LEIBNIZ-INSTITUT
FÜR ARBEITSFORSCHUNG
AN DER TU DORTMUND

fakultät für chemie
und chemische biologie

Stage-dependent evaluation of cytotoxicity, biomarker expression and self-organisation in iPSC-derived neuroectoderm allows accurate and interpretable assessment of developmental toxicity

DISSERTATION

ZUR ERLANGUNG DES AKADEMISCHEN GRADES DES DOKTORS DER NATURWISSENSCHAFTEN (DR. RER. NAT.)

FAKULTÄT FÜR CHEMIE UND CHEMISCHE BIOLOGIE DER TECHNISCHEN UNIVERSITÄT DORTMUND

VORGELEGT VON

ANDREAS SCHOLTZ-ILLIGENS, M.Sc.

DORTMUND 2026

1. GUTACHTER: PROF. DR. JAN G. HENGSTLER

2. GUTACHTER: PROF. DR. JÖRG RAHNENFÜHRER

Eidesstattliche Versicherung (Affidavit)

Name, Vorname
(Surname, first name)

Matrikel-Nr.
(Enrolment number)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden, § 63 Abs. 5 Hochschulgesetz NRW.

Die Abgabe einer falschen Versicherung an Eides statt ist strafbar.

Wer vorsätzlich eine falsche Versicherung an Eides statt abgibt, kann mit einer Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft werden, § 156 StGB. Die fahrlässige Abgabe einer falschen Versicherung an Eides statt kann mit einer Freiheitsstrafe bis zu einem Jahr oder Geldstrafe bestraft werden, § 161 StGB.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offence can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offences of this type is the chancellor of the TU Dortmund University. In the case of multiple or other serious attempts at deception, the candidate can also be unenrolled, Section 63, paragraph 5 of the Universities Act of North Rhine-Westphalia.

The submission of a false affidavit is punishable.

Any person who intentionally submits a false affidavit can be punished with a prison sentence of up to three years or a fine, Section 156 of the Criminal Code. The negligent submission of a false affidavit can be punished with a prison sentence of up to one year or a fine, Section 161 of the Criminal Code.

I have taken note of the above official notification.

Ort, Datum
(Place, date)

Unterschrift
(Signature)

Titel der Dissertation:
(Title of the thesis):

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel selbstständig und ohne unzulässige fremde Hilfe angefertigt habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht.
Die Arbeit hat in gegenwärtiger oder in einer anderen Fassung weder der TU Dortmund noch einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegen.

I hereby swear that I have completed the present dissertation independently and without inadmissible external support. I have not used any sources or tools other than those indicated and have identified literal and analogous quotations.

The thesis in its current version or another version has not been presented to the TU Dortmund University or another university in connection with a state or academic examination.*

*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the PhD thesis is the official and legally binding version.

Ort, Datum
(Place, date)

Unterschrift
(Signature)

Table of Contents

Summary

Today, evaluation of developmental toxicity largely depends on animal testing. However, chemical risk assessment *in vivo* is associated with ethical considerations, high costs, questionable human relevance and limited accuracy, raising an urgent need for alternatives. Here, induced pluripotent stem cell-based methods are considered a promising option, as they enable mimicking key processes of human embryonic development. Especially directed neuroectodermal differentiation has been an important focus. The UKN1 assay as a prominent example of this concept previously enabled highly accurate classification of developmental toxicants based on whole transcriptome profiling. However, high costs and limited mechanistic interpretability constrain applicability, while restricting the model to early neural induction alone may limit sensitivity towards compounds exhibiting developmental toxicity at later developmental stages. In this thesis, both limitations were addressed by first developing a targeted RNA sequencing approach based on the selection of 190 transcriptional biomarkers to provide efficient and mechanistically interpretable assessment of disturbed early neural induction, and second, extending the differentiation model for evaluation of perturbation of self-organisation into neural rosettes, representing a developmental hallmark comparable to neural tube formation *in vivo*. After exposure to a panel of 18 validation compounds with known developmental toxicity status, Targeted RNA Expression Measurement (TREx) revealed distinct modes of perturbed differentiation, including impaired acquisition of neuroectodermal identity and activation of divergent transcriptional programs. Integration of differential biomarker expression and cytotoxicity in a random forest-based classification model yielded moderately accurate prediction of *in vivo* developmental toxicity based on early-stage neuroectodermal differentiation. Comparative cytotoxicity analysis revealed pronounced stage-dependent sensitivity patterns across the validation compounds, and evaluation of neural rosette formation during the late-stage model unravelled exposure-induced disruption of cellular self-organisation. Notably, it was found that some compounds, which exclusively affected biomarker expression during early neuroectodermal differentiation, did not influence self-organisation and vice versa, underscoring the importance of considering different endpoints for assessing developmental perturbation. Integration of disrupted self-organisation and late-stage cytotoxicity in a random forest-based prediction model was shown to elevate predictive performance and enabled identification of a developmental toxicant which could not be detected solely based on early neural induction. Finally, it was demonstrated that developmental stage-aware modelling of early embryonic neurodevelopment based on directed neuroectodermal differentiation of pluripotent stem cells enabled identification of distinct modes of perturbed differentiation, providing accurate, efficient, and mechanistically interpretable prediction of developmental toxicity.

Zusammenfassung

Bis heute basiert die Evaluierung von Entwicklungstoxizität weitgehend auf Tierversuchen. Die ethischen Bedenken, hohen Kosten sowie eingeschränkte Genauigkeit und fragwürdige Human-Relevanz dieses Verfahrens führen jedoch zu einem hohen Bedarf an Alternativmethoden. Hierbei werden auf pluripotenten Stammzellen basierende Methoden, sowie insbesondere gerichtete neuroektodermale Differenzierung, als vielversprechende Option betrachtet, da so essentielle Prozesse der humanen Embryonalentwicklung nachgebildet werden können. Der UKN1 Assay, ein prominentes Beispiel dieses Konzepts, hat in der Vergangenheit basierend auf Transkriptom-Analysen eine hochakkurate Klassifizierung von Entwicklungstoxizität ermöglicht. Hohe Kosten und eingeschränkte Interpretierbarkeit limitieren jedoch die Anwendbarkeit dieses Verfahrens, während der alleinige Fokus auf frühe Stadien der neuronalen Induktion die Sensitivität für Substanzen begrenzt, welche in späteren Entwicklungsstadien Toxizität aufweisen. In der vorliegenden Dissertation wurden beide Limitationen adressiert, einerseits durch die Entwicklung eines gezielten RNA-Sequenzierungsverfahrens basierend auf 190 Biomarker-Genen, um gestörte frühe neurale Induktion effizient und interpretierbar beurteilen zu können, andererseits durch eine Erweiterung des Differenzierungsmodells zur Evaluierung gestörter Selbstorganisation in neuronalen Rosetten, einem Prozess der vergleichbar zur Bildung des Neuralrohrs *in vivo* ist. Nach Exposition gegenüber 18 Substanzen mit bekanntem Toxizitäts-Status offenbarte gezielte RNA-Expressionsmessung unterscheidbare Modi gestörter Differenzierung, inklusive beeinträchtigter Übernahme von neuroektodermaler Identität sowie Aktivierung abweichender transkriptioneller Programme. Die Integration von Biomarker Expression und Zytotoxizität in einem Random-Forest-basierten Klassifizierungsmodell ermöglichte eine moderat genaue Toxizitätsvorhersage basierend auf früher neuroektodermaler Differenzierung. Vergleichende Zytotoxizitätsanalysen zeigten ausgeprägte Stadium-abhängige Sensitivitätsmuster einzelner Substanzen, und die Evaluierung der Bildung von neuralen Rosetten offenbarte expositionsinduzierte Störung zellulärer Selbstorganisation. Bemerkenswerterweise konnte gezeigt werden, dass einzelne Substanzen, welche die Biomarker-Expression beeinflussten, keinen Effekt auf Selbstorganisation im späten Stadium hatten und umgekehrt. Dies unterstreicht die Bedeutung, verschiedene Endpunkte zur Evaluierung von Entwicklungstoxizität zu erfassen. Es konnte gezeigt werden, dass die Integration von gestörter Selbstorganisation und Zytotoxizität die Vorhersagegüte erhöhte und die korrekte Klassifizierung einer toxischen Substanz ermöglichte, welche auf Grundlage des frühen Stadiums nicht richtig erkannt werden konnte. Abschließend wurde demonstriert, dass Entwicklungsstadium-bewusste Modellierung von früher embryonaler Neuroentwicklung basierend auf gerichteter Differenzierung pluripotenter Stammzellen die Identifizierung verschiedener Modi gestörter Differenzierung ermöglicht und so genaue, effiziente und mechanistisch interpretierbare Vorhersage von Entwicklungstoxizität bietet.

Abbreviations

4pLL	Four-parameter log-logistic
AI	artifical intelligence
AIC	Akaike Information Criterion
ALEC	Absolute lowest effective concentration
ALK	Activin Receptor-Like Kinase
ALOEC	Absolute lowest observed effective concentration
ATRA	all-trans retinoic acid
AUC	area under the curve
BMP	Bone Morphogenetic protein
bp	base pair
CDH1	Cadherin-1
CDH2	Cadherin-2
cDNA	complementary DNA
c_{max}	maximum maternal plasma concentration
c-MYC	cellular MyeloCytomatosis
CTB	CellTiter-Blue
DART	Developmental and Reproductive Toxicity
DEG	Differentially expressed gene
DevTox	Developmental Toxicity
DiPaC	Differentiation Pattern Clustering
DLHP	Dorsolateral hinge points
DPG	Differentiation Pattern Group
DPPA5	Developmental Pluripotency Associated 5
EC	Effective concentration
EFD	Embryo-Fetal Development
ESC	Embryonic Stem Cell
ESS	Endpoint Sensitivity Shift
FEED	Fertility and Early Embryonic Development
FGF2	Fibroblast Growth Factor 2
FGF8	Fibroblast Growth Factor 8
HDAC	Histone deacetylase
HPLC-MS	High Performance Liquid Chromatography coupled Mass Spectrometry
ICH	International Council for Harmonisation
iPSC	induced pluripotent stem cells
IS	interpretable summary
KLF4	Krüppel-like factor 4
$LL.2_{bin}$	Binomial two-parameter log-logistic
log2FC	log2 Fold Change
LOOCV	Leave-One-Out cross validation
mEST	Mouse embryonic stem cell test

MHP	Median Hinge Point
MSX1	Msh homeobox 1
MSX2	Msh homeobox 2
NAM	New Approach Methodology
NI1	Neural induction medium 1
NI2	Neural induction medium 2
NPLC	Neuroepithelial progenitor-like cells
NPro	Neural progenitor medium
NTD	Neural Tube Defect
OECD	Organisation for Economic Co-operation and Development
OTX2	Orthodenticle Homeobox 2
PAX3	Paired box 3
PAX6	Paired box 6
PBPK	Physiologically based Pharmacokinetic
PC	principal component
PCA	Principal component analysis
POU5F1	POU domain, class 5, transcription factor 1
PPND	Pre- and Postnatal Development
PS	Probe set
qRT-PCR	Quantitative Reverse Transcription Polymerase Chain Reaction
RFM	Rosette formation medium
ROC	Receiver operating characteristic
rpm	rounds per minute
RROA	Rosette (Re-)Organisation Assay
R-SMAD	Receptor regulated SMAD
SHH	Sonic Hedgehog
SIX3	SIX homeobox 3
SMAD	Small Mothers Against Decapentaplegic
SOX1	SRY-Box Transcription Factor 1
SOX17	SRY-Box Transcription Factor 17
SOX2	SRY-Box Transcription Factor 2
TBXT	T-box transcription factor T
TFAP2A	Transcription Factor AP-2 Alpha
TFAP2B	Transcription Factor AP-2 Beta
TPBG	Trophoblast Glycoprotein
TREx	Targeted RNA Expression Measurement
TSI	Toxicity Separation Index
ZIC	Zinc finger of the cerebellum

1. Introduction

Evaluation of developmental toxicity (DevTox) is a central component of chemical safety assessment. To date, testing compounds for DevTox still largely depends on animal testing, especially in the context of regulatory toxicology. However, *in vivo* testing not only raises ethical concerns but is associated with low throughput, insufficient accuracy, questionable human relevance and substantial cost. Therefore, new approach methodologies (NAM) that can provide sufficiently reliable, accurate and human relevant DevTox assessment are urgently needed. Here, employing pluripotent stem cells is considered a promising option, as they enable mimicking key mechanisms of embryonic development in an experimentally traceable *in vitro* platform. Yet, although stem cell-based methods have been developed for decades, they have not been widely adopted for regulatory decision-making. Major challenges that currently hinder regulatory acceptance include limited predictive accuracy and low biological interpretability, thereby constraining the reliability of stem cell-based DevTox prediction. In this sense, recent studies with the goal of developing more accurate and mechanistically informative assays suggested that especially modelling early neurodevelopment may improve predictive performance and interpretability. *In vivo*, the earliest stages of neurodevelopment are characterised by a complex interplay of extrinsic signalling cues, intrinsic transcriptional responses and the resulting cellular self-organisation. Importantly, stem cell-based approaches that use directed neuroectodermal differentiation can recapitulate many of the involved key mechanisms, thereby offering a promising perspective for reliable, accurate and human relevant DevTox assessment.

1.1 Developmental Toxicity

An essential foundation of reliable DevTox assessment is to define the broad concept of developmental toxicity and the various subcategories that emphasize distinct characteristics and challenges.

A particularly broad definition is provided by the Organisation for Economic Co-operation and Development (OECD) (1). Within the OECD framework, DevTox is defined as part of the broader class of developmental and reproductive toxicity (DART). In this context, developmental toxicity refers to adverse effects on the developing offspring, including pre-, peri- and post-natal, structural or functional disorders. In contrast, reproductive toxicity refers to adverse effects on male or female reproductive function.

Within the concept of DevTox, various subcategories narrow the focus on different aspects of developmental toxicity. The term “Teratogen”, for example, indicates that a compound can produce or increase the frequency of congenital malformations or defects in the progeny (2), specifying the type of adverse effect. “Embryotoxicity”

covers adverse effects during embryonic development (3), specifying the timing of the insult, whereas the term "Developmental Neurotoxicity" (DNT) can be used to specify the affected organ system (4). These subcategories are therefore not just semantic, but map to concrete toxicological concepts.

In many cases, adverse outcomes can be classified under multiple subcategories, such as neural tube defects (NTD) which can arise e.g. from developmental exposure to anticonvulsants (5). NTD are severe congenital malformations of the central nervous system that originate during embryonic development. They arise when the neural tube, a precursor of the nervous system, fails to close properly (6). The most common type of NTD is spina bifida, where the spinal column remains open as a consequence of failed closure of the caudal end of the neural tube (7). As this represents a congenital malformation that originates during embryonic development and manifests in the nervous system, chemical exposures that induce this adverse outcome can be categorized as teratogenic, embryotoxic as well as developmentally neurotoxic. Notably, NTD represent one of the most prevalent congenital malformations with approximately 18.6 cases per 10,000 live births (8), highlighting that evaluation of developmental toxicity is a central component of chemical risk assessment.

Importantly, a fundamental concept of developmental toxicity is that adverse effects are dose-dependent (9). Furthermore, it is well established that developmental toxicity can vary widely between different species (10). A compound that produces severe defects in one species may do so only at much higher dosages in another or may not elicit developmental effects in that species at all. All of this cumulated in the belief that any chemical can be considered a developmental toxicant if administered in the right dosage, at the right time and in the right species (2).

Overall, the various subcategories reflect key characteristics of developmental toxicants: They can perturb embryonic development in a developmental stage-, species-, dosage- or organ-specific manner, including different combinations thereof. This complexity prompts specific requirements for safety assessment in this field – like broad assessment of different organs, developmental stage sensitivity and selection of suitable models.

1.2 Current assessment of Developmental Toxicity *in vivo*

Various *in vivo* test guidelines for pharmaceuticals and industrial chemicals incorporate measures to address the challenges of DevTox assessment outlined above, including temporal sensitivity and endpoint variety. In toxicology, an 'endpoint' refers to a measurable outcome used to detect or quantify adverse effects. Because developmental toxicity can manifest as structural malformations, functional impairments, or organ-specific defects, reliable assessment requires multiple endpoints rather than reliance on a single readout.

For pharmaceuticals, assessment of DART and, more specifically, DevTox is conducted following the International Council for Harmonisation (ICH) guideline ICH S5 (R3) (11). Here, the focus lies on covering multiple developmental stages by designing up to three individual *in vivo* studies, respecting the importance of developmental stage sensitivity in DevTox. The individual designs address Fertility and Early Embryonic Development (FEED), Embryo-Fetal Development (EFD) as well as Pre- and Postnatal Development (PPND), and the guideline stresses the importance of ensuring complete evaluation by leaving no gaps between developmental stages.

For industrial chemicals, the REACH regulation links the required developmental toxicity data to the annual production volume of a substance. These regulatory data requirements are most widely fulfilled by application of suitable OECD test guidelines. For DevTox assessment, the OECD Test No. 443 (12) is especially interesting: In the so-called Extended One-Generation Reproductive Toxicity Study, developmental effects that occur as a result of pre- and postnatal chemical exposure are evaluated by detailed examination of a wide range of endpoints. To ensure a complete assessment of potentially adverse effects, the guideline describes three types of Cohorts: 1) General reproductive and developmental endpoints, 2) potential impact on the developing nervous system and 3) potential impact on the developing immune system. By this extended evaluation of a broad range of effects, the study design stresses the importance of endpoint variability in DevTox.

Taken together, existing *in vivo* testing guidelines consider the specific challenges of DevTox assessment by taking measures to cover a range of developmental stages and endpoints. While scientifically comprehensive, these study designs come at a cost: They require large numbers of animals, which are even exaggerated when multiple species are tested to evaluate interspecies differences. The relevancy of this becomes obvious when considering the number of animals used for different types of safety assessment in the EU and Norway over the years 2018 – 2022: Whereas for the most types of safety assessments, the number of animals used was steady or even shrinking, the number of animals used for DevTox assessment has been increasing over the course of four years and by 2022 needed more animals than any other toxicity testing (13) (**Figure 1**). This demonstrates that the specific challenges of DevTox testing make this type of safety assessment a major and growing contributor to scientific animal use and highlights the need for alternative approaches – such as *in vitro* models that mimic embryonic development.

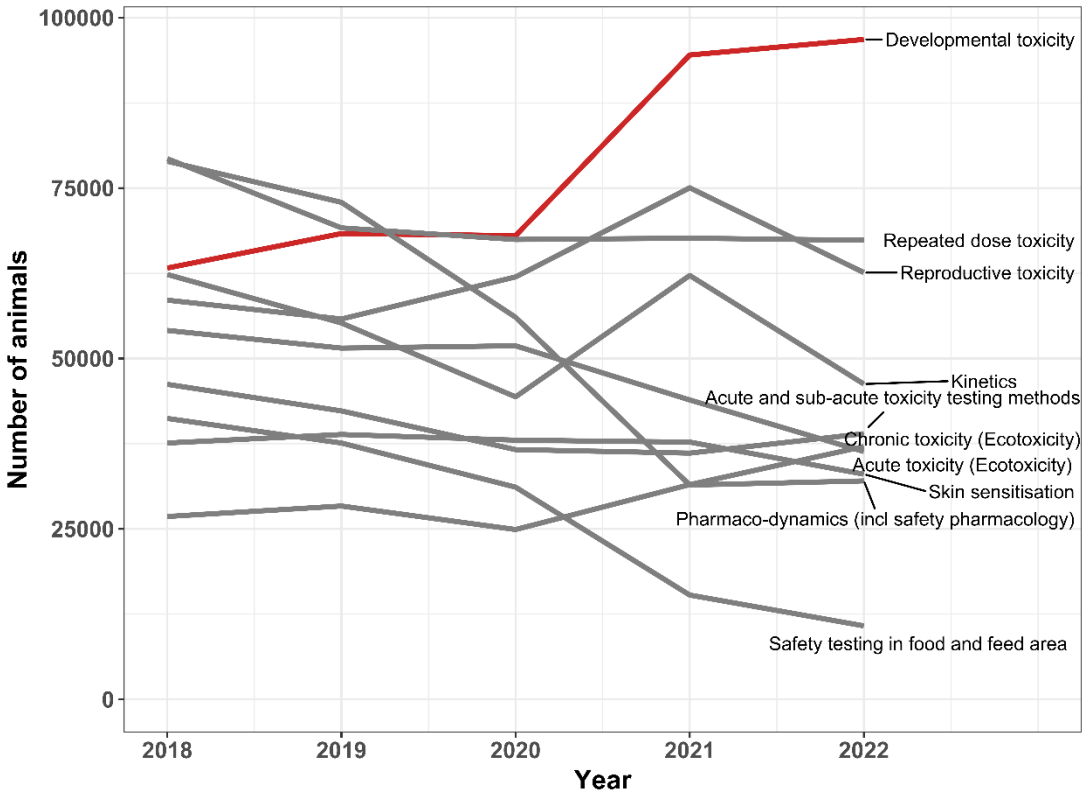


Figure 1: Numbers of animals used for different types of safety assessment in the EU and Norway in 2022. *Assessment of developmental toxicity (indicated in red) has required an increasing number of animals throughout the years 2018-2022, and needed more animals than any other type of safety assessment (indicated in grey) in 2022.*

1.3 Stem cell-based approaches for developmental toxicity testing

To enable reliable assessment of DevTox, alternative methods must provide accurate and mechanistically interpretable prediction of adverse effects on the developing organism. Here, stem cell-based models are a promising approach, as they mirror key processes of early human development.

In the broadest sense, stem cells are defined by their capacity for self-renewal and for generating differentiated cell types (14). For modelling early human development, embryonic stem cells (ESC) have played a central role in the development of alternative testing strategies for DevTox. ESC originate from the inner cell mass of the blastocyst (15) and are considered pluripotent for their capability of generating all three germ layers and - by further differentiation - any cell type of the human body (16). Differentiation refers to the process by which cells acquire more specialised identities, progressing from a pluripotent or multipotent state towards defined lineages. In the context of pluripotent stem cells, spontaneous differentiation occurs when cells are cultured without specific signals, leading to the emergence of derivatives of all three germ layers (17). In contrast, directed differentiation uses defined external cues to guide cells along selected developmental trajectories (18). This distinction is fundamental for applying PSC in

developmental toxicity testing, as it enables controlled modelling of specific developmental processes.

The pluripotency of ESC is maintained by a network of transcription factors, where POU5F1, NANOG and SOX2 play a central role (19). It was shown that impaired expression of POU5F1 and NANOG leads to loss of pluripotency and differentiation (20), (21). Cooperatively, SOX2, POU5F1 and NANOG regulate the expression of genes involved in self-renewal and pluripotency (22), and maintain the potential required for lineage specification during development.

Although this makes ESC a particular interesting model for DevTox, the usage of ESC is restricted in Germany based on ethical concerns, limiting availability under current regulations (23). Here, induced pluripotent stem cells (iPSC) are a promising alternative: These cells are generated by reprogramming somatic cells, such as human fibroblasts, through overexpression of four transcription factors: POU5F1, SOX2, KLF4 and c-MYC (24). Induced pluripotent stem cells closely resemble ESC in morphology and proliferation, as well as in the expression of pluripotency markers and the ability to form teratomas. This demonstrates that, like ESC, iPSC are capable of giving rise to cell types of all three germ layers. Moreover, since they are obtained from adult cells that are readily accessible, their usage mitigates the ethical concerns that arise from sacrificing human embryos required for ESC isolation.

In summary, the capacity of pluripotent stem cells to differentiate into all lineages makes stem cell-based approaches attractive candidates for modelling early embryonic development. Yet, applying such systems in a meaningful alternative assay requires tools that translate *in vitro* effects into predictions of *in vivo* outcomes, and the accuracy of these predictive approaches must be critically evaluated.

