available	47.7	no hit	no hit	no hit
136	PFOA (Perfluorooctanoic acid)	335-67-1	in vivo positive	42.8	738	no hit	no hit
138	Potassium perfluorooctanesulfonate	2795-39-3	in vivo positive	16	667	no hit	no hit
139	Pymetrozine	123312-89-0	in vivo data available	no hit	no hit	13.9	no hit
140	Rotenone	83-79-4	model compound	0.1	0.1	0.8	1.2
151	Sodium valproate*	1069-66-5	in vivo positive	5159	8141	6562	9058
159	tert-Butylphenyl diphenyl phosphate	56803-37-3		no hit	no hit	15.8	no hit
163	Thiamethoxam	153719-23-4	in vivo data available	no hit	no hit	16	no hit
167	Tri-allate	2303-17-5		12.6	no hit	no hit	no hit
169	Tributyltin chloride	1461-22-9	in vivo positive	0.1	0.1	0.3	0.3
171	Trichlorfon	52-68-6	in vivo positive	no hit	no hit	4	5.8
172	Triethyl-tin bromide	2767-54-6	in vivo positive	0.5	5	0.2	0.2

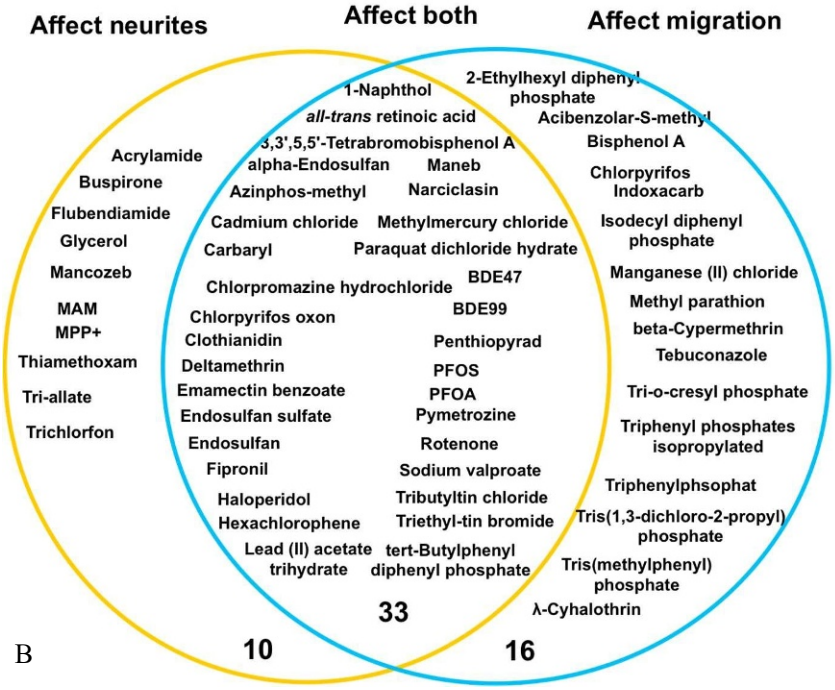
Next, we performed the same analysis for the migration assays (Table 21). Migration assays cover NCC migration (UKN2), radial glia migration (NPC2a), neuronal migration (NPC2b) and oligodendrocyte migration (NPC2c). Here, 49 compounds (41% of the 119 substances) produced a hit in at least one migration assay yielding 63% of all hits of the DNT *in vitro* testing battery. Only 7 compounds affect migration in all assays (TBBPA, CdCl, chlorpromazine, emamectin, hexachlorophene, tributyltin chloride,

triethyl-tin bromide) supporting the non-redundancy of migration assays. 21 compounds affect this endpoint with BMC <10 µM displaying the importance of these assays. Compounds' effects on migration contribute to assay performance by effects lying in the MSE 3-fold range in 33 cases for scenario (a) and in 10 cases for scenario (b). The only migration assay never serving as a MSE is NPC2b, neuronal migration. It seems to be redundant for this set of compounds.

Table 21 Comparison of migration assays for the 49 out of the 119 compounds producing hits in at least one of the assays (µM).

Nr	compound name	CAS #	Selection criteria	UKN2 (NCC migration)	NPC2a72h (radial glia migr.)	NPC2a120h (radial glia migr.)	NPC2b (neuron migration)	NPC2c (oligo. migr.)
3	1-Naphthol	90-15-3		40.4	11.8	12.4	no hit	no hit
4	2-Ethylhexyl diphenyl phosphate	1241-94-7	NTP	4.5	no hit	no hit	no hit	no hit
5	3,3',5,5'-Tetrabromobisphenol A	79-94-7	in vivo data available	1.2	1.1	2.1	4	3
15	Acibenzolar-S-methyl	135158-54-2	in vivo data available	no hit	no hit	17.6	no hit	no hit
19	all-*trans*-Retinoic acid	302-79-4	in vivo positive	4.7	no hit	no hit	no hit	no hit
21	alpha-Endosulfan	959-98-8		9	no hit	no hit	no hit	no hit
28	Azinphos-methyl	86-50-0		16.2	16.3	13.6	no hit	no hit
33	Bisphenol A	80-05-7	in vivo data available	11.9	no hit	no hit	no hit	no hit
37	Cadmium chloride	10108-64-2	human positive // in vivo positive	0.5	1.7	2.2	11.9	7.6
41	Carbaryl	63-25-2	in vivo data available	9.6	8.6	9.2	no hit	no hit
44	Chlorpromazine hydrochloride*	69-09-0	in vivo positive	9.4	17.6	7	9.4	4.4
45	Chlorpyrifos	2921-88-2	human positive // in vivo data available	38.4	no hit	no hit	no hit	no hit
46	Chlorpyrifos oxon	5598-15-2	metabolite/related	no hit	no hit	12.5	no hit	no hit
48	Clothianidin	210880-92-5	in vivo data available	no hit	no hit	14.4	no hit	no hit
59	Deltamethrin	52918-63-5	in vivo positive	18.4	no hit	16.3	no hit	no hit
73	Emamectin benzoate	155569-91-8	in vivo data available	1.1	2.5	1.7	2.1	1.5
74	Endosulfan	115-29-7	in vivo data available	7	no hit	15.1	no hit	no hit
75	Endosulfan sulfate	1031-07-8		3.9	12	no hit	no hit	5.2
81	Fipronil	120068-37-3	in vivo data available	16.8	13.3	13.6	no hit	19.7
88	Haloperidol	52-86-8	in vivo positive	no hit	12.7	11.9	no hit	13.2
91	Hexachlorophene	70-30-4	human positive	0.4	0.1	0.8	0.4	0.2
94	Imidacloprid	138261-41-3	in vivo data available	no hit	no hit	13.6	no hit	no hit
95	Indoxacarb	173584-44-6	in vivo data available	16.6	no hit	no hit	no hit	no hit
96	Isodecyl diphenyl phosphate	29761-21-5	NTP	11	no hit	no hit	no hit	no hit
99	Lead (II) acetate trihydrate	6080-56-4	human positive	28.9	no hit	no hit	no hit	no hit
104	Maneb	12427-38-2	in vivo positive	11.4	no hit	no hit	no hit	no hit
105	Manganese (II) chloride	7773-01-5	human positive	29.3	no hit	no hit	no hit	no hit
114	Methyl parathion	298-00-0	in vivo positive // in vivo uncertainty	no hit	no hit	15.1	no hit	no hit
116	Methylmercury chloride*	115-09-3	human positive	5.1	0.9	0.5	no hit	2.7
122	Narciclasin	29477-83-6	model compound	0.22	0.01	0.01	no hit	0.05
127	Paraquat dichloride hydrate	1910-42-5	in vivo positive	309	no hit	10.9	no hit	no hit
130	PBDE 47	5436-43-1	human positive	54.3	no hit	no hit	no hit	no hit
131	PBDE 99	60348-60-9	human positive	27.3	no hit	no hit	no hit	no hit
134	Penthiopyrad	183675-82-3	in vivo data available	38.6	16.3	16.2	no hit	no hit
138	Potassium perfluorooctanesulfonate	2795-39-3	in vivo positive	44	no hit	14.7	no hit	no hit
139	Pymetrozine	123312-89-0	in vivo data available	no hit	no hit	7.3	no hit	no hit
140	Rotenone	83-79-4	model compound	0.03	0.3	0.2	no hit	1.9
151	Sodium valproate*	1069-66-5	in vivo positive	no hit	no hit	5474	no hit	no hit
153	beta-Cypermethrin	1224510-29-5	in vivo data available	no hit	no hit	no hit	no hit	14.1
156	Tebuconazole	107534-96-3	in vivo positive	no hit	no hit	16.5	no hit	no hit
159	tert-Butylphenyl diphenyl phosphate	56803-37-3		4.5	no hit	10.5	no hit	8.9
169	Tributyltin chloride	1461-22-9	in vivo positive	0.1	0.2	0.2	0.3	0.3
172	Triethyl-tin bromide	2767-54-6	in vivo positive	1.3	0.3	0.4	0.4	0.4
174	Tri-o-cresyl phosphate	78-30-8		3.8	no hit	no hit	no hit	no hit
175	Triphenyl phosphates isopropylated	68937-41-7	NTP	6.1	no hit	18.2	no hit	17.3
176	Triphenylphsophat	115-86-6	NTP	10	no hit	no hit	no hit	no hit
178	Tris(1,3-dichloro-2-propyl) phosphate	13674-87-8	NTP	no hit	no hit	13.7	no hit	15.7
180	Tris(methylphenyl) phosphate	1330-78-5	NTP	7.8	no hit	no hit	no hit	no hit
183	λ-Cyhalothrin	91465-08-6	in vivo data available	no hit	no hit	17.9	no hit	no hit

A



B

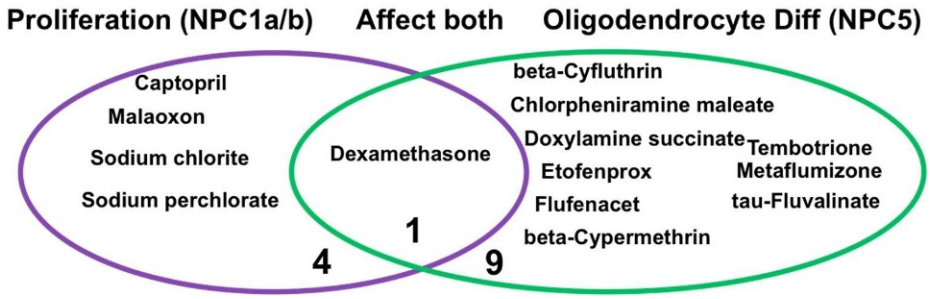


Figure 60: Evaluation of compounds affecting neurite morphology (UKN4, UKN5, NPC4) and any type of migration (UKN2, NPC2). Chemicals affecting just any neurite or migration endpoint or at least one of each endpoints were identified (a). Additionally, endpoints affected by compounds that do not affect either neurite morphology or migration are assigned to NPC proliferation or oligodendrocyte differentiation including their overlap (b). As 64% of all compounds affect oligodendrocyte differentiation (NPC5), a variety of compounds depicted under a) are also identified by NPC5.

Taken the data from Table 20 and Table 21 together, produced the Venn diagram shown in Figure 60. While 10 and 16 compounds affected neurite morphology independently of migration and migration independently of neurite morphology, respectively, the majority of 33 compounds affected both. Substances (14) not interfering with either affected NPC proliferation (NPC1) or oligodendrocyte differentiation (NPC5). Here it is to note that NPC1 and NPC5 produced more hits, yet not in the absence of any migration or neurite morphology hit. Here NPC5 provides in most cases the MSE.

In summary, the correlation analyses as well as the comparison of similar endpoint assays clearly demonstrate the non-redundancy of all assays measuring similar or completely different endpoints in the DNT *in vitro* testing battery. This outcome clearly supports the notion that all assays currently present in the battery are needed. Neurite morphology, migration, proliferation and oligodendrocyte differentiation contribute all to the DNT *in vitro* battery outcome by (i) finding "unique" compounds for their endpoint and (ii) identifying high alert compounds (e.g. chlorpromazine affects all four neurodevelopmental processes).

However, some analyses, like neuronal migration might be superfluous. As information on this endpoint is generated automatically during data analysis of NPC2-5, this is not an endpoint that leads to a more time- or cost-efficient battery.

10.2. How is the data handled?

All raw data handling is done concordance with the FAIR principles meaning that the data should be stored and documented in way that allows their **F**indability, **A**ccessibility, **I**nteroperability, and **R**eusability (Wilkinson *et al*, 2016). To comply with the FAIR principles and allow a full transparency and traceability of data handling, all experimental raw data and meta data should be documented with sufficient detail to understand, re-use and interpret the data. Therefore, the use of consistent layout sheets for experimental raw and meta data is recommended. These sheets should contain all meta data for one experiment and allow the storage of numbered raw data. Pre-defined selection options and a consistent documentation format minimize possible user error and allow a consistent documentation that is especially important if automated processing of the raw and meta data is applied. First data visualization in this layout sheet further allows an initial inspection of the plausibility of the data. Annex Q shows an example of the experiment layout sheet used for the documentation of NPC2-5 for this Project.

In addition, all data handling steps in which the data is processed to obtain the overall test results should be described:

1) How is the data generated?

Data generation depends on the test method and the endpoint that is assessed. Examples are the manual or automatic identification of cell types based of ICC images, fluorescence readouts from a micro-plate reader, C_t values in a PCR cycler or electrical activity on a multielectrode array. The process of data generation should be described in sufficient detail. It should further be considered how a user bias can be prevented in the method of data generation. In the assays performed for this report image analysis and data analysis in performed automatically reducing the possibility for accidental interference with the outcome. However, there are several data handling steps that cannot be automatized. This is the case if an endpoint is measured manually (e.g. migration distance of radial glia after 72 h) or if data must be copied from the output of a measuring device (e.g. micro plate reader). Here above-mentioned layout experiment sheets with consistent form for data transfer and regular control for plausibility limit the risk for user bias.

2) What are the acceptability criteria for the test method and how are outliers identified?

Acceptability parameters need to be defined for each endpoint in the respective assay to inform on the validity and reliability of the *in vitro* data. Acceptability criteria can for example be defined a) based on the response of an endpoints specific control (e.g. PP2 as a positive control for cell migration in NPC2 should reduce radial glia migration to 0-20 % of the solvent control), b) based on the variation in the assay (e.g. the mean variation within the solvent control of one experiment should not exceed a certain variation) or c) based on a defined range of acceptable raw values (e.g. neuronal differentiation of 5 day differentiated NPC should result in 2-20% of neurons). It is important that if any of the pre-defined acceptability criteria is not met, the affected replicate, condition or whole experiment needs to be excluded from data evaluation. At what point one replicate value, a complete testing condition or the complete experiment will be excluded depends on the acceptability criteria, how it is applied and how it is defined for the assay and endpoints. If the acceptability criteria is not met for only one replicate, it might be acceptable that this replicate is removed, as long as sufficient replicates are present for an appropriate statistical analysis of the data. The same applies to the exclusion of single conditions in the data set. Here it might be acceptable that only 5

conditions are used if they offer enough information to assess the concentrations response relationship for a test chemical.

If data is excluded because it did not fit the acceptability criteria or was identified as an outlier, this needs to be documented in the data handling process.

3) What are mathematical pre-processing steps of the raw data

Pre-processing steps include for example the subtraction of background or the calculation of a ratio or a gradient. These steps are highly dependent on the test method and the endpoints that is assessed. In general, it is recommended to document all data in a raw non-processed format to allow full traceability of any processing that is performed.

4) How is the data normalized?

For normalization of the data it is important to consider the overall variation of the assay. The question here is if the raw data can be used or if the data should be normalized. In general, a normalization of the raw values to a negative control such as the solvent control is recommended as is allows a better comparison between different endpoints.

For NPC1-5 we gathered all raw solvent control data for all assays performed under this procurement to visualize the deviation of controls for individual assays (Figure 27). These graphs show that some endpoints are rather stable with little variation in raw values between experiments, e.g. migration distance radial glia, while neuronal or oligodendrocyte differentiation are more variable and show the need to perform normalization to compare different experiments with each other. Such visualization of assay variation is recommended for a DNT *in vitro* battery to give full transparency over the variation expected in the system.

5) What measure is used as overall result?

For deriving a reference point (RP) or point of departure (Pod) the Benchmark Dose (BMD) approach is recommended by the EFSA Scientific Committee (Hardy et al, 2017b). The BMD or for in vitro toxicity testing Benchmark Concentration (BMC) is derived from a concentration response information and makes use of all data points of the fitted concentration response curve. Thereby the BMC is defined as the concentration that is associated with a specific change in response, the Benchmark Response (BMR). The BMR is a value of effect size and should be defined as an effect size that is higher than the general variability of the measured endpoint. It is therefore recommended to define the BMR based on the overall variability of the respective endpoint. See 8.3.8 for an example on how the BMR was defined in this project.

6) How is the concentration response modelling performed?

In general, one of the following three approaches should be applied to model the concentration response data.

1) one mathematical model (e.g the logistic model) is used to describe the data. For this approach the fit of one model is re-evaluated over and over until the curve fit best describes the data.
2) a selection of models including logistic and exponential models (with either three or four parameters) are used to describe the data. This 'best fit' approach repeats the process mentioned in 1 for each of the models selected and selects the model with those parameters that best fit the data.
3) in this 'model averaging' approach, a selection of models is used but instead of selecting the model with the best fist, the BMC of all model is combined to calculate the final BMC. Thereby models that fit the data better get a higher weight in the overall BMC calculation.

7) How is the uncertainty analyzed?

The uncertainty for the overall result (e.g the BMC values) should be given as upper and lower confidence limits around the calculated BMC. For the data presented here the bootstrapping method was used to

quantify the uncertainty. For each sample, a new dataset was derived from the original one by parametric replacement. This procedure results in e.g. 200 artificial datasets, where for each dataset a new fit is calculated, resulting in a new BMC. The uncertainty is than represented as 95% confidence intervals, where the upper and lower limits are defined as the respective 95% and 5% quantile of the e.g. 200 BMCs. This method can be combined with any of the three concentration-response modelling approaches described above. If the bootstrapping approach is combined with model averaging, the uncertainty represents not only the uncertainty based on variation between experiments but also uncertainty based on the selection of different fitting models.

Although the data evaluation pipeline developed in this project offers all three concentration-response modelling options in combination with the bootstrapping method, we selected the first option for the evaluation of the data in this report. The reason is that there are still some improvements that need to be made to combine the 'best fit' or the 'model averaging' approach with the bootstrap method. However, once these improvements are made it should be studied in more detail how the selection of different approaches for concentration-response fitting and uncertainty calculation affects the overall hazard identification based on the *in vitro* test battery data.

Each of the above-mentioned points on data handling is described in Annex C-G (toxtemplate) for the UKN and NPC assays used in this report and can be used as a more detailed example for documentation on data handling (Krebs *et al*, 2019).

10.3. How are the results interpreted?

Interpretation of test battery data differs from single assay evaluation. The assays of the test battery are based on KNDPs and as such weigh equally when disturbed. This means, if NPC proliferation or differentiation is interrupted, both will produce an adverse outcome. Thus, for identifying the MSE across the battery all endpoints have to be evaluated. With the MSE identified, information can be gained on the respective benchmark concentrations (BMC; for details see WP5b) across the whole testing battery and on the test method/neurodevelopmental process affected by the compound as exemplified with a very small set of compounds and endpoints (Baumann *et al*, 2016). Due to a general uncertainty in testing results, we did not restrict the MSE to one endpoint, but we decided that all endpoints within a 3-fold range of the lowest BMC across all assays are equally toxicologically important (see 8.4 Testing results). All endpoints affected by the compounds within this 3-fold range of the MSE were plotted in Figure 33 for scenario (a) and in Figure 34 for scenario (b).

For different battery use scenarios, DNT *in vitro* battery data is interpreted differently. If one wants to use the battery for screening and prioritization, analyses of results in a heatmap format similar to the evaluation of the compound classes (WP5b, 8.4.3) are warranted. Here, comparison between effects, e.g. within one compound class, can inform the researcher on compounds' potencies. For one, the number of endpoints changed across the battery and secondly the magnitude of BMCs of the different assays are two important parameters. Thereby a combination of both can be considered the worst with the effect concentration driving the hazard characterization. One example in the current testing results is again narciclasine with effects in the < 1µM range across a variety of endpoints compared to MAM affecting also multiple endpoints, yet at much higher concentrations. Similarly, domoic acid, which is highly toxic, alters only one endpoint at a very low concentration, while etofenprox also affects one endpoint, yet at a much higher concentration. Here, one would prioritize the multiple endpoint hits at low BMC>single endpoint hit at low BMC>multiple endpoint hit at high BMC>single endpoint hit at high BMC. It is to note, that this procedure is pure *in vitro* hazard identification and does not inform on any risk.

For risk assessment with follow-up targeted *in vivo* testing data interpretation needs to be based on a higher certainty level than for screening and prioritization. Here, completeness of the DNT *in vitro* testing battery is crucial. In addition, information on the individual endpoint receives more weight. A mechanistic understanding of the individual endpoint modulations is crucial. Here an issue is a lack of data for most compounds regarding their *in vivo* DNT MoA. In the Results section discussing effects of compound classes (8.4.3) we tried to assemble some information on compound classes and individual compounds' MoA to relate them to the battery results. This was rather difficult due to an overall lack of data. In addition, for many substances, the battery results revealed 'off-target' effects of compounds within the same compound class. If such pose risk to humans beyond pure *in vitro* hazard needs to be further investigated.

For translating these data from the results of the *in vitro* test in the testing battery into a toxicological statement, one needs to consider biokinetics of the test compounds. These considerations can take three directions, all starting from the nominal test concentration that was found to be effective (nominal PoD = nPoD): (1) relating the nPoD to intracellular (target) concentrations; (2) relating the nPoD to the corresponding free test compound concentration (fPoD); (3) relating the nPoD, or better the fPoD derived therefrom to a corresponding *in vivo* dose.

Concerning (1): This could be done experimentally or by modelling. The experimental approach is on the one hand extremely resource-intensive, on the other hand, it is linked to huge uncertainties as to which compartment should be modelled, especially, when the target is not known. There is no convincing data available that suggest that the average cellular concentration is a better predictor than the medium concentration. This is plausible as the concentration at a target within the endoplasmic reticulum may be largely different from the concentration in the cytosol or from the average cell concentration. It would be possible to model average cell concentrations, based on available physicochemical parameters, but we do not see an immediate benefit. It would also be possible to measure cellular contents of few selected compounds, but given the remaining uncertainties in relation to the large cost, we also do not consider this beneficial.

Concerning (3): as discussed elsewhere in this report, translation of the nPoD to an external dose requires *in vitro*-to-*in vivo* extrapolation (IVIVE). This requires knowledge on barrier passage, metabolism, and excretion. In the end, data would only be useful, if reliable exposure estimates are available to compare them. In the absence of all this background information, IVIVE is not adding value to the data already available. A potential hazard, and a relative ranking of potencies can be done directly with the test summary data (benchmark concentrations).

Concerning (2): One fundamental input to IVIVE, but also for the comparison of test data amongst different test systems (with different medium constituents) is the free test compound concentration (not bound to protein or lipid). Often, it is claimed that data interpretations have too high uncertainties in the absence of such information. Therefore, we examined this. First, we adopted from the literature an approximation formula (Fisher *et al*, 2019) that allows estimates of free drug concentrations (Figure 61). This formula uses logP as predictor for lipid and protein binding, so that no further experimental data are required. Based on this, we calculated the ratio of free drug and nominal drug for three well-characterized compounds known to be high, medium and low protein binding (paracetamol, colchicine, tolbutamide). We did this for all media used in this study (assuming a drug concentration of 1 µM), and the difference of nPoD and fPoD was usually < 3%, never larger than 14% (UKN5 protein-rich medium with strongly protein binding test compound). We conclude from this exercise, that using nPoD (instead of the calculated fPoD) is a reasonable and sufficiently exact solution (Krebs *et al*, 2020b).

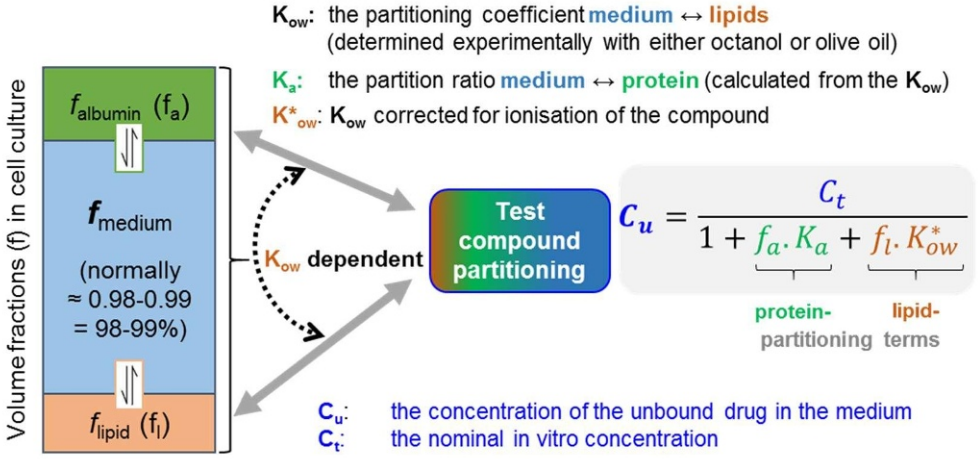


Figure 61: Approximation formula for estimating free drug concentrations *in vitro* (Krebs *et al*, 2020b; Fisher et al. 2019).

Arriving at a PoD for risk assessment

We have dealt extensively with the issue of deriving a point-of-departure from *in vitro* testing data. We followed the principles discussed extensively for the benchmark dose concept *in vivo*:

- All data points (and their uncertainties) should be used for the estimate.
- The concentration-response relationship should be modelled and the PoD derived from the model.
- The PoD should always be accompanied by a measure of uncertainty (confidence interval).
- The PoD was chosen here as a benchmark concentration, as opposed to the benchmark dose, which is more appropriate for *in vivo* studies.

Accordingly, an algorithm was developed, and used throughout the data evaluation. The background to this algorithm has been compiled for publication (Krebs *et al*, 2020a). It is publicly available both as R-script and as graphical user interface. All mathematical background for the data fitting has been checked by statisticians and is transparently explained in the publication.

Concentration-response modelling as such has also been re-visited for this study. It was found that BMC estimates (e.g. BMC20) sensitively depend on the modelling of the upper asymptote of response curves. This is often a problem with real-life screen data, and an algorithm has been developed that ensures correct positioning of upper curve asymptotes. This has been explained in detail in an accompanying publication (Krebs *et al*, 2018).

10.4. How to arrive at relevant *in vivo* exposure levels

Once a PoD has been obtained, this needs to be translated to relevant *in vivo* exposure. The usual procedure for this is an *in vitro*-to-*in vivo* extrapolation (IVIVE; Basketter *et al*, 2012), as we have illustrated earlier. Details on this process are covered elsewhere by experts, but in the context of DNT, for instance the blood-placenta barrier requires special consideration, and general guidance will be given on the process.

10.5. How can be dealt with species-specific effects?

As outlined before, the species aspect should be considered when testing proceeds from human *in vitro* to rodent *in vivo* testing. It is now widely accepted that humans are no 70 kg rats (Leist & Hartung, 2013b). Therefore, species aspects have to be carefully considered. One way of understanding species aspects is given by comparative species-overarching *in vitro* models. This approach allows studying compounds' toxicodynamics in a relatively simple and inexpensive way. For the human DNT *in vitro* testing battery there are comparative rodent models available for most of the human assays. Human NPC1-5 is copied by rat or mouse NPC1-5 (Gassmann *et al*, 2010, 2012; Baumann *et al*, 2015, 2016), rat cortical synaptogenesis (Harrill et al, 2018) and rat NNF (Shafer, 2019; Frank et al, 2018) can be accompanied by a human synaptogenesis assay (Pistollato et al, 2014) and a human NNF assay (Bartmann, Masjosthusmann, Fritsche, unpublished). As an example, responses of the endpoint-specific control, the PKC inhibitor Bis-I can be compared between the rat and human NNF assay (Figure 62). IC50 values derived from respective experiments reveal higher sensitivity of the rat compared to the human networks towards the network inhibitory effect of Bis-I. However, besides the mean firing rate, the BMC_{50} for human endpoints were within the 3-fold range of the rat BMC. That the rat NNF assay, at least for this endpoint, is more sensitive than its human counterpart is an interesting observation because rat networks were cultured for 12 and human networks for 35 days *in vitro.* However, due to the different species and the primary (rNNF) versus stem cell-derived (hNNF) nature of the cultures, these rat and human cells might also differ in differentiation/maturation stages. In addition, some microglia are present in the rat cultures that are not present in the human NNF test system. Hence, the difference in effect of Bis-I might have a toxicodynamic rather than a kinetic rationale. Unknowns here are e.g. the magnitude of expression of the proteins involved in the PKC pathway contributing to neuronal network formation.

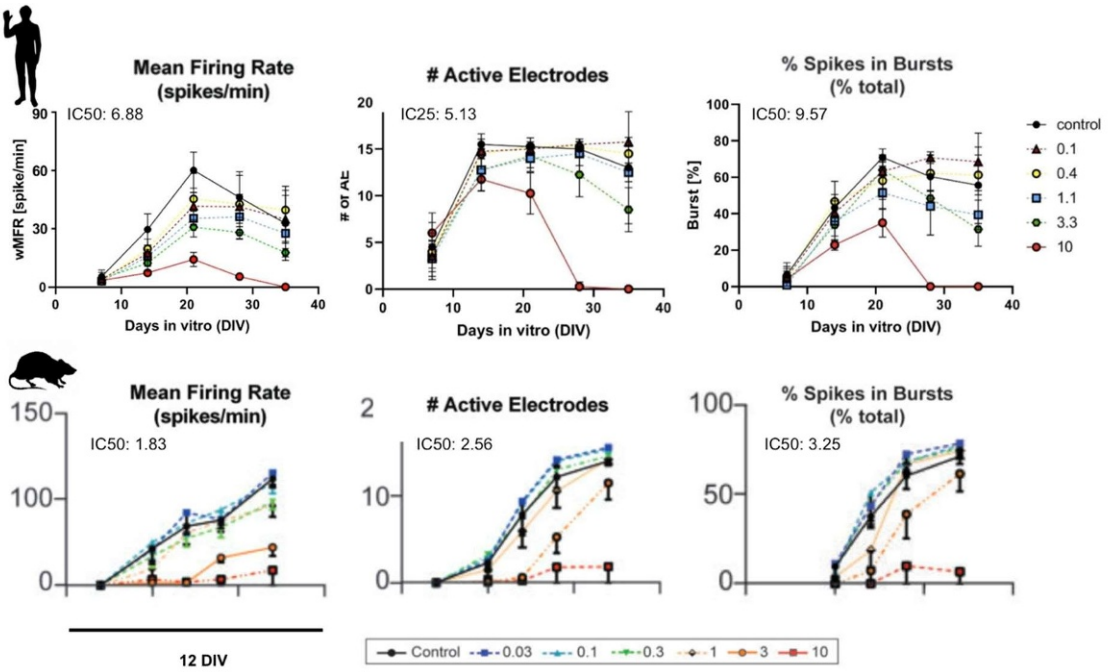


Figure 62: Species comparison of the NNF assay. Upper row: results of the effects of the PKC inhibitor BisI on human NNF. The NeuCyte SynFire Kit containing excitatory and inhibitory neurons as well as astrocytes was used for studying neuronal network formation. Network activity was monitored over a course of 35 days and measured weekly. Lower row: effects of BisI on rat primary cortical neuronal network formation (Brown et al, 2016). Both studies were performed using the Axion multi-well MEA platform.

By studying species-overarching KNDP upon chemical exposure, targeted *in vivo* testing or interpretation of existing *in vivo* data with regards to human relevance will be highly facilitated.

10.6. Uncertainties for regulatory use of data

At this point uncertainties are collected, which encompass general aspects of use of *in vitro* data for regulatory purposes and thus move beyond assay-related, scientific uncertainties. Yet they have to be addressed when moving with the use and application of *in vitro* data such as exemplified by this DNT *in vitro* battery into the regulatory arena.

First, the definition of what is called an assay response differs between labs and assays. LOAEC, NOAEC, IC/EC/benchmark responses (BMR) at certain response levels (most often 10, 20, 25, 30 or 50) are amongst the most common measures for assay response evaluation. While for individual assays, this is not an issue, caution has to be taken when individual assays are assembled in a battery where assay responses are compared to each other directly. As an example, the MSE concept of this study compares benchmark concentrations of different bench mark responses with each other, e.g. NPC2a (radial glia migration) has a BMR of 10, while NPC5 (oligodendrocyte differentiation) is assessed with a BMR of 30. These different BMR are based on assay variability as comprehensively described in chapter 8.3.8.. An uncertainty based on this approach is that the final data interpretation might be biased by the magnitude of the BMR. However, in this study this does not seem to be the case as NPC5 has a much higher hit rate than NPC2a. Here, biology rather than assay response values seems to rule the final data interpretation. If this is also the case for a larger or different compound set needs to be assessed in the future.

Second, what is called a specific *in vitro* hit is managed in a lot of different ways in the *in vitro* field. Cytotoxicity/viability assays might be used, nuclei numbers might be a measure of culture health without performing viability assays, or even none of the former is being executed. Moreover, determination of differences between a compound's effect on cytotoxicity/viability and the specific endpoint can be done in a multitude ways. Every procedure in the end gains its rectification from the individual test system used and is generally scientifically correct. However, when the application of an *in vitro* testing battery moves into the focus of application, these individual procedures might produce an overall uncertainty. In the test methods used in this study, DNT-specific hits are also defined in an assay-specific manner and distinct cytotoxicity and/or viability assays are performed. These data is used inconsistently for DNT-specificity definition: e.g., while UKN determines an x-fold ratio between the DNT BMC and the respective viability BMC, the IUF determines the overlap of their confidence intervals. If such different DNT-specificity hit calls adversely influence final battery outcomes or if they are necessary due to being integral parts of respective test methods need to be statistically assessed in the future.

This report is a source of data to be used for different DNT regulatory problem formulations as proposed in the different case studies. The detail of the uncertainty analysis is therefore dependent on available additional data and the nature of the regulatory problem formulation; more uncertainty can be accepted for screening and prioritization vs. single substance hazard characterization. In addition, this report should be used as a scientific background for development of international regulatory guidance and/or development of an *in vitro* assay DNT test guidance.

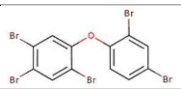
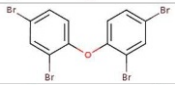
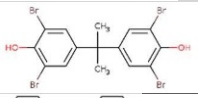
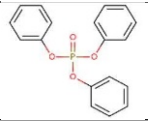
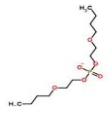
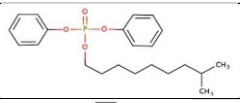
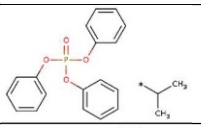
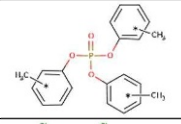
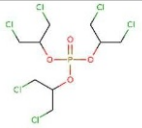
11. WP8: Use of the Test Battery within an IATA Framework

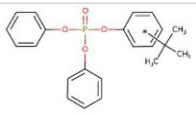
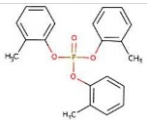
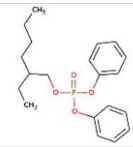
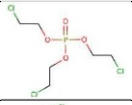
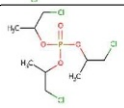
In this WP we developed two case studies to demonstrate **how the battery test data can be used within an IATA for regulatory decisions.** The case study compounds are part of the library of WP5a.

11.1. Screening type case study, type A:

We selected a set of 14 data-poor compounds belonging to the compound class of flame retardants (FR), containing some old and a variety of new substances. A list of FR compounds, their abbreviations and chemical structures is given in Table 22.

Table 22 CAS numbers, chemical names, compound IDs and their chemical structures of the 14 FR used as a case study.

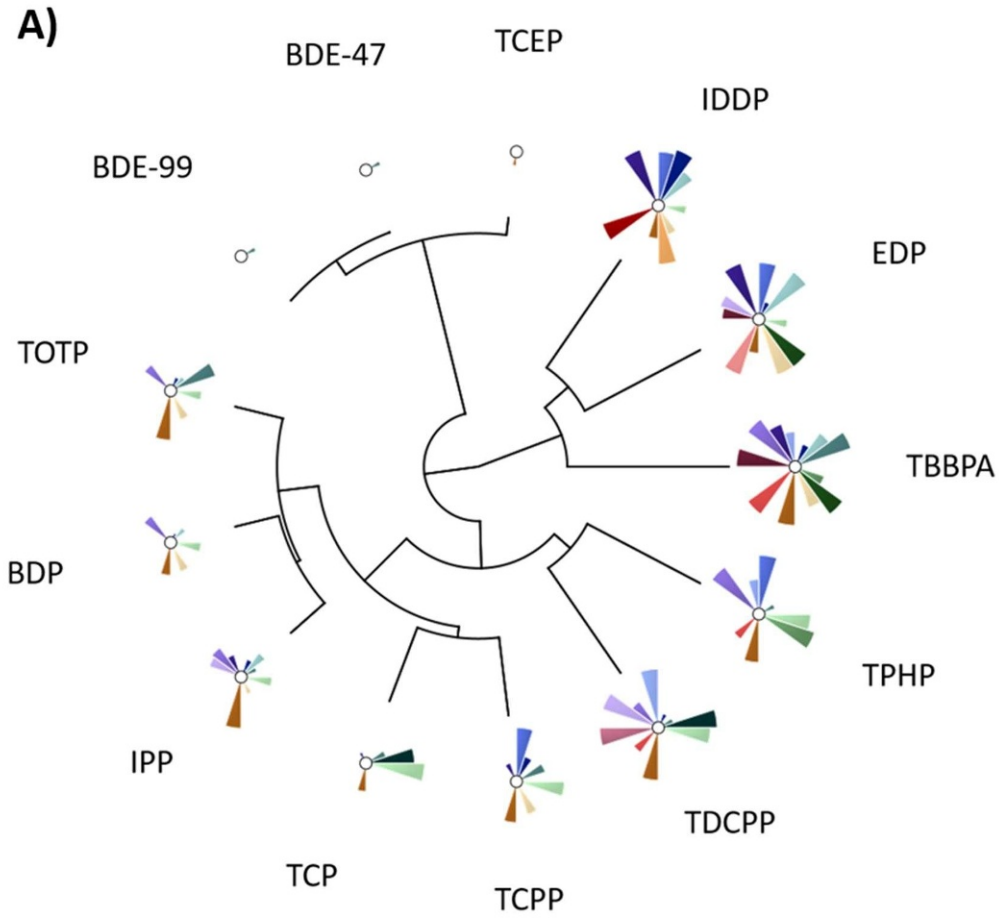
CAS number	Chemical name	Chemical ID	Structure
60348-60-9	2,2',4,4',5-Pentabromdiphenylether	BDE-99	
5436-43-1	2,2',4,4'-Tetrabromdiphenylether	BDE-47	
79-94-7	Tetrabromobisphenol A	TBBPA	
115-86-6	Triphenyl phosphate	TPHP	
Metabolite	Bis (2-butoxyethyl) phosphate	BBOEP	
29761-21-5	Isodecyl diphenyl phosphate	IDDP	
68937-41-7	Isopropylated phenyl phosphate (3:1)	IPP	
1330-78-5	Tricresyl phosphate	TCP	
13674-87-8	Tris (1,3-dichloro-2-propyl) phosphate	TDCPP	