1.4 Prediction of *in vivo* toxicity based on *in vitro* measurements

Although iPSC-based systems can recapitulate key processes of embryonic development, translating *in vitro* effects into quantitative predictions of *in vivo* developmental toxicity remains a major challenge. To achieve this translation, a spectrum of biostatistical models can be applied, ranging from simple rule-based threshold criteria to multivariate machine-learning approaches. In practice, most prediction models are developed using a set of reference compounds with known *in vivo* toxicity (training set). The compound-induced effects on the *in vitro* model provide the basis for constructing a prediction model and, because true classifications are known, enable direct evaluation of the model's performance through metrics such as accuracy, sensitivity, and specificity. Accuracy reflects the fraction of all compounds that are correctly classified, whereas sensitivity and

specificity correspond to the proportion of toxic and non-toxic compounds that are correctly classified, respectively.

The arguable simplest approach to predict *in vivo* toxicity from *in vitro* data is to apply a single threshold to one measured endpoint. For example, the OECD 3T3 NRU phototoxicity test (25) classifies compounds as phototoxic when a viability-based index exceeds a defined cut-off value. Notably, this guideline also introduces an intermediate range in which results are considered inconclusive. Although conceptually simple and easy to interpret, manually selecting a cut-off value can lead to thresholds that specifically fit the particular dataset and fail to generalise to new compounds—a process referred to as overfitting. Such loss of generalisability occurs because the model captures not only the true signal but also random noise specific to the training set (26). To evaluate how well a single *in vitro* endpoint can, in principle, distinguish toxic from non-toxic compounds – independent of arbitrary selected cut-offs – AUC-based approaches evaluate prediction performance across all possible thresholds. In this context, Albrecht et al (27) introduced the Toxicity Separation Index (TSI), defined as the area under the Receiver operating characteristic (ROC) curve obtained from testing every potential threshold. This approach does not yield a classification threshold or mitigate overfitting; rather, it solely quantifies the endpoint's overall discriminatory potential.

Furthermore, approaches that rely on a single threshold are often limited in their ability to integrate multiple measured endpoints into a unified toxicity prediction, even though such multi-endpoint data are common in *in vitro* assays. To address this limitation, more advanced statistical approaches can be applied. For example, Worth and Cronin (28) used logistic regression to integrate results from the 3T3 NRU test with multiple endpoints from an additional assay to predict chemically induced eye irritation. While this demonstrates the ability of logistic regression to handle multidimensional data, such models offer limited interpretability with respect to the contribution of individual endpoints.

In this context, rule-based systems like random forest models are a promising option. Based on decision trees that are trained on different subsets of the data, this machine learning-approach enables accurate predictions, integration of heterogeneous endpoints and good interpretability. Consistent with this, Stolte et al. (29) systematically compared several statistical learning approaches for hepatotoxicity prediction and showed that tree-based methods, including random forests, outperformed other models when integrating cytotoxicity and gene expression endpoints.

While bootstrapping of the data inherent to random forests already provides a certain reduction of overfitting, cross-validation approaches can further minimize overfitting and improve generalisability. A simple cross-validation strategy is leave-

one-out cross-validation (LOOCV) (30). Here, a prediction model is trained on all available compounds except one, which is then used as an independent test case. Repeating this procedure for each compound provides an estimate of how well the model generalises when predicting *in vivo* toxicity from *in vitro* endpoints.

Besides high prediction accuracy, which can be provided by advanced statistical models, biological relevance and mechanistic interpretability are increasingly considered additional prerequisites for applying NAMs as alternatives to animal testing. This idea is illustrated by a multi-stakeholder workshop held in 2017 that included groups from academia, industry and regulatory bodies (31). In this workshop with the goal to define readiness criteria for regulatory use of alternative methods, it was stated that reliability and human relevance are critical factors that define test readiness. Notably, the list of recommended test criteria that resulted from this workshop included both performance benchmarks – like accuracy, sensitivity and specificity – as well as the biological relevance of the test system and the measured endpoints.

Taken together, a broad range of biostatistical approaches for predicting *in vivo* toxicity from *in vitro* data has been established. Importantly, this methodological diversity reflects the iterative development of alternative testing strategies over time, as different assays have required different solutions to quantify predictive performance.

1.5 Current PSC-based models in developmental toxicity testing

Building on these methodological considerations, stem cell–based *in vitro* testing methods have been explored as alternative test systems since the 1990s and have undergone continuous refinement.

1.5.1 Early Stem cell-based Developmental Toxicity Tests

The first reported stem cell-based assay for testing DevTox was the mouse embryonic stem cell test (mEST) (32). Here, mouse ESCs are allowed to spontaneously differentiate into a variety of different cell types, including beating cardiomyocytes, while being exposed to test chemicals in the culture medium. In parallel, fully differentiated 3T3 mouse fibroblasts are cultured and exposed under equivalent conditions to distinguish effects specific for differentiating cells. After 10 days of differentiation and exposure, three endpoints are assessed: Cytotoxicity in mouse ESC, cytotoxicity in fibroblasts as well as inhibition of differentiation by microscopically assessing in which conditions beating cardiomyocytes have formed. Based on these endpoints, prediction of embryotoxicity was conducted by employing a linear discriminant analysis. Using this biostatistical model, test compounds were classified into one of three categories: non-embryotoxic, weak and strong embryotoxic. Formal validation of the mEST in the early 2000s (33) demonstrated highly accurate prediction of *in vivo* developmental toxicity using

20 test substances, especially for chemicals classified as strongly embryotoxic. However, expanding the number of tested compounds in the following years revealed insufficient accuracy, suggesting that further improvements are needed to ensure reliable prediction (34). This started a series of refinements of the mEST, including using more informative endpoints (such as changes on RNA and protein level) to gain deeper understanding of disturbed cell differentiation (35,36). Overall, this process shaped the landscape of alternative methods that continue to be developed today.

In more recent development of alternative test methods, advanced cell culture techniques enable generation of physiologically relevant *in vitro* models that more closely resemble key events of human embryonic development. In parallel, the emergence of high-content analytical platforms, including -omics technologies or high throughput image analysis, provides the means to examine how chemicals perturb cellular processes at unprecedented depth. Through these developments, modern DevTox testing no longer relies on simple viability and morphological measures solely but increasingly includes mechanistically informed evaluation of differentiation and morphogenesis.

1.5.2 Spontaneous Differentiation-based Alternative Assays

A prominent example of this development is the so-called "devTOX quickPredict" assay (37). Here, ESC are cultivated in a pluripotent state while being exposed to chemicals through the culture medium for three days. In an untargeted metabolomics approach, subsequent analysis of small molecules by High Performance Liquid Chromatography coupled Mass Spectrometry (HPLC-MS) in the spent culture medium suggested highly promising potential to discriminate developmental toxicants based on amino acid concentration (38). A follow-up study demonstrated highly accurate prediction of DevTox by simply evaluating concentrations of the amino acids Ornithin and Cystin in the spent culture medium (39). Although this technically simple procedure was demonstrated to allow both efficient and accurate DevTox testing, the mechanisms by which developmental toxicants perturbed cellular metabolism remained elusive and the prediction of *in vivo* DevTox was solely based on data driven optimisation without anchorage to a biologically interpretable outcome.

To gain a deeper understanding of how chemicals interfere with cellular processes, other approaches utilized transcriptomics in differentiating cells as a more foundational readout, which allows clearer mechanistic interpretation. This is reflected in the TeraTox assay (40). Here, iPSC undergo multilineage differentiation into all three germ layers while being exposed to test chemicals. After seven days of differentiation and exposure, expression of a set of 1215 genes was measured. This panel of genes was designed to cover pathway reporter genes, genes that are specifically modulated by pathway perturbation as well as germ-layer specific

marker genes, in order to capture both general pathway activity modulation as well as germ-layer specific effects. Here, prominent examples of context-dependent pathways essential for a variety of developmental processes are included, such as the Wnt or Hedgehog signalling pathway, providing a possible explanation of how a specific endpoint without temporal resolution can still identify toxicants that act on various stages of embryonic development. In a validation study on 45 chemicals with known *in vivo* data, an accuracy of 69% positioned the predictive potential of the TeraTox Assay above the mEST and comparable to the quickPredict assay. This accuracy was achieved by a systematic comparison of candidate biostatistical models, where Random Forest classification yielded the best predictive performance – consistent with findings from comparative studies in hepatotoxicity (29). Interestingly, the authors reported that the most accurate prediction was achieved by a combination of the TeraTox and Quick Predict assay, underlining the importance of model and endpoint variability. Furthermore, the mechanistically interpretable read-out of the TeraTox Assay enabled deeper biological insights – revealing that across all germ layers, ectodermal differentiation was the most sensitive to teratogenic disruption.

1.5.3 Directed Differentiation-based Alternative Assays

Conversely, other studies reported promising results in identifying teratogens using directed neuroectodermal differentiation as a model for early embryonic neurodevelopment (41). Despite their narrower developmental coverage compared to multilineage models, directed differentiation enables targeted modelling of specific developmental stages and thereby offers an experimentally traceable framework to assess perturbations at defined stages of embryonic development. This is taken up in the UKN1 assay, where directed differentiation of neuroectodermal progenitor cells for six days in combination with whole-transcriptome micro array analysis showed very promising separation of teratogens and non-teratogens (42).

To extend the lineage specificity of directed differentiation beyond individual developmental trajectories, some approaches have combined multiple models representing different germ layers. This concept is taken up in the ReproTracker Assay (43), which employs directed differentiation of iPSC into cardiomyocytes and hepatocytes combined with biomarker expression measurement. A validation study with 21 chemicals of known toxicity reported high accuracy, suggesting that combination of different germ-layer models enhances prediction of DevTox. Similarly, a recent study showed that combination of the UKK2 Assay (44) and the UKN1 assay, which utilize mesodermal and ectodermal differentiation, respectively, achieved a higher accuracy compared to each assay alone (42).

Overall, various stem cell-based approaches have been proposed which employ diverse strategies to model embryonic stages critical for DevTox assessment.

Interestingly, both multilineage assays (such as the TeraTox assay) as well as directed differentiation models (such as the UKN1 assay) consistently report high sensitivity of ectodermal differentiation to teratogenic disruption. This is in line with the observation that common congenital malformations, like NTD, are linked to disruption of ectodermal – and particularly neuroectodermal - differentiation. It is therefore of high interest to understand the mechanisms and gene regulatory networks that control neuroectodermal development *in vivo*.

1.6 Neuroectodermal development *in vivo*

In vivo, pluripotent stem cells give rise to the neuroectoderm in a process called primary neurulation. This developmental stage comprises neural induction, formation of the neural plate originating from early ectoderm, bending and folding of the neural fold and closure of the neural tube. During this process, a spatially and temporally tightly regulated interplay of extrinsic signals and intrinsic transcriptional responses ensures proper differentiation and morphogenesis.

Pluripotent stem cells *in vivo* are located in the inner cell mass of the blastocyst. These cells give rise to the epiblast, a transient embryonic tissue that subsequently forms the three germ layers from which all further embryonic tissues arise (45). In this process, referred to as gastrulation, activation of key pathways, including the TGFβ and Wnt/β-catenin family, determines the differentiation trajectory (46). Mouse studies demonstrated that polarity in the embryo is specified by signals originating from extraembryonic tissue. By this, concentration gradients of differentiation cues, including bone morphogenetic protein (BMP), are established, which consequently stimulate or inhibit the TGFβ-related protein Nodal and induce Wnt3 expression (47). Subsequently, high to intermediate Nodal/Wnt levels lead to endo- or mesodermal differentiation, whereas the remaining epiblast cells acquire an ectodermal identity, forming the dorsal layer of the embryo. Therefore, sufficient inhibition of meso- and endoderm specifying agents like TGFβ signalling is essential for ectodermal lineage specification (48).

The subsequent specification of neural and non-neural ectoderm has been subject to numerous extensive studies that unravelled the mechanisms controlling this stage of differentiation. This led to the proposal of the so-called “default model”: According to this theory, ectodermal progenitor cells acquire a neural identity by default, unless specific signalling cues promote a different lineage – especially BMP signalling, which originates from the underlying mesoderm and drives the acquisition of an epidermal phenotype at the lateral borders of the ectoderm (49). Conversely, the primitive node near the midline of the ectoderm secrets BMP inhibitors like Noggin, Chordin and Follistatin. Thereby, a mediolateral BMP gradient is generated, where weak BMP signalling in the median ectoderm induces a population of neuroectodermal progenitor cells, the neural plate (50). Interestingly,

even before the neural plate can be observed visually, a first hallmark of neural induction can be demonstrated on a transcriptional level by upregulation of the earliest anterior neuroectoderm marker gene OTX2 (51). This is followed by induction of SOX2, a transcription factor required for cells to adopt neural identity. In cooperation with transcription factors from the ZIC family, SOX2 activates neural genes while repressing BMP target genes, forming a feedback loop that reinforces the cell identity primed by the BMP signalling gradient (52,53). This reflects a fundamental concept: Although a morphogen gradient provides continuous spatial information, transcriptional feedback loops can convert these into discrete, self-reinforcing states. Once a differentiation trajectory is established, this feedback mechanisms leads to lineage restriction that creates sharp spatial boundaries between cell populations with distinct developmental identities (54). In addition, proper development requires meticulous temporal regulation, as exemplified by the role of Fibroblast Growth Factor 2 (FGF2): Early in induction, FGF2 promotes neural differentiation by repressing BMP signalling. However, timely downregulation of FGF2 signalling is required for activation of the master regulator PAX6, a second hallmark of progressing neural induction (55). Upon its upregulation, PAX6 binds with SOX2 and OTX2 to cooperatively maintain neural identity, sustain a proliferative progenitor state, and regulate epithelial polarity (56).

The specification of neural progenitor cells in the median ectoderm, referred to as the neural plate, is followed by a process of bending and folding, that eventually gives rise to the neural groove, neural fold and neural tube. Importantly, this morphogenesis is controlled by key signalling gradients that need to be considered in the context of a well-orchestrated sequence of morphological changes, as illustrated in (**Figure 2**).

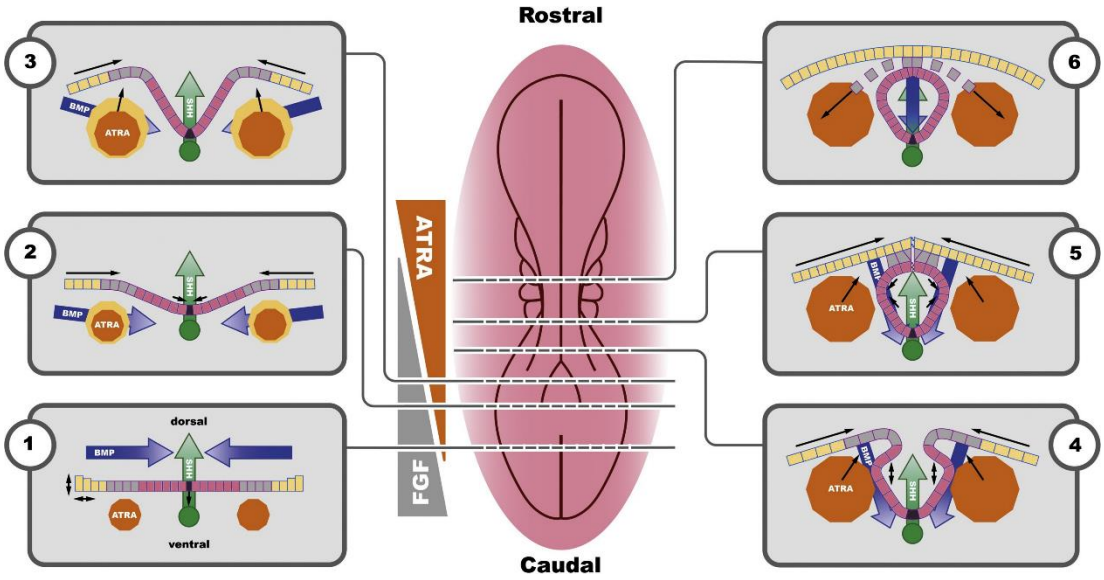


Figure 2: Primary neurulation and formation of the neural tube. *During early embryonic development, signalling gradients of all-trans retinoic acid (ATRA) (orange), SHH (green), and BMP (dark blue) interact with neuroectodermal progenitor cells (pink), neural crest-progeny (grey), and non-neural ectoderm (yellow) to progressively fold the neural plate, create the median hinge point (MHP) (black) and eventually form the neural tube. Adapted with modifications from (57).*

1) Signalling gradients in the neural plate. Morphogenesis of the neural plate is controlled by key signalling pathways, including a dorsoventral gradient of Sonic Hedgehog (SHH) (58) originating from the notochord as well as all-trans retinoic acid (ATRA) originating from the paraxial mesoderm (59). Furthermore, neuroectodermal cells express transcription factors of the ZIC family, which support cytoskeletal organisation and reinforce neural identity by inducing expression of the BMP inhibitor Noggin (58). Together with BMP secretion originating from the lateral non-neural ectoderm, a mediolateral BMP gradient is thereby established.

2) Formation of the median hinge point. It was proposed that SHH mediated inhibition of ZIC in the midline of the neural plate promotes cytoskeletal reorganisation that leads to contraction of cells on the apical side of the median neural plate, forming the median hinge point (MHP, indicated in black) (57,60). By this process, the neural plate bends and creates the neural groove, inducing a shift in the relative direction of the BMP- to the SHH gradient that will progress with further folding.

3) Non-neural ectoderm promotes folding of the neural groove. Following the formation of the MHP, cell proliferation in the non-neural ectoderm provides an additional driving force by pushing the ridges of the neural groove to the midline. Consequently, the BMP gradient originating from the non-neural ectoderm further changes its relative direction by becoming increasingly dorsoventral (61).

4) Formation of the dorsolateral hinge points. Subsequently, the now opposing BMP and SHH gradients promote the formation of the dorsolateral hinge points (DLHP), bending the neural groove further to form the neural fold (57).

5) Positional changes of signalling gradients further drive neurulation. In consequence to the folding of the neural groove, positional changes lift BMP-mediated repression of ATRA signalling. It has been proposed that the consequently increasing levels of ATRA effectively reinforce BMP signalling by suppression of the BMP inhibitor Noggin, creating a feedback loop that further drives folding of the neuroectoderm (62).

6) Closure of the neural tube. Finally, the neuroectodermal switch of CDH1 to CDH2 elicits structural cellular remodelling to fuse the ridges of the neural fold and close the neural tube, while the CDH1-expressing non-neural surface ectoderm seals on the dorsal side (6,63).

An especially interesting cell population is found on the very border between neural and non-neural ectoderm: The neural crest cells. Upon closure of the neural tube, BMP signalling from the neighbouring non-neural ectoderm induces expression of MSX1, which inhibits ZIC-mediated activation of neural genes, including CDH2 mediated cytoskeletal reshaping. Contrarily, MSX1-activated transcriptional programs induce CDH7/11 expression and thereby promote epithelial-to-mesenchymal transition in neural crest cells which consequently delaminate (64,65). This process is of particular interest for DevTox modelling, as some studies suggest a role in the pathogenesis of congenital malformations like NTD: It has been reported that histone deacetylase inhibitors such as the drug Valproic Acid disturb neural crest delamination, contributing to adverse outcomes associated with failed neural tube closure in spina bifida (66).

In conclusion, understanding the interactions between the various signalling cues, transcriptional responses and resulting morphological changes outlines the molecular mechanisms that need to be reflected in an *in vitro* system for modelling neuroectodermal development. It is therefore of high interest to which extent stem cell-based alternative methods allow mimicking this complex interplay.

1.7 Directed neuroectodermal differentiation *in vitro*

Establishing directed neuroectodermal differentiation in pluripotent stem cell models has long been a central goal in developmental biology. Over the last 15 years, dual-SMAD inhibition has emerged as an established standard for neural induction *in vitro*. This approach enables robust and reproducible neural induction across a broad range of different embryonic and induced pluripotent stem cell lines. Moreover, it enables recapitulating key hallmarks of neural induction *in vivo*, making

it a particularly interesting model for early embryonic neurodevelopment and assessing developmental toxicity (**Figure 3**).

Given the central role of BMP signalling in embryonic neurodevelopment, its inhibition was thought to be critical for neural induction *in vitro* as well. Consequently, studies that employed the BMP inhibitor Noggin demonstrated that impaired BMP signalling in ESC promotes the emergence of $SOX1^+$ cells, suggesting an early neural progenitor state (67). Under physiological conditions, proteins of the BMP family, such as BMP4 and BMP7, bind to their cellular receptors ALK1, ALK2, ALK3 and ALK6. After binding BMP, these serine/threonine kinase receptors phosphorylate the receptor regulated SMAD proteins (R-SMADS) SMAD1, SMAD5 and SMAD8, which subsequently bind the co-SMAD SMAD4 to cooperatively activate BMP target genes (68). The BMP inhibitor Noggin directly binds BMP4 and BMP7, blocking the ALK-receptor binding epitope and thereby inhibiting SMAD-transduced BMP signalling (69). Furthermore, since Nodal and Activin signalling are well known to be essential for meso- and endodermal fates *in vivo*, it was suggested that sufficient inhibition of these pathways might promote directed ectodermal differentiation. In line with this theory, employing the drug SB-431542 to inhibit Activin/Nodal/TGFβ signalling enhanced neural induction in embryoid body-based ESC differentiation protocols (70). Similar to BMP, signalling of the TGFβ/Activin/Nodal family under normal conditions is transduced by R-SMAD phosphorylation: Binding of TGFβ to the ALK5 receptor or binding of Activin or Nodal to the ALK4 and ALK7 receptors leads to phosphorylation of SMAD2 and SMAD3, which after binding to SMAD4 cooperatively activate their target genes (68). SB-431542 inhibits this pathway by blocking phosphorylation of the ALK4, ALK5 and ALK7 receptors (71). However, although both Noggin and SB-431542 improved the efficacy of neural induction *in vitro*, neither treatment alone was sufficient to effectively restrict differentiation to the neuroectodermal lineage (72).

In 2009, Chambers et al demonstrated in an elegant study that the combination of Noggin and SB-431542 greatly increased the efficiency of neural induction (72). Considering PAX6 expression a hallmark of neural lineage commitment, it was demonstrated that combined treatment led to more than 80 % $PAX6^+$ cells compared with fewer than 10 % $PAX6^+$ cells when Noggin or SB-431542 were used alone. This process of inhibiting SMAD-transduced signalling in two pathways was named dual-SMAD inhibition. Several mechanisms were proposed to explain the synergistic effect of Noggin and SB-431542, including destabilization of the Activin and NANOG-mediated pluripotency network, suppression of BMP induced non-neural ectoderm as well as of meso- and endodermal fates by inhibiting Activin signalling (72). Importantly, the tightly regulated temporal patterns of gene expression during neural induction *in vivo* were closely recapitulated in the dual-SMAD inhibition protocol – early OTX2 and SOX1 upregulation, concurrent with

downregulation of the pluripotency marker POU5F1, was followed by subsequent upregulation of PAX6 and ZIC1. In parallel, it was shown that typical endodermal and mesodermal markers such as SOX17 and TBXT were not induced, suggesting ectodermal lineage restriction.

Moreover, PAX6^{+} neuroectodermal progenitor cells derived from dual-SMAD inhibition acquire the competence to self-organize into characteristic morphological patterns called neural rosettes (**Figure 3**). This morphological process is considered a hallmark of neural fate commitment, as neural rosettes share key similarities with the embryonic neural tube. These similarities include morphological, functional and transcriptional features, such as radial cellular organisation around a central apical lumen established by tight junctions, as demonstrated by expression of the tight junction marker ZO1. Moreover, neural rosettes reflect the polarity and interkinetic nuclear migration that is observed in the neural tube and leads to cell division on the apical side. Neural rosettes express markers typical of neuroepithelial progenitor cells in the neural tube, such as PAX6, OTX2, SOX1, CDH2 and NESTIN (73). This transcriptional state is thought to be essential for further lineage commitment: Like cells of the neural tube, neural rosettes exhibit responsiveness to dorsoventral and anteroposterior patterning cues like Fibroblast Growth Factor 8 (FGF8), ATRA and SHH, enabling further differentiation into more specialized neurons (74).

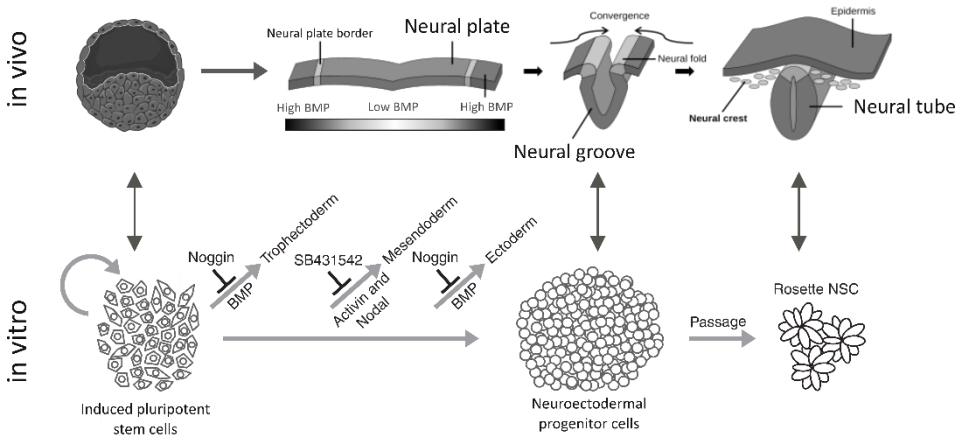


Figure 3: Key mechanisms of neuroectodermal development reflected in directed neuroectodermal differentiation of iPSC. *Central components of early embryonic neurodevelopment in vivo (top) including regulation by BMP signalling, differentiation of cells with neuroectodermal identity and folding of the neural tube can be mimicked by directed neuroectodermal differentiation in vitro (bottom). Adapted with modifications from (72).*

Interestingly, it was shown that the competence to self-organize into neural rosettes can be disrupted by exposure to developmental toxicants during neural induction, making this morphological feature a promising endpoint for DevTox assessment *in vitro* (75). Importantly, it was demonstrated in this study that disruption of self-organization depended strongly on the exact exposure time window: Whereas exposure to the HDAC inhibitor Trichostatin A during the first three days of

differentiation led to nearly complete absence of neural rosettes, later exposure had no significant effect. This timing specific effect in an *in vitro* system reflects a key characteristic of developmental toxicity *in vivo* and underlines the relevance of neural induction models for risk assessment in this field.

Overall, neuroectodermal differentiation models not only have the potential for *in vitro* prediction of DevTox but also closely recapitulate key hallmarks of embryonic neurodevelopment. Therefore, they are considered a promising option for not only highly accurate but also human relevant and mechanistically interpretable alternative test approaches.

1.8 Aim of this work

Since animal-based assessment of developmental toxicity is associated with ethical considerations, questionable human relevance and limited predictive accuracy, alternative methods in this field are urgently needed. Here, stem cell-based *in vitro* approaches and especially directed neuroectodermal differentiation have been an important focus, as they provide focussed and experimentally traceable models to study key steps of embryonic neurodevelopment, such as cellular self-organisation and activation of essential transcriptional networks. The UKN1 assay, as a prominent example of this concept, showed promising predictive performance in developmental toxicity testing based on whole-transcriptome microarray analysis. However, its applicability remains limited by high analytical cost, low mechanistic interpretability, and incomplete consideration of key developmental toxicity features such as timing sensitivity.

In this work, it was hypothesized that replacing microarray analysis with targeted assessment of transcriptional signatures indicative of disturbed neuroectodermal differentiation could enable more efficient and mechanistically interpretable identification of developmental toxicants while retaining predictive accuracy. It was further hypothesized that integration of a morphological endpoint combined with expanding exposure time windows could improve the sensitivity of DevTox prediction. Therefore, the goal of this work was to develop an advanced test system that employs early and late directed neuroectodermal differentiation combined with transcriptional, morphological and viability-based endpoints. Subsequently, the capability of this platform to accurately predict developmental toxicity at human relevant concentrations while providing mechanistic interpretability of chemically-induced perturbed differentiation will be tested.

The specific goals were:

1) Establishing and characterizing an iPSC-based directed neuroectodermal differentiation protocol.

2) Development of a UKN1-based targeted RNA sequencing assay by data-driven selection of biomarker genes representative for chemical-induced disruption of early neuroectodermal differentiation.
3) Implementation of Random Forest–based prediction models for assessment of developmental toxicity using *in vitro* benchmark concentrations derived from multiple endpoints.
4) Development of a rosette re-organisation assay to assess adverse effects on late neuroectodermal development by measuring the disruption of self-organisation of neuroepithelial progenitor-like cells into neural rosettes.
5) Comparison of these models to evaluate the added mechanistic and predictive value of combining distinct endpoints and developmental stages.

2 Material and Methods

2.1 Material

2.1.1 Equipment

Table 1: Equipment

Item	Manufacturer
7500 Real-Time PCR-System	Thermo Fisher Scientific
AF 100 ice machine	Scotsman Ice Systems
Agilent 2100 Bioanalyzer	Agilent Technologies
All-in-One Fluorescence Microscope BZ-X800	Keyence
Autoclave Systec VX-150	Systec
Axiovert A1	Carl Zeiss Microscopy
Balance EW	Kern & Soh
Biosafe SC-smart	Cryotherm
BVC professional Fluid aspiration system	Vacuubrand
Centrifuge 5418	Eppendorf SE
Centrifuge 5427R	Eppendorf SE
Centrifuge 5810R	Eppendorf SE
Chemical Safety Cabinet Herasafe TM	Heraeus
CO2 incubator CB-S 170	Binder
Combi-Spin Minicentrifuge, SR-16 rotor	BioSan
Fridge/Freezer KD33EAI40	Siemens
Hemocytometer Neubauer improved	Marienfeld
Infinite M200 PRO	Tecan
Lab Armor Bead Bath	Lab Armor
Miniature Shaker KM CO2	Edmund Bühler
Mr Frosty	Nalgene
MSC-Advantage Class II Biological Safety Cabinet	Thermo Fisher Scientific
NanoDrop One	Thermo Fisher Scientific
NextSeq 550 System	Illumina
Pipetus	Hirschmann Laborgeräte
Qubit 4 Fluorometer	Thermo Fisher Scientific
Research Plus pipettes, 1-channel	Eppendorf SE
Tecan Infinite M200 Pro Plate Reader	Tecan
Thermocycler T-Gradient	Biometra
Ultra-Low Temperature Freezer MDF-DU702VX	PHC Holding
Vortex	VWR
Waterbath WTB	Memmert
Xplorer Plus pipettes, 1-channel	Eppendorf SE
Xplorer Plus pipettes, 8-channel	Eppendorf SE

2.1.2 Consumables

Table 2: Consumables

Item	Manufacturer
Biosphere Filtered Tip (2.5, 10, 20, 100, 200, 1000 µl)	Sarstedt
Cell culture flask, filter cap (T25, T75, T175)	Sarstedt
Cell culture plates, flat base (96W, 24W, 12W, 6W)	Sarstedt
CryoPure microtube 1 ml	Sarstedt
Filtropur V50 Vacuum Filtration Unit	Sarstedt
MicroAmp Optical 96-well Reaction Plate	Thermo Fisher Scientific
MicroAmp Optical Adhesion Film	Thermo Fisher Scientific
Multiply-µStrip Pro-strip	Eppendorf SE
Parafilm M	Bemis Company
Pipette tips (10 µL, 200 µL, 1250 µL, 5000 µL)	Sarstedt
SafeSeal reaction tube (0.5 mL, 1.5 mL, 2 mL and 5 mL)	Sarstedt
Screw cap micro tube (2 mL)	Sarstedt
Screw cap tube (15 mL and 50 mL)	Sarstedt
Serological pipettes (5 mL, 10 mL, 25 mL and 50 mL)	Sarstedt
Syringe filter, Filtropur S	Sarstedt
Syringe Omnifix Luer Solo 10ml	Braun

2.1.3 Reagents

Table 3: Reagents

Item	Manufacturer
1 N NaOH, sodium hydroxide solution molecular biology-grade	Merck
2-Mercaptoethanol (50 mM)	Thermo Fisher Scientific
2-Propanol	Carl Roth
Accutase Cell detachment solution	STEMCELL Technologies
Agencourt AMPure XP	Beckman Coulter Genomics
Agilent DNA 1000 Kit	Agilent Technologies
AmpliSeq CD Indexes, Set A-D for Illumina	Illumina
AmpliSeq cDNA Synthesis for Illumina	Illumina
AmpliSeq Custom RNA Panel for Illumina	Illumina
AmpliSeq Library PLUS for Illumina	Illumina
Ascorbic acid	Sigma Aldrich
Bovine serum albumin	Sigma-Aldrich
Cellartis DEF-CS 500 culture system	Takara
Cellartis DEF-CS COAT-1	Takara
CellTiter-Blue Cell Viability Assay	Promega
DAPI	Thermo Fisher Scientific
DEPC-treated water	Thermo Fisher Scientific
DMSO (cell culture grade)	PanReac AppliChem

Item	Manufacturer
DNAZap DNA and RNA Decontamination Solution	Thermo Fisher Scientific
Dorsomorphin Dihydrochloride	Cell Guidance Systems
DPBS with Ca2+ and Mg2+	Thermo Fisher Scientific
DPBS without Ca2+ and Mg2+	Thermo Fisher Scientific
Dulbecco's Modified Eagle Medium (DMEM)	Thermo Fisher Scientific
Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F-12)	Thermo Fisher Scientific
Ethanol absolute for analysis	Sigma-Aldrich
FBS Good	PAN-Biotech
GlutaMAX (100x)	Thermo Fisher Scientific
High-capacity cDNA Reverse Transcription Kit	Thermo Fisher Scientific
KnockOut Dulbecco's Modified Eagle Medium	Thermo Fisher Scientific
KnockOut Serum Replacement	Thermo Fisher Scientific
MEM Non-Essential Amino Acids (100x)	Thermo Fisher Scientific
N-2 Supplement (100X)	Thermo Fisher Scientific
NextSeq 500/550 Hi Output KT v2.5	Illumina
Normal Donkey Serum	Jackson ImmunoResearch
Penicillin/Streptomycin	PAN Biotech
QuantiNova SYBR Green PCR Kit	Qiagen
Qubit 1× dsDNA HS Assay Kit	Thermo Fisher Scientific
Qubit RNA BR Assay Kit	Thermo Fisher Scientific
Recombinant Human FGF-basic (154 a.a.)	Peprotech
Recombinant Human Noggin Protein	R&D Systems
Recombinant Human Sonic Hedgehog/Shh (C24II) N-Terminus	R&D Systems
RNase-Free DNase Set	Qiagen
RNaseZap RNase Decontamination Solution	Thermo Fisher Scientific
RNeasy Mini Kit	Qiagen
ROTI Histofix 4%	Carl Roth
SB431542	Cell Guidance Systems
STEM-CELLBANKER EX	Amsbio
Tris-HCl Buffer 1 M pH7.0	Jena Bioscience
Tris-HCl Buffer 1 M pH8.5	Jena Bioscience
Triton X-100	Sigma-Aldrich
Trypan Blue Solution	Thermo Fisher Scientific
TrypLE™ Select Enzyme (1X), no phenol red	Thermo Fisher Scientific
Tween20	Sigma-Aldrich
Y-27632 (Dihydrochloride)	STEMCELL Technologies

Table 4: Validation compounds

Item	Manufacturer
(+/-)-Warfarin	Cayman Chemicals
5,5-Dimethyl-2,4-Oxazolidindion (Dimethadione)	Sigma Aldrich
Aciclovir	Sigma Aldrich
Bosentan hydrate	Sigma Aldrich
Caffeine	Sigma Aldrich
D-Penicillamine	Cayman Chemicals
Hydroxyurea	Hycultec
Imatinib	Sigma Aldrich
Isoniazid	Sigma Aldrich
Isotretinoin	Sigma Aldrich
Metformin (hydrochloride)	Cayman Chemicals
Methoxyacetic acid	Hycultec
Metoclopramide hydrochloride	Sigma Aldrich
Oseltamivir (phosphate)	Hycultec
Phthalic acid mono-2-ethyl hexyl ester (MEHP)	Sigma Aldrich
Saccharin	Sigma Aldrich
Thiamine hydrochloride	Sigma Aldrich
Topiramate	Hycultec
Valproic acid	Sigma Aldrich
Zidovudine	Sigma Aldrich

2.1.4 Antibodies and qPCR primers

Table 5: Primary antibodies

Antigen	Host	Identifier	Manufacturer
POU5F1	Rabbit	94310	Cell signaling technology
GM130	Rabbit	ab52649	abcam
SSEA4	Mouse	43782	Cell signaling technology
TRA-1-60	Mouse	61220	Cell signaling technology
TRA-1-81	Mouse	83321	Cell signaling technology
ZO1	Mouse	33-9100	Thermo Fisher Scientific

Table 6: Secondary antibodies

Antigen	Host	Conjugate	Identifier	Manufacturer
Mouse IgG	Donkey	Alexa Fluor® 488	715-545-150	Jackson ImmunoResearch Laboratories, West Grove, US
Rabbit IgG	Donkey	Alexa Fluor® 488	711-545-152	Jackson ImmunoResearch Laboratories, West Grove, US
Rabbit IgG	Donkey	Cyanine Cy™3	711-165-152	Jackson ImmunoResearch Laboratories, West Grove, US

Table 7: Proprietary qPCR primer

Target	ID	Manufacturer
CDH2	Hs_CDH2_1_SG_QuantiTect Primer Assay	Qiagen
CDH1	Hs_CDH1_1_SG_QuantiTect Primer Assay	Qiagen

Table 8: Primer with known sequence, manufactured by Eurofins Scientific

Target	Type	Sequence (5' --> 3')
PAX6	Forward	CCGCCTATGCCCAGCTTCAC
PAX6	Reverse	AAGTGGTGCCCGAGGTGCCC
OTX2	Forward	GTTCAGAGTCCTTGGTGGGT
OTX2	Reverse	CCCTCACTCGCCACATCTAC
POU5F1	Forward	GCAAAGCAGAAACCCTCGTGC
POU5F1	Reverse	ACACTCGGACCACATCCTTCTCG
NANOG	Forward	GGTGAAGACCTGGTTCCAGAAC
NANOG	Reverse	CATCCCTGGTGGTAGGAAGAGTAAAG
MSX1	Forward	GGATCAGACTTCGGAGAGTGAACT
MSX1	Reverse	GCCTTCCCTTTAACCCTCACA
TBP	Forward	GGGCACCACTCCACTGTATC
TBP	Reverse	GCAGCAAACCGCTTGGGATTATATTCG

2.1.5 Cells

Caco2 cells were purchased from ATCC. The induced pluripotent stem cell line SBAD2 was obtained from the Chair of in-vitro Toxicology and Biomedicine Dept inaugurated by the Doerenkamp-Zbinden foundation at the University of Konstanz, originally generated within the StemBANCC project (76) by reprogramming dermal fibroblasts from a 51-year old male Caucasian using the non-integrating Sendai virus CytoTune-iPS reprogramming Kit (Thermo Fisher Scientific).

2.2 Methods

2.2.1 Cultivation of iPSC

Cultivation of iPSC was conducted using the Cellartis DEF-CS 500 culture system (Takara). Culture vessels were prepared with 100 µl/cm2 COAT-1 solution (Takara) diluted 1:20 in DPBS containing Ca^{2+} and Mg^{2+} (hereafter DPBS+/+), followed by incubation for at least 20 min at 37 °C and 5% CO_2 before transferring cells to the culture ware. Cells were thawed in a 37° C water bath and 1 ml cell suspension was slowly added to 9 ml warm DEF-CS medium containing GF-1 (1:333), GF-2 (1:1000) and GF-3 (1:1000) (hereafter DEF-CS 123). After centrifugation (300g, 5 min), supernatant was discarded, and the cell pellet was resuspended in 7 ml warm DEF-CS 123. Coating solution was discarded immediately before transferring the cell suspension to the culture vessels. Cultures were maintained at 37° C and 5% CO_2, with daily medium changes to fresh DEF-CS medium containing GF-1 (1:333) and GF-2 (1:1000) (hereafter DEF-CS 12). Cultures were inspected daily for morphological signs indicative of differentiation and expression of the pluripotency markers POU5F1 and NANOG was assessed regularly by qRT-PCR.

For passaging, confluent and densely arranged cells were washed with DPBS without Ca^{2+} and Mg^{2+} (hereafter DPBS-/-) before adding 20 µl/cm2 TrypLE Select (Thermo Fisher Scientific). After incubation for 5 min at 37 °C and 5% CO_2, 6 volumes of DEF-CS 123 were added and the culture vessel was flushed multiple times. Cells were counted using a Neubauer hemocytometer (Marienfeld). For further cultivation, 13,333 or 33,333 cells/cm2 were reseeded to achieve confluency after 3 or 4 days, respectively. For cryopreservation, the single cell suspension was centrifuged (300g, 5 min), supernatant was discarded and cells were resuspended in ice-cold stem cellbanker (Amsbio). Aliquots of 1 ml cell suspension containing 1.5 million cells were transferred to cryo-tubes and stored at -80 °C in a Mr. Frosty freezing container (Nalgene) to ensure a linear cooling rate of -1 °C/min. After at least 24 h, cells were transferred to liquid nitrogen for long-term storage.

2.2.2 Cultivation of Caco2 cells

Caco2 cells were thawed in a 37° C water bath and transferred to 5 ml prewarmed Dulbecco's Modified Eagle Medium (DMEM) (Thermo Fisher Scientific). The suspension was centrifuged (300g, 5 min), the supernatant discarded, and the cell pellet was resuspended in 10 ml fresh DMEM. Cells were counted using a Neubauer hemocytometer and seeded at 50,000 cells/cm2. Cultures were maintained at 37° C and 5% CO_2 with medium changes every third day, and passaged before reaching 80% confluency.

2.2.3 Neural induction of iPSC

For neural induction of iPSC in monolayer culture, culture vessels were coated with DEF-CS COAT-1 as described above. On day of differentiation (DoD) -2,

24,000 cells/cm2 were seeded in DEF-CS 123. On DoD -1, medium was replaced with fresh DEF-CS 12.

In order to control readiness for neural induction on DoD 0, cell confluency was assessed via phase contrast microscopy. If cultures were at least 80% confluent, cells were washed with DPBS+/+ and medium was replaced with neural induction medium 1 (NI1), consisting of Knockout-DMEM supplemented with Knockout Serum Replacement, GlutaMax, Non-essential amino acids, 50 µM Mercaptoethanol (all Thermo Fisher Scientific), 21.6 µM SB431542 (Cell Guidance Systems), 0.64 µM Dorsomorphin (Cell Guidance Systems) and 35 ng/ml Noggin (R&D Systems). Medium was changed to fresh NI1 on DoD 1 and 2. On DoD 4, medium was replaced with a 3:1 mixture of NI1 and neural induction medium 2 (NI2), consisting of DMEM/F12 (Thermo Fisher Scientific) supplemented with GlutaMax, N2 Supplement (Thermo Fisher Scientific), 21.6 µM SB431542, 0.64 µM Dorsomorphin and 35 ng/ml Noggin. After six days of differentiation, cells were ready for downstream analysis.

For neural induction of iPSC in organoid suspension culture, 300,000 cells/ml were seeded into uncoated 6W plates (2 ml/well) in DEF-CS 123 on DoD-2 and cultivated at 37° C and 5% CO_2 on a shaker (70 rounds per minute (rpm)). If not stated otherwise, medium changes were performed by tilting the plate, allowing spheroids to sink to the bottom edge of the well before carefully aspirating spend medium. On DoD -1, medium was replaced with DEF-CS 12.

On DoD 0, spheroids were collected, centrifuged (300g, 3 min), washed once with NI1, centrifuged again, and resuspended in NI1. Medium was changed to fresh NI1 on DoD 1 and 2 (2 ml/well). On DoD 4, 6, and 8, spheroids received 3 ml/well of NI1:NI2 mixtures in ratios of 3:1, 2:2, and 1:3, respectively. On DoD10, medium was changed to 3 ml/well fresh NI2.

On DoD 11, spheroids were collected, centrifuged (300g, 3 min), medium was aspirated and spheroids were resuspended in 3 ml/well neural progenitor medium (NPro), containing DMEM/F12 supplemented with GlutaMax, N2 Supplement, 20 ng/ml FGF-2 (Peprotech) and 20 µM Ascorbic acid (Sigma Aldrich). 6W plates were coated with DEF-CS COAT-1 as described above and 2 ml/well spheroid suspension was seeded into the culture plates. Cultures were maintained at 37° C and 5% CO_2 without shaking to allow spheroid attachment and cell outgrowth. On DoD13, medium was changed to fresh NPro. Cells were used for downstream application, including the Rosette (Re-) Organisation Assay on DoD15.

2.2.4 UKN1 Assay

To assess the effect of chemicals on neural induction of iPSC, test compounds were added to the culture medium from DoD 0 to DoD 4 during differentiation in monolayer culture. Compound stocks were prepared either as 1000× solutions in

DMSO or as 10× solutions in NI1 containing 1% DMSO. Exposure medium was freshly prepared on DoD 0 immediately before starting neural induction. On DoD 4, compound exposure was terminated by washing cells with DPBS+/+ and replacing the medium with a 3:1 mixture of NI1 and NI2 without test compounds. After six days of neural induction, cells were either collected for RNA collection or cell viability was measured using CellTiter-Blue (CTB) (Promega) assay.

In order to evaluate compound effects in non-differentiating cells, Caco2 cells were seeded at 40,000 cells/cm2 in DMEM and maintained in confluent culture for 21 days with medium changes every third day to allow complete differentiation. On day 21, compound exposure was initiated by adding test compounds to the medium for four days. On day 25, Caco2 cells were washed with DPBS+/+ and medium was changed to DMEM without test compounds. On day 27, cells were collected for RNA isolation and gene expression analysis.

2.2.5 Rosette Re-Organisation Assay

Cytotoxicity and disruption of rosette (re-) self-organisation was assessed using cells on DoD 15 of neural induction in spheroid culture. Cells were reseeded by washing once with DPBS-/- before adding 500 µl Accutase (STEMCELL Technologies) per well of a 6W plate. After 15 min incubation at 37° C, wells were flushed and single cells were collected using 500 µl rosette formation medium (RFM), consisting of DMEM/F12 supplemented with GlutaMax, N2 Supplement, FGF-2 (20 ng/ml), ascorbic acid (20 µM) and Sonic Hedgehog (20 ng/ml) (R&D Systems). The cell suspension was centrifuged (300g, 3 min), medium was aspirated, and cells were resuspended in RFM supplemented with 10 µM Y-27632 (STEMCELL Technologies).

For cytotoxicity analysis, COAT-1 coated 96W plates were prepared as described above and 175,000 cells/cm2 were seeded into the inner 60 wells. Plates were left under the bench for 15 min to promote uniform attachment before cultivation at 37° C and 5% CO_2.

For the Rosette (Re-)Organisation Assay (RROA), cells were reseeded by “spotting”-procedure adapted from (77) with modifications. 12W plates were coated by adding three 10 µl drops of DEF-CS COAT-1 (1:20 in DPBS+/+) per well and incubating for 20 min at 37° C. Immediately before seeding, coating drops were carefully aspirated without touching the coated surface. Without allowing the coated surface to dry out, 10 µl drops of cell suspension containing 1,500 cells/µl were placed on the coated areas and incubated for 45 min at 37° C and 5% CO_2. Wells were then gently filled with 1 ml RFM containing 10 µM Y-27632.

Medium changes, including test compound exposure, were carried out on DoD 16 and DoD 17 using RFM. On DoD 19, either cell viability was measured by CTB assay or cells were fixed for immunostaining by washing twice with DBPS+/+, 15 min

incubation with 4% ROTI Histofix 4% formaldehyde (Carl Roth) and washing three times with DPBS+/+. Fixed cells were stored at 4° C until downstream analysis.