56803-37-3	Tert-butylphenyl diphenyl phosphate	BDP	
78-30-8	Tri-O-Tolyl phosphate	TOTP	
1241-94-7	2-Ethylhexyl diphenyl phosphate	EDP	
115-96-8	Tris (2-chloroethyl) phosphate	TCEP	
13674-84-5	Tris (2-chloroisopropyl) phosphate	TCPP	

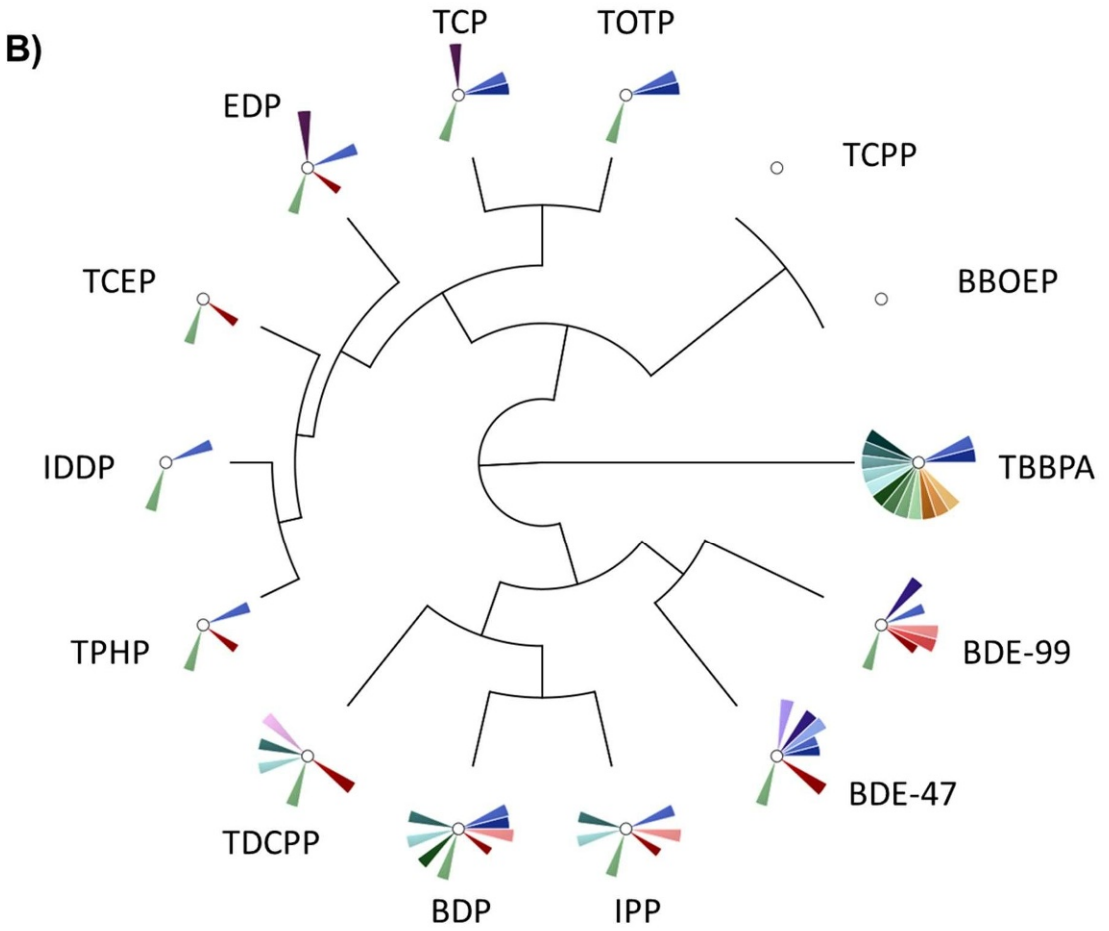
For these compounds, we generated ToxPies with the existing ToxCast data using the Toxicological Prioritization Index (ToxPi) Graphical User Interface (GUI) tool version 2.3. Here it is important to know that the size of the Pie pieces does not reflect the actual BMC, but relates the BMC for the studied compound with the BMC of this endpoint across the highest and lowest values of the data set irrespective of the values by distributing them between 0 and 1. Hence, it is a relative, not an absolute value. Therefore, Pie slice sizes of different endpoints for one compound do not reflect magnitude of BMC values and cannot be absolutely compared to each other.

The ToxCast data (Figure 63A) demonstrates the general toxicity of these compounds across a large variety of non DNT-specific endpoints and thus gives an initial idea on the general toxicity/potency of these FR. Endpoints assessed for these FR include effects on assays measuring cell cycle, cell adhesion, cell morphology, cytokines, growth factors, ion channels, nuclear receptors and others.

These data are compared to DNTPies generated with the same Pie-Tool by using the data from the DNT *in vitro* testing battery (Figure 63B). Here, instead of the ToxCast data, data generated with the NPC, UKN as well as the rNNF assay from the US-EPA (missing for TOTP und TCPP) were used. The ToxPi tool then hierarchically clusters the compounds within the ToxPies and the DNTPies according to their potency and assay hit patterns. Comparison of clustering between ToxPies and DNTPies gives the first indication whether a substance class produces a DNT-specific effect or if general toxicity of these compounds predicts DNT toxicity well.



intended target family	Weight	Metrics	Color	intended target family	Weight	Metrics	Color
-	1 (5,0%)	2		ion channel	1 (5,0%)	1	
background measurement	1 (5,0%)	12		kinase	1 (5,0%)	1	
cell adhesion molecules	1 (5,0%)	7		malformation	1 (5,0%)	4	
cell cycle	1 (5,0%)	17		misc protein	1 (5,0%)	1	
cell morphology	1 (5,0%)	3		nuclear receptor	1 (5,0%)	29	
cyp	1 (5,0%)	5		phosphatase	1 (5,0%)	1	
cytokine	1 (5,0%)	20		protease	1 (5,0%)	7	
mixed	1 (5,0%)	6		protease inhibitor	1 (5,0%)	1	
gpcr	1 (5,0%)	4		steroid hormone	1 (5,0%)	6	
growth factor	1 (5,0%)	2		transporter	1 (5,0%)	2	



DNT endpoint	Weight	Metrics	Color	DNT endpoint	Weight	Metrics	Color
viability (UKN2)	1 (4,0%)	1		migration distance oligodendrocytes (NPC2c)	1 (4,0%)	1	
migration (UKN2)	1 (4,0%)	1		neuronal differentiation (NPC3)	1 (4,0%)	1	
viability (UKN4)	1 (4,0%)	1		neurite length (NPC4)	1 (4,0%)	1	
neurite area (UKN4)	1 (4,0%)	1		neurite area (NPC4)	1 (4,0%)	1	
viability (UKN5)	1 (4,0%)	1		oligodendrocyte differentiation (NPC5)	1 (4,0%)	1	
neurite area (UKN5)	1 (4,0%)	1		cell number (NPC2-5)	1 (4,0%)	1	
proliferation BrdU (NPC1a)	1 (4,0%)	1		cytotoxicity (72 NPC2-5)	1 (4,0%)	1	
proliferation area (NPC1b)	1 (4,0%)	1		cytotoxicity (120h; NPC2-5)	1 (4,0%)	1	
viability (NPC1)	1 (4,0%)	1		viability (120h; NPC2-5)	1 (4,0%)	1	
cytotoxicity (NPC1)	1 (4,0%)	1		network formation (NNF)	1 (4,0%)	1	
migration distance radial glia (NPC2a; 72h)	1 (4,0%)	1		viability (NNF)	1 (4,0%)	1	
migration distance radial glia (NPC2a; 120h)	1 (4,0%)	1		cytotoxicity (NNF)	1 (4,0%)	1	
migration distance neurons (NPC2b)	1 (4,0%)	1					

Figure 63: ToxPies for A) general and B) DNT-specific toxicities using the ToxCast data and the results of the DNT *in vitro* battery, respectively. Graphs were produced with the Toxicological Prioritization Index (ToxPi) Graphical User Interface (GUI) tool version 2.3. Size of pie slices represents relative strength of effect on respective endpoint. FR present in B but not A were not tested in the ToxCast assay

These data illustrate that number of effects of each Pie and subsequently the Pie clustering is very different between the general and the DNT-specific Pies. For example, the two historical FR BDE-47 and -99 are almost negative for all assays in the ToxCast assays (caution! not all FR compounds were tested in the ToxCast assays, there is no ToxCast data for BBOEP). In contrast, they provoke multiple responses in the DNT assays. Similarly, e.g. TCPP gives alerts in the general ToxPi, whereas there is no effect showing in the DNTPies. These data clearly illustrate that for this subset of compounds, the 14 FR, general ToxCast data are no good predictors for DNT.

In addition to the graphical illustration as Pies, the program creates toxicity rankings of compounds (Table 23). The ranking numbers indicates that in comparison to the general toxicity ranking, DNT-specific ranking across the DNT *in vitro* testing battery strongly differs. This means, similar to the graphical Pie visualization, that results of the the ToxCast data do not predict effects of DNT assays well.

Table 23 Ranking of FR according to ToxCast (ToxPi) and DNT *in vitro* battery (DNTPi) data. Rankings were produced with the Toxicological Prioritization Index (ToxPi) Graphical User Interface (GUI) tool version 2.3.

Ranking	DNTPie		Ranking	ToxPi	
	ToxPi Score	Name		ToxPi Score	Name
1	0.560	TBBPA	**1**	0.474	TBBPA
2	0.279	BDP	**10**	0.114	BDP
3	0.252	BDE-47	**12**	0.009	BDE-47
4	0.211	BDE-99	**11**	0.009	BDE-99
5	0.203	IPP	**7**	0.195	IPP
6	0.183	TDCPP	**3**	0.355	TDCPP
7	0.138	TCP	**9**	0.132	TCP
8	0.135	EDP	**2**	0.420	EDP
9	0.108	TOTP*	**8**	0.172	TOTP
10	0.094	TPHP	**5**	0.295	TPHP
11	0.068	IDDP	**4**	0.350	IDDP
12	0.059	TCEP	**13**	0.008	TCEP
13	0.000	BBOEP*#	**-**	-	-
14	0.000	TCPP*	**6**	0.201	TCPP

*compounds not tested in the rNNF assay
compound not tested in ToxCast

In addition to ToxPi ranking, one can also rank the compounds with regards to their MSE (Figure 65). The latter produces different results from the ToxPi ranking (Table 23) and is an alternative way to analyze the data. While pure *in vitro* potency drives the ranking according to the MSE (Figure 65), ToxPi ranking also takes e.g. the number of altered endpoints into consideration. The latter might or might not make sense as alteration of only one endpoint can possibly strongly affect brain development. Hence, both methods can be viewed as complementary for determining potency and performing prioritization within a screening effort.

Next, we investigated for the case study A how the data produced by the DNT *in vitro* testing battery can be used in a weight-of-evidence (WoE) approach to prioritize the compounds for further testing. Therefore, we followed the procedure published as an EFSA scientific opinion by Hardy et al. (2017). The strategy followed is visualized in Figure 64.

The **problem formulation** for this case study is: 'screening for prioritization for further testing'. The goal is here to generate DNT-specific alerts that will subsequently be followed up. For the WoE assessment, the assembly of the evidence is a) information on toxicokinetics and b) on toxicodynamics (from the *in vitro* testing battery). This leads to the three steps in the WoE assessment:

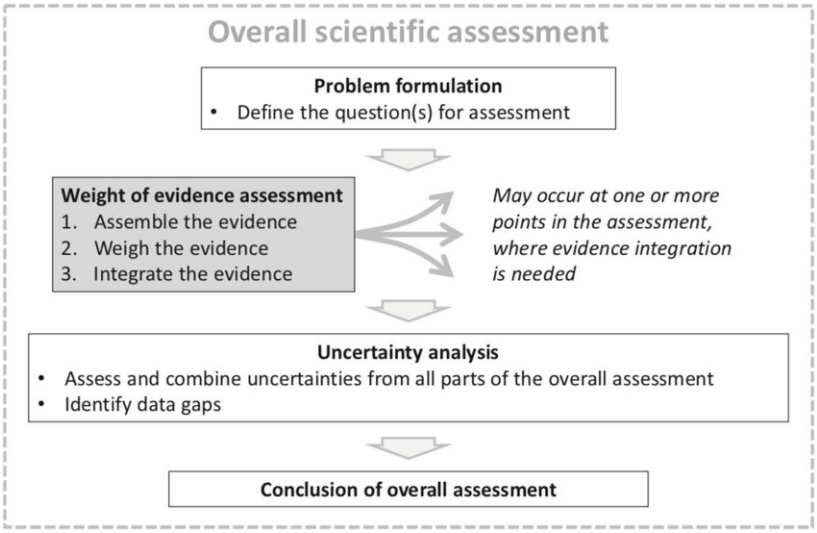


Figure 64: Diagrammatic illustration of weight of evidence assessment as a 3-step process according to (Hardy *et al*, 2017a).

11.1.1. WoE Assessment

1. Assembly of evidence:

a) Toxicokinetics: there is perinatal exposure data of humans towards some of the FR available (Tang & Zhai, 2017; Cariou et al, 2008). A thorough exposure analysis is still necessary and will be performed in collaboration with Mamta Behl from the National Toxicology Program of the National Institute of Environmental Health Sciences, USA. Currently available human exposure data from published data is summarized in Table 24. However, there is a huge data gap on exposure data for most of the compounds leaving us with such data for BDE-47, BDE-99 and TBBPA. For the substances where human internal concentrations were measured as ng/g fat, we used a fat content of 3.1 g/l and 44 g/l for blood and breast

milk, respectively, for calculating molarities for presence of these FR in respective body compartments Table 24. These data serve as the basis for making the comparison of *in vitro* BMC to *in vivo* exposure possible. However, such a comparison is very crude and does not account for *in vitro* kinetics or for actual fetal brain concentration *in vivo.* Advanced kinetic modelling is needed to perform proper IVIVE (see WP7).

Table 24 Exposure data collected from published FR measures in different human body fluids (Tang & Zhai, 2017; Cariou et al, 2008).

	Breast milk						Cord blood					
	BDE-99		BDE-47		TBBPA		BDE-99		BDE-47		TBBPA	
	ng/g lw	µM	ng/g lw	µM	ng/g lw	µM	ng/g lw	µM	ng/g lw	µM	ng/g lw	µM
Korea	54.0	0.042	31.0	0.028	-	-	19.0	0.001	36.0	0.002	-	-
China	10.8	0.008	27.5	0.025	-	-	3.45	0.0002	8.49	0.0005	-	-
USA	6.40	0.005	29.7	0.027	-	-	23.3	0.0003	4.60	0.0015	-	-
France	0.53	0.001	1.15	0.0004	4.1	0.003	7.43	0.0004	-	-	103	0.006
Germany	0.18	0.0001	0.45	0.0004	-	-	-	-	-	-	-	-
UK	0.80	0.0006	2.70	0.0025	-	-	-	-	-	-	-	-
Sweden	0.48	0.0004	2.28	0.0021	-	-	0.22	0.00001	3.4	0.0002	-	-
Spain	0.51	0.0004	0.54	0.0005	-	-	4.3	0.00023	3.3	0.00023	-	-

b) Toxicodynamics: results of FR testing in the DNT *in vitro* battery as BMCs are assembled in Annex A4. DNTPies reveal that MSE are scattered across the different FR. However, some MSE patterns can be observed by the hierarchical clustering (Figure 63B). Two FR (TCPP and BBOEP) exert no effects. Therefore, these compounds have the lowest toxicity profile. This is different for TCPP in the ToxPies from the ToxCast data, where TCPP produces a variety of hazards in different assays. Another DNTPie cluster involves TPHP, IDDP, TCEP, EDP, TCP and TOTP. Here, in general, 2 to 4 DNT endpoints were affected and the potencies were not that high (TCP effects on UKN2 and NPC5 with a BMC or 3.8 and 3.9 µM are the lowest BMCs within this group). The third clustered group contains BDE-47, BDE-99, TDCPP, BDP and IPP. In this group, ≥ 4 endpoints were affected and thus a higher hazard proposed. The sole compound clustering by itself is TBBPA because TBBPA affects almost all endpoints tested (not rNNF) at BMCs that are low in relation to the BMCs of the other FR for these endpoints. As all DNTPie slices have maximum size, all endpoints affected by TBBPA are amongst the lowest BMCs compared to the other FR compounds used in this case study. One endpoint is altered across all FR used (besides the cluster of non-DNT toxic FR containing TCPP

and BBOEP), i.e. oligodendrocyte formation. This is the most striking common result within this compound class. This points to a common MoA for this substance class. However, molecular studies that will more in-depth evaluate the nature of oligodendrocyte toxicity for the individual FR will help understanding if the MoA leading to this *in vitro* phenotype based on NPC5 is common or differs between the individual compounds. Oligodendrocyte differentiation can be modulated by a large variety of signaling pathways (Figure 50, WP6). Therefore, different MoA of individual FR on oligodendrocyte differentiation can be envisioned.

2. Weigh the evidence: Reliability of the evidence was assessed by analyzing the data according to the Klimisch criteria. This was done by using the ToxRTool. The outcome of the ToxRTool indicates high reliability of all assays of the battery with 17 to 18 ToxRTool points (

Table **25**). This reflects the high readiness of the Test Methods published earlier (Bal-Price *et al*, 2018).

Table 25 Evaluation of NPC and UKN DNT Test Methods with the ToxRTool. More detailed analyses was published in Bal-Price et al. 2018.

DNT battery IUF Düsseldorf endpoints	ToxRTool points	ToxRTool category	
		A	B
NPC1a	18	1	1
NPC1b	18	1	1
NPC1 (cytotoxicity)	18	1	1
NPC1 (viability)	18	1	1
NPC2a (migration distance radial glia 72h)	18	1	1
NPC2a (migration distance radial glia 120h)	18	1	1
NPC2b (migration distance neurons)	17	1	3
NPC2c (migration distance oligodendrocytes)	17	1	3
NPC3 (neuronal differentiation)	18	1	1
NPC4 (neurite area)	17	1	3
NPC4 (neurite length)	17	1	3
NPC5 (oligodendrocyte differentiation)	18	1	1
NPC2-5 (cell number)	17	1	3
NPC2-5 (cytotoxicity 72h)	18	1	1
NPC2-5 (cytotoxicity 120h)	18	1	1
NPC2-5 (viability)	18	1	1
UKN2 (NCC migration	18	1	1
UKN4 (neurite area)	18	1	1
UKN5 (neurite area)	18	1	1

Variability of data was taken as another potential uncertainty here. Therefore, CI for all BMC of the respective MSE of individual FR (where available) were plotted in relation to their 3-fold BMC range and the BMCs of the less sensitive endpoints (Figure 65A). Here, all endpoints, including vi-ability/cytotoxicity are shown. The CI for the NNF were not provided in the publications (Shafer et al, 2019; Frank et al, 2017) and thus were not plotted. BDE-47, TDCPP and TBBPA were the three compounds with the lowest BMC < 1µM for unspecific hits, thus candidates for priorisation according to the MSE (TBBPA, BDE-99, BDP according to ToxPi). In addition, we analyzed the DNT-specific endpoints in the same manner (Figure 65B). Here, data points are less in number due to the missing viability/cytotoxicity data. Considering only the

specific hit changes some of the FR ranking, yet not the first three most potent compounds BDE-47, TDCPP and TBBPA (BMC < 1 µM). For the ranking according to MSE and DNTPies see Table 23.

In both graphs, CI are shown for the MSE only to make the graph easy to perceive. Almost all CI are smaller than the 3-fold range we defined around the BMC for identifying endpoints that are equally important for hazard characterization. Only NPC proliferation due to EDP exposure has a larger CI. This is due to flat slope of the CR-curve and the larger SEM around the BMR_{30} limit. How slope and SEMs contribute to FR CI is depicted in Figure 66. Here, EDP and TBBPA both have relatively flat slopes, yet SEMs for TBBPA around the BMR are smaller resulting in smaller CI for TBBPA than for EDP. In contrast, the slope of IDP around the BMR is steeper and despite a similar magnitude of SME than EDP, thus produces a smaller CI than EDP for the BMC. This indicates that the CI is a measure of certainty for data variability, yet also the slope of the underlying CR-curves contribute to CI of the BMC.

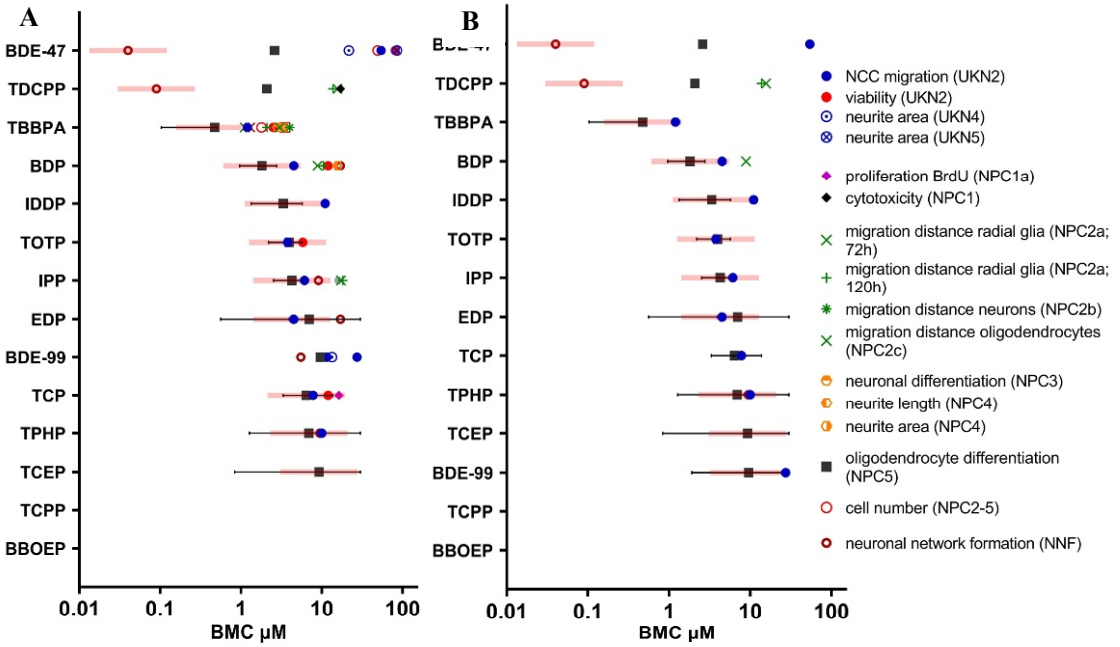


Figure 65: FR MSE including the 3-fold MSE limit and CI for the MSE. The endpoint NNF does not provide CIs (Shafer et al, 2019; Frank et al, 2017). A MSE using hits regardless of DNT-specificity, i.e. DNT-specific and viability-related endpoints. B Only DNT-specific hits.

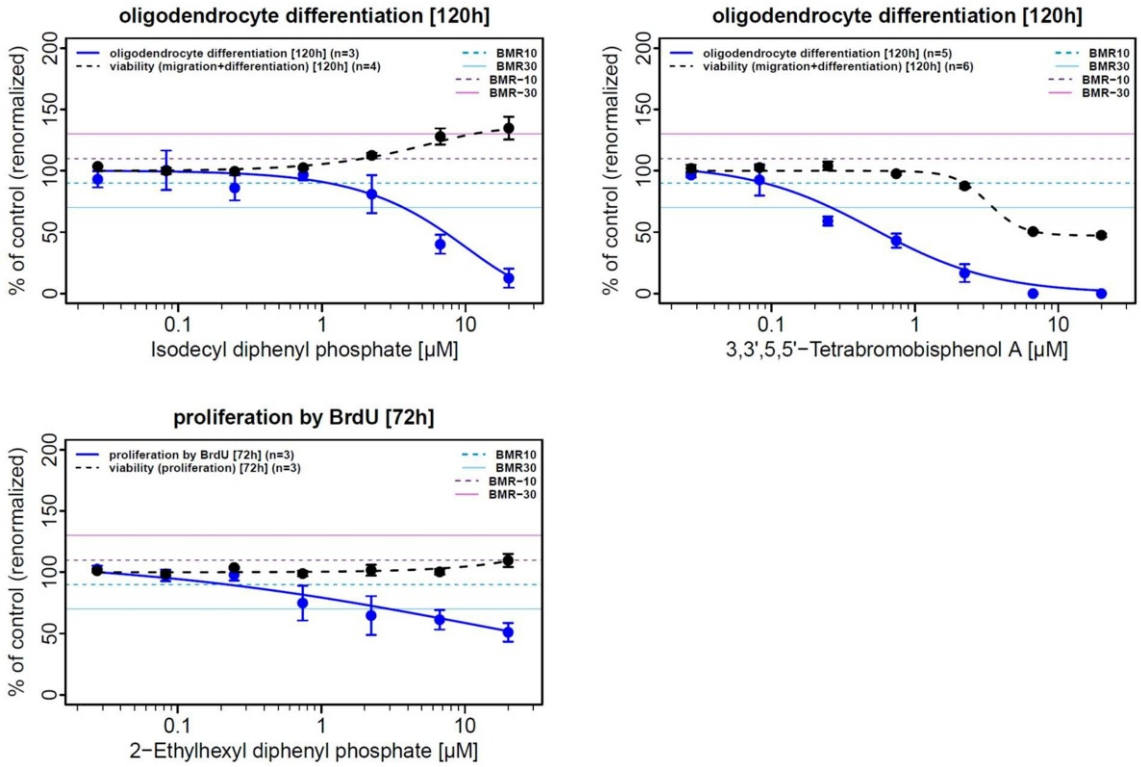


Figure 66: Examples of CR-curves with different slopes and SEM. Shown are examples for MSE for respective compounds identified within the MSE ranking Figure 65. Curves are taken from all CR-curves of the data set and can be found in the Annex N.

Hence, CI calculated for the case study of FR indicate an overall high certainty as they generally are smaller than the 3-fold range around respective BMC.

Another measure to consider might be the number of endpoints altered by individual FR that gather within the 3-fold range of the MSE. As discussed for the DNTPie versus MSE ranking above, this should not be anticipated in the first line. However, as a second line information it might increase certainty. Examples, where this might be of relevance, are EDP, IDP, TOTP, TPI, BDE-99 and TPP. For TBBPA, however, the majority of endpoints are affected at BMC values > 1 µM with the exception of oligodendrocyte differentiation that is altered with a BMC of 0.47 µM. Here, the regulation would consider the lowest BMC although the majority of endpoints gather at almost one order of magnitude higher.

3. Integrate the evidence: One endpoint consistently affected by all FR besides TCPP and BBOEP (10 out of 14) is oligodendrocyte differentiation (NPC5). In addition, NCC migration (UKN2) is an endpoint altered in 10 out of the 13 FR. NNF is affected by 8 out of the 11 that were tested. Hence, these endpoints mainly drive FR hazard profiles. Here, the MSE is produced by effects on NPC5 (7 times), NNF (3 times) UKN2 (1 times) and NPC1 (1 time; data table FR, Annex A4).
FR were also studied in the context of redundant migration and neurite morphology assays. Data (Table 26) indicates that indeed migration assays are not redundant for the compound class of FR. Here, NCC migration (UKN2) is the most sensitive migration assay and is altered by a majority of FR in a specific

manner, while only one compound (TDCPP) specifically affects radial glia and oligodendrocyte migration without altering NCC migration. This pattern supports non-redundancy of assays and the different profiles of altered migration across compounds supports their distinct MoA. In addition, this specificity in pattern increases confidence in assays as it reflects differences in compounds' MoA as a definitive strength of the battery.

Table 26 Effects of FR on migration assays.

Nr	compound name	CAS #	Selection criteria	UKN2 (NCC migration)	NPC2a72h (radial glia migr.)	NPC2a120h (radial glia migr.)	NPC2b (neuron migration)	NPC2c (oligo. migr.)
4	2-Ethylhexyl diphenyl phosphate	1241-94-7	NTP	4.5	no hit	no hit	no hit	no hit
5	3,3',5,5'-Tetrabromobisphenol A	79-94-7	in vivo data available	1.2	1.1	2.1	4	3
31	Bis (2-butoxyethyl)phosphat	-		no hit	no hit	no hit	no hit	no hit
96	Isodecyl diphenyl phosphate	29761-21-5	NTP	11	no hit	no hit	no hit	no hit
130	PBDE 47	5436-43-1	human positive	54.3	no hit	no hit	no hit	no hit
131	PBDE 99	60348-60-9	human positive	27.3	no hit	no hit	no hit	no hit
159	tert-Butylphenyl diphenyl phosphate	56803-37-3		4.5	no hit	10.5	no hit	8.9
174	Tri-o-cresyl phosphate	78-30-8		3.8	no hit	no hit	no hit	no hit
175	Triphenyl phosphates isopropylated	68937-41-7	NTP	6.1	no hit	18.2	no hit	17.3
176	Triphenylphsophat	115-86-6	NTP	10	no hit	no hit	no hit	no hit
177	Tris (2-chloroethyl)phosphat	115-96-8	NTP	no hit	no hit	no hit	no hit	no hit
178	Tris(1,3-dichloro-2-propyl) phosphate	13674-87-8	NTP	no hit	no hit	13.7	no hit	15.7
179	Tris(2-chloroisopropyl)phosphate	13674-84-5	NTP	no hit	no hit	no hit	no hit	no hit
180	Tris(methylphenyl) phosphate	1330-78-5	NTP	7.8	no hit	no hit	no hit	no hit

specific hit
unspecific hit

Similar is true for the neuronal morphology assays (Table 27). One FR affects the morphology of UKN4 and UKN5 neurons, without altering NPC4 neuronal morphology, while two FR reduce neurite length of NPC4, while not changing UKN4 or 5 neurite morphology. Hence, although the majority of assays is negative for the endpoint neuronal morphology the battery identifies different profiles possibly due to time of exposure, cell type and developmental timepoint (discussed in WP7).

Table 27 Effects of FR on neurite morphology assays.

Nr	compound name	CAS #	Selection criteria	UKN4 neurite area	UKN5 peripheral neurite area	NPC4 neurite length	NPC4 neurite area
4	2-Ethylhexyl diphenyl phosphate	1241-94-7	NTP	no hit	no hit	no hit	no hit
5	3,3',5,5'-Tetrabromobisphenol A	79-94-7	in vivo data available	no hit	no hit	3.3	3.4
31	Bis (2-butoxyethyl)phosphat	-		no hit	no hit	no hit	no hit
96	Isodecyl diphenyl phosphate	29761-21-5	NTP	no hit	no hit	no hit	no hit
130	PBDE 47	5436-43-1	human positive	21.6	85.1	no hit	no hit
131	PBDE 99	60348-60-9	human positive	13.4	no hit	no hit	no hit
159	tert-Butylphenyl diphenyl phosphate	56803-37-3		no hit	no hit	15.8	no hit
174	Tri-o-cresyl phosphate	78-30-8		no hit	no hit	no hit	no hit
175	Triphenyl phosphates isopropylated	68937-41-7	NTP	no hit	no hit	no hit	no hit
176	Triphenylphsophat	115-86-6	NTP	no hit	no hit	no hit	no hit
177	Tris (2-chloroethyl)phosphat	115-96-8	NTP	no hit	no hit	no hit	no hit
178	Tris(1,3-dichloro-2-propyl) phosphate	13674-87-8	NTP	no hit	no hit	no hit	no hit
179	Tris(2-chloroisopropyl)phosphate	13674-84-5	NTP	no hit	no hit	no hit	no hit
180	Tris(methylphenyl) phosphate	1330-78-5	NTP	no hit	no hit	no hit	no hit

specific hit
unspecific hit
borderline

Integrating BMC of MSE with exposure levels of BDE-47 and TBBPA will attempt moving from hazard characterization towards risk assessment. Here, uncertainties are very high and thus these data need to be taken with caution. However, cord blood values for BDE-99, BDE-47 and TBBPA were 0.001, 0.002 and 0.006 µM in a Korean (BDEs) and French (TBBPA) cohort, respectively (Table 24). Breast milk concentrations were 0.042 and 0.028 µM for BDE-99 and BDE-47 in Korea and 0.003 for TBBPA in France. Assuming a breast milk intake of 1liter/day, exposure to the three compounds would be 42, 28 and 3 nmol/day. BMC for the MSE for these three compounds are 40 nM for BDE-47, 5.5 µM for BDE-99 and 200 nM for TBBPA. While the effects seen for BDE-99 at 5.5 µM is far away from the estimated daily intake (42 nmol/day) and measured cord blood concentrations (0.001 µM), BDE-47 effects at 40 nM are fairly close to the estimated daily intake (28 nmol/day) and calculated cord blood concentrations with just one order of magnitude apart (0.002 µM). For TBBPA the situation is somewhat in-between. The MSE is affected at 200 nM and the exposure estimated to 3 nmol/day with a cord blood concentration of 0.006 µM measured. More than one order of magnitude lies in-between effective *in vitro* concentrations and measured exposures. However, uncertainties due to no IVIVE calculations do not allow a move from hazard to risk here. A second uncertainty is the effect of mixtures. Even if single compounds seem to be at a too low exposure for posing a risk, adding up exposures very quickly moves into relevant exposure levels. This is especially true for endpoints that are affected by multiple compounds like oligodendrocyte differentiation, NNF and NCC migration. In addition, compounds hitting different endpoints at similar concentrations might increase toxicity in comparison to effects on just one endpoint. Here TBBPA is a good example as it is the compound of the case study with the most affected endpoints.

11.1.2. Uncertainty analysis

From the WoE assessment as well as expert knowledge, there are points of uncertainty that are summarized below:

- incomplete DNT *in vitro* testing battery
- missing *in vitro* kinetics instead working with nominal medium concentrations
- two compounds not evaluated with the NNF assay
- different ranking outcome with the MSE or the DNTPi evaluation

11.1.3. Conclusion and overall assessment

Within a compound set for screening and prioritization, the DNT *in vitro* testing battery is well-suited for identifying the 'bad guys' with high hazard potential specifically in the DNT context. Hits within the DNT *in vitro* testing battery strongly differ from ToxCast hazard identification for FR providing valuable data for prioritization for further testing. However, moving from hazard to risk needs thorough kinetic consideration that to this end was not possible for this case study.

Toxicodynamic considerations reveal rankings of FR within this compound class (Figure 65). Such rankings produced with physiologically-relevant assays for DNT strongly help assessing and prioritizing individual compounds within such a case study. Combining the DNTPi ranking with the MSE ranking reveals the following ranking for the FR (here, the MSE with specific hits is set to the first priority and the DNTPi ranking to the second priority; Table 28). This ranking can be applied for further testing. It will be interesting to see how this ranking changes when e.g. zebrafish data or additional DNT *in vitro* battery assays are added to the evaluation matrix.

Specifically, from the current database we recommend further testing for the compounds producing specific DNT hits. First, compounds should be tested on positive endpoints in equivalent rat test systems (in case the MSE was not found in rat cells). If effects can be reproduced in rat cells, targeted *in vivo* DNT testing is recommended. Here, the type of *in vivo* investigation depends on the respective MSE (and other endpoints within its 3-fold BMC range). It is thought that endpoints *in vivo* should correspond to the *in vitro* MSE. In case rat cells do not show the effects seen in human cells *in vitro,* there might be a species-specific effect due to different toxicodynamics in human and rat brain cells. In this case, more certainty will be gained by follow-up studies in human and rat cells identifying the human-specific target. In case the hazard is clearly human-specific, regulation on *in vitro* in combination with kinetic data might be considered in the future.

Table 28 Combined toxicity ranking of FR using the combined MSE and ToxPi methods.

Ranking	DNTPi		MSE		DNTPi & MSE
	ToxPi Score	Name	Ranking	Name	Ranking*
1	0.560	TBBPA	1	BDE-47	**BDE-47**
2	0.279	BDP	2	TDCPP	**TDCPP**
3	0.252	BDE-47	3	TBBPA	**TBBPA**
4	0.211	BDE-99	4	BDP	**BDP**
5	0.203	IPP	5	IDDP	**BDE-99**
6	0.183	TDCPP	6	TOTP	**IPP**
7	0.138	TCP	7	IPP	**TCP**
8	0.135	EDP	8	EDP	**EDP**
9	0.108	TOTP	9	TCP	**TOTP**
10	0.094	TPHP	10	TPHP	**TPHP**
11	0.068	IDDP	11	TCEP	**IDDP**
12	0.059	TCEP	12	BDE-99	**TCEP**
13	0.000	BBOEP	13	TCPP	**TCPP**
14	0.000	TCPP	14	BBEOP	**BBEOP**

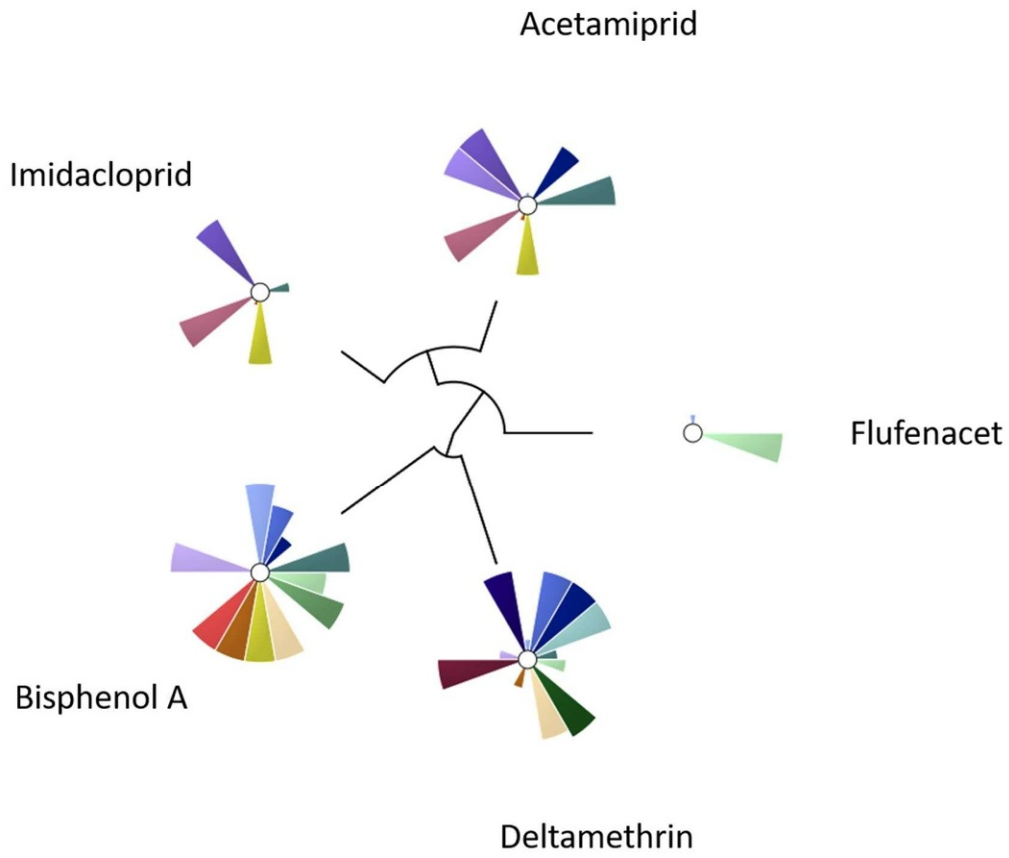
*first priority was given to MSE, in the second line ToxPi ranking was considered, e.g. for compounds with similar MSEs (starting from number 4 in the MSE analysis (Figure 65), due to overlapping 3-fold ranges for the MSE, ToxPi ranking is used).

11.2. Hazard characterization case study, type B:

As a hazard characterization case study, deltamethrin and flufenacet were chosen. These compounds have been evaluated with regards to their *in vitro* and *in vivo* hazards as well as their effects on human brain development by an EFSA working group and are thus very data-rich. Therefore, they are highly suitable as case studies for DNT *in vitro* battery outcomes.