2.2.6 CellTiter Blue Assay

For cytotoxicity analysis, each test condition was prepared in cell culture triplicates, with at least 6 untreated-cell controls (no compound but same solvent concentration) and at least 6 no-cell controls. Viability was measured using the CellTiter-Blue (CTB) kit according to the instructions from the manufacturer.

On the day of analysis, CTB medium was prepared in the dark by adding 1 part CTB reagent to 4 parts culture medium. Spent culture medium was discarded, cells were washed with DPBS+/+, and 100 µl CTB medium was added to each well in the dark, including no-cell controls. Cells were incubated 30-90 minutes until a colour shift in the untreated-cell controls was visible. Then, fluorescence was measured in an Infinite M200 Pro plate reader (Tecan) at excitation/emission wavelength of 540/594 nm, respectively.

Measured plates were accepted for downstream analysis if the mean raw fluorescence value of the untreated-cell controls were at least fivefold higher than that of the no-cell controls. For data processing, fluorescence values of each condition were normalized by subtracting the mean no-cell control value. Relative viability of each condition was then calculated as the ratio of the normalized condition value and the normalized mean untreated-cell control value.

2.2.7 RNA Isolation

All steps were performed using RNase-free material and consumables. Total RNA was isolated using the RNeasy Mini Kit (Qiagen) according to the instruction of the manufacturer.

In brief, cells were washed once with DPBS+/+ and lysed in Buffer RLT (350 µl/well of a 12W plate). After incubation for 1 min at room temperature, wells were rinsed several times, lysates were transferred to 1.5 ml tubes and stored at -80° C until further processing. In order to isolate total RNA, thawed cell lysates were mixed with 1 volume 70% ethanol and transferred to RNAeasy mini spin columns. Columns were centrifuged (12,000g, 30 s) and flow-through was discarded.

DNA was removed using the RNase-Free DNase Set (Qiagen) according to the instructions of the manufacturer. In brief, after washing once with Buffer RW1, mini-columns were incubated with 80 µl DNase for 15 min at room temperature. Columns were then washed once with Buffer RW1 and twice with Buffer RW2.

RNA was eluted by placing columns in a collection tube and adding 50 µl RNase-free water. After 1 min incubation, collection tubes with the mini-spin columns were centrifuged (12,000g, 1 min) and total RNA concentration was measured using a

NanoDrop One UV/VIS spectrometer (Thermo Fisher Scientific) before samples were stored at -80° C until further use.

2.2.8 Quantitative RT-PCR

Synthesis of cDNA for qRT-PCR was conducted using the cDNA Reverse Transcription Kit (Thermo Fisher Scientific) according to the instructions from the manufacturer, with 0.7 µg RNA input per reaction.

The qRT-PCR was performed on a 7500 Real-Time PCR System (Thermo Fisher Scientific), using the QuantiNova SYBR Green PCR Kit (Qiagen) with 17.5 ng cDNA input and the respective QuantiTect Primer Assay (Qiagen). Cycling conditions were: 50° C for 2 min, 95° C for 4 min, 40 cycles of 95° C for 10 s and 60° C for 30 s. Melt curve analysis was performed with 95° C for 15 s, 60° C for 20 s and a ramp to 95° C at the lowest available rate (1%).

Relative gene expression was calculated using the ΔΔCt method (78) with TBP as the housekeeping gene.

2.2.9 Biomarker panel selection

A data-driven panel of biomarker genes was compiled to assess transcriptional signatures of exposure-induced disruption during neuroectodermal differentiation. Biomarker selection was based on whole-transcriptome microarray data from neuroectodermal progenitor cells exposed to 24 known developmental toxicants (42) and was done in close cooperation with Franziska Kappenberg from the Technical University of Dortmund. Three complementary selection strategies were applied, based on Limma Analysis-derived log2 fold changes (log2FC), limma t-test derived FDR-adjusted p-values and normalized gene expression values.

Univariate selection: All probe sets (PS) were ranked by the number of toxicants that induced differential expression (absolute log2FC > 1; FDR adjusted p-value < 0.05). In case of ties, PS were re-ranked by the number of toxicants inducing differential expression with an absolute log2FC greater than the smallest absolute log2FC observed among the tied sets. This procedure was repeated iteratively until all PS were assigned a unique rank. The top 150 ranked PS were retained, and every gene annotated to at least one specific PS of these was included.

Multivariate selection: The multivariate selection was based on the Top-1000-Classifier presented in (42). In a leave-one-out cross validation scheme, the 1000 PS with the highest variance concerning the normalized expression values across all samples were used to train an L1-regularized logistic regression-based classifier. Based on this analysis, every gene annotated to at least one specific PS where L1-penalization did not result in a coefficient of 0 in at least one iteration of the Leave-One-Out classifier was selected.

Top10 selection: For each toxicant, the lowest concentration inducing at least one differential PS (absolute log2FC $>$ 1; adjusted p-value $<$ 0.05) was identified. From these conditions, the 10 genes annotated to the PS with the largest absolute log2FC were selected.

The final biomarker panel comprised all genes selected by at least one of these strategies.

To confirm relevance for embryonic development, Gene Ontology (GO) Enrichment analysis was performed using the topGO R package (79). The analysis was restricted to biological process terms and was carried out using the classic algorithm and fisher's exact test.

In addition to the panel of biomarker genes, 40 housekeeping genes were included for normalization. In a primary selection, 70 candidates were identified by literature research (80–84). From this list, final housekeeping genes were manually chosen based on 1) low standard deviation across all conditions in the microarray data set and 2) moderate absolute expression in order to optimize read distribution in targeted RNA sequencing.

2.2.10 Targeted RNA Sequencing

Library preparation and targeted RNA sequencing was done in close cooperation with Katharina Derksen. To prepare amplicon libraries for Targeted RNA Expression Measurement (TREx), cDNA was synthesized using the AmpliSeq cDNA Synthesis for Illumina Kit (Illumina) with 10 ng RNA input according to the instructions from the manufacturer. Target enrichment was performed using the AmpliSeq Library Plus for Illumina Kit together with an AmpliSeq for Illumina Custom Panel (Illumina), designed by the manufacturer to cover the selected 230 transcripts.

In brief, AmpliSeq HiFi Mix and Custom RNA Panel primers were added to the cDNA, followed by PCR to enrich the targeted transcripts with 17 cycles of 99° C for 15 s and 60° C for 4 min. After partial digestion, AmpliSeq CD Indexes Set A-D (Illumina) were ligated to the amplicons. Libraries were finalized by amplification and clean-up.

All libraries were quantified using the Qubit 1× dsDNA HS Assay Kit (Thermo Fisher Scientific) and analysed for size distribution using the Agilent Bioanalyzer with the DNA 1000 Kit (Agilent Technologies) according to the instructions of the manufacturer.

All libraries were sequenced on a NextSeq 550 sequencing platform (Illumina) using the NextSeq 500/550 High Output Kit v2.5 (Illumina) according to the instructions from the manufacturer. After library normalization, denaturation, pooling and dilution to a concentration of 1.025 to 1.05 pM, targeted RNA sequencing was carried out as single end (1×75 bp).

2.2.11 Targeted RNA Sequencing data analysis

Basic processing of raw Targeted RNA Expression Measurement data was performed using the Local Run Manager RNA Amplicon Module (Illumina), including demultiplexing of indexed reads, aligning reads against amplicons sequences using the BWA Aligner and counting amplicon fragments to generate raw read counts.

To assess robustness between individual experiments, raw read counts were correlated among replicates of each condition. Pearson correlation coefficients were summarized, and an individual experiment was only accepted if all correlations exceeded 0.85. Genes with fewer than 10 counts in at least four samples were excluded from further analysis.

Integration of multiple sequencing batches for principal component analysis (PCA) of variance stabilizing transformed raw read counts was achieved using ComBat-Seq (85), with three independent sequencing runs specified as batch variable. Importantly, toxic and non-toxic compounds were equally distributed across sequencing runs to minimize risk of confounding.

For normalisation of gene expression data, size factor estimation in the DESeq2 analysis pipeline is based on the assumption that the majority of genes show similar expression across samples. As targeted assessment of biomarkers selected for their responsiveness to neuroectodermal differentiation would substantially violate this assumption, differentially expressed biomarker genes were identified using DESeq2 (86) with size factors estimated from the 40 housekeeping genes. Comparisons were made against the solvent control condition. Biomarker genes were considered differentially expressed if the absolute log2FC was ≥ 1 and the adjusted p-value was ≤ 0.05.

Supervised clustering of biomarker genes was performed as described in (87). In brief, biomarkers were sorted into biologically interpretable differentiation pattern groups (DPG) based on the measured log2FC between exposed cells and solvent control cells in relation to the log2FC between undifferentiated and differentiated cells. Cutoff values were chosen to for integrated analysis of PCA loading patterns in the context of DiPaC-based supervised clustering, loading values of a specific principal component were visualised as colour scale in the DiPaC plot.

2.2.12 Immunofluorescence staining

Cells were fixed for immunostaining by washing twice with DPBS+/+ and incubation with 4% ROTI Histofix for 15 min. After fixation, cells were washed three times and stored at 4° C until downstream analysis.

Fixed cells were permeabilized in Triton X-100 (0.3% in DPBS+/+) for 10 min, washed three times with DPBS+/+ and blocked with 5% normal donkey serum (Jackson ImmunoResearch) in DPBS+/+ for 1 h. Cells were then incubated for 1 h

with primary antibody solution containing 0.5% normal donkey serum and the respective primary antibody in a dilution of 1:400 in DPBS+/+. After three washes, cells were incubated for 1 h with secondary antibody solution containing 0.5% normal donkey serum and the respective secondary antibody in a dilution of 1:200 in DPBS+/+. Nuclei were counterstained with DAPI (Thermo Fisher Scientific) (1:5000 in DPBS+/+) following two additional washes. After three final washes, plates were stored at 4° C until imaging.

2.2.13 High-throughput image analysis

Detection of neural rosettes was adapted from (75) with modifications. Image analysis was done in R using the EBImage package (88).

Neural rosettes were visualized by immunostaining of GM130 and ZO1 (**Figure 4A**), and images of whole spots were acquired on a BZ-X800 microscope (Keyence) by capturing 3×3 overlapping fields of view, which were stitched using the BZ-X800 Analyzer software (Keyence). Segmentation of ZO1+ objects was achieved by Otsu thresholding after applying a difference of Gaussian blur filter (**Figure 4B**). Objects <50 pixel were excluded from further analysis. To account for substantially varying GM130 signal intensities depending on local cell density, binarization of the GM130 channel was conducted by adaptive local thresholding.

For primary identification of putative neural rosettes, Voronoi segmentation of the binary GM130 channel using ZO1+ objects as seeding points was employed (**Figure 4C**). Putative rosettes touching image borders or not exceeding a size of 300 pixel were excluded from further analysis. Putative lumina were defined as ZO1+ objects inside primary rosettes, and a second Voronoi segmentation, using putative lumina as seeding points, was conducted to detect secondary rosettes (**Figure 4D**).

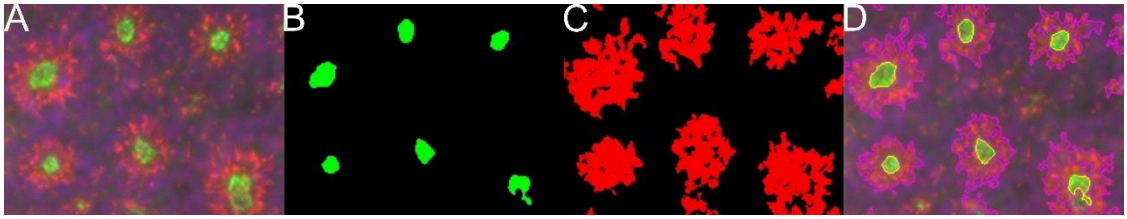


Figure 4: Image processing for neural rosette formation analysis. *A) Neural rosettes were visualized by immunofluorescence staining for ZO1 (green), GM130 (red) and DAPI (blue). B) Segmentation of putative lumen positive for ZO1 was conducted by Otsu thresholding. C) Segmentation of putative neural rosettes was achieved by Voronoi segmentation of the GM130 channel using putative lumen as seeding points. D) Mark up of segmented neural rosettes after a second Voronoi segmentation based on putative lumen inside GM130 positive objects as seeding points.*

For each putative lumen and secondary rosette, area, circularity, eccentricity and mean ZO1/GM130 intensity were extracted. Based on this, a random forest classifier, trained on 346 manually assigned objects, was generated using the R package randomForest (89) to increase detection accuracy. Objects with a prediction score ≥ 0.8 were classified as neural rosettes.

2.2.14 Curve fitting

All curve fittings were performed in R using the package drc (90) if not stated otherwise. The analysis was done in close cooperation with Franziska Kappenberg from the Technical University of Dortmund.

For cytotoxicity analysis, a four-parameter log-logistic (4pLL) model and a flat model were fitted to the viability data of each compound. The 4pLL model was based on the assumption that cytotoxicity induced by a test compound can be described by a sigmoidal curve, where for a concentration x and four parameters b, c, d, e, the response is given by:

$$f(x|b,c,d,e) = c + \frac{d-c}{1+exp\big(b(\log(x)-\log(e))\big)}$$

Where c and d represent the lower and upper asymptote of the curve, e represent the EC_{50} value (and the inflection point of the curve) and b represents the slope of the curve.

The flat model was based on the assumption that compound exposure induced no relevant changes to relative viability, fitting the mean viability across all concentrations to the data. For each 4pLL model, the quality of each curve fit was assessed by calculating a goodness-of-fit statistic as done in (27), and only curves with a goodness-of-fit > 0.85 were accepted for further analysis. Then, the Akaike Information Criterion (AIC) (91) was calculated for each of the two models and the model with the lower AIC was retained.

To ensure that the left asymptote attains a value of 100%, a refit-procedure was conducted. In case of a 4pLL model, all values were divided by the value of the left asymptote after a first fit and were multiplied by 100 to obtain percentages before refitting a second 4pLL model. In case of a flat model, the mean response was refitted to 100%.

EC_k values were calculated as the concentration where the curve attains the value (100 – k) %. For each compound, the ratio of the highest and the lowest EC_{10} value across all experiment was calculated. Depending on this ratio, three, four, or five individual experiment per compound were conducted ($< 3, \geq$ 3-5, ≥ 5, respectively.)

To evaluate disruption of rosette re-organisation, each spot was classified positive (1) if ≥ 3 rosettes were detected or negative (0) otherwise. Then, a flat model as described above and a binomial two-parameter log-logistic ($LL.2_{bin}$) model were fitted to the data, where for a concentration x and the parameters b and e, the response is given by:

$$f(x|b,e) = \frac{1}{1+\exp(b(\log(x)-\log(e)))}$$

Where b represents the slope of the curve and e represents the EC_{50} value. As described above, model selection was based on AIC and for $LL.2_{bin}$ models, EC_k values were calculated.

For each concentration of a compound, the estimated probability of rosette formation on spot-level including Wilson 95% confidence intervals were reported, and an independent experiment was excluded from curve fitting if the probability of rosette formation in the control samples was < 0.9.

2.2.15 Performance parameters

To evaluate the predictive performance of a classification model, the parameters accuracy, sensitivity, specificity, AUC and Brier score were determined. Accuracy was calculated as the sum of correctly classified compounds divided by the sum of all compounds. Sensitivity was calculated as the number of correctly classified toxicants (true positives, TP) divided by the sum of true positives and falsely classified toxicants (false negatives, FN). Conversely, specificity was calculated as the number of correctly classified non-toxicants (true negatives, TN) divided by the sum of true negatives and falsely classified non-toxicants (false positives, FP).

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

The AUC was calculated as the area under the ROC curve to assess a model's ability to discriminate between toxic and non-toxic compounds across all possible classification thresholds. The Brier score was calculated as the mean squared difference between the predicted probabilities of toxicity and the true class (1 for toxic, 0 for non-toxic.

2.2.16 Toxicity Separation Index and Endpoint Sensitivity Shift Index

To quantify how well an *in vitro* benchmark concentration can discriminate developmental toxicants and non-toxicants, Toxicity Separation Index (TSI) analysis was conducted as described in (27). In brief, for each compound, the difference between the respective *in vivo* reference concentration and the *in vitro* benchmark concentration was calculated on a log10 scale. The differences for each compound were sorted in ascending order, and for each interval between two consecutive differences, a candidate threshold was defined to classify each compound with values above the interval as toxic and those below as non-toxic. Then, for each threshold, the classification was compared to the true toxicity statues to derive sensitivity and specificity for each threshold. The R package pROC (92) was used to

calculate the area under the curve (AUC) of the ROC curve generated by plotting 1-specificity against sensitivity for each threshold, and the TSI was defined as the resulting AUC. For systematic assessment of the difference between two *in vitro* benchmark concentrations, the Endpoint Sensitivity Shift Index (ESS) was defined for each compound as the difference between the respective concentrations on a log2 scale.

2.2.17 Random Forest Classification

For prediction of developmental toxicity, random forest models were generated using the R package Rforestry (93). All steps of analysis were conducted in close cooperation with Annalena Weissert and Jörg Rahnenführer from the Technical University of Dortmund.

To reduce dimensionality, *in vitro* measurements were summarized to interpretable summaries (IS) as described in (29) with modifications. More specifically, cytotoxicity measurements were summarized as median EC_{10}, EC_{20} and EC_{50} values across the different individual experiments. For cases in which model selection during curve fitting favoured a constant model, EC_k values were imputed as 5× the maximum tested concentration. Gene expression measurements were summarized as four types of IS: Absolute lowest observed effective concentration (ALOEC), absolute lowest effective concentration (ALEC), concentration inducing differentially expressed genes (DEG), and the maximum number of differentially expressed genes (nrdegs_max). The ALOEC of a specific gene was defined as the lowest tested concentration that induced an absolute log2FC greater than 1, and the ALOEC-based IS were summarized as the minimum (ALOEC.min), the 25%-quantile (ALOEC.Q25), and the median (ALOEC.Med) of all available ALOEC. The ALEC of a specific gene was defined as the lowest concentration at which a 4pLL model fitted to variance stabilizing transformed data attained an absolute log2FC of 1, provided that the goodness of fit of the model was > 0.4 and that the calculated ALEC did not exceed five times the highest tested concentration. Otherwise, the ALEC of a gene was set to NA. ALEC-based IS (ALEC.min, ALEC.Q25, ALEC.Med) were derived analogously to the ALOEC-based IS. Furthermore, penalty variables (Penalty.ALEC.Min, Penalty.ALEC.Q25, Penalty.ALEC.Med) were defined as 0 if at least one ALEC could be calculated for a compound. If no ALEC could be calculated for a compound, the penalty variables were set to 1, and the ALEC-based IS were imputed as 10^100. DEG-based IS were defined as the lowest tested concentration at which at least one (DEG1), two (DEG2), or five (DEG5) biomarker genes were differentially expressed with an absolute log2FC > 1 and an FDR-adjusted p-value $<$ 0.05. Analogous to the ALEC-based IS, penalty variables (Penalty.DEG1, Penalty.DEG2, Penalty.DEG5) were defined as 0 if at least 1, 2, or 5 DEG were detected, or set to 1 otherwise, and DEG-based IS were imputed according to the ALEC-based IS. The nrdegs_max IS was defined as the number of differentially

expressed biomarker genes (absolute log2FC $>$ 1 and adjusted p-value $<$ 0.05) at the highest tested concentration. Disruption of rosette formation was summarized as the EC_{10}, EC_{20} and EC_{50} values of the respective compound probability-concentration curve. For compounds where model selection favoured a constant model, a penalty variable (Penalty.Ros) was set to 1 and rosette EC_k values were imputed as 5× the maximum tested concentration. To anchor summaries of *in vitro* measurements to *in vivo* relevant concentrations, all concentration-based IS were expressed relative to the respective *in vivo*-reference concentration (c_{max} for training; 1000 log-spaced concentrations between 1 nM and 10 mM for concentration-dependent prediction).

Prediction models were trained and validated in a Leave-One-Out cross validation (LOOCV) scheme, where for each compound, prediction was based on a model that was generated using all other compounds as training data. Random forest models used different sets of IS as features, and all models were trained with monotonic constraints for all IS that represented benchmark concentrations to ensure monotonic concentration-response patterns. The predicted probability for each compound was defined as the DevTox score, and evaluation of predictive performance was conducted in two approaches: For binary classification, a threshold that maximises accuracy for discrimination of toxicants and non-toxicants based on all iterations of the LOOCV scheme was determined. If more than one possible threshold yielded maximal accuracy, the threshold that maximises sensitivity was chosen. In the three-category framework, compound-exposures were classified as positive in the case of a DevTox score $>$ 0.8, negative in the case of DevTox score $<$ 0.2, and uncertain in the case of 0.2 $\leq$ DevTox score $\leq$ 0.8.

2.2.18 Usage of artificial intelligence (AI)

In the present thesis, ChatGPT (GPT-5.3 and GPT-5.4, OpenAI) was employed for proofreading, including spelling and grammar correction, as well as assistance in literature research. No image generation, data analysis, or scientific reasoning was conducted using AI.

3 Results

3.1 Morphological characteristics and transcriptional signatures confirm directed neuroectodermal differentiation by dual-SMAD inhibition

As a prerequisite for employing directed neuroectodermal differentiation to evaluate developmental toxicity, basic characterisation of morphological and transcriptional features was conducted. The undifferentiated iPSC exhibited morphological patterns characteristic for pluripotent stem cells: Shortly after passaging, cells showed an elongated morphology with high nucleus-to-cytoplasm ratio (**Figure 5A**). During further cultivation, colonies became progressively denser (**Figure 5B**), until a compact confluent monolayer without visible cell borders was established after three days in culture (**Figure 5C**). To verify maintenance of an undifferentiated iPSC state prior to differentiation, cells were immunostained for established pluripotency markers (POU5F1, SSEA4, TRA-1-60, and TRA-1-81). All markers showed clear expression, exemplified by POU5F1 staining (**Figure 6**). The full set of stainings is provided in (Supplementary figure 1).

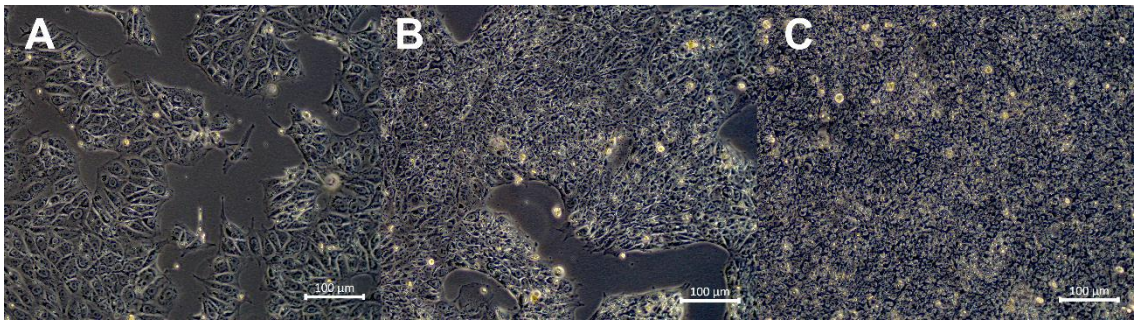


Figure 5: Light microscopy of representative iPSC. *Images were taken immediately after passaging (A), after two days (B) and after three days (C) of cultivation. Scale bar represents 100 µM.*

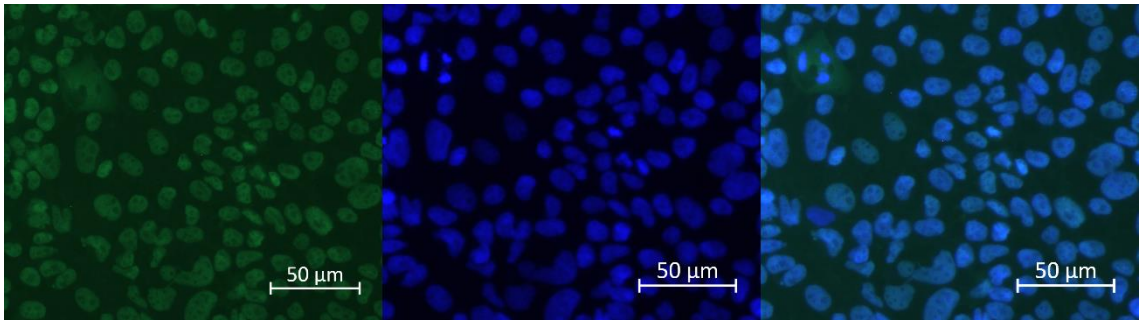


Figure 6: Immunofluorescence staining of POU5F1. *Undifferentiated cells were stained for the pluripotency marker POU5F1 (green), and nuclei were stained with DAPI (blue). Scale bar represents 50 µM.*

Next, cells were subjected to directed neuroectodermal differentiation for six days (**Figure 7**). Whereas cells exhibited an iPSC-like morphology immediately after plating (**Figure 8A**), characteristic morphological changes were observed in the

course of the differentiation, which indicated adoption of a neuroectodermal phenotype. Soon after initiating neural induction, continuous cell proliferation led to a dense confluent monolayer (**Figure 8B**). With progressing differentiation, increasing polarization was observed, reflected in parallel alignment of elongated cells characteristic of ectodermal differentiation. After six days of neural induction, a sheet of putative ectodermal progenitor cells had formed (**Figure 8C**).

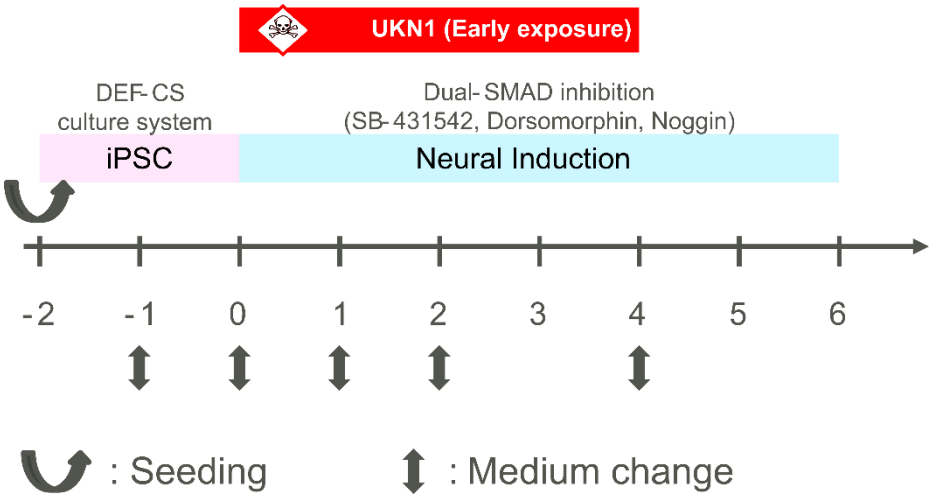


***Figure 7: Experimental design early neural induction**. Schematic of the iPSC to early neuroectodermal progenitor cells differentiation protocol. Straight arrows indicate medium changes, curved arrow indicates seeding. On DoD6, cells were used for downstream analysis.*

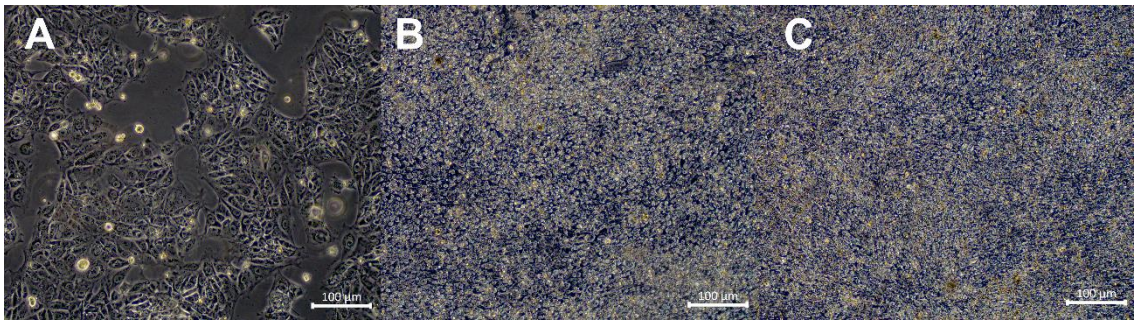


***Figure 8: Light microscopy of iPSC during early neural induction**. Images were taken on DoD0, DoD4 and DoD6. Scale bars represent 100 µM.*

To confirm adoption of neuroectodermal identity, expression of established marker genes was measured by q-RT-PCR (Table 9).

Table 9: Marker genes associated with different cell types

Marker Gene	Associated cellular identity	Function	Reference
OTX2	Neuroectodermal progenitor cells	Earliest pioneer of ectodermal differentiation	(51)
PAX6		Master regulator confirming neuronal identity	(56)
CDH2		Cell adhesion protein essential for structural development of neuroectodermal progenitor cells	(63)
NANOG	Undifferentiated stem cells	Essential transcription factor for maintenance of pluripotency in undifferentiated stem cells	(21)
POU5F1		Essential transcription factor for maintenance of pluripotency in undifferentiated stem cells	(20)
CDH1		Cell adhesion protein essential for structural organisation of undifferentiated stem cells	(63)
MSX1	Neural crest cells	Transcriptional repressor essential for neural crest cell development	(65)

This analysis showed that neural induction for six days led to a significant increase of the neuroectodermal marker genes PAX6 and OTX2 (**Figure 9**). Conversely, expression of the pluripotency marker genes POU5F1 and NANOG was significantly downregulated, indicating loss of an undifferentiated stem cell identity. Notably, there was no significant change of CDH1 or CDH2 expression, consistent with early stage neuroectodermal differentiation in which transcription factor regulation precedes the cadherin-associated switch required for subsequent structural remodelling. Furthermore, dual-SMAD inhibition induced a modest yet significant upregulation of the neural crest-associated gene MSX1. In combination with the observed morphological patterns, these transcriptional changes demonstrated successful acquisition of an early neuroectodermal identity after six days of differentiation, thus supporting suitability of the model to reflect key features of embryonic neurodevelopmental essential for assessing early-stage developmental toxicity by subsequent exposure experiments.

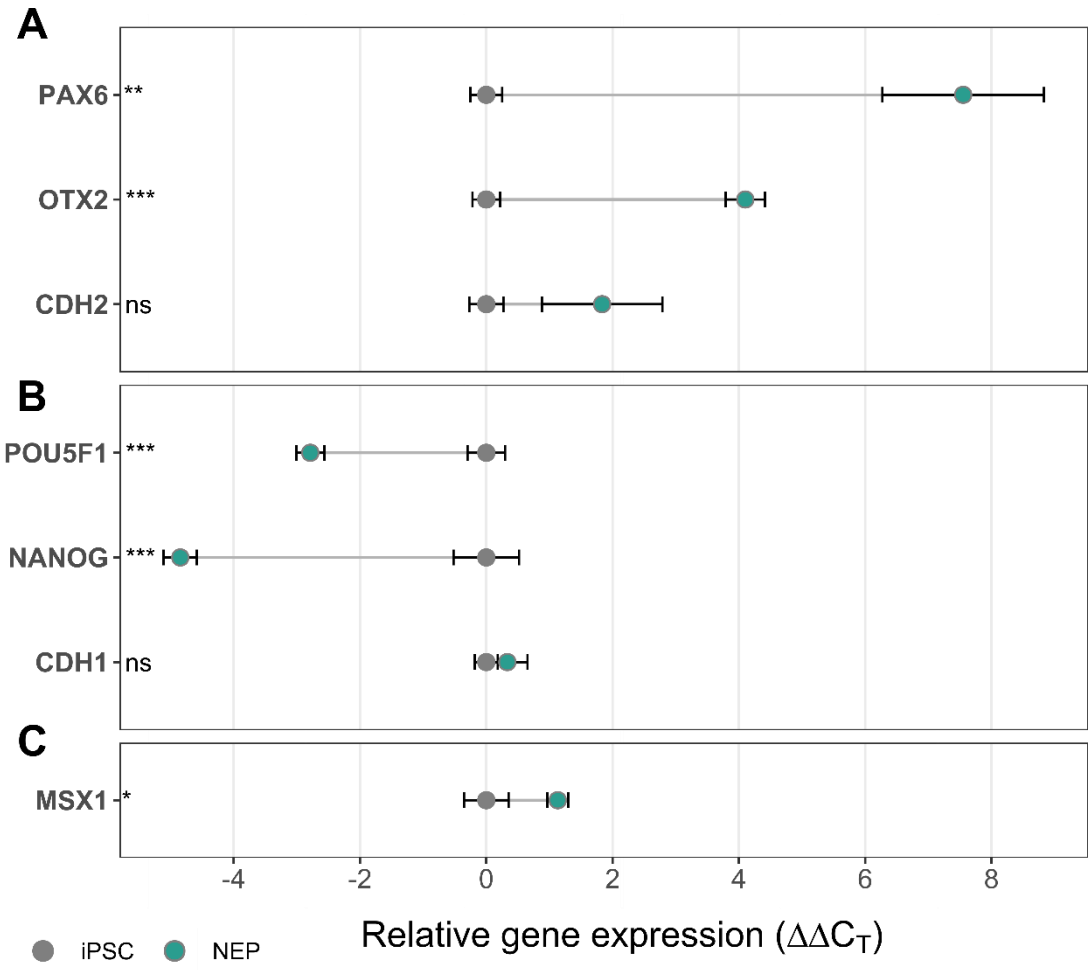


Figure 9: Relative gene expression of established markers after neural induction of iPSC. *The panels (A), (B), and (C) depict expression of established marker genes associated with neuroectodermal progenitor cells, pluripotent stem cells, and neural crest cells, respectively. Grey data points represent expression in undifferentiated iPSC, blue data points represent expression in neuroectodermal progenitor cells. ns: not significant, *: p<0.05, **: p<0.01, ***p<0.001.*

3.2 Cytotoxicity induced by exposure to a validation panel enabled moderately accurate separation of toxicants and non-toxicants

For assessing exposure-induced disruption of differentiation, a panel of validation compounds with known developmental toxicity was required. For this purpose, a literature-based selection was conducted to compose suitable developmental toxicants as well as non-toxic reference compounds. The resulting validation panel used for subsequent exposure experiments comprised 18 compounds with known developmental toxicity, including nine positive and nine negative reference substances (Table 10) which have so far not been evaluated in the setting of the UKN1 assay. Importantly, as DevTox is inherently dose-dependent, every compound can in principle be considered developmentally toxic or non-toxic depending on the administered dosage and the resulting internal concentration. Thus, interpretation of exposure-induced developmental perturbation requires placing the observed effects in the context of *in vivo*-relevant concentrations. Although concentrations in

embryonic tissue would provide the most direct analogue to those in the culture medium of stem cell-based models, such data is rarely available. Because early embryonic exposure is largely governed by passive diffusion from the maternal circulation, the maximum maternal plasma concentration (c_{max}) was therefore considered the best available surrogate for embryonic exposure. Accordingly, maternal c_{max} values associated with either adverse developmental outcomes or their absence were identified for each compound of the validation panel.

Table 10: Validation compound panel including developmental toxicity, in vivo relevant concentrations and EC10 values of cytotoxicity during early neural induction

Compound name	Classification	Reference classification	c_{max}	Reference c_{max}	EC_{10}
Bosentan	Toxic	(94,95)	2.9 µM	(96)	167.7 µM
Dimethadione	Toxic	(97–99)	5.4 mM	(100)	5.0 mM
D-Penicillamine	Toxic	(101,102)	27.5 µM	(103)	1.9 mM
Hydroxyurea	Toxic	(104,105)	565.4 µM	(105)	53.7 µM
Imatinib	Toxic	(106,107)	5.3 µM	(108)	9.3 µM
MEHP	Toxic	(109,110)	146 µM	(111)	1.0 mM
Methoxyacetic acid	Toxic	(112,113)	696 µM	(114)	2.0 mM
Topiramate	Toxic	(115,116)	19.9 µM	(117)	1.3 mM
Warfarin	Toxic	(118,119)	2.2 µM	(120)	372.8 µM
Aciclovir	Non- Toxic	(121)	3.0 µM	(122)	709.1 µM
Caffeine	Non- Toxic	(123,124)	19.3 µM	(125)	245.8 µM
Isoniazid	Non- Toxic	(126)	52.2 µM	(127)	1.3 mM
Metformin	Non- Toxic	(128,129)	12.2 µM	(130)	211.6 µM
Metoclopramide	Non- Toxic	(131)	0.18 µM	(132)	256.5 µM
Oseltamivir	Non- Toxic	(133–135)	0.33 µM	(136)	394.4 µM
Saccharin	Non- Toxic	(137)	3.07 µM	(138)	2.5 mM
Thiamine	Non- Toxic	(139,140)	35.3 nM	(141)	7.6 mM
Zidovudine	Non- Toxic	(142)	2.0 µM	(143)	735.9 mM

With the validation panel established, exposure experiments were conducted to examine compound-induced perturbations of the differentiation process. As cell death represents the most fundamental disruption of an *in vitro* system, the first objective was to measure exposure-induced viability decrease by CTB assay (**Figure 7**). Across the validation panel, all compounds exhibited concentration-dependent cytotoxicity, and concentration-cytotoxicity curves for all compounds are provided in (Supplementary figure 2, Supplementary figure 3). For instance, the developmental toxicant MEHP reduced cell viability to near-complete loss at approximately 3000 µM (**Figure 10A**). On the other hand, the negative reference compound Oseltamivir required substantially lower concentrations to induce complete loss of viable cells (**Figure 10B**), highlighting the importance to interpret effective *in vitro*-concentrations in relation to relevant *in vivo* reference concentrations.

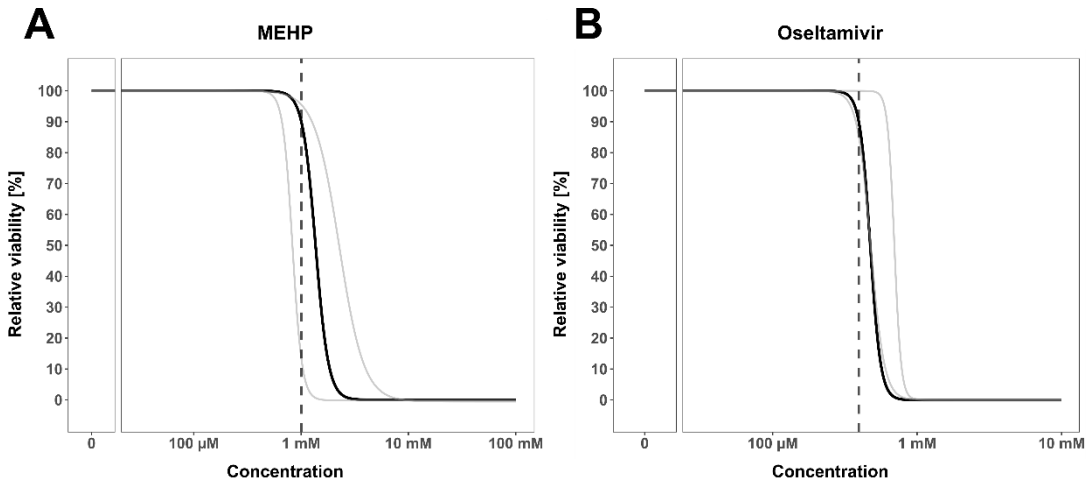


Figure 10: Concentration-viability curves of MEHP and Oseltamivir. *Four parameter log-logistic models were fitted to the data of three independent experiments. The bold line represents the experiment from which the median EC10 value was obtained. The dashed line indicates the median EC10.*

Accordingly, cytotoxicity curves were normalized to literature-derived c_{max} values for each compound (**Figure 11A**), revealing that several developmental toxicants induced cytotoxicity at concentrations close to their *in vivo* relevant range. Conversely, various negative controls did not induce cytotoxicity up until 1,000× of their *in vivo* relevant concentration. However, a substantial overlap observed at 10 – 100× c_{max} indicated limited capability to discriminate toxicants and non-toxicants based on cytotoxicity alone. This was further supported by systematic assessment of the predictive potential of cytotoxicity during early neural induction through Toxicity Separation Index (TSI) analysis based on median EC_{10} values extracted from each curve. As illustrated in (**Figure 11B**), the resulting TSI of 0.852 demonstrated partial separation of toxicants and non-toxicants. Consistent with the patterns observed across the normalized cytotoxicity curves, these results suggested that cytotoxicity during early neuroectodermal differentiation can only provide limited discriminatory power of developmental toxicants and non-toxicants, prompting the exploration of endpoints capturing disrupted differentiation rather than general toxicity.

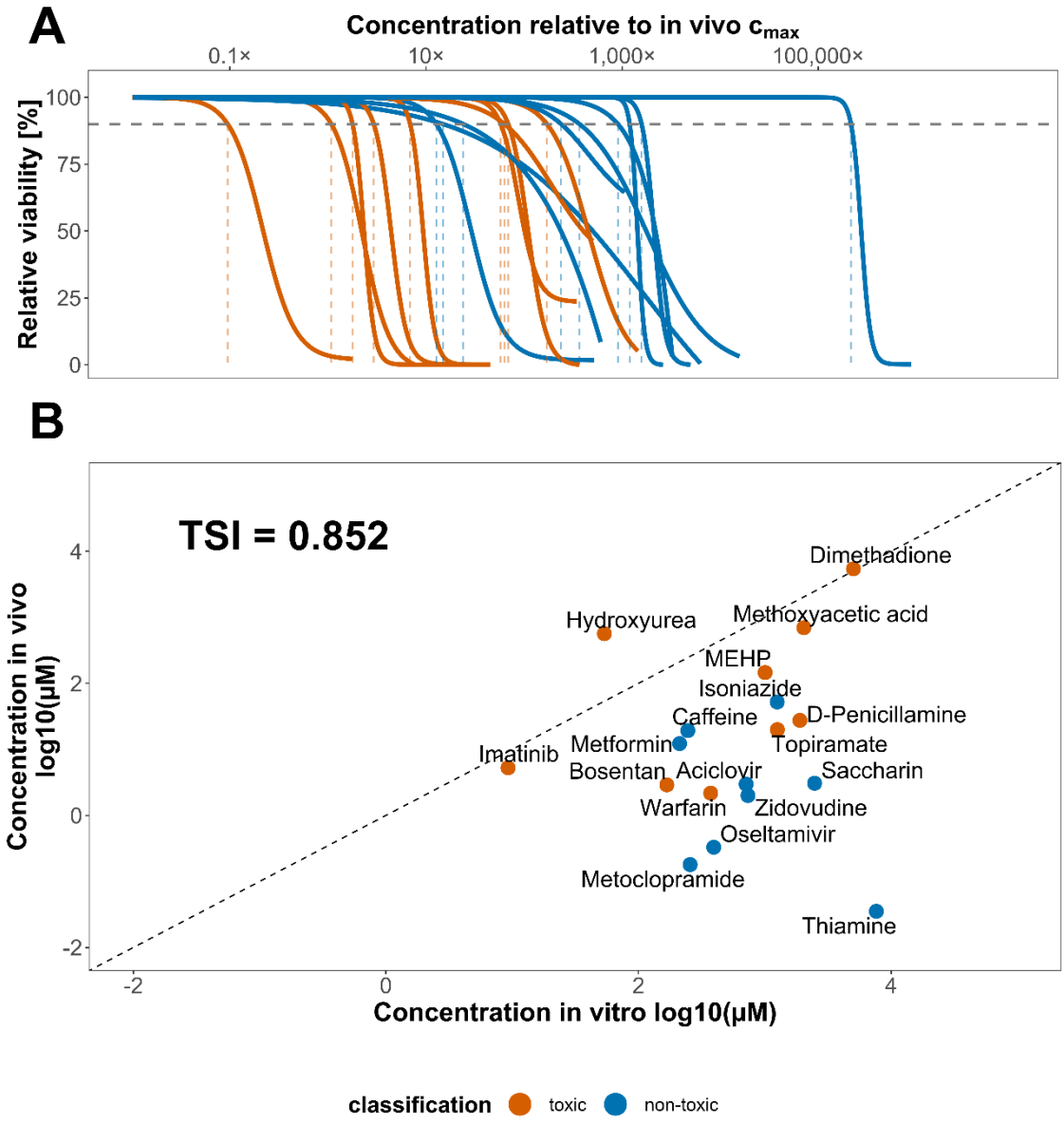


Figure 11: Discrimination of developmental toxicants and non-toxic reference compounds based on cytotoxicity in the early-stage model. *A) Concentration-viability curves of all validation compounds were normalized to the c_{max} of the respective compound. Blue lines indicate non-toxic compounds, orange lines indicate toxicants. B) Toxicity Separation Index analysis using median EC_{10} values to define positive in vitro concentrations and maternal c_{max} values as in vivo reference concentrations. Blue dots represent non-toxicants, orange dots represent toxicants.*

3.3 Integration of selection strategies defines a transcriptional biomarker panel for detecting disturbed early differentiation

In earlier studies on the UKN1 assay, combining cytotoxicity with whole-transcriptome microarray analysis enabled highly accurate discrimination of 39 developmental toxicants and non-toxicants at *in vivo* relevant concentrations (42). However, high costs and substantial workload limit feasibility of whole-transcriptome assessment for high and moderate throughput *in vitro* systems. Instead, focussing gene expression analysis on a curated panel of biomarker

genes tailored to capture transcriptional signatures of disrupted neuroectodermal differentiation represented an attractive strategy for efficient and scalable DevTox assessment. Here, the previously generated whole-transcriptome profiling provided a powerful foundation for data-driven identification of gene expression changes indicative of exposure-induced disruption of early neural induction. Thus, differential gene expression reported in this dataset was used to compile a panel of toxicologically relevant transcriptional biomarkers by integration of three different approaches, where the key difference between these lay in the level of aggregation: Biomarkers were selected either based on individual or across multiple compounds, as well as on individual or collective gene (sets) level (**Figure 12**). Two strategies (Top10 and Univariate approach) focused on individual gene responses, where the Top10 approach was based on effect-size of individual compound exposure-induced responses and the Univariate approach prioritized DEG quantity across all compounds. The third strategy (Multivariate approach) selected biomarkers based on the level of collective gene sets and across all compounds according to a penalized logistic regression-based classifier (see [2.2.9 Biomarker panel selection] for a detailed description).