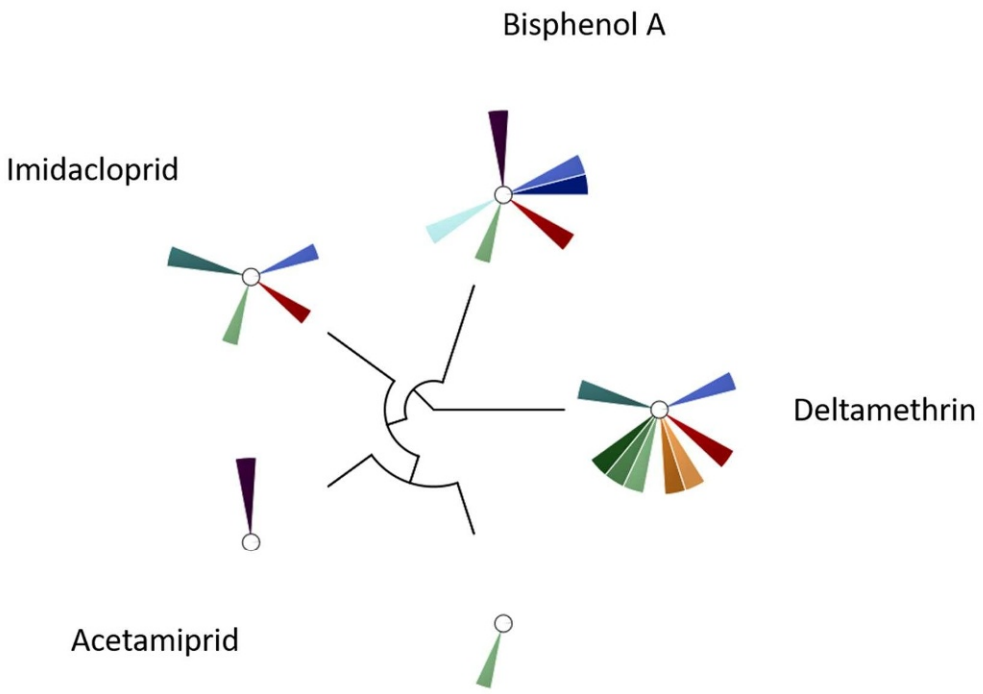
For hazard identification, we again created ToxPis and DNTPies. Here, we included compounds from other case studies because the Toxicological Prioritization Index (ToxPi) Graphical User Interface (GUI) tool needs more than two data sets for comparison. These additional compounds (neonicotinoids and bisphenol A) will be addressed in additional case studies. For deltamethrin and flufenacet we compared the effects that led to a BMC in the different assays across the whole battery and applied the different CM for the data (Figure 25 and Figure 26). Here again, it is important to know that the size of the Pie pieces does not reflect the actual BMC, but relates the BMC for the studied compound with the BMC of this endpoint across the highest and lowest values of the data set irrespective of the values by distributing them between 0 and 1. Hence, it is a relative, not an absolute value. Therefore, Pie slice sizes of different endpoints for one compound do not reflect magnitude of BMC values and cannot be absolutely compared to each other (such data are found in tables of this report).

A)



intended target family	Weight	Metrics	Color	intended target family	Weight	Metrics	Color
background measurement	1 (5,6%)	7		growth factor	1 (5,6%)	2	
cell adhesion molecules	1 (5,6%)	4		ion channel	1 (5,6%)	2	
cell cycle	1 (5,6%)	5		malformation	1 (5,6%)	2	
cell morphology	1 (5,6%)	2		nuclear receptor	1 (5,6%)	29	
cyp	1 (5,6%)	20		oxidoreductase	1 (5,6%)	3	
cytokine	1 (5,6%)	7		protease	1 (5,6%)	2	
dna binding	1 (5,6%)	2		protease inhibitor	1 (5,6%)	1	
esterase	1 (5,6%)	2		steroid hormone	1 (5,6%)	2	
gpcr	1 (5,6%)	5		transporter	1 (5,6%)	4	

B)



DNT endpoint	Weight	Metrics	Color	DNT endpoint	Weight	Metrics	Color
viability (UKN2)	1 (4,0%)	1		migration distance oligodendrocytes (NPC2c)	1 (4,0%)	1	
migration (UKN2)	1 (4,0%)	1		neuronal differentiation (NPC3)	1 (4,0%)	1	
viability (UKN4)	1 (4,0%)	1		neurite length (NPC4)	1 (4,0%)	1	
neurite area (UKN4)	1 (4,0%)	1		neurite area (NPC4)	1 (4,0%)	1	
viability (UKN5)	1 (4,0%)	1		oligodendrocyte differentiation (NPC5)	1 (4,0%)	1	
neurite area (UKN5)	1 (4,0%)	1		cell number (NPC2-5)	1 (4,0%)	1	
proliferation BrdU (NPC1a)	1 (4,0%)	1		cytotoxicity (72 NPC2-5)	1 (4,0%)	1	
proliferation area (NPC1b)	1 (4,0%)	1		cytotoxicity (120h; NPC2-5)	1 (4,0%)	1	
viability (NPC1)	1 (4,0%)	1		viability (120h; NPC2-5)	1 (4,0%)	1	
cytotoxicity (NPC1)	1 (4,0%)	1		network formation (NNF)	1 (4,0%)	1	
migration distance radial glia (NPC2a; 72h)	1 (4,0%)	1		viability (NNF)	1 (4,0%)	1	
migration distance radial glia (NPC2a; 120h)	1 (4,0%)	1		cytotoxicity (NNF)	1 (4,0%)	1	
migration distance neurons (NPC2b)	1 (4,0%)	1					

Figure 67: ToxPies for A) general and B) DNT-specific toxicities using the ToxCast data and the results of the DNT *in vitro* battery, respectively. Graphs were produced with the Toxicological Prioritization Index (ToxPi) Graphical User Interface (GUI) tool version 2.3. Size of pie slices represents strength of effect on respective endpoint

In both, ToxPies and DNTPies for deltamethrin and flufenacet similar potencies of the compounds were observed. While deltamethrin is a compound producing multiple hits in both Pies, flufenacet is much less potent altering one endpoint each (Figure 67). Table 29 specifies the altered endpoints for both compounds. Here, all available data with additional test methods from the US-EPA were taken into consideration: deltamethrin affects with the highest potency rNNF and oligodendrocyte formation (NPC5), while effects of deltamethrin on human NNF (hNNF) and NCC migration (UKN2) were seen at concentrations one or two orders of magnitude higher. These *in vitro* hits are DNT-specific, i.e. not caused by cytotoxicity of the compound. Deltamethrin also affects rat synaptogenesis, rat neurite maturation, human neurite length (NPC4) and area (NPC4, UKN5) and radial glia migration (NPC2a). These endpoints are also affected at higher deltamethrin concentrations and not specific, i.e. cannot be distinguished from effects on cell viability. In contrast, the BMC was not reached by flufenacet in any assay but the NPC5 assay, yet also unspecific for DNT.

Table 29 Summary of BMC across the positive assays of the DNT testing battery. Specific hits according to respective CM are marked in red. Numbers are given in µM.

	Deltamethrin	Flufenacet
BMC_{25} migration (UKN2)	18.4^{s}	> 100
BMC_{25} neurite are (UKN5)	112.8^{ns}	>100
BMC_{10} migration radial glia (120h; NPC2a)	16.3^{ns}	> 20
BMC_{30} neurite length (NPC4)	14.9^{ns}	> 20
BMC_{30} neurite area (NPC4)	15.9^{ns}	> 20
BMC_{30} oligodendrocyte differentiation (NPC5)	0.6^{s}	17.8^{ns}
BMC_{50} rat neuronal network formation (rNNF)	0.5^{s}	> 20
BMC_{50} human neuronal network formation (hNNF)	3.9^{s}	> 20
BMC_{30} neurite maturation (rat cortical neurons 2)	9.8^{ns}	-
BMC_{30} synaptogenesis (rat cortical neurons 2)	8.6^{ns}	-

s = specific
ns= not specifi
Beside UKN2/5 and NPC2-5, the table also includes data that was not generated as part of the report. BMC30 values for neurite maturation and synaptogenesis (rat cortical neurons 2) were derived from Harrill et al. (2018). Flufenacet was not tested in these assays. Data for deltamethrin assessed in the rat neuronal network formation assay was taken from Frank et al. (2017) and flufenacet data from Shafer et al. (2019). hNNF values originate from unpublished data generated at the IUF. Specific hits according to respective CM are marked in red and numbers are given in µM.

For a specific overview of the underlying CR-curves for the BMC of deltamethrin and flufenacet please see Figure 68 and Figure 69.

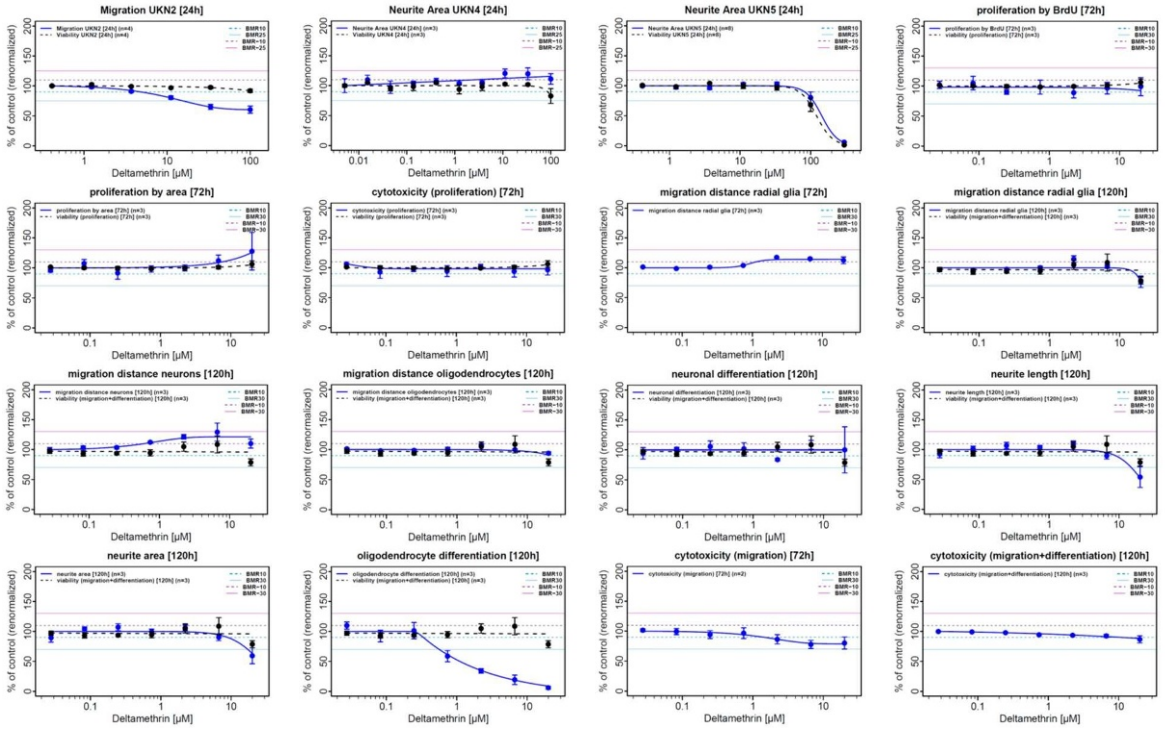


Figure 68: Concentration-response curves for effects of deltamethrin on individual neurodevelopmental endpoints assessed in NPC/UKN assays. Respective BMCs are displayed in Table 29.

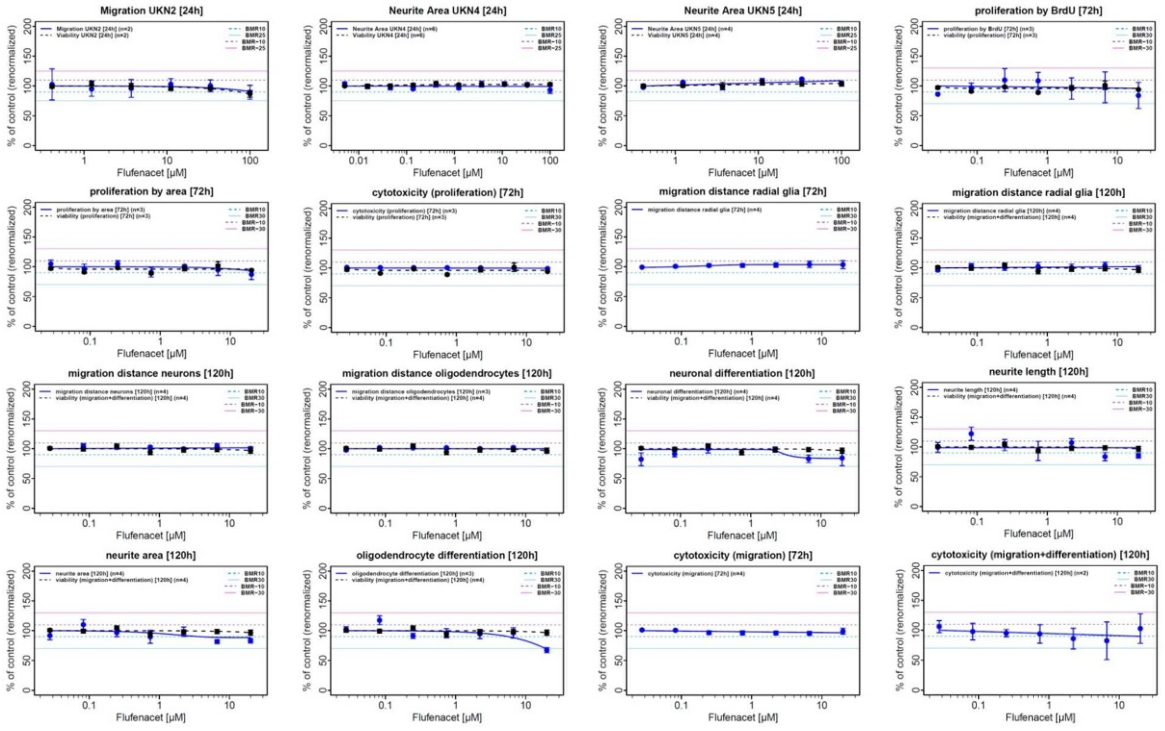


Figure 69: Concentration-response curves for effects of flufenacet on individual neurodevelopmental endpoints assessed in NPC/UKN assays. Respective BMCs are displayed in Table 29 were derived.

In addition to the data produced with the DNT *in vitro* testing battery that was already discussed extensively within this report, we added data of an assay that had not been established previously, the human NNF (hNNF) assay, and is thus unpublished. The establishment of this assay was sponsored by the Danish EPA and we were able to generate the data for deltamethrin and flufenacet with this assay (note: this is not the NNF assay (rat) from the EPA!). This is of particular relevance as due to this assay we are now able to compare the effects of these pesticides directly between the two species, human and rat.

The hNNF assay is based on human pre-differentiated hiPSC-based excitatory and inhibitory neurons as well as primary astroglia commercially offered as the Synfire Kit (Figure 70A; NeuCyte, USA). Human iPS cells are converted into functional iN cells in less than two weeks by forced expression of single transcription factors. Neurogenin-2 (Ngn2) overexpression rapidly transforms hiPSCs into neurons, which form spontaneous excitatory synaptic networks and express typical markers for excitatory cortical neurons like Brn2 and FoxG1 (Zhang *et al*, 2013). The transient expression of the transcription factors Asc1 and dlx2 in hiPSCs induces the generation of GABAergic neurons that exhibit various subtype-specific markers like pan-DLX proteins, GABA, GAD1/2 and vGAT (Yang *et al*, 2017). The Synfire Kit was recently used for investigating acute effects of polyfluoroalkyl substances on spontaneous neuronal network activity and the GABAA receptor. Additionally, comparison between the Synfire iNS co-culture and other neuronal co-cultures revealed a higher activity and excitability of the Synfire iNS (Tukker *et al*, 2020).

In the Assay the cells are plated in 48-well MEA plates (Axion Biosystem, USA) in a defined ratio and differentiated for 7 days. At DIV 7 we perform a baseline recording followed by the start of compound exposure in 7 concentrations. This baseline recording at DIV 7 ensures that the well is active, as few, single spikes are recorded at this timepoint (Figure 70C). However, at this time activity is restricted to single spikes without bursting or network synchronicity. An example of the matured network on DIV35 is given in Figure 70D. Cells are then exposed to the compound DIV 7 to DIV 35 with a weekly recording of the electrical activity and weekly cytotoxicity assessment. To minimize possible acute effects of compounds in the measurements, a washout is performed one day before each recording day. This weekly procedure is repeated until DIV 35 (Figure 70B).

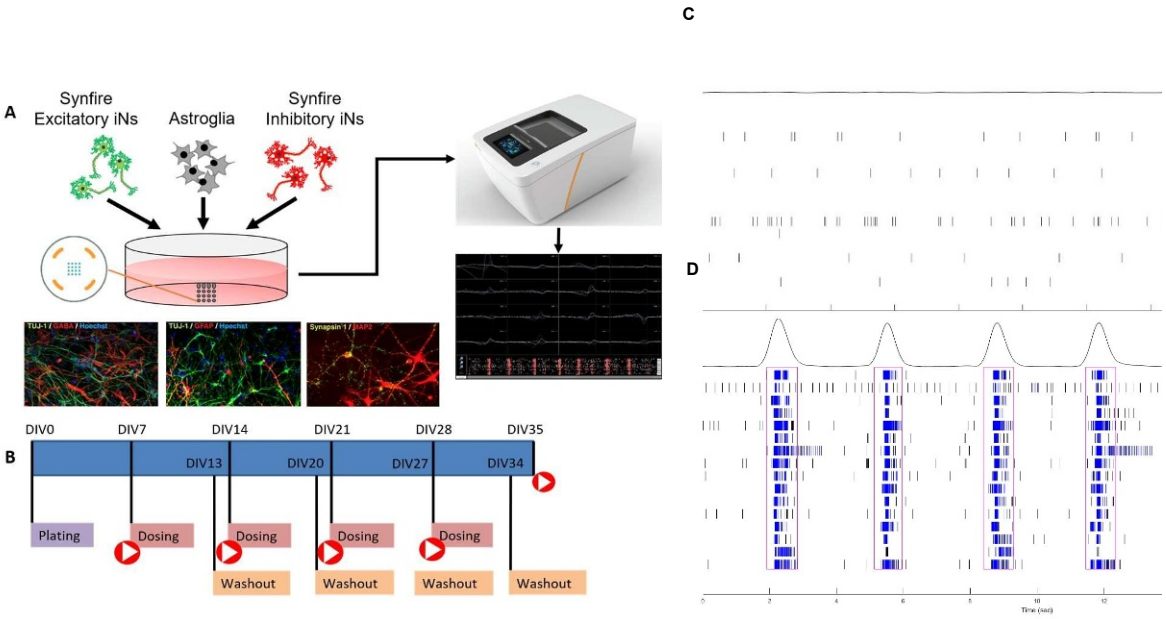


Figure 70: Set-up of the human neuronal network formation (hNNF) assay. A) Cellular and measuring set-up (Synfire Kit, NeuCyte, Maestro Pro, Axion). B) Experimental set-up, treatment, medium change and measuring scheme.

C) Spike-raster plot of cell measurements after 7 DIV with time on the x-axis and individual electrodes on the y-axis. Each line represents a single electrical event. D) Spike raster plots of the same cells than in C) after 35 DIV. Bursts are marked in blue, red boxes indicate network synchronicities.

The AxIS Navigator software (Axion Biosystem, USA) records the network activity analyses data according to different network parameters (e.g. number of spikes, number of active electrodes, burst duration, network burst frequency and different synchronicity parameters).

To obtain an effective concentration (e.g. IC or BMC) for the compounds, we calculated the area under the curve (AUC) for each concentration over time and used these AUCs normalized to the solvent control to generate concentration response curves.