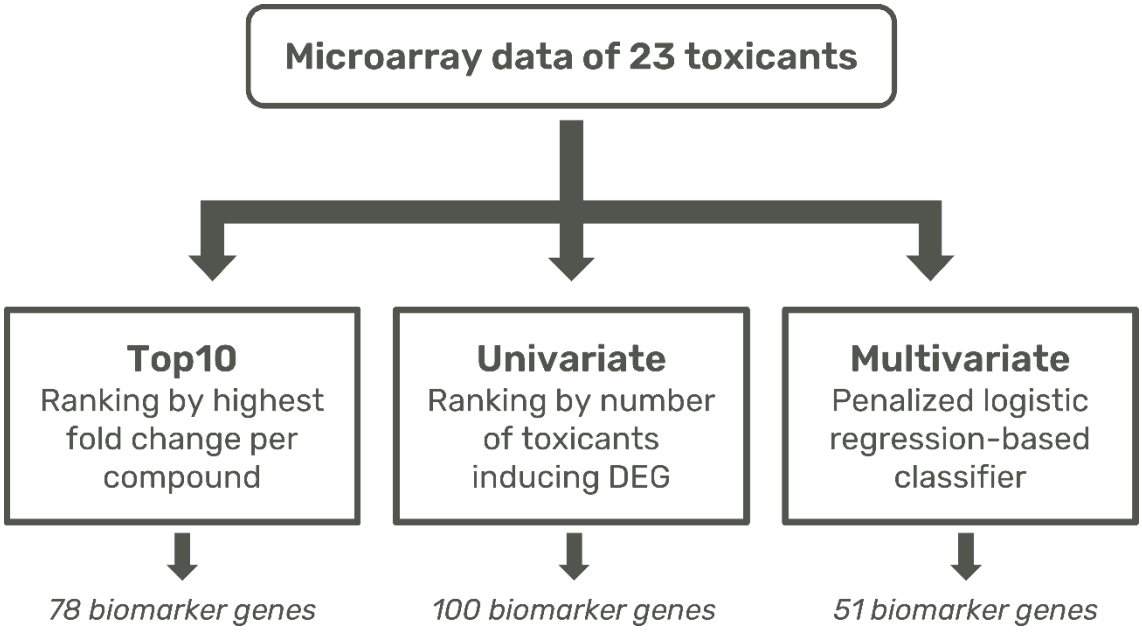


Figure 12: Biomarker panel selection scheme. *Biomarker genes indicative of disrupted neuroectodermal differentiation were selected based on three complementary strategies, yielding 78, 100, and 51 biomarker genes, respectively. The final biomarker panel compiled all genes selected by at least one selection strategy and contained 190 biomarker genes overall.*

Application of these strategies to the microarray data set yielded 78, 100 and 51 candidate genes, respectively. Importantly, all selection strategies yielded largely distinct gene sets, with only a few genes selected by more than one strategy and none by all of them, suggesting that each approach captured a different subset of transcriptional responses (**Figure 13A**). All genes selected by at least one strategy

constituted the final panel of 190 putative biomarkers representative of exposure-induced gene expression changes in neuroectodermal differentiation.

To evaluate whether the selected transcriptional biomarkers indeed represented genes relevant to embryonic development rather than general stress response markers, the panel was subjected to GO enrichment analysis, revealing significant enrichment of biological process GO terms associated with development, differentiation and morphogenesis (**Figure 13B**). Based on these findings, the biomarker panel was employed to investigate whether integration of differential gene expression can improve the discrimination of developmentally toxic and non-toxic compounds based on cytotoxicity alone.

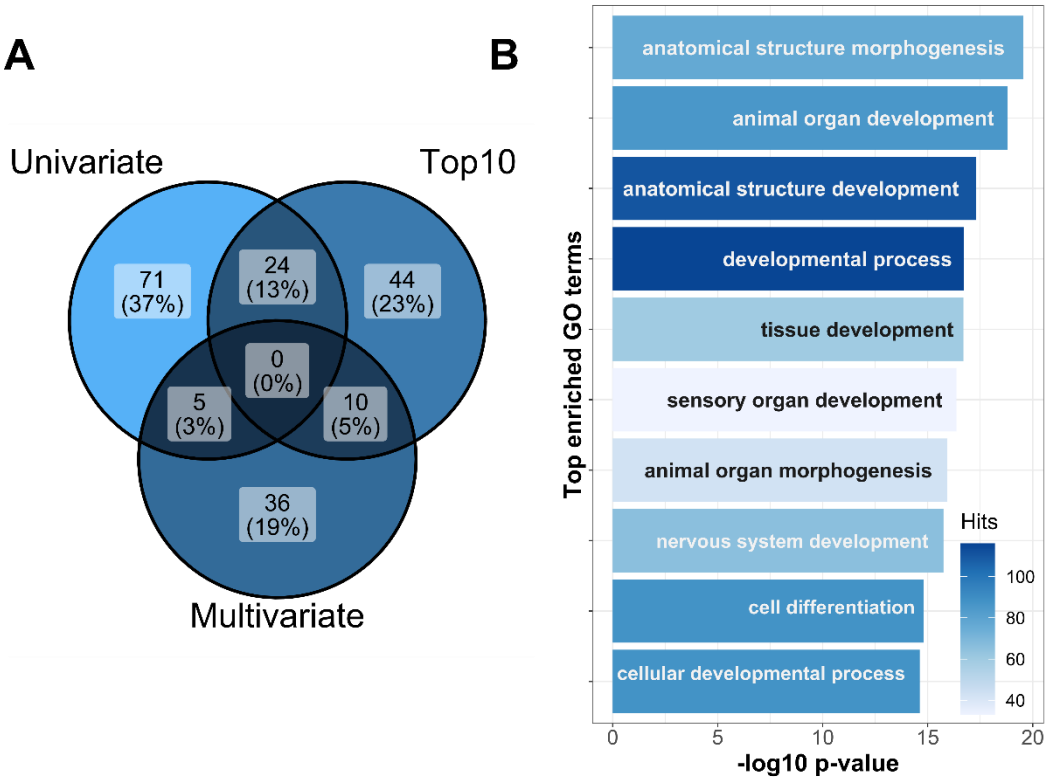


Figure 13: Selection strategy overlap and GO enrichment analysis. *A) Venn diagram of biomarker genes obtained by different selection strategies. B) GO enrichment analysis for the final biomarker panel. The colour scale indicates the number of genes annotated to the respective term. Analysis was restricted to biological process terms.*

3.4 Targeted RNA Expression Measurement reveals distinct modes of differential biomarker expression induced by compound exposure

To assess whether the transcriptional biomarkers capture differentiation-specific perturbations beyond cytotoxicity, biomarker expression changes were analysed following compound exposure during neural induction at four non-cytotoxic concentrations defined relative to the respective EC_{10} (1×, 0.3×, 0.1×, 0.03× EC_{10}). To explore compound-induced variability of biomarker expression, principal

component analysis (PCA) was applied. PCA revealed that the majority of exposure conditions clustered closely with the solvent control. Accordingly, (**Figure 14**) focuses on compounds inducing a shift from this cluster, while the fully annotated PCA is provided in (Supplementary figure 4). The most pronounced separation in the PCA was observed between undifferentiated iPSC and the differentiated solvent control along the first principal component (PC1), capturing a substantial fraction of the transcriptional changes induced during undisturbed differentiation. In comparison, only a minor shift between undifferentiated iPSC and the solvent control was observed along PC2. This enabled evaluation of exposure-induced shifts not only as deviations from the solvent control, but also in relation to the transcriptional changes associated with undisturbed differentiation.

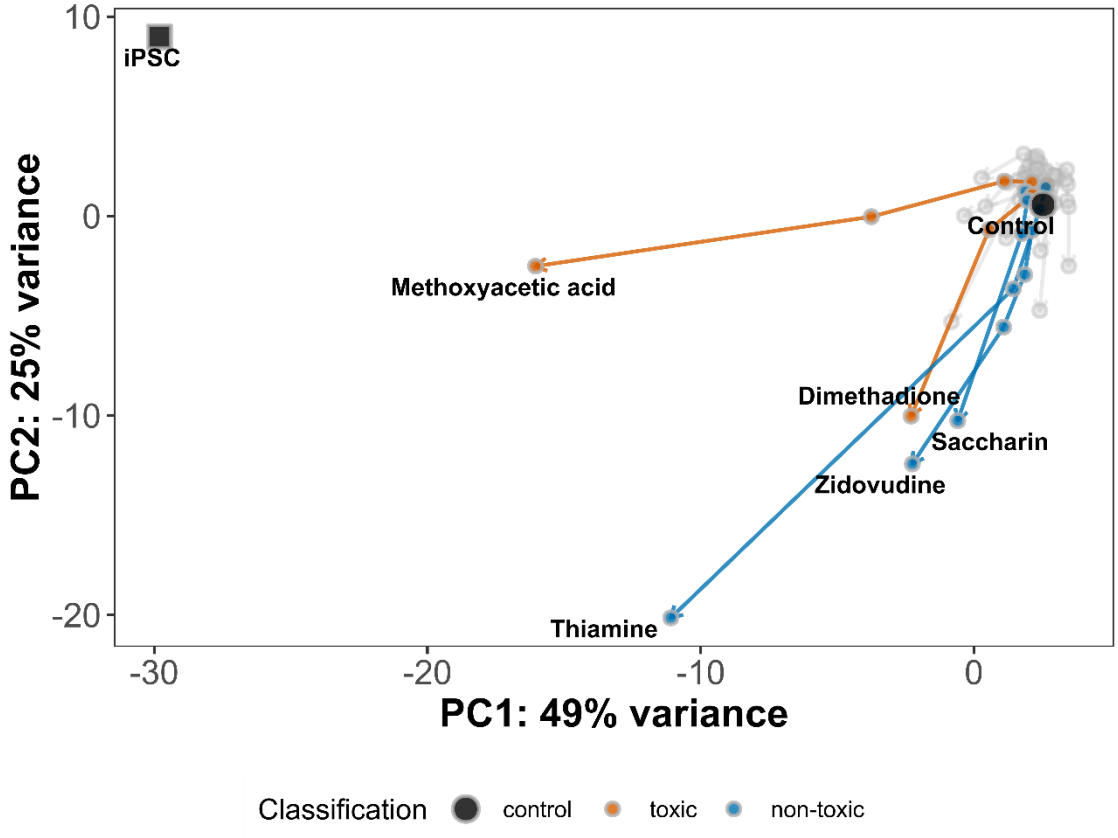


Figure 14: PCA of validation compounds after early neural induction. *Biomarker expression was assessed by Targeted RNA Expression Measurement. Dots indicate neuroectodermal progenitor cells, square indicates undifferentiated iPSC. Arrows indicate increasing concentration. Black data points indicate controls, orange data points and arrows indicate toxicants, blue data points and arrows indicate non-toxicants. Grey data points and arrows represent samples clustering closely with the solvent control.*

For instance, exposure to Methoxyacetic acid induced a pronounced shift along PC1 towards the iPSC state. As changes on PC1 captured a substantial fraction of transcriptional variation associated with undisturbed differentiation, this shift indicated impaired transition from an undifferentiated stem cell-state towards neuroectodermal progenitor cells. In contrast, other compounds (such as

Dimethadione) predominantly induced shifts along PC2 that extended beyond the transcriptional variation observed between iPSC and solvent control, accounting for 25% of the overall transcriptional variance. Thus, these shifts reflected biomarker expression changes distinct from those introduced during undisturbed differentiation. To further resolve whether the observed PCA shifts reflected impaired differentiation or divergent transcriptional programs, principal component loadings were analysed to identify biomarkers with the highest contribution to PC1 and PC2 (**Figure 15**). Analysis of PC1 loadings revealed that the strongest contributors to a shift towards the solvent control included neuroectoderm-associated genes such as PAX6, TPBG, and SIX3, whereas genes contributing to a shift towards the iPSC state included genes associated with pluripotency, such as NANOG and DPPA5. Consistent with the PCA geometry, these loading patterns supported the interpretation that PC1 predominantly captures the variance introduced by transition from an undifferentiated iPSC state towards a neuroectodermal progenitor identity. In contrast, analysis of PC2 loadings showed that the strongest contributors to shifts extending beyond the iPSC-solvent control separation, as observed for Dimethadione, included genes associated with neural crest-identity, such as TFAP2A, TFAP2B and MSX2.

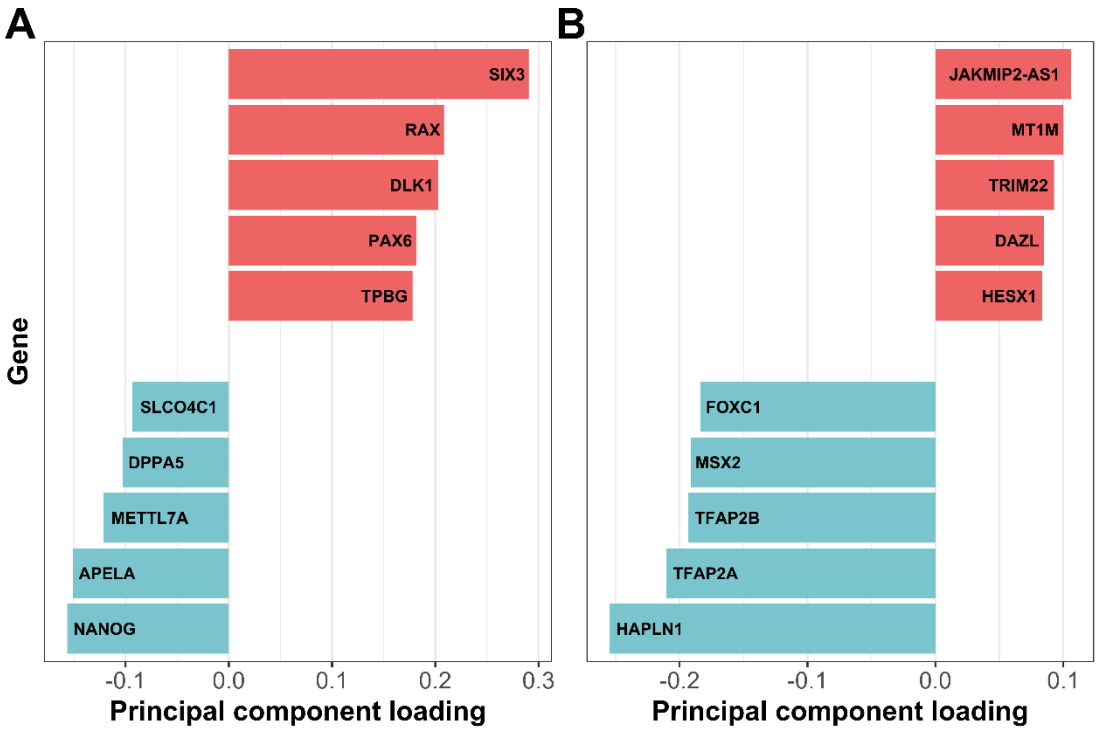


Figure 15: Principal component loadings. *A) shows genes with the highest contribution to PC1, B) shows genes with the highest contribution to PC2. Red bars represent the top five genes contributing to positive principal component shifts, blue bars represent the top five genes contributing to negative principal component shifts.*

Building on the PCA-derived patterns, differentially expressed genes (DEG) were determined to identify specific biomarker expression changes that contribute to the observed variability. Consistent with the PCA, differential expression analysis

demonstrated that exposure to a subset of compounds induced substantial numbers of differentially expressed biomarkers, such as Methoxyacetic acid, which deregulated 105 of 190 biomarkers at the highest concentration (1× EC_{10}) (**Figure 16**). Importantly, differential biomarker expression was largely restricted to near-cytotoxic exposure levels. At 1× EC_{10}, 14 of 18 compounds induced more than three DEG, compared to 7 compounds at 0.3× EC_{10} and only 2 compounds at 0.1× EC_{10}. At lower concentrations, only single biomarkers were differentially expressed. Furthermore, although the three biomarker selection strategies comprised largely distinct gene sets, all subsets displayed similar concentration-dependent patterns, with most changes occurring at near-cytotoxic exposure levels (Supplementary figure 5). Notably, TREx-based biomarker analysis in a non-differentiating Caco2 control population revealed substantially fewer DEG in response to the same compounds, supporting that the observed biomarker expression changes were specific to differentiating cells (**Figure 16**, right).

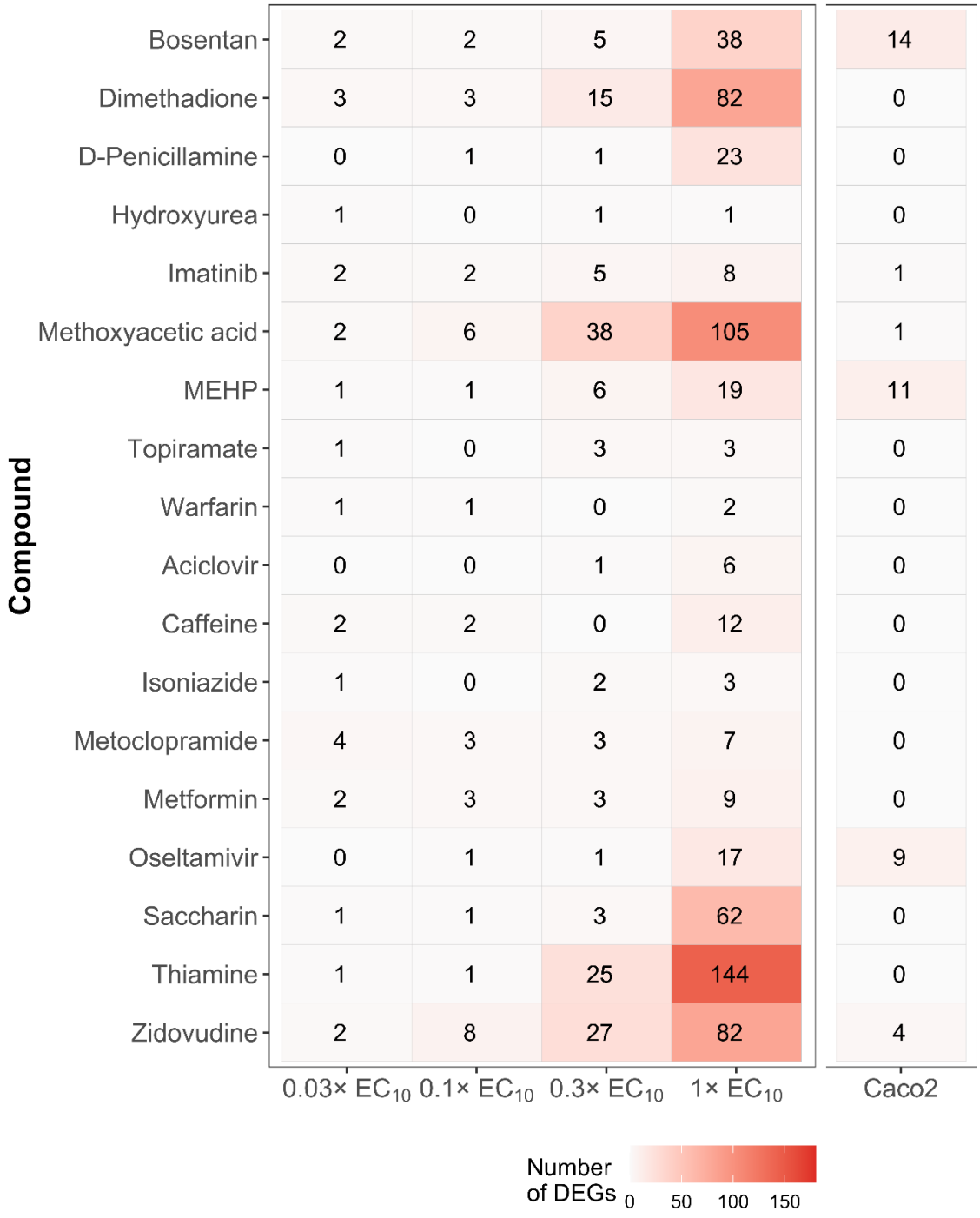


Figure 16: Heat map of differentially expressed biomarker genes. *Concentrations are given relative to the EC10 derived from cytotoxicity analysis during early neural induction. Differential biomarker expression in Caco2 cells was assessed after exposure to the highest compound concentration tested during neural induction of iPSC (1× EC_{10}).*

To further substantiate the interpretations derived from the PCA, differentiation pattern clustering (DiPaC) (87) was applied: By contextualizing compound-induced gene expression changes relative to differential gene expression introduced during undisturbed differentiation, this supervised clustering approach enabled

summarizing differentially expressed biomarkers in biologically interpretable groups. Consistent with observations from the PCA, DiPaC analysis enabled distinction of two modes of disturbed differentiation, which was best exemplified by Methoxyacetic acid, which primarily induced a shift along PC1 towards the iPSC state, and Dimethadione, which predominantly induced a shift along PC2 beyond the iPSC-solvent control separation.

Consistent with patterns observed in the PCA, DiPaC analysis revealed that exposure to Methoxyacetic acid in the highest tested concentration primarily induced differential expression of biomarkers assigned to the Differentiation Pattern Groups (DPG) 1 and 2 (**Figure 17**). These DPG comprise expression changes representing insufficient or completely blocked upregulation of genes that are normally induced during undisturbed neural induction, including genes associated with neuroectodermal identity such as SOX2 or PAX6. In alignment with the shift on PC1, this suggested that exposure to Methoxyacetic acid inhibits acquisition of neuroectodermal properties. This interpretation was directly reflected in the distribution of PC1 loadings across DPG obtained for Methoxyacetic acid exposure (**Figure 18**). Genes contributing to a shift along PC1 towards the differentiated state were predominantly assigned to DPG1 and DPG2, reflecting insufficient or blocked upregulation under Methoxyacetic acid exposure. Conversely, genes contributing to a shift towards the undifferentiated iPSC state were primarily assigned to DPG6 and DPG7, indicating insufficient or blocked downregulation. In comparison to this, the distribution of PC2 loadings across DPG obtained for Methoxyacetic acid exposure showed less specific patterns. Together, the integrated analysis consistently supports the notion that exposure to Methoxyacetic acid impairs the transition from an undifferentiated iPSC state towards neuroectodermal progenitor cells.

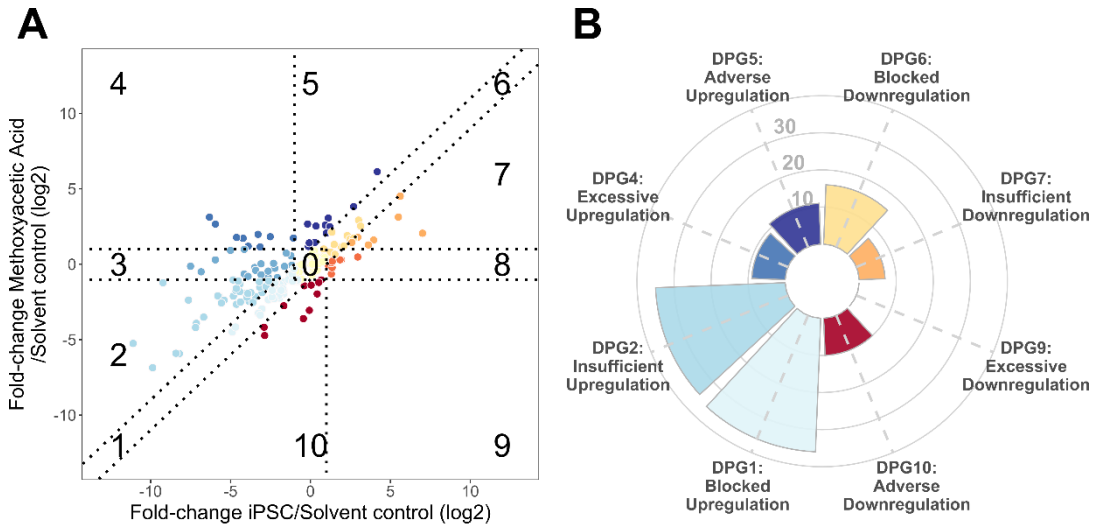


Figure 17: Differentiation Pattern Clustering after exposure to Methoxyacetic acid. *A) DiPaC plot showing gene expression changes induced by the highest tested concentration of Methoxyacetic acid against gene expression changes during undisturbed differentiation. Dotted lines represent cut off values that define DPGs. B) Histogram indicating the number of biomarker genes summarized in each DPG except DPG0, DPG3, and DPG8.*

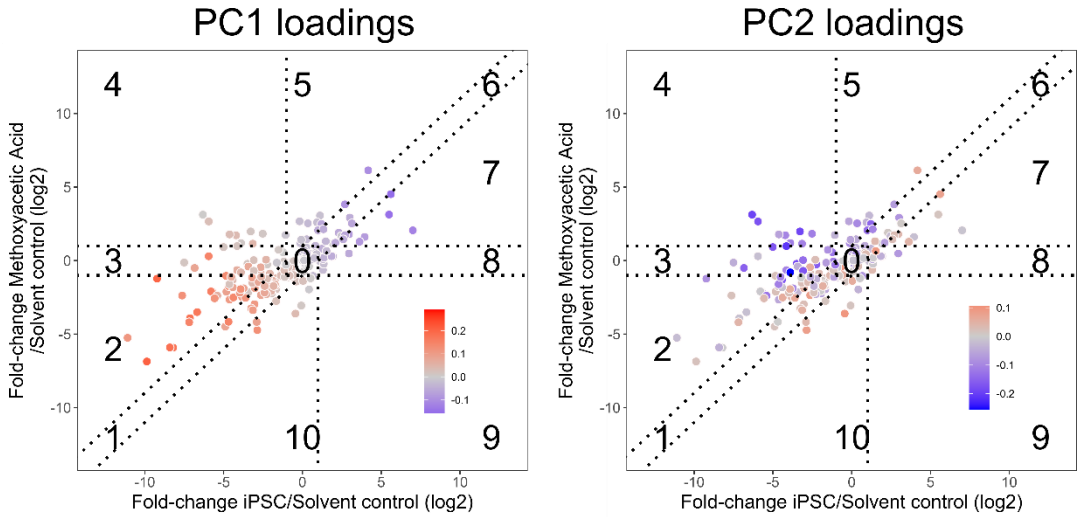


Figure 18: Principal component loadings in the context of DiPaC analysis after exposure to Methoxyacetic acid. *Fold changes used to classify biomarker genes in DPGs were obtained after exposure to Methoxyacetic acid in the highest tested concentration.*

Conversely, exposure to Dimethadione in the highest tested concentration induced a substantial number of differentially expressed biomarkers in DPG 4 and 5 – groups, which comprise expression changes that represent adverse upregulation of genes which are not induced during undisturbed differentiation or excessive upregulation of genes not induced to the extent observed under compound exposure (**Figure 19**). Consistent with the shift on PC2, this observation suggested that exposure to Dimethadione induced divergent transcriptional programs, which may reflect alternate differentiation trajectories. Accordingly, PC2 loadings were analysed in the context of the DiPaC results, analogous to the approach used for Methoxyacetic acid (**Figure 20**). This analysis revealed that genes with the highest contribution to a shift

along PC2 are predominantly assigned to DPG4 and DPG5 under Dimethadione exposure. Consistent with the interpretation that exposure to Dimethadione induced divergent transcriptional programs, biomarkers assigned to DPG4 and DPG5 included genes associated with neural crest identity, such as PAX3, TFAP2A, TFAP2B and MSX2, suggesting that that the divergent transcriptional program induced by Dimethadione exposure may involve activation of neural crest-related gene expression. On the other hand, in comparison to Methoxyacetic acid exposure, PC1 loading distribution showed a stronger enrichment in DPG3 and DPG8, which comprise gene expression changes unaffected by compound exposure. Importantly, these results suggest a perturbed differentiation pattern distinct from the impaired differentiation progression observed for Methoxyacetic acid, illustrating that assessment of differential biomarker expression can unravel distinct modes of disturbed differentiation beyond simple cytotoxicity. For all remaining compounds, DiPaC analysis results are provided in Supplementary figure 6 and Supplementary figure 7.

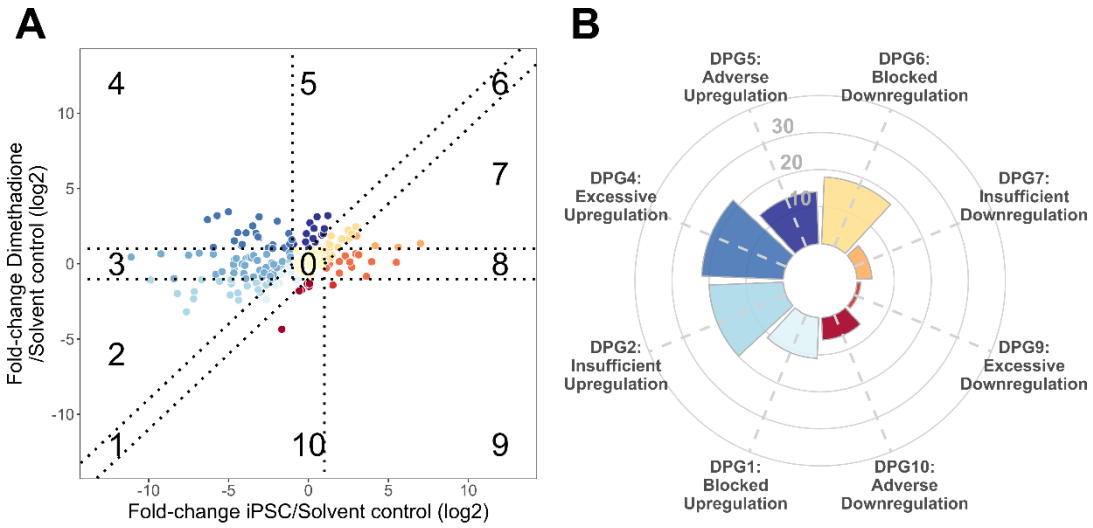


Figure 19: Differentiation Pattern Clustering after exposure to Dimethadione. *A) DiPaC plot showing gene expression changes induced by the highest tested concentration of Dimethadione against gene expression changes during undisturbed differentiation. Dotted lines represent cut off values that define DPGs. B) Histogram indicating the number of biomarker genes summarized in each DPG except DPG0, DPG3, and DPG8.*

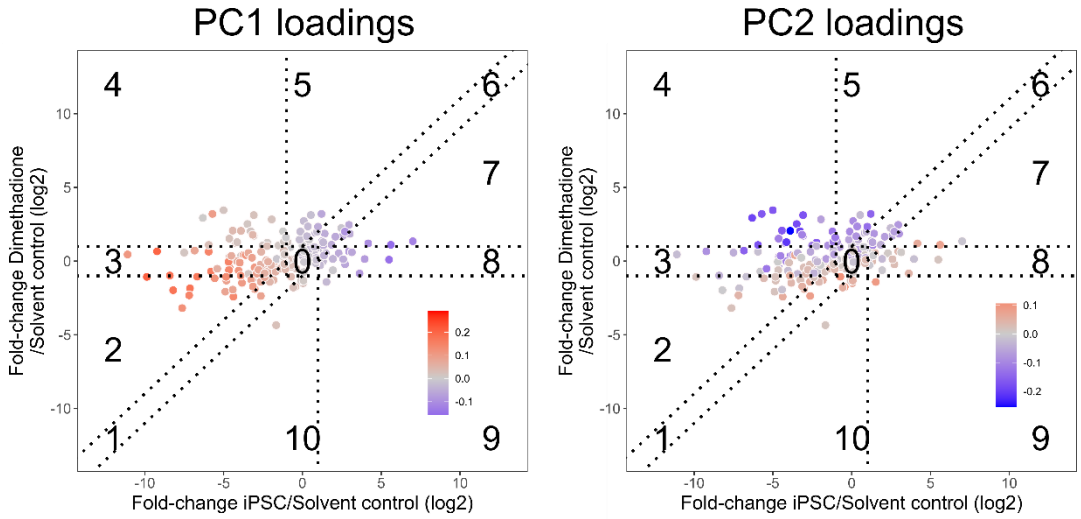


Figure 20: Principal component loadings in the context of DiPaC analysis after exposure to Dimethadione. *Fold changes used to classify biomarker genes in DPGs were obtained after exposure to Dimethadione in the highest tested concentration.*

3.5 Random forest-based prediction provides moderately accurate prediction of developmental toxicity by early neural induction

To derive developmental toxicity prediction based on effects on viability and gene expression, a random forest model was employed. Although random forests are capable of integrating heterogeneous types of data, high-dimensional data can impair interpretability. Therefore, compound-induced cytotoxicity and differential gene expression were aggregated into 19 interpretable summaries (IS), as previously reported in (29) (Table 11). Out of these, 12 IS represented benchmark concentrations, where a predefined benchmark response was observed (e.g., the concentration of a compound that induced at least one differential biomarker). Importantly, every compound can in principle be considered a developmental toxicant depending on its respective concentration. Thus, to anchor the prediction model to *in vivo* relevant concentrations, these 12 benchmark concentration-IS were normalized to the c_{max} of the respective compound.

Table 11: Interpretable summaries obtained from cytotoxicity and gene expression measurements after early neural induction

Interpretable Summary	Explanation	Group	Type
medianEC10	Concentration where a fitted concentration-viability curve attains a value of 0.9, 0.8 or 0.5.	Cytotoxicity	Benchmark concentration
medianEC20			Benchmark concentration
medianEC50			Benchmark concentration
ALOEC.Min	Summary values of the absolute lowest observed effective concentration of a specific gene at which the absolute log2 fold change exceeds 1.	Gene expression	Benchmark concentration
ALOEC.Q25			Benchmark concentration
ALOEC.Med			Benchmark concentration
ALEC.Min	Summary values of the absolute lowest effective concentration of a specific gene at which a fitted concentration-expression curve attains an absolute log2 fold change of 1.		Benchmark concentration
ALEC.Q25			Benchmark concentration
ALEC.Med			Benchmark concentration
DEG1	The minimal observed concentration at which 1, 2, or 5 genes are differentially expressed (adj. p-Value <0.05, abs log2 fold change >1).		Benchmark concentration
DEG2			Benchmark concentration
DEG5			Benchmark concentration
nrdegs_max	The maximum number of DEGs (adj. p-Value <0.05, abs log2 fold change >1).		Number of genes
Penalty.ALEC.Min	Penalty if ALEC summary values cannot be determined.		Binary penalty variable
Penalty.ALEC.Q25			Binary penalty variable
Penalty.ALEC.Med			Binary penalty variable
Penalty.DEG1	Penalty if DEG1, DEG2 and DEG5 values cannot be determined.		Binary penalty variable
Penalty.DEG2			Binary penalty variable
Penalty.DEG5			Binary penalty variable

Importantly, random forest-based prediction models do not inherently assume that lower benchmark concentrations relative to the corresponding *in vivo* reference concentration reflect higher developmental toxicity risk. This flexibility may result in the model learning global non-monotonic concentration–response patterns that do not reflect the experimental data of individual compounds, thereby complicating biological interpretation and violating basic toxicological principles (Supplementary figure 8). To address this limitation, the random forest models for DevTox prediction were trained with monotonic constraints, enforcing one-sided

thresholding for all benchmark concentration IS to ensure that increasing *in vitro* potency could not be interpreted as reduced toxicity. Furthermore, Leave-One-Out cross validation (LOOCV) was employed to restrain overfitting of the model.

IS-based LOOCV random forests with monotonic constraints were then employed to predict the developmental toxicity risk of each compound. The resulting output, referred to as the DevTox score, represented a continuous probability scale from 0 (no developmental toxicity expected) to 1 (highest predicted risk). Overall, the classifier separated around half of the developmental toxicants and non-toxicants with high accuracy, yielding DevTox scores close to 1 or 0 in the case of toxicants or non-toxicants, respectively. The remaining compounds however exhibited intermediate DevTox scores, leading to limited discriminatory potential due to a substantial overlap of toxicants and non-toxicants (**Figure 21**). Subsequent binary classification by applying a threshold to maximise classification accuracy yielded an overall prediction accuracy of 72.2%, a sensitivity of 55.6% and a specificity of 88.9%.

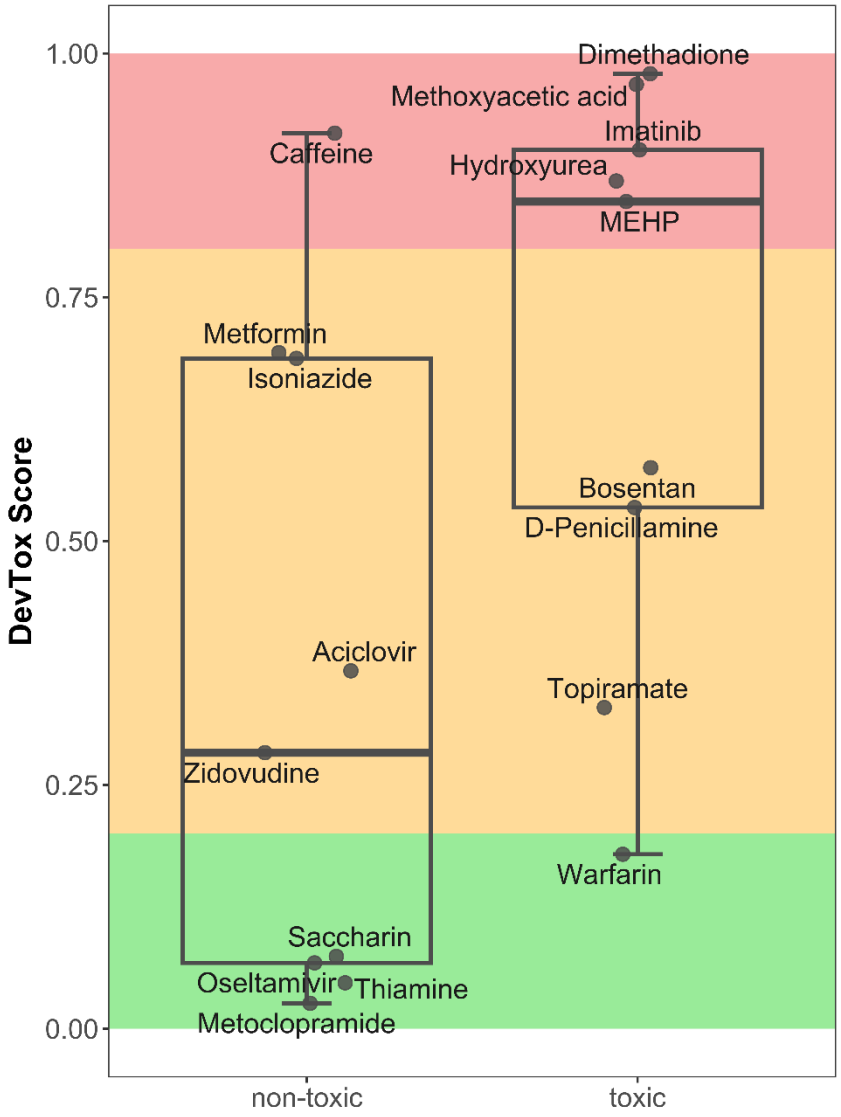


Figure 21: Prediction of developmental toxicity after early neural induction. *DevTox scores for each compound were derived from Leave-One-Out cross validated random forest prediction. The colour scheme represents a three-category framework for compound classification, where data points in the red area represent compounds classified as toxic, data points in the yellow area are classified as uncertain and data points in the green area are classified as non-toxic.*

Alternatively, the observation that classifier predictions clustered near the extremes for some compounds – either close to 1 for toxicants or close to 0 for non-toxicants – while other compounds yielded intermediate scores, motivated categorizing prediction results into three classes: toxic (DevTox score > 0.8), non-toxic (DevTox score < 0.2) and uncertain (DevTox score 0.2 – 0.8). This framework enabled compound classification within a defined concentration range: By limiting interpretation to predictions categorized as either toxic or non-toxic, the model yields a more reliable classification with an accuracy of 81.8%. This comes at the cost of leaving an intermediate DevTox score range in which developmental toxicity cannot be predicted with confidence. For instance, Zidovudine, a known non-

toxicant at an *in vivo* c_{max} of 2 µM, was confidently classified as non-toxic up to approximately 1 µM, while higher concentrations yielded uncertain classification, including the *in vivo* c_{max} (**Figure 22B**). Conversely, Bosentan, a known developmental toxicant at an *in vivo* c_{max} of 2.9 µM, was classified as toxic only at concentrations above 15 µM, corresponding to roughly fivefold the *in vivo* exposure level (**Figure 22A**). Importantly, two compounds, Caffeine and Warfarin, were misclassified even within the three-category framework, being assigned to the opposite class rather than to the uncertain range.

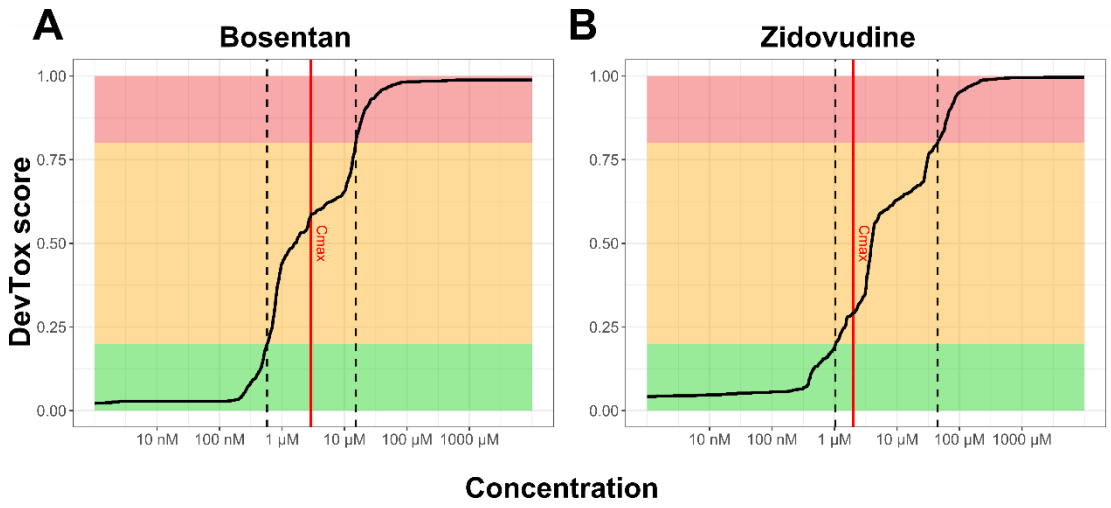


Figure 22: Concentration dependent developmental toxicity prediction after early neural induction*. A) Prediction of the developmental toxicant Bosentan, B) Prediction of the non-toxic reference compound Zidovudine. Red, yellow and green areas indicate whether a prediction leads to toxic, uncertain or non-toxic classification. Dashed lines indicate maximum safe concentration and the minimal toxic concentration. The red lines indicate the cmax of the respective compound.*

Besides assessing overall prediction performance, it was an essential goal to evaluate whether including biomarker expression as an endpoint can improve prediction accuracy compared to cytotoxicity alone. For this purpose, the AUC of the random forest classification was compared to the AUC-based TSI, which used the EC_{10} as an indicator of the earliest cytotoxic response to separate toxicants and non-toxicants. Although cross-validation generally yields more conservative AUC estimates than direct two-class separations, the observed LOOCV-based classifier AUC (0.704) compared with the TSI (0.852) nevertheless indicated that inclusion of biomarker expression did not improve prediction accuracy. To confirm this, a second LOOCV-based random forest classifier was trained solely on cytotoxicity-based IS. This classifier yielded an AUC of 0.852, demonstrating that in the case of the 18 validation compounds, integration of biomarker expression did not improve prediction accuracy.

Taken together, integrating cytotoxicity and gene expression by IS-based random forest classification after exposure during early neural induction yielded moderately accurate prediction of DevTox. Interpretation of prediction values in a three-

category framework enabled definition of more reliable concentration ranges. Intriguingly, although assessing differential biomarker expression was shown to provide mechanistically informed insights into disturbed differentiation, integration of gene expression did not improve prediction accuracy compared to cytotoxicity alone. This indicated that a subset of compounds, such as Bosentan, may not be reliably identified as developmentally toxic despite in-depth evaluation of disturbed early neural induction – prompting whether expansion of the differentiation model is necessary to improve prediction accuracy.

3.6 Organoid-based prolonged neural induction yields neuroepithelial progenitor-like cells competent to self-organise into neural rosettes

To explore whether integration of later differentiation stages could elevate prediction accuracy, a prolonged neural induction protocol was established to advance cells towards neuroepithelial progenitor-like cells (NPLC) competent to self-organise into neural rosettes (**Figure 23**). However, while 2D monolayer culture was preferred for early stages of neuroectodermal development, extending cultivation beyond six days led to substantial overgrowth and detachment of cells, resulting in highly variable outcomes of differentiation. In order to overcome this limitation, prolonged neural induction was conducted in suspension culture. After seeding into uncoated dishes, the undifferentiated cells quickly formed 50 – 100 µm diameter aggregates over two days of cultivation (**Figure 24A**). Subsequently, initiation of neural induction led to progressive growth of the assembled organoids to diameter sizes up to around 200 µm on DoD11 (**Figure 24B**).

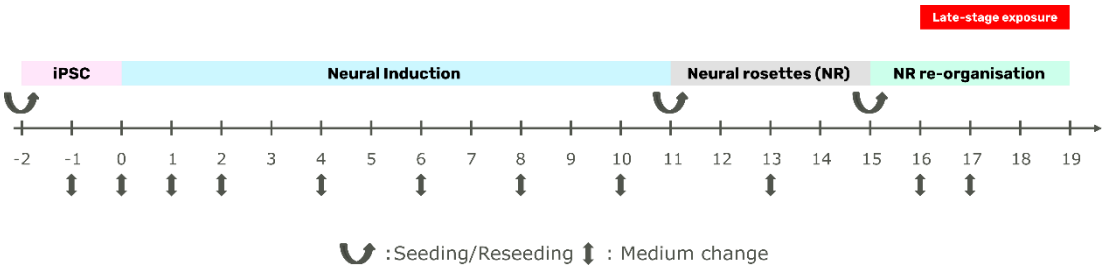


Figure 23: Experimental design prolonged neural induction*. Schematic of the iPSC to neuroepithelial progenitor-like cells differentiation protocol. Straight arrows indicate medium changes, curved arrow indicates seeding. On DoD15, cells were used for evaluation of rosette formation.*

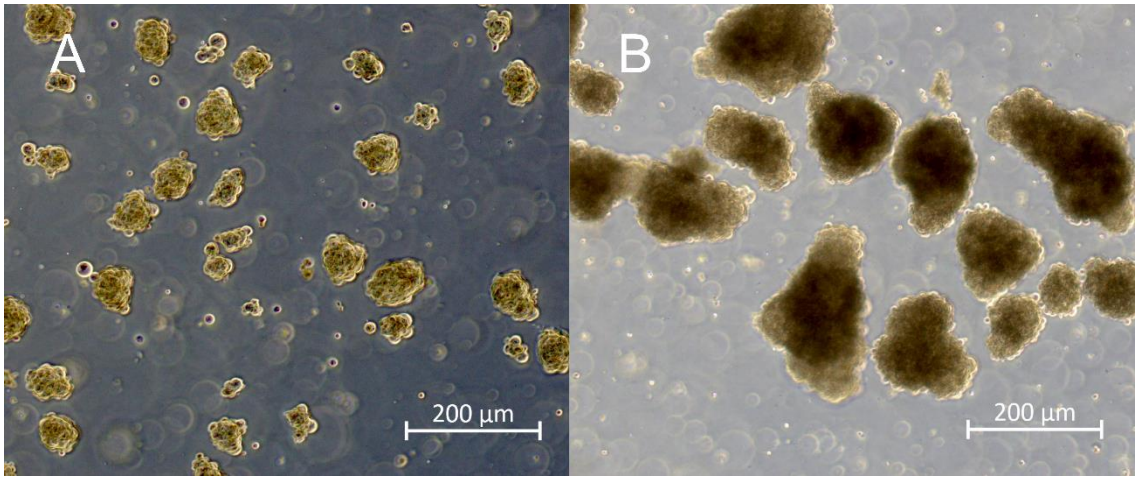


Figure 24: Light microscopy of representative organoids during neural induction. *Images were taken two days after seeding (A) and after eleven days of differentiation (B). Scale bar represents 200 µM.*

To assess whether organoid-based neural induction leads to adoption of a neuroepithelial progenitor-like identity comparable to monolayer differentiation at the transcriptional level, expression of established marker genes was analysed by qPCR on DoD6 and DoD11 (**Figure 25**). Progressive downregulation of the pluripotency marker genes NANOG and POU5F1 comparable between monolayer- and organoid-based neural induction protocols demonstrated loss of an undifferentiated stem cell identity. Likewise, monolayer- and organoid-based differentiation led to similar upregulation of the neuroectodermal markers PAX6 and OTX2. Importantly, whereas PAX6 was further upregulated from DoD6 to DoD11, the maximum upregulation of OTX2 was already observed after six days of differentiation, underlining its role as a pioneer of neuroectodermal differentiation. Intriguingly, prolonging neural induction led to progressive downregulation of the pluripotency-associated cadherin CDH1 alongside upregulation of the neuroectodermal cadherin CDH2, exhibiting a transcriptional signature compatible with subsequent structural organisation associated with cadherin switching. Notably, the observed CDH1-to-CDH2 switch was more pronounced in organoid- compared to monolayer-based differentiation.