Exposing the human neural cells during network formation starting from DIV 7 allows the assessment of adverse effects of compounds on hNNF. Data for deltamethrin and flufenacet was evaluated with the same machine (Axion Maestro 768-channel amplifier) than data from the rNNF assay. Beside 'mutual information' (same R-package than rNNF assay) all endpoints of the hNNF assay were analyzed using the Axion Software. Due to the striving of keeping as many parameters as possible similar between the test methods with the two species, we aim at reducing uncertainty. Figure 71 shows the results of the MEA testing in the hNNF assay. As a comparison, rat BMC_{50} for the same parameters are added in red. The MSE in the hNNF was 'number of spikes per networkburst per channel'. In contrast, deltamethrin produced a reduction in 'mutual information', a network parameter, as the MSE in the rNNF assay.

For flufenacet, we did not measure any effect in any of the parameters assessed with the MEAs. Examples for the curves are given in Figure 72.

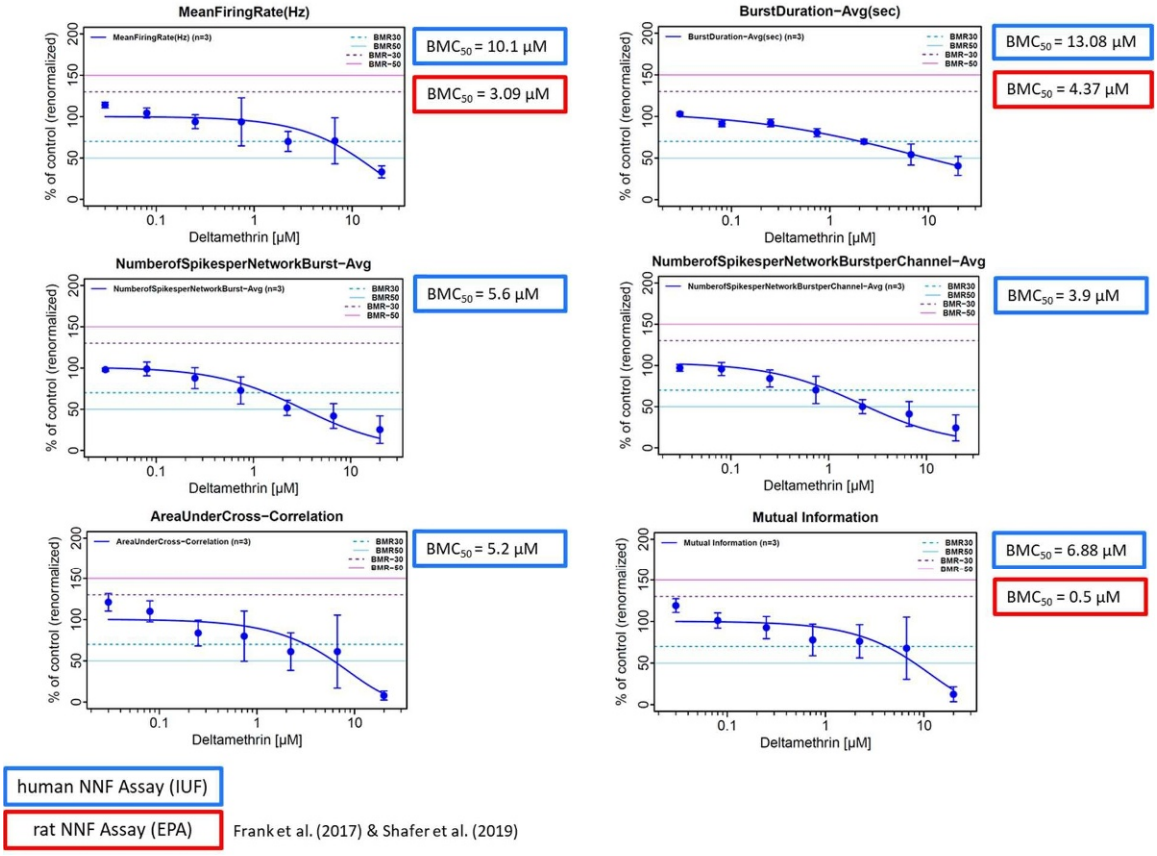


Figure 71: MEA testing results of the hNNF formation assay. Plotted is the effect on 6 out of 72 parameters assessed. Deltamethrin produced a reduction in all of the parameter shown here with BMC50 values between 3.9 µM and 13.08 µM. Some parameters e.g. number of active electrodes, Inter-Burst-Interval and Burst-Frequency were not affected (not shown here).

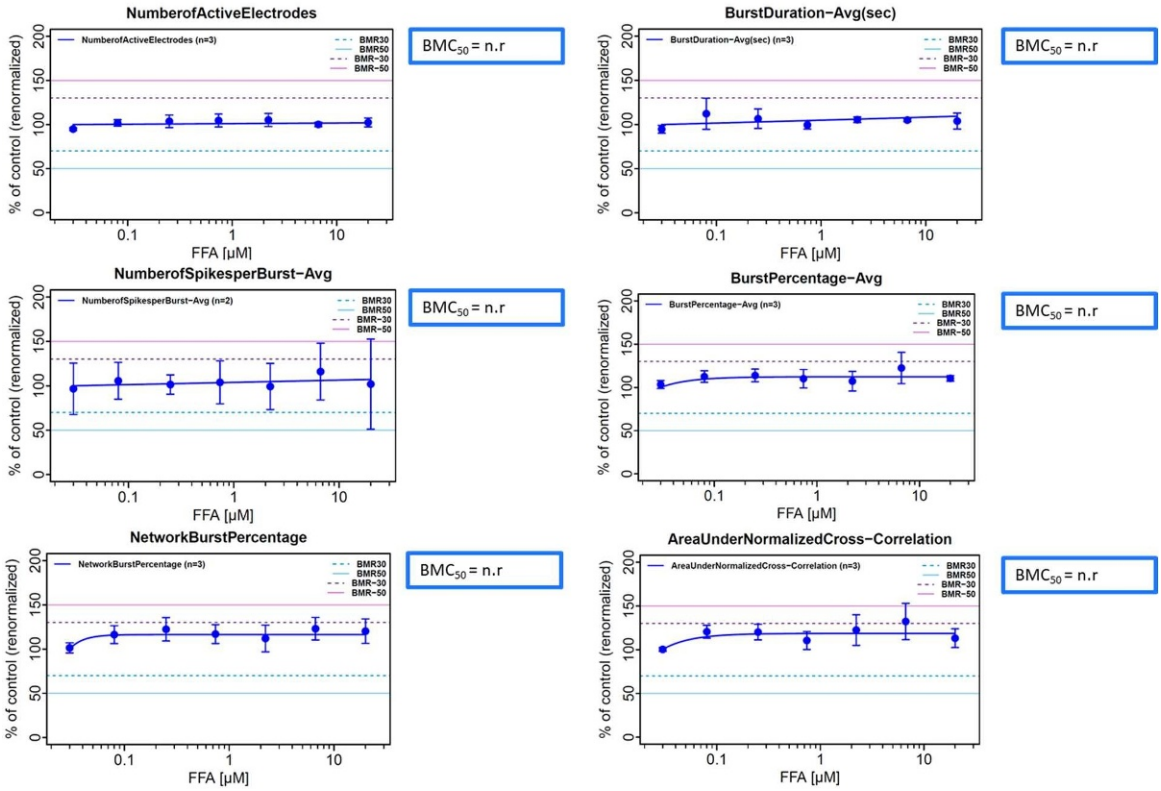


Figure 72: MEA testing results of the hNNF formation assay. Flufenacet did not affect any of the network parameters. Representative negative paramaters are shown here.

Similar to the BMC_{50} of the deltamethrin treated rNNF (0.5 µM), also NPC5 was affected for oligodendrocyte differentiation by this compound, being unaffected by flufenacet (Table 29). Hence, for deltamethrin, the rNNF and human oligodendrocyte differentiation (NPC5) were the two MSE. What could be a possible mechanistic rationale behind the effects on young oligodendrocytes?

The MoA for deltamethrin is interference with closing of voltage-gated sodium channels (VGSC). Therefore, we asked ourselves if also oligodendrocytes might be sensitive to interference with VGSC. Studying the literature revealed that oligodendrocytes, and here mainly O4+ oligodendrocyte precursor cells (OPCs) express VGSC during their maturation (Paez *et al*, 2009; Sontheimer *et al*, 1989; Barres *et al*, 1990). Moreover, these channels are active (Káradóttir *et al*, 2008; Paez *et al*, 2009) and can be modulated by TTX (Káradóttir *et al*, 2008) in OPC of a variety of different species including humans (Livesey *et al*, 2016). It is well known that during development OPC, in contrast to mature oligodendrocytes, are especially sensitive. They have a high sensitivity towards free radicals, excitotoxicity, and pro-inflammatory stimuli. This is the reason why in premature infants, periventricular leukomalacia (PVL), a form of cerebral white matter injury, is the predominant pathology. Pre-OLs are the predominant form of the oligodendroglial lineage in human cerebral white matter during the premature period, with until 28 weeks of gestation, the O4+ cell counting for 90% of oligodendrocyte lineage cells in the brain. 25–50% of premature infants with very low birth weight exhibit cognitive/attentional/behavioral/socialization defects later in life that significantly impair their quality of life. These adverse outcomes in children are caused by lack in oligodendrocytes due to PVL. Hence, there is a mechanistic rationale in humans that links lack of oligodendrocytes in children caused by reduced numbers of OPC during development due to PVL to an adverse outcome. These data are supported by observations in rodents (work on PVL comprehensively

summarized by (Volpe *et al*, 2011b). From these published data, a hypothetical AOP can be assembled that links exposure to compounds that interfere with VGSC to a reduction in OPC *in vitro* and a reduced number of OPC *in vivo* to a human adverse outcome (Figure 73A). Currently, there is no AOP for interference of compounds with VGSC leading to altered NNF. However, the AOP 'Binding of Pyrethroids to Voltage-gated Sodium Channels induces acute neurotoxicity' is already published in the AOP Wiki. Hyperexcitability, however, might also interfere with neuronal network formation and maturation during brain development. Hence, from the same MIE, i.e. 'interference of compound with VGSC', two hypothetical AOPs can be generated. One, involving neurons and the formation of neuronal networks, and a second one involving OPCs (Figure 73B).

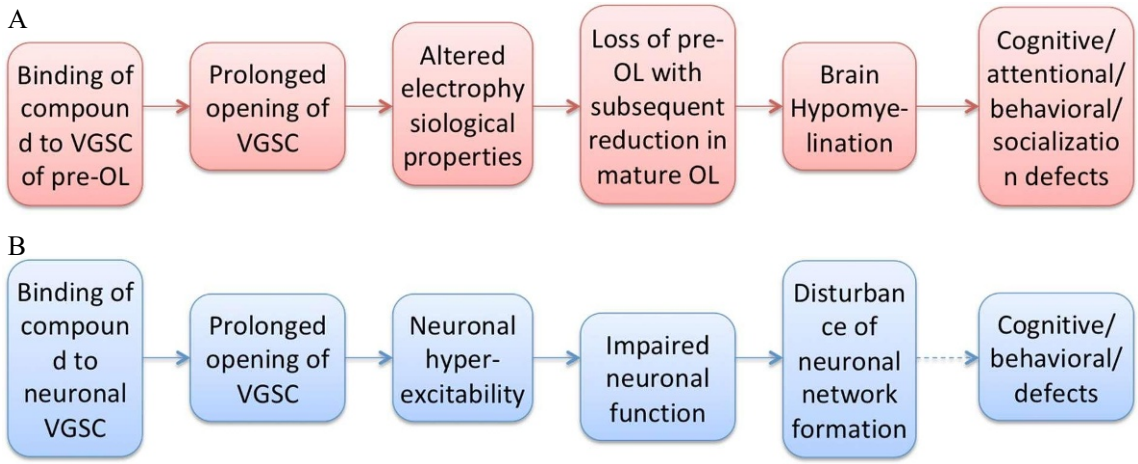


Figure 73: Hypothetical AOPs for compounds binding to VGSC of A) oligodendrocyte precursor cells and B) neurons.

These data will now proceed into an integrated analysis for testing and assessment (IATA) by EFSA. Together with already existing *in vivo* and *in vitro* data the results from the *in vitro* DNT testing battery will contribute to risk management in an IATA context.

In this context, the AOPs provide the toxicity hypothesis, and the *in vitro* testing data provide the threshold concentration at which the AOP is triggered. In a risk assessment context, the BMC identified here needs to be translated to a dose that would lead to the respective fetal brain concentration. This can be determined by reverse physiology-based kinetic modelling (PBK), leading to a quantitative *in vitro* to *in vivo* extrapolation (qIVIVE). The internal dose, determined by qIVIVE can be translated into an external dose, depending on the exposure scenario (e.g. exposure by food, or exposure on the skin) and by modelling the respective barriers (oral bioavailability, dermal bioavailability). With such data known, external exposures can be defined that would lead to the toxic threshold dose, and by using safety factors, an acceptable daily intake (or other measure of exposure) can be defined. Alternatively, known human exposures may be used as starting point (e.g. modelled from average food consumption, or measured in plasma in monitoring studies). The internal plasma concentrations may then be converted by PBK models to fetal brain concentrations, and these can be compared to the threshold values for adversity determined here.

12. Summary and Conclusion

In this project we set up a human cell-based DNT *in vitro* testing strategy that is based on test methods with high readiness and data generated therefrom. The methods underwent a fit-for-purpose evaluation that considered four key elements: 1. The test system, 2. the exposure scheme, 3. the assay and analytical endpoint(s) and 4. the classification model. This testing battery was challenged with 119 chemicals for which rich toxicological information was available (for some of them also on their DNT hazard). Testing was performed in 5 test systems measuring 10 DNT-specific endpoints and additional 9 viability/cytotoxicity-related parameters. For approximately half of the compounds, additional and complementary data from DNT *in vitro* tests was added by the US-EPA. This extended battery was also evaluated.

Testing results revealed that the test methods of this current DNT *in vitro* battery are reliable and reproducible. The endpoints had to a large extent low redundancy. Battery performance, as assessed with compounds well-characterized for DNT hazard had a sensitivity of 82.7 % and a specificity of 88.2 % (sensitivity of 90.5 % and specificity of 100 % for the extended battery). Gap analyses suggested that radial, astro- and microglia as well as myelination endpoints may be added to the battery. Two case studies, one for screening and prioritization of 14 flame retardants, and one on hazard characterization of 2 pesticides, were presented. Hypothetical AOPs were developed based on the latter case study.

In conclusion, the DNT testing strategy explored here is a very promising first approach for DNT hazard identification and characterization. The performance is encouraging and may be improved by inclusion of further tests. Some uncertainties in DNT *in vitro* battery testing outcomes could be reduced by incorporating test data and modelling approaches related to *in vitro* and *in vivo* toxicokinetics of test compounds.

13. References

Abreu-Villaça Y & Levin ED (2017) Developmental neurotoxicity of succeeding generations of insecticides. *Environ. Int.* **99:** 55–77

Aldridge JE, Seidler FJ, Meyer A, Thillai I & Slotkin TA (2003) Serotonergic systems targeted by developmental exposure to chlorpyrifos: Effects during different critical periods. *Environ. Health Perspect.* **111:** 1736–1743

Alépée N, Bahinski A, Daneshian M, De Wever B, Fritsche E, Goldberg A, Hansmann J, Hartung T, Haycock J, Hogberg H, Hoelting L, Kelm JM, Kadereit S, McVey E, Landsiedel R, Leist M, Lübberstedt M, Noor F, Pellevoisin C, Petersohn D, et al (2014) State-of-the-art of 3D cultures (organs-on-a-chip) in safety testing and pathophysiology. *ALTEX* **31:** 441–77 Available at: http://www.ncbi.nlm.nih.gov/pubmed/25027500 [Accessed May 31, 2017]

Andreas K (1993) Efficacy of cerebroprotective substances in the management of functional disorders induced by the cytotoxic brain oedema-producing substance hexachlorophene. *Naunyn. Schmiedebergs. Arch. Pharmacol.* **347:** 79–83

Anelli T, Cardarelli S, Ori M, Nardi I, Biagioni S & Poiana G (2013) 5-Hydroxytryptamine 1A and 2B serotonin receptors in neurite outgrowth: Involvement of early growth response protein 1. *Dev. Neurosci.* **35:** 450–460

Aschner M, Ceccatelli S, Daneshian M, Fritsche E, Hasiwa N, Hartung T, Hogberg HT, Leist M, Li A, Mundy WR, Padilla S, Piersma AH, Bal-Price A, Seiler A, Westerink RH, Zimmer B & Lein PJ (2017) Reference compounds for alternative test methods to indicate developmental neurotoxicity (DNT) potential of chemicals: Example lists & criteria for their selection & use. In *Altex*

Bal-Price A, Crofton KM, Leist M, Allen S, Arand M, Buetler T, Delrue N, FitzGerald RE, Hartung T, Heinonen T, Hogberg H, Bennekou SH, Lichtensteiger W, Oggier D, Paparella M, Axelstad M, Piersma A, Rached E, Schilter B, Schmuck G, et al (2015a) International STakeholder NETwork (ISTNET): creating a developmental neurotoxicity (DNT) testing road map for regulatory purposes. *Arch. Toxicol.* **89:** 269–

287
Bal-Price A, Crofton KM, Sachana M, Shafer TJ, Behl M, Forsby A, Hargreaves A, Landesmann B, Lein PJ, Louisse J, Monnet-Tschudi F, Paini A, Rolaki A, Schrattenholz A, Su C, Thriel C Van, Whelan M, Fritsche E, Suñol C, van Thriel C, et al (2015b) Putative adverse outcome pathways relevant to neurotoxicity. *Crit. Rev. Toxicol.* **8444:** 83–91 Available at: http://www.ncbi.nlm.nih.gov/pubmed/25605028%5Cnhttp://informahealthcare.com/doi/abs/10.3109/10408444.2014.981331

Bal-Price A, Hogberg HT, Crofton KM, Daneshian M, FitzGerald RE, Fritsche E, Heinonen T, Hougaard Bennekou S, Klima S, Piersma AH, Sachana M, Shafer TJ, Terron A, Monnet-Tschudi F, Viviani B, Waldmann T, Westerink RHSS, Wilks MF, Witters H, Zurich M-GG, et al (2018) Recommendation on test readiness criteria for new approach methods in toxicology: Exemplified for developmental neurotoxicity. *ALTEX* **35:** 306–352 Available at: http://www.altex.org/index.php/altex/latest_articles [Accessed March 12, 2018]

Bal-Price A, Lein PJ, Keil KP, Sethi S, Shafer T, Barenys M, Fritsche E, Sachana M & Meek MEB (2016) Developing and applying the adverse outcome pathway concept for understanding and predicting neurotoxicity. *Neurotoxicology* Available at: http://linkinghub.elsevier.com/retrieve/pii/S0161813X16300882

Bal-Price AK, Coecke S, Costa L, Crofton KM, Fritsche E, Goldberg A, Grandjean P, Pamela J. Lein, Abby Li 8 RL, Mundy WR, Padilla S, Persico AM, Seiler AEM & Kreysa J (2012) Conference Report: Advancing the Science of Developmental Neurotoxicity (DNT): Testing for Better Safety Evaluation. *ALTEX* **29:** 202–15

Barenys M, Gassmann K, Baksmeier C, Heinz S, Reverte I, Schmuck M, Temme T, Bendt F, Zschauer TC, Rockel TD, Unfried K, Watjen W, Sundaram SM, Heuer H, Colomina MT & Fritsche E (2016) Epigallocatechin gallate (EGCG) inhibits adhesion and migration of neural progenitor cells in vitro. *Arch Toxicol* Available at: http://www.ncbi.nlm.nih.gov/pubmed/27116294 [Accessed January 5, 2017]

Barres BA, Koroshetz WJ, Swartz KJ, Chun LLY & Corey DP (1990) Ion channel expression by white matter glia: The O-2A glial progenitor cell. *Neuron* **4:** 507–524

Basketter DA, Clewell H, Kimber I, Rossi A, Blaauboer B, Burrier R, Daneshian M, Eskes C, Goldberg A, Hasiwa N, Hoffmann S, Jaworska J, Knudsen TB, Landsiedel R, Leist M, Locke P, Maxwell G, McKim J, McVey EA, Ouédraogo G, et al (2012) A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing - t4 report*. *ALTEX* **29:** 3–91

Baud A, Wessely F, Mazzacuva F, McCormick J, Camuzeaux S, Heywood WE, Little D, Vowles J, Tuefferd M, Mosaku O, Lako M, Armstrong L, Webber C, Cader MZ, Peeters P, Gissen P, Cowley SA & Mills K (2017) Multiplex High-Throughput Targeted Proteomic Assay to Identify Induced Pluripotent Stem Cells. *Anal. Chem.* **89:** 2440–2448

Baumann J, Barenys M, Gassmann K, Fritsche E, Baumann J, Barenys M, Gassmann K & Fritsche E (2014) Comparative Human and Rat "Neurosphere Assay" for Developmental Neurotoxicity Testing. In *Current Protocols in Toxicology* pp 12.21.1-12.21.24. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Baumann J, Dach K, Barenys M, Giersiefer S, Goniwiecha J, Lein PJ & Fritsche E (2015) Application of the Neurosphere Assay for DNT Hazard Assessment: Challenges and Limitations. *Methods Pharmacol. Toxicol.***:** 1–29 Available at: http://link.springer.com/10.1007/7653_2015_49 [Accessed November 28, 2016]

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:** Available at: http://www.ncbi.nlm.nih.gov/pubmed/26216354

Baumann N & Pham-Dinh D (2001) Biology of Oligodendrocyte and Myelin in the Mammalian Central Nervous System. *Physiol. Rev.* **81:** Available at: http://physrev.physiology.org/content/81/2/871.long [Accessed May 6, 2017]

Bell SM, Chang X, Wambaugh JF, Allen DG, Bartels M, Brouwer KLR, Casey WM, Choksi N, Ferguson SS, Fraczkiewicz G, Jarabek AM, Ke A, Lumen A, Lynn SG, Paini A, Price PS, Ring C, Simon TW, Sipes NS, Sprankle CS, et al (2018) In vitro to in vivo extrapolation for high throughput prioritization and decision making. *Toxicol. Vitr.* **47:** 213–227 Available at: http://www.ncbi.nlm.nih.gov/pubmed/29203341 [Accessed March 18, 2019]

Bengoa-Vergniory N & Kypta RM (2015) Canonical and noncanonical Wnt signaling in neural stem/progenitor cells. *Cell. Mol. Life Sci.* **72:** 4157–4172

Bennett D, Bellinger D, Health LB-E & 2016 U (2016) Project TENDR: Targeting environmental neuro-developmental risks. the TENDR consensus statement. *Environ. Health Perspect.* **124:** A118–A122 Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4937840/ [Accessed January 31, 2018]

Béranger R, Hardy EM, Binter AC, Charles MA, Zaros C, Appenzeller BMR & Chevrier C (2020) Multiple pesticides in mothers' hair samples and children's measurements at birth: Results from the French national birth cohort (ELFE). *Int. J. Hyg. Environ. Health* **223:** 22–33

Black JA & Waxman SG (2013) Noncanonical roles of voltage-gated sodium channels. *Neuron* **80:** 280–291

Borrell V & Götz M (2014) Role of radial glial cells in cerebral cortex folding. *Curr. Opin. Neurobiol.* **27:** 39–46 Available at: http://linkinghub.elsevier.com/retrieve/pii/S0959438814000427 [Accessed January 31, 2018]

Briz V, Molina-Molina JM, Sánchez-Redondo S, Fernández MF, Grimalt JO, Olea N, Rodríguez-Farré E & Suñol C (2011) Differential estrogenic effects of the persistent organochlorine pesticides dieldrin, endosulfan, and lindane in primary neuronal cultures. *Toxicol. Sci.* **120:** 413–427

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

Brummelte S, Mc Glanaghy E, Bonnin A & Oberlander TF (2017) Developmental changes in serotonin signaling: Implications for early brain function, behavior and adaptation. *Neuroscience* **342:** 212–231

Bye N, Christie KJ, Turbic A, Basrai HS & Turnley AM (2016) Rho kinase inhibition following traumatic brain injury in mice promotes functional improvement and acute neuron survival but has little effect on neurogenesis, glial responses or neuroinflammation. *Exp. Neurol.* **279:** 86–95

Cameron HA, Hazel TG & McKay RDG (1998) Regulation of neurogenesis by growth factors and neurotransmitters. *J. Neurobiol.* **36:** 287–306

Cariou R, Antignac JP, Zalko D, Berrebi A, Cravedi JP, Maume D, Marchand P, Monteau F, Riu A, Andre F & bizec B Le (2008) Exposure assessment of French women and their newborns to tetrabromobisphenol-A: Occurrence measurements in maternal adipose tissue, serum, breast milk and cord serum. *Chemosphere* **73:** 1036–1041

Carusi A, Davies MR, De Grandis G, Escher BI, Hodges G, Leung KMY, Whelan M, Willett C & Ankley GT (2018) Harvesting the promise of AOPs: An assessment and recommendations. *Sci. Total Environ.* **628**–**629:** 1542–1556

Chambers SM, Qi Y, Mica Y, Lee G, Zhang XJ, Niu L, Bilsland J, Cao L, Stevens E, Whiting P, Shi SH & Studer L (2012) Combined small-molecule inhibition accelerates developmental timing and converts human pluripotent stem cells into nociceptors. *Nat. Biotechnol.* **30:** 715–720

Chang X, Mansouri K, Rai R, Hermes F, Phillips J, Bell S, Allen D, Casey W & Kleinstreuer N (2019) An Open-source , Generalized Workflow for IVIVE Analysis. **:** 1–12

Choi BH (1989) The effects of methylmercury on the developing brain. *Prog Neurobiol* **32:** 447–470

Crofton KM, Mundy WR, Lein PJ, Bal-Price A, Coecke S, Seiler AEM, 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 Available at: http://www.ncbi.nlm.nih.gov/pubmed/21311847 [Accessed July 2, 2017]

Crofton KM, Mundy WR & Shafer TJ (2012) Developmental neurotoxicity testing: A path forward. *Congenit. Anom. (Kyoto).* **52:** 140–146 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22925214 [Accessed March 24, 2017]

Dach K, Bendt F, Huebenthal U, Giersiefer S, Lein PJ, Heuer H & Fritsche E (2017) BDE-99 impairs differentiation of human and mouse NPCs into the oligodendroglial lineage by species-specific modes of action. *Sci. Rep.* **7:** 44861 Available at: http://www.ncbi.nlm.nih.gov/pubmed/28317842 [Accessed April 6, 2017]

Delp J, Gutbier S, Klima S, Hoelting L, Pinto-Gil K, Hsieh J-H, Aichem M, Klein K, Schreiber F, Tice RR, Pastor M, Behl M & Leist M (2018) A high-throughput approach to identify specific neurotoxicants/ developmental toxicants in human neuronal cell function assays. *ALTEX***:** 235–253 Available at: https://www.altex.org/index.php/altex/article/view/311 [Accessed July 4, 2018]

Devakumar D, Bamford A, Ferreira MU, Broad J, Rosch RE, Groce N, Breuer J, Cardoso MA, Copp AJ, Alexandre P, Rodrigues LC & Abubakar I (2018) Infectious causes of microcephaly: epidemiology,

pathogenesis, diagnosis, and management. *Lancet Infect. Dis.* **18:** e1–e13
Dorman DC, Allen SL, Byczkowski JZ, Claudio L, Fisher JE, Fisher JW, Harry GJ, Li AA, Makris SL, Padilla S, Sultatos LG & Mileson BE (2001) Methods to identify and characterize developmental neurotoxicity for human health risk assessment. III: pharmacokinetic and pharmacodynamic considerations. *Environ. Health Perspect.* **109 Suppl 1:** 101–11 Available at: http://www.ncbi.nlm.nih.gov/pubmed/11250810 [Accessed February 17, 2018]
Dotti CG, Esteban JA & Ledesma MD (2014) Lipid dynamics at dendritic spines. *Front. Neuroanat.* **8:** 76
Druwe I, Freudenrich TM, Wallace K, Shafer TJ & Mundy WR (2016) Comparison of Human Induced Pluripotent Stem Cell-Derived Neurons and Rat Primary Cortical Neurons as In Vitro Models of Neurite Outgrowth . *Appl. Vitr. Toxicol.* **2:** 26–36
Edoff K, Raciti M, Moors M, Sundström E & Ceccatelli S (2017) Gestational Age and Sex Influence the Susceptibility of Human Neural Progenitor Cells to Low Levels of MeHg. *Neurotox. Res.* **32:** 683–693
Ehrlich DE & Josselyn SA (2016) Plasticity-related genes in brain development and amygdala-dependent learning. *Genes, Brain Behav.* **15:** 125–143
Eskes C, Honegger P, Juillerat-Jeanneret L & Monnet-Tschudi F (2002) Microglial reaction induced by noncytotoxic methylmercury treatment leads to neuroprotection via interactions with astrocytes and IL-6 release. *Glia* **37:** 43–52
Fisher C, Siméon S, Jamei M, Gardner I & Bois YF (2019) VIVD: Virtual in vitro distribution model for the mechanistic prediction of intracellular concentrations of chemicals in in vitro toxicity assays. *Toxicol. Vitr.* **58:** 42–50
Frank CL, Brown JP, Wallace K, Mundy WR & Shafer TJ (2017) From the Cover: Developmental Neurotoxicants Disrupt Activity in Cortical Networks on Microelectrode Arrays: Results of Screening 86 Compounds During Neural Network Formation. *Toxicol. Sci.* **160:** 121–135
Frank CL, Brown JP, Wallace K, Wambaugh JF, Shah I & Shafer TJ (2018) Defining toxicological tipping points in neuronal network development. *Toxicol. Appl. Pharmacol.* **354:**
French HM, Reid M, Mamontov P, Simmons RA & Grinspan JB (2009) Oxidative stress disrupts oligodendrocyte maturation. *J. Neurosci. Res.* **87:** 3076–3087
Fritsche E (2016) Report on integrated testing strategies for the identi cation and evaluation of chemical hazards associ- ated with the developmental neurotoxicity (DNT):In Report of the OECD/EFSA workshop on developmental neurotoxic- ity (DNT): The use of non-animal test. *OECD Environ. Heal. Saf. Pub- lications Ser. Test. Assess.* **242:**
Fritsche E, Alm H, Baumann J, Geerts L, Hakansson H, Masjosthusmann S & Witters H (2015) Literature review on in vitro and alternative Developmental Neurotoxicity (DNT) testing methods. *EFSA Support. Publ.* **778:** 1–186
Fritsche E, Cline JE, Nguyen N-H, Scanlan TS & Abel J (2005) Polychlorinated biphenyls disturb differentiation of normal human neural progenitor cells: clue for involvement of thyroid hormone receptors. *Environ. Health Perspect.* **113:** 871–876
Fritsche E, Crofton KM, Hernandez AF, Hougaard Bennekou S, Leist M, Bal-Price A, Reaves E, Wilks MF, Terron A, Solecki R & Gourmelon A (2017) OECD/EFSA workshop on developmental neurotoxicity (DNT): The use of non-animal test methods for regulatory purposes. *ALTEX* **34:** 311–315 Available at: http://www.ncbi.nlm.nih.gov/pubmed/28407175 [Accessed May 2, 2017]
Fu Y, Wu X, Lu J & Huang J (2012) Presynaptic GABAB receptor regulates activity-dependent maturation and patterning of inhibitory synapses through dynamic allocation of synaptic vesicles. *Front. Cell. Neurosci.* **6:** 57
Fuller LC, Cornelius SK, Murphy CW & Wiens DJ (2002) Neural crest cell motility in valproic acid. *Reprod. Toxicol.* **16:** 825–839
Fürst R (2016) Narciclasine - an Amaryllidaceae Alkaloid with Potent Antitumor and Anti-Inflammatory Properties. *Planta Med.* **82:** 1389–1394
Garcia SJ, Seidler FJ & Slotkin TA (2003) Developmental neurotoxicity elicited by prenatal or postnatal chlorpyrifos exposure: Effects on neurospecific proteins indicate changing vulnerabilities. *Environ. Health Perspect.* **111:** 297–303
Gassmann K, Abel J, Bothe H, Haarmann-Stemmann T, Merk HF, Quasthoff KN, Rockel TD, Schreiber T & Fritsche E (2010) Species-specific differential ahr expression protects human neural progenitor cells against developmental neurotoxicity of PAHs. *Environ. Health Perspect.* **118:** 1571–1577
Gassmann K, Baumann J, Giersiefer S, Schuwald J, Schreiber T, Merk HF & Fritsche E (2012) Automated

neurosphere sorting and plating by the COPAS large particle sorter is a suitable method for high-throughput 3D in vitro applications. *Toxicol. Vitr.* **26:** 993–1000 Available at: http://dx.doi.org/10.1016/j.tiv.2012.04.025

Gąssowska M, Baranowska-Bosiacka I, Moczydłowska J, Frontczak-Baniewicz M, Gewartowska M, Strużyńska L, Gutowska I, Chlubek D & Adamczyk A (2016) Perinatal exposure to lead (Pb) induces ultrastructural and molecular alterations in synapses of rat offspring. *Toxicology* **373:** 13–29

Giordano G, Kavanagh TJ & Costa LG (2009) Mouse cerebellar astrocytes protect cerebellar granule neurons against toxicity of the polybrominated diphenyl ether (PBDE) mixture DE-71. *Neurotoxicology* **30:** 326–329

Grandjean P & Landrigan PJ (2006) Developmental neurotoxicity of industrial chemicals. *Lancet* **368:** 2167–2178 Available at: http://ac.els-cdn.com/S0140673606696657/1-s2.0-S0140673606696657-main.pdf?_tid=cd7f1e8a-c3ec-11e4-bda3-00000aacb35d&acdnat=1425638435_f9ffc36f1f4a4da8284df71e68992714

Grandjean P & Landrigan PJ (2014) Neurobehavioural effects of developmental toxicity. *Lancet Neurol.* **13:** 330–338 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24556010 [Accessed August 18, 2016]

Gyamfi C, Wapner RJ & D'Alton ME (2009) Thyroid dysfunction in pregnancy: The basic science and clinical evidence surrounding the controversy in management. *Obstet. Gynecol.* **113:** 702–707

Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen HK, More S, Naegeli H, Noteborn H, Ockleford C, Ricci A, Rychen G, Schlatter JR, Silano V, Solecki R, Turck D, Benfenati E, Chaudhry QM, Craig P, Frampton G, Greiner M, et al (2017a) Guidance on the use of the weight of evidence approach in scientific assessments. *EFSA J.* **15:** Available at: http://doi.wiley.com/10.2903/j.efsa.2017.4971 [Accessed December 13, 2019]

Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen KH, More S, Mortensen A, Naegeli H, Noteborn H, Ockleford C, Ricci A, Rychen G, Silano V, Solecki R, Turck D, Aerts M, Bodin L, Davis A, Edler L, Gundert-Remy U, et al (2017b) Update: use of the benchmark dose approach in risk assessment. *EFSA J.* **15:** 1–41

Harrill JA, Freudenrich T, Wallace K, Ball K, Shafer TJ & Mundy WR (2018) Testing for developmental neurotoxicity using a battery of in vitro assays for key cellular events in neurodevelopment. *Toxicol. Appl. Pharmacol.* Available at: http://www.ncbi.nlm.nih.gov/pubmed/29626487 [Accessed April 18, 2018]

Harrill JA, Robinette BL & Mundy WR (2011) Use of high content image analysis to detect chemical-induced changes in synaptogenesis in vitro. *Toxicol. Vitr.* **25:** 368–387

Hartung T (2007) Food for thought ... on cell culture. *ALTEX* **24:** 143–147

Hartung T (2010) Evidence-Based Toxicology - The Toolbox of Validation for the 21 st Century? *ALTEX* **27:** 253–263

Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, Gribaldo L, Halder M, Hoffmann S, Roi AJ, 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:** 467–72 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15656771 [Accessed July 3, 2017]

Haydar TF, Wang F, Schwartz ML & Rakic P (2000) Differential modulation of proliferation in the neocortical ventricular and subventricular zones. *J. Neurosci.* **20:** 5764–5774

Hevner RF (2015) Brain overgrowth in disorders of RTK-PI3K-AKT signaling: A mosaic of malformations. *Semin. Perinatol.* **39:** 36–43

Hoelting L, Klima S, Karreman C, Grinberg M, Meisig J, Henry M, Rotshteyn T, Rahnenführer J, Blüthgen 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:** 476–87 Available at: http://www.ncbi.nlm.nih.gov/pubmed/26933043 [Accessed April 4, 2017]

Hofrichter M, Nimtz L, Tigges J, Kabiri Y, Schröter F, Royer-Pokorac B, Hildebrandt B, Schmuck M, Epanchintsev A, Theiss S, Adjaye J, Egly J-M, Krutmann J & Fritsche E (2017) Comparative performance analysis of human iPSC-derived and primary neural progenitor cells (NPC) grown as neurospheres in vitro. *Stem Cell Res.***:** under revision

Huang X & Saint-Jeannet J-P (2004) Induction of the neural crest and the opportunities of life on the edge. *Dev. Biol.* **275:** 1–11

Iglesias-González A, Hardy EM & Appenzeller BMR (2020) Cumulative exposure to organic pollutants of French children assessed by hair analysis. *Environ. Int.* **134:**

Ikonomidou C, Bittigau P, Koch C, Genz K, Hoerster F, Felderhoff-Mueser U, Tenkova T, Dikranian K & Olney JW (2001) Neurotransmitters and apoptosis in the developing brain. *Biochem. Pharmacol.* **62:** 401–405

Imayoshi I, Shimojo H, Sakamoto M, Ohtsuka T & Kageyama R (2013) Genetic visualization of notch signaling in mammalian neurogenesis. *Cell. Mol. Life Sci.* **70:** 2045–2057

Kang DS, Yang YR, Lee C, Kim SB, Ryu SH & Suh PG (2016) Roles of phosphoinositide-specific phospholipase Cγ1 in brain development. *Adv. Biol. Regul.* **60:** 167–173

Kanno T, Sasaki S, Yamada N, Kawasako K & Tsuchitani M (2012) Hexachlorophene and cuprizone induce the spongy change of the developing rat brain by different mechanisms: The role of 2′, 3′-cyclic nucleotide 3′-phosphodiesterase (CNPase). *J. Vet. Med. Sci.* **74:** 837–843

Káradóttir R, Hamilton NB, Bakiri Y & Attwell D (2008) Spiking and nonspiking classes of oligodendrocyte precursor glia in CNS white matter. *Nat. Neurosci.* **11:** 450–456

Kaufmann W (2003) Current status of developmental neurotoxicity: an industry perspective. *Toxicol. Lett.* **140**–**141:** 161–9 Available at: http://www.ncbi.nlm.nih.gov/pubmed/12676463 [Accessed February 17, 2018]

Kirmse K, Hübner CA, Isbrandt D, Witte OW & Holthoff K (2018) GABAergic Transmission during Brain Development: Multiple Effects at Multiple Stages. *Neuroscientist* **24:** 36–53 Available at: http://www.ncbi.nlm.nih.gov/pubmed/28378628 [Accessed December 13, 2019]

Kiryushko D, Korshunova I, Berezin V & Bock E (2006) Neural cell adhesion molecule induces intracellular signaling via multiple mechanisms of Ca2+ homeostasis. *Mol. Biol. Cell* **17:** 2278–2286

Komuro H & Rakic P (1993) Modulation of neuronal migration by NMDA receptors. *Science (80-.).* **260:** 95–97

Krebs A, Nyffeler J, Karreman C, Schmidt BZ, Kappenberg F, Mellert J, Pallocca G, Pastor M, Rahnenführer 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 Available at: https://pubmed.ncbi.nlm.nih.gov/31833558/ [Accessed July 1, 2020]

Krebs A, Nyffeler J, Rahnenführer J & Leist M (2018) Normalization of data for viability and relative cell function curves. *ALTEX* **35:** 268–271 Available at: http://www.ncbi.nlm.nih.gov/pubmed/29984806 [Accessed March 4, 2019]

Krebs A, van Vugt-Lussenburg BMA, Waldmann T, Albrecht W, Boei J, ter Braak B, Brajnik M, Braunbeck T, Brecklinghaus T, Busquet F, Dinnyes A, Dokler J, Dolde X, Exner TE, Fisher C, Fluri D, Forsby A, Hengstler JG, Holzer AK, Janstova Z, et al (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 Available at: https://doi.org/10.1007/s00204-020-02802-6 [Accessed August 27, 2020]

Krebs A, Waldmann T, Wilks MF, Van Vugt-Lussenburg BMA, Van der Burg B, Terron A, Steger-Hartmann T, Ruegg J, Rovida C, Pedersen E, Pallocca G, Luijten M, Leite SB, Kustermann S, Kamp H, Hoeng J, Hewitt P, Herzler M, Hengstler JG, Heinonen T, 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

Krug AK, Balmer N V., Matt F, Schonenberger F, Merhof D, Leist M, Schönenberger F, Merhof D & Leist M (2013) Evaluation of a human neurite growth assay as specific screen for developmental neurotoxicants. *Arch Toxicol* **87:** 2215–2231

Lafourcade CA, Lin T V., Feliciano DM, Zhang L, Hsieh LS & Bordey A (2013) Rheb activation in subventricular zone progenitors leads to heterotopia, ectopic neuronal differentiation, and rapamycin-sensitive olfactory micronodules and dendrite hypertrophy of newborn neurons. *J. Neurosci.* **33:** 2419–2431

Lampert P, O'Brien J & Garrett R (1973) Hexachlorophene encephalopathy. *Acta Neuropathol.* **23:** 326–333

Lee DY (2015) Roles of mTOR Signaling in Brain Development. *Exp. Neurobiol.* **24:** 177–185

Lee G, Papapetrou EP, Kim H, Chambers SM, Tomishima MJ, Fasano CA, Ganat YM, Menon J, Shimizu F, Viale A, Tabar V, Sadelain M & Studer L (2009) Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature* **461:** 402–406

Lee I, Eriksson P, Fredriksson A, Buratovic S & Viberg H (2015) Developmental neurotoxic effects of two pesticides: Behavior and neuroprotein studies on endosulfan and cypermethrin. *Toxicology* **335:** 1–10

Lefranc F, Sauvage S, Van Goietsenoven G, Mégalizzi V, Lamoral-Theys D, Debeir O, Spiegl-Kreinecker S, Berger W, Mathieu V, Decaestecker C & Kiss R (2009) Narciclasine, a plant growth modulator, activates Rho and stress fibers in glioblastoma cells. *Mol. Cancer Ther.* **8:** 1739–1750

Lein P, Silbergeld E, Locke P & Goldberg AM (2005) In vitro and other alternative approaches to developmental neurotoxicity testing (DNT). In *Environmental Toxicology and Pharmacology* pp 735–744. Available at: https://www.sciencedirect.com/science/article/pii/S1382668904003060 [Accessed January 31, 2018]

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 (2013a) Reprint: Inflammatory findings on species extrapolations: Humans are definitely no 70-kg mice1. *ALTEX* **30:** 227–230

Leist M & Hartung T (2013b) Inflammatory findings on species extrapolations: humans are definitely no 70-kg mice. *Arch. Toxicol.* **87:** 563–567

Leist M, Hasiwa N, Daneshian M & Hartung T (2012) Validation and quality control of replacement alternatives - Current status and future challenges. *Toxicol. Res. (Camb).* **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 TB, McKim J, Harris G, Pamies D & Hartung T (2014) Consensus report on the future of animal-free systemic toxicity testing. *ALTEX* **31:** 341–56 Available at: http://www.ncbi.nlm.nih.gov/pubmed/25061899 [Accessed July 4, 2017]

Leontaridou M, Urbisch D, Kolle SN, Ott K, Mulliner DS, Gabbert S & Landsiedel R (2017) The borderline range of toxicological methods: Quantification and implications for evaluating precision. *ALTEX* **34:** 525–538 Available at: http://www.ncbi.nlm.nih.gov/pubmed/28230889 [Accessed March 4, 2019]

Lewandowski TA (2003) Effect of Methylmercury on Midbrain Cell Proliferation during Organogenesis: Potential Cross-Species Differences and Implications for Risk Assessment. *Toxicol. Sci.* **75:** 124–133 Available at: https://academic.oup.com/toxsci/article-lookup/doi/10.1093/toxsci/kfg151 [Accessed December 13, 2019]

Livesey MR, Magnani D, Cleary EM, Vasistha NA, James OT, Selvaraj BT, Burr K, Story D, Shaw CE, Kind PC, Hardingham GE, Wyllie DJA & Chandran S (2016) Maturation and electrophysiological properties of human pluripotent stem cell-derived oligodendrocytes. *Stem Cells* **34:** 1040–1053

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

Masjosthusmann S, Becker D, Petzuch B, Klose J, Siebert C, Deenen R, Barenys M, Baumann J, Dach K, Tigges J, Hübenthal U, Köhrer K & Fritsche E (2018) A transcriptome comparison of time-matched developing human, mouse and rat neural progenitor cells reveals human uniqueness. *Toxicol. Appl. Pharmacol.*

Masjosthusmann S, Siebert C, Hübenthal U, Bendt F, Baumann J & Fritsche E (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 Available at: http://www.ncbi.nlm.nih.gov/pubmed/31272005 [Accessed September 23, 2019]

Menegola E, Broccia ML, Di Renzo F, Massa V & Giavini E (2005) Craniofacial and axial skeletal defects induced by the fungicide triadimefon in the mouse. *Birth Defects Res. Part B Dev. Reprod. Toxicol.* **74:** 185–195

Moors M, Bose R, Johansson-Haque K, Edoff K, Okret S & Ceccatelli S (2012) Dickkopf 1 mediates glucocorticoid-induced changes in human neural progenitor cell proliferation and differentiation. *Toxicol. Sci.* **125:** 488–495 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22048647 [Accessed September 30, 2016]

Moors M, Cline JE, Abel J & Fritsche E (2007) ERK-dependent and -independent pathways trigger human neural progenitor cell migration. *Toxicol. Appl. Pharmacol.* **221:** 57–67

Moors M, Rockel TD, Abel J, Cline JE, Gassmann K, Schreiber T, Shuwald J, Weinmann N & Fritsche E (2009) Human neurospheres as three-dimensional cellular systems for developmental neurotoxicity testing. *Environ. Health Perspect.* **117:** 1131–1138

Mundy WR, Padilla S, Breier JM, Crofton KM, Gilbert ME, Herr DW, Jensen KF, Radio NM, Raffaele KC & Schumacher K (2015) Expanding the test set: Chemicals with potential to disrupt mammalian brain development. *Neurotoxicol Teratol* **52:** 25–35 Available at: http://ac.els-

cdn.com/S0892036215300362/1-s2.0-S0892036215300362-main.pdf?_tid=57125766-83b4-11e5-a437-00000aab0f6c&acdnat=1446724808_76af14fa89f23883930600e652c46d76

Murthy S, Niquille M, Hurni N, Limoni G, Frazer S, Chameau P, Van Hooft JA, Vitalis T & Dayer A (2014) Serotonin receptor 3A controls interneuron migration into the neocortex. *Nat. Commun.* **5:** 1–10

Nam SM, Cho IS, Seo JS, Go TH, Kim JH, Nahm SS, Chang BJ & Lee JH (2019) Ascorbic Acid Attenuates Lead-Induced Alterations in the Synapses in the Developing Rat Cerebellum. *Biol. Trace Elem. Res.* **187:** 142–150

Neal AP & Guilarte TR (2010) Molecular neurobiology of lead (Pb2+): Effects on synaptic function. *Mol. Neurobiol.* **42:** 151–160

Neal AP, Stansfield KH, Worley PF, Thompson RE & Guilarte TR (2010) Lead exposure during synaptogenesis alters vesicular proteins and impairs vesicular release: Potential role of NMDA receptor-dependent BDNF signaling. *Toxicol. Sci.* **116:** 249–263

Nguyen L, Rigo JM, Rocher V, Belachew S, Malgrange B, Rogister B, Leprince P & Moonen G (2001) Neurotransmitters as early signals for central nervous system development. *Cell Tissue Res.* **305:** 187–202

Ni M, Li X, Rocha JBT, Farina M & Aschner M (2012) Glia and methylmercury neurotoxicity. In *Journal of Toxicology and Environmental Health - Part A: Current Issues* pp 1091–1101. Taylor and Francis Inc.

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 *Neuromethods* pp 141–168. Humana Press Inc.

Nyffeler J, Karreman C, Leisner H, Kim YJ, 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 (2014) OECD Series on Testing and Assessment. Number 211. Guidance Document for describing non-guideline in vitro test methods. ENV/JM/MONO(2014)35. Available at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2014)35&doclanguage=en [Accessed February 17, 2018]

OECD (2018) Guidance Document on Good In Vitro Method Practices (GIVIMP) for the Development and Implementation of In Vitro Methods for Regulatory Use in Human Safety Assessment. **2017:** 1–156 Available at: http://www.oecd.org/env/ehs/testing/OECD Final Draft GIVIMP.pdf

Ojeda J & Ávila A (2019) Early Actions of Neurotransmitters During Cortex Development and Maturation of Reprogrammed Neurons. *Front. Synaptic Neurosci.* **11:** 33

Paez PM, Fulton D, Colwell CS & Campagnoni AT (2009) Voltage-operated Ca2+ and Na+ channels in the oligodendrocyte lineage. *J. Neurosci. Res.* **87:** 3259–3266

Pamies D, Bal-Price A, Simeonov A, Tagle D, Allen D, Gerhold D, Yin D, Pistollato F, Inutsuka T, Sullivan K, Stacey G, Salem H, Leist M, Daneshian M, Vemuri MC, McFarland R, Coecke S, Fitzpatrick SC, Lakshmipathy U, Mack A, et al (2016) Good Cell Culture Practice for stem cells and stem-cell-derived models. *ALTEX* **34:** Available at: https://doi.org/10.14573/altex.1607121 [Accessed February 23, 2017]

Park SY, Kang MJ & Han JS (2018) Interleukin-1 beta promotes neuronal differentiation through the Wnt5a/RhoA/JNK pathway in cortical neural precursor cells. *Mol. Brain* **11:** 39

Pistollato F, De Gyves EM, Carpi D, Bopp SK, Nunes C, Worth A & Bal-Price A (2020) Assessment of developmental neurotoxicity induced by chemical mixtures using an adverse outcome pathway concept. *Environ. Heal. A Glob. Access Sci. Source* **19:** 23

Pistollato F, Louisse J, Scelfo B, Mennecozzi M, Accordi B, Basso G, Gaspar JA, Zagoura D, Barilari M, Palosaari T, Sachinidis A & Bremer-Hoffmann S (2014) Development of a pluripotent stem cell derived neuronal model to identify chemically induced pathway perturbations in relation to neurotoxicity: Effects of CREB pathway inhibition. *Toxicol. Appl. Pharmacol.* **280:** 378–388

Pizzurro DM, Dao K & Costa LG (2014) Diazinon and diazoxon impair the ability of astrocytes to foster neurite outgrowth in primary hippocampal neurons. *Toxicol. Appl. Pharmacol.* **274:** 372–382

Pugh PC & Berg DK (1994) Neuronal acetylcholine receptors that bind α-bungarotoxin mediate neurite retraction in a calcium-dependent manner. *J. Neurosci.* **14:** 889–896

Qiao D, Seidler FJ, Abreu-Villaça Y, Tate CA, Cousins MM & Slotkin TA (2004) Chlorpyrifos exposure during neurulation: Cholinergic synaptic dysfunction and cellular alterations in brain regions at adolescence and adulthood. *Dev. Brain Res.* **148:** 43–52

Quasthoff S & Hartung HP (2002) Chemotherapy-induced peripheral neuropathy. *J. Neurol.* **249:** 9–17
Rogers SW, Gregori NZ, Carlson N, Gahring LC & Noble M (2001) Neuronal nicotinic acetylcholine receptor expression by O2A/oligodendrocyte progenitor cells. *Glia* **33:** 306–313
Roth AD & Núñez MT (2016) Oligodendrocytes: Functioning in a delicate balance between high metabolic requirements and oxidative damage. *Adv. Exp. Med. Biol.* **949:** 167–181
Roy TS, Andrews JE, Seidler FJ & Slotkin TA (1998) Chlorpyrifos elicits mitotic abnormalities and apoptosis in neuroepithelium of cultured rat embryos. *Teratology* **58:** 62–68
Sachana M, Rolaki A & Bal-Price A (2018) Development of the Adverse Outcome Pathway (AOP): Chronic binding of antagonist to N-methyl-D-aspartate receptors (NMDARs) during brain development induces impairment of learning and memory abilities of children. *Toxicol. Appl. Pharmacol.* **354:** 153–175 Available at: /pmc/articles/PMC6095943/?report=abstract [Accessed August 24, 2020]
Salazar P, Velasco-Velázquez MA & Velasco I (2008) GABA effects during neuronal differentiation of stem cells. *Neurochem. Res.* **33:** 1546–1557
Schantz SL & Widholm JJ (2001) Cognitive effects of endocrine-disrupting chemicals in animals. *Environ. Health Perspect.* **109:** 1197–1206
Schmidt CW (2013) Beyond uncertainty factors: Protecting the tails of the bell curve. *Environ. Health Perspect.* **121:** a26–a29 Available at: http://ehp.niehs.nih.gov/121-A26 [Accessed March 23, 2017]
Schmuck MR, 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:** 1–12 Available at: http://www.ncbi.nlm.nih.gov/pubmed/27722930 [Accessed February 6, 2017]
Scholz D, Pöltl 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
Schreiber T, Gassmann K, Goetz C, Huebenthal U, Moors M, Krause G, Merk HF, Nguyen NH, Scanlan TS, Abel J, Rose CR & Fritsche E (2010) Polybrominated diphenyl ethers induce developmental neurotoxicity in a human in vitro model: Evidence for endocrine disruption. *Environ. Health Perspect.* **118:** 572–578
Shafer TJ (2019) Application of Microelectrode Array Approaches to Neurotoxicity Testing and Screening. In *Advances in neurobiology* pp 275–297. Available at: http://www.ncbi.nlm.nih.gov/pubmed/31073941 [Accessed May 13, 2019]
Shafer TJ, Brown JP, Lynch B, Davila-Montero S, Wallace K & Friedman KP (2019) Evaluation of Chemical Effects on Network Formation in Cortical Neurons Grown on Microelectrode Arrays. *Toxicol. Sci.* **169:** 436–455
Shah R, Courtiol E, Castellanos FX & Teixeira CM (2018) Abnormal serotonin levels during perinatal development lead to behavioral deficits in adulthood. *Front. Behav. Neurosci.* **12:**
Sheets LP, Li AA, Minnema DJ, Collier RH, Creek MR & Peffer RC (2015) A critical review of neonicotinoid insecticides for developmental neurotoxicity. *Crit. Rev. Toxicol.* **46:** 153–190
Silva MH (2020) Effects of low-dose chlorpyrifos on neurobehavior and potential mechanisms: A review of studies in rodents, zebrafish, and Caenorhabditis elegans. *Birth Defects Res.* **112:** 445–479
Slotkin TA, Cho H & Whitmore WL (1987) Effects of prenatal nicotine exposure on neuronal development: Selective actions on central and peripheral catecholaminergic pathways. *Brain Res. Bull.* **18:** 601–611
Sontheimer H, Perouansky M, Hoppe D, Lux HD, Grantyn R & Kettenmann H (1989) Glial cells of the oligodendrocyte lineage express proton-activated Na+ channels. *J. Neurosci. Res.* **24:** 496–500 Available at: http://doi.wiley.com/10.1002/jnr.490240406 [Accessed July 1, 2020]
Stiegler N V, Krug AK, 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
Tang H, Hammack C, Ogden SC, Wen Z, Qian X, Li Y, Yao B, Shin J, Zhang F, Lee EM, Christian KM, Didier RA, Jin P, Song H & Ming G (2016) Zika Virus Infects Human Cortical Neural Progenitors and Attenuates Their Growth. *Cell Stem Cell* **18:** 587–590
Tang J & Zhai JX (2017) Distribution of polybrominated diphenyl ethers in breast milk, cord blood and placentas: a systematic review. *Environ. Sci. Pollut. Res.* **24:** 21548–21573
Tegenge MA, Rockel TD, Fritsche E & Bicker G (2011) Nitric oxide stimulates human neural progenitor cell migration via cGMP-mediated signal transduction. *Cell. Mol. Life Sci.* **68:** 2089–99 Available at:

http://www.ncbi.nlm.nih.gov/pubmed/20957508
Tena A, Peru E, Martinetti LE, Cano JC, Loyola Baltazar CD, Wagler AE, Skouta R & Fenelon K (2019) Long-term consequences of early postnatal lead exposure on hippocampal synaptic activity in adult mice. *Brain Behav.* **9:**
Threadgill R, Bobb K & Ghosh A (1997) Regulation of dendritic growth and remodeling by Rho, Rac, and Cdc42. *Neuron* **19:** 625–634
Tofighi R, Wan Ibrahim WN, Rebellato P, Andersson PL, Uhlén P & Ceccatelli S (2011) Non–Dioxin-like Polychlorinated Biphenyls Interfere with Neuronal Differentiation of Embryonic Neural Stem Cells. *Toxicol. Sci.* **124:** 192–201
Towfighi J (1980) Hexachlorophene. *Experi- Ment. Clin. Neurotoxicology (Spencer, P. S. Scha- umburg, H. H. eds.), Williams Wilkins, Balt.* **2:** 440–455
Tsuji R & Crofton KM (2012) Developmental neurotoxicity guideline study: Issues with methodology, evaluation and regulation. *Congenit. Anom. (Kyoto).* **52:** 122–128
Tukker AM, Bouwman LMS, van Kleef RGDM, Hendriks HS, Legler J & Westerink RHS (2020) Perfluorooctane sulfonate (PFOS) and perfluorooctanoate (PFOA) acutely affect human α1β2γ2L GABAA receptor and spontaneous neuronal network function in vitro. *Sci. Rep.* **10:**
Volpe JJ, Kinney HC, Jensen FE & Rosenberg PA (2011a) The developing oligodendrocyte: Key cellular target in brain injury in the premature infant. *Int. J. Dev. Neurosci.* **29:** 423–440
Volpe JJ, Kinney HC, Jensen FE & Rosenberg PA (2011b) The developing oligodendrocyte: Key cellular target in brain injury in the premature infant. *Int. J. Dev. Neurosci.* **29:** 423–440
Wang J, Ou SW & Wang YJ (2017) Distribution and function of voltage-gated sodium channels in the nervous system. *Channels* **11:** 534–554
Whitney KD, Seidler FJ & Slotkin TA (1995) Developmental neurotoxicity of chlorpyrifos: Cellular mechanisms. *Toxicol. Appl. Pharmacol.* **134:** 53–62
Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, et al (2016) Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3:** 1–9 Available at: http://figshare.com [Accessed August 26, 2020]
Wong CT, Ahmad E, Li H & Crawford DA (2014) Prostaglandin E2 alters Wnt-dependent migration and proliferation in neuroectodermal stem cells: Implications for autism spectrum disorders. *Cell Commun. Signal.* **12:** 1–18
Xing L & Huttner WB (2020) Neurotransmitters as Modulators of Neural Progenitor Cell Proliferation During Mammalian Neocortex Development. *Front. Cell Dev. Biol.* **8:** 391
Yang N, Chanda S, Marro S, Ng YH, Janas JA, Haag D, Ang CE, Tang Y, Flores Q, Mall M, Wapinski O, Li M, Ahlenius H, Rubenstein JL, Chang HY, Buylla AA, Südhof TC & Wernig M (2017) Generation of pure GABAergic neurons by transcription factor programming. *Nat. Methods* **14:** 621–628
Zhang Y, Pak CH, Han Y, Ahlenius H, Zhang Z, Chanda S, Marro S, Patzke C, Acuna C, Covy J, Xu W, Yang N, Danko T, Chen L, Wernig M & Südhof TC (2013) Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78:** 785–798