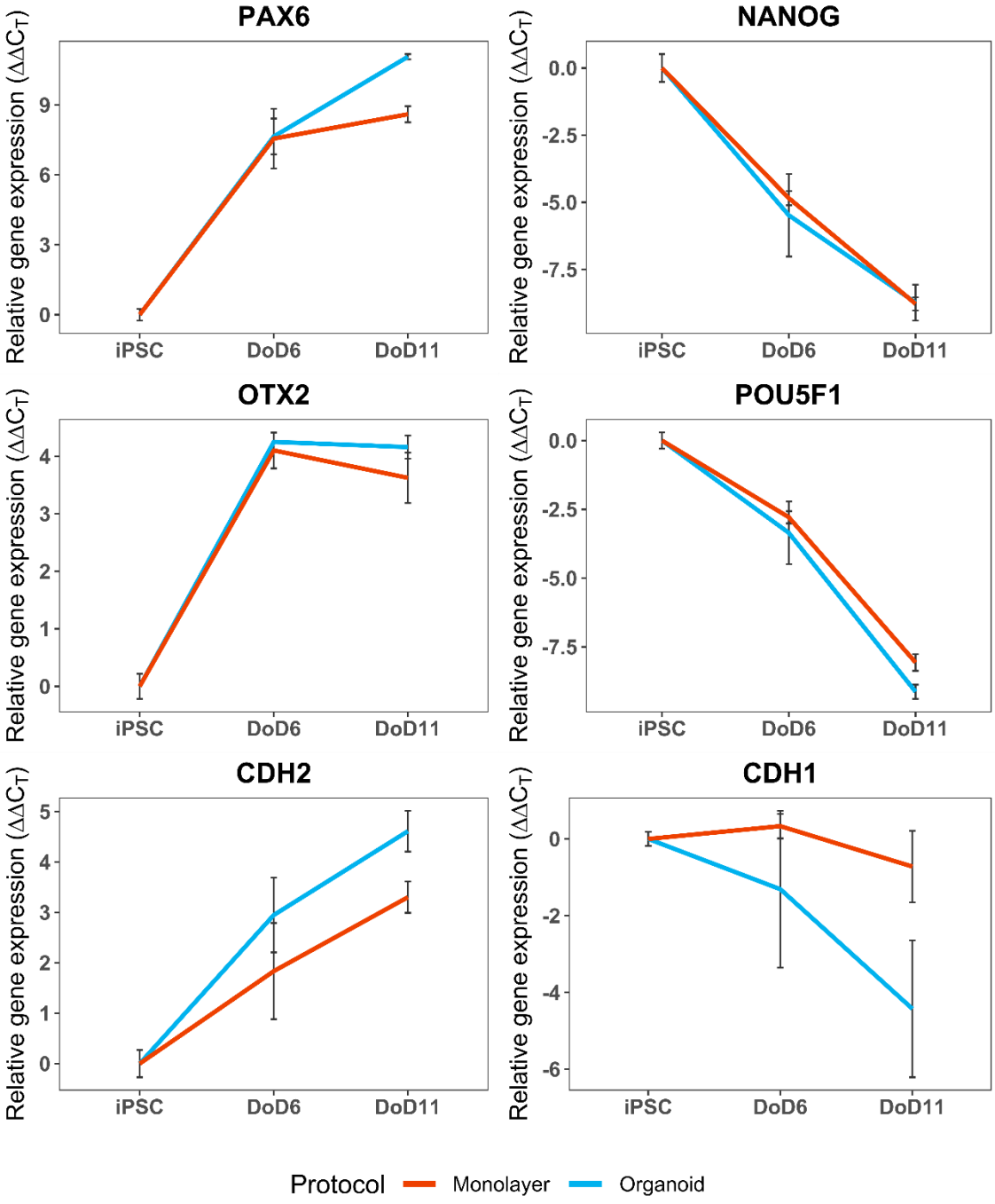


Figure 25: Relative gene expression changes during monolayer and organoid-based neural induction. *The time of analysis is indicated on the x-axis (undifferentiated iPSC, 6 or 11 days after initiating neural induction). Orange lines indicate monolayer-based differentiation, blue lines indicate organoid-based differentiation.*

Given the transcriptional signatures indicative of a neuroepithelial progenitor-like identity, it was next assessed whether the differentiated cells were competent to self-organise into neural rosettes by transferring the organoids to a coated dish. During cultivation without shaking, the transferred organoids attached and formed outgrowth that resulted in a confluent monolayer after four more days. On DoD15, radially organised cell clusters with elongated cells were observed, indicating

formation of neural rosettes (**Figure 26A**). Importantly, the competence to self-organise into these clusters was retained after disassembling the putative neural rosettes: After detachment and reseeding as single cell suspension, the NPLC re-organised into rosettes within the following four days (**Figure 26B**).

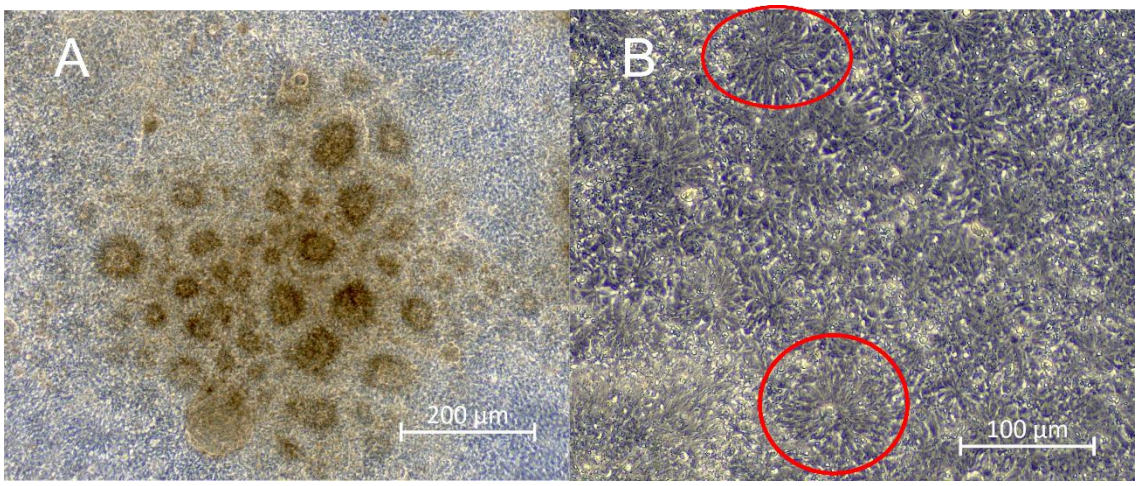


Figure 26: Light microscopy of putative neural rosettes before and after reseeding. *Images were taken after 15 days (A) or 19 days (B) or neural induction. Red circles indicate putative neural rosettes that re-organised after reseeding as single cell suspension. Scale bar represents 200 µM (A) or 100 µM (B).*

3.7 Differentiation stage-dependent sensitivity to compound-induced cytotoxicity suggests improved prediction capabilities of the late-stage model

To explore whether compound-induced effects on later differentiation stages can enhance DevTox prediction, cells that underwent neural induction were exposed to the validation compounds during re-organisation into neural rosettes (**Figure 23**). As cytotoxicity represents the most fundamental perturbation of an *in vitro* system that can substantially disrupt self-organisation, the first goal was to measure exposure-induced viability decrease by CTB assay. Exposure to all developmental toxicants and 7 out of 9 non-toxic reference compounds induced concentration-dependent cytotoxicity. In the case of the non-toxicants Saccharin and Aciclovir, AIC-based model selection favoured a constant model, indicating no concentration-dependent cytotoxicity up to the highest soluble concentration. In order to interpret compound-induced viability changes in relation to relevant *in vivo* concentrations, concentration-viability curves of all compounds that induced cytotoxicity were normalized to the respective c_{max} (**Figure 27A**). This analysis revealed a similar general pattern compared to the cytotoxicity observed after exposure during early neural induction: Although some toxicants induced cytotoxicity in concentrations near their respective *in vivo* relevant concentration, other compounds required substantially higher concentrations, resulting in an overlap of toxic and non-toxic validation compounds and potentially limiting DevTox prediction.

In order to systematically assess the predictive potential of cytotoxicity induced by exposure during the late stage, EC_{10} values were determined from each concentration-viability curve and used for TSI analysis (**Figure 27B**). Thereby, the resulting TSI of 0.937 demonstrated better separation based on cytotoxicity induced during the late exposure time window compared to the early stage ($TSI_{early} = 0.852$). This result prompted a closer examination of differences between cytotoxicity induced by specific compounds during early and late stages.

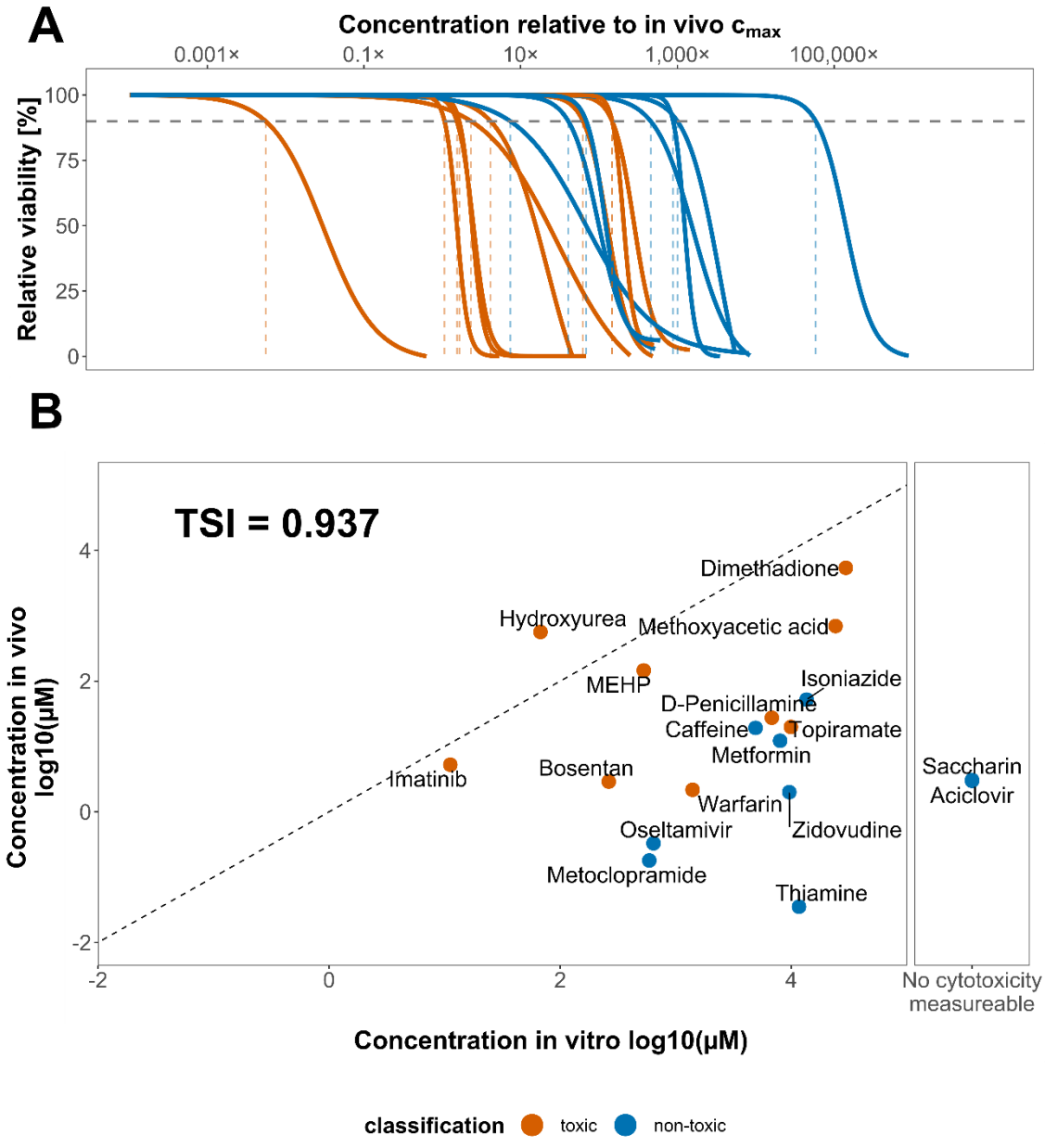


Figure 27: Discrimination of developmental toxicants and non-toxic reference compounds based on cytotoxicity in the late-stage model. *A) Concentration-viability curves of all validation compounds were normalized to the c_{max} of the respective compound. Blue lines indicate non-toxic compounds, orange lines indicate toxicants. B) Toxicity Separation Index analysis using median EC_{10} values to define positive in vitro concentrations and maternal c_{max} values as in vivo reference concentrations. Blue dots represent non-toxicants, orange dots represent toxicants.*

Here, a specifically interesting compound was MEHP (**Figure 28**), where direct comparison of concentration-viability curves derived from the different exposure schedules revealed substantially lower effective concentrations during the late stage. For instance, the EC_{10} of around 1 mM during early neural induction was approximately 4 times higher compared to the late exposure time window, demonstrating a stage-dependent sensitivity to compound-induced cytotoxicity.

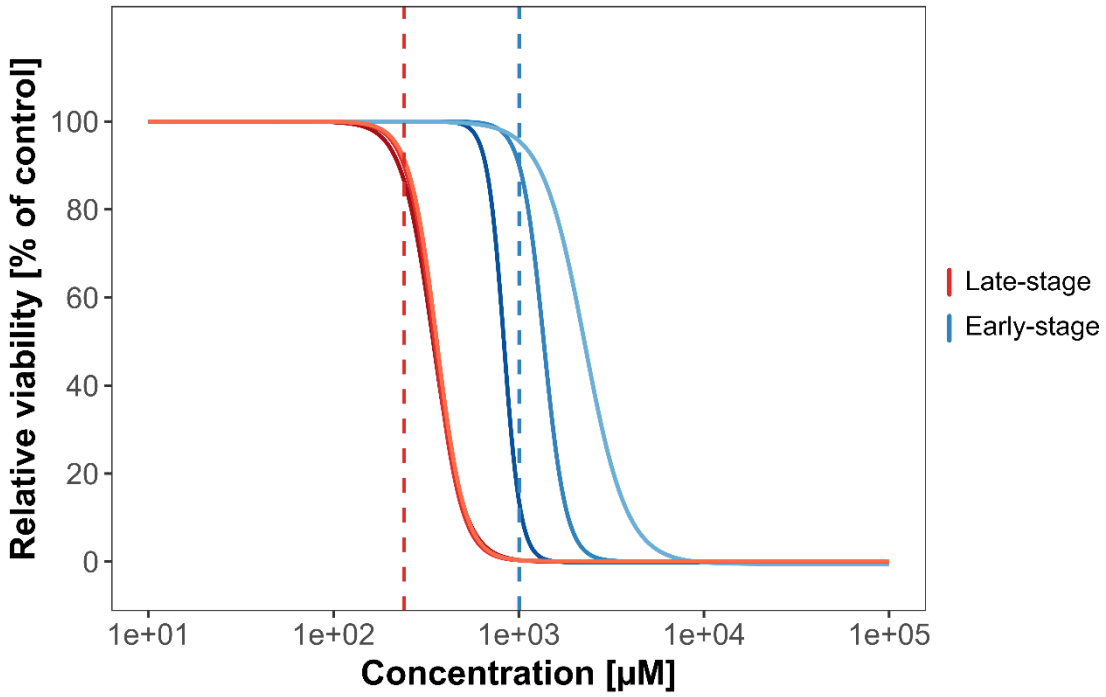


Figure 28: Concentration-viability curves after exposure to MEHP during different stages of neural induction. *Blue lines indicate individual experiments during the early-stage, orange lines indicate the individual experiments during the late-stage. Dotted lines indicate the median EC_{10} for the early (blue) and the late (orange) stage.*

This intriguing result prompted further exploration of stage-dependent sensitivity to other validation compounds. Systematic analysis of EC_{10} values derived from early and late stages was performed based on the log2 ratio of two benchmark concentrations (here, early and late EC_{10} values), which henceforth will be referred to as Endpoint Sensitivity Shift (ESS) index (**Figure 29**). This analysis revealed pronounced stage-dependent cytotoxicity in a subset of compounds. Notably, while some compounds exhibited lower EC_{10} values in the late stage (positive ESS, such as Hydroxyurea), other compounds were more cytotoxic during early neural induction (negative ESS, such as Aciclovir). Finally, some compounds exhibited ESS values close to zero, indicating similar sensitivity of early and late exposure (such as Warfarin or Zidovudine). Interestingly, some compounds which were falsely classified by early neural induction exhibited stage-dependent sensitivity patterns that might indicate improved prediction capabilities by integrating late exposure into the test system. For instance, Bosentan, a developmental toxicant which was previously classified false-negative, was cytotoxic in lower concentrations when

exposed during the late stage. Conversely, Caffeine, a false-positive classified non-toxicant was less cytotoxic compared to early neural induction. Together, these patterns provided a promising perspective for improving DevTox prediction by integrating the additional exposure time window, and prompted further exploration of how compound exposure interferes with neuronal differentiation during this stage.

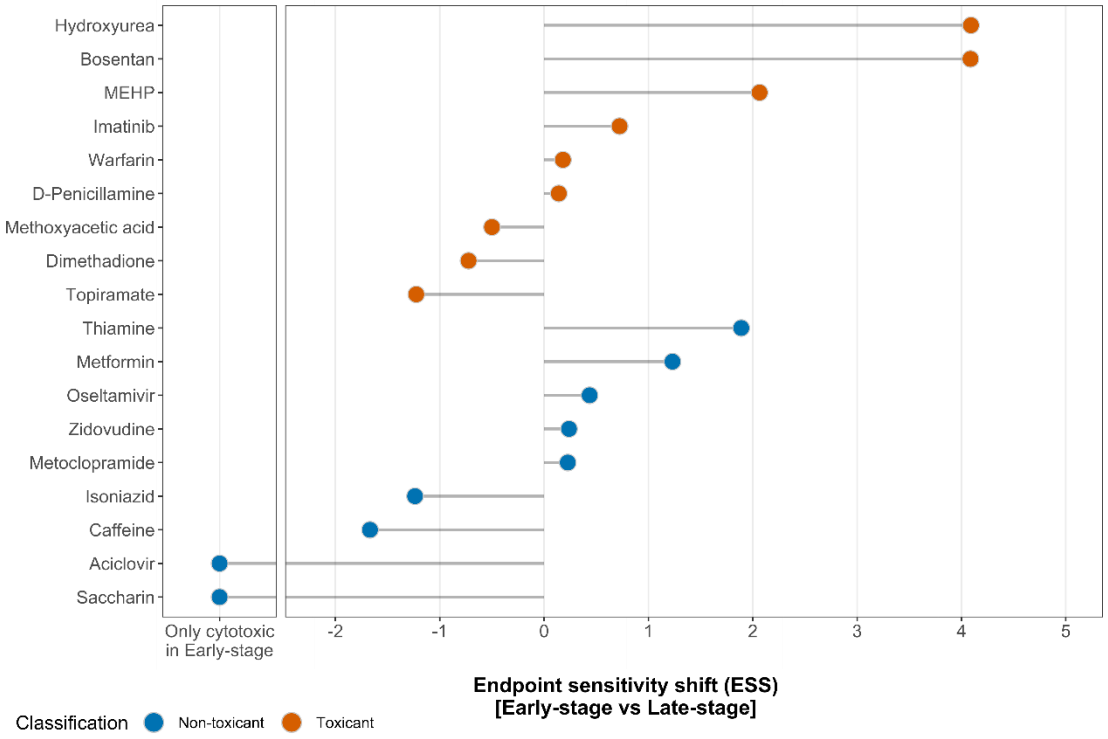


Figure 29: Endpoint sensitivity shift index (ESS) of early and late-stage cytotoxicity. *The ESS was defined as the log2 ratio of median EC10 cytotoxicity values from early and late-stage neural induction. Positive ESS values indicate higher sensitivity during the late stage, negative ESS values indicate higher sensitivity during the early stage. Orange data points indicate developmental toxicants, blue data points indicate non-toxic reference compounds.*

3.8 Exposure to a subset of validation compounds disrupts neural rosette formation at non-cytotoxic concentrations

To assess exposure-induced disruption of cellular self-organisation beyond general cytotoxicity, compound effects on rosette formation were examined. Reseeding rosette-forming NPLC as single cells in spots, as established previously in (77), yielded discrete, radially organised cell clusters which provided confined microenvironments with consistent local cell density. Re-organisation into neural rosettes was confirmed by immunostaining of ZO1 and GM130, establishing the baseline for subsequent exposure experiments (**Figure 23**). Immunostaining of spots exposed to the validation compounds demonstrated that while re-organisation into rosette remained largely intact under many conditions, exposure to a subset of compounds disrupted neural rosette formation. For instance, exposure

to the highest tested concentration of MEHP (1× EC_{10}) resulted in near-complete loss of neural rosettes, highlighting rosette formation as a sensitive endpoint for assessing exposure-induced disruption of self-organisation (**Figure 30**).

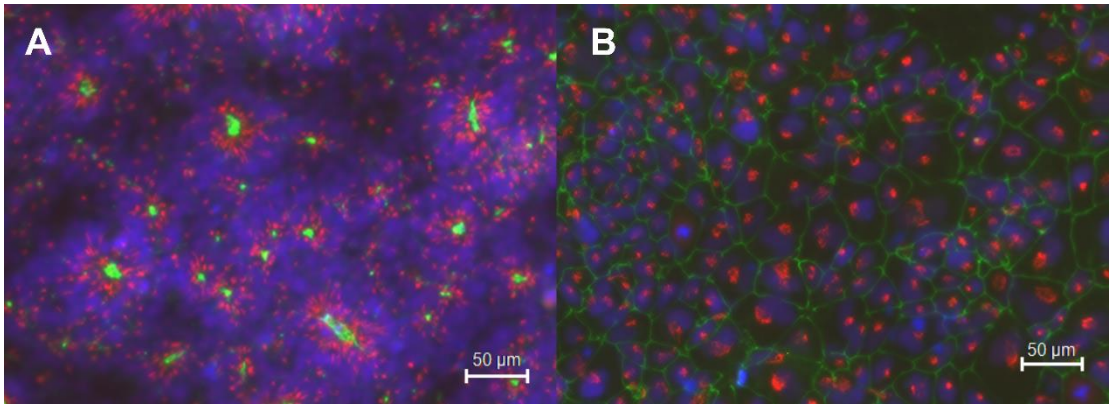


Figure 30: Visualisation of neural rosettes by immunostaining of ZO1 and GM130 reveals disrupted self-organisation by MEHP exposure. *Images were taken on DoD19. Cells were stained for ZO1 (green), GM130 (red) and DAPI (blue). Panel (A) shows the solvent control, panel (B) shows cells exposed to MEHP at a concentration of 1× EC_{10}. Scale bar represents 50 µm.*

To determine effective concentrations that inhibit neural rosette formation, high-throughput image analysis, as established in (75), was used to identify exposed spots lacking detectable neural rosettes. Exposure to a subset of compounds induced a concentration-dependent decrease of neural rosette formation probability per spot, as revealed by binomial concentration-response modelling based on the fraction of non-rosette-forming spots. For instance, concentration-response modelling of MEHP revealed a pronounced decline in rosette formation probability, attaining a value of 0.9 at around 40 µM (**Figure 31A**). In contrast, exposure to Methoxyacetic acid resulted in a constant model, indicating no concentration-dependent inhibition of rosette formation within the non-cytotoxic range (**Figure 31B**). All concentration-response curves are provided in Supplementary figure 13 and Supplementary figure 14.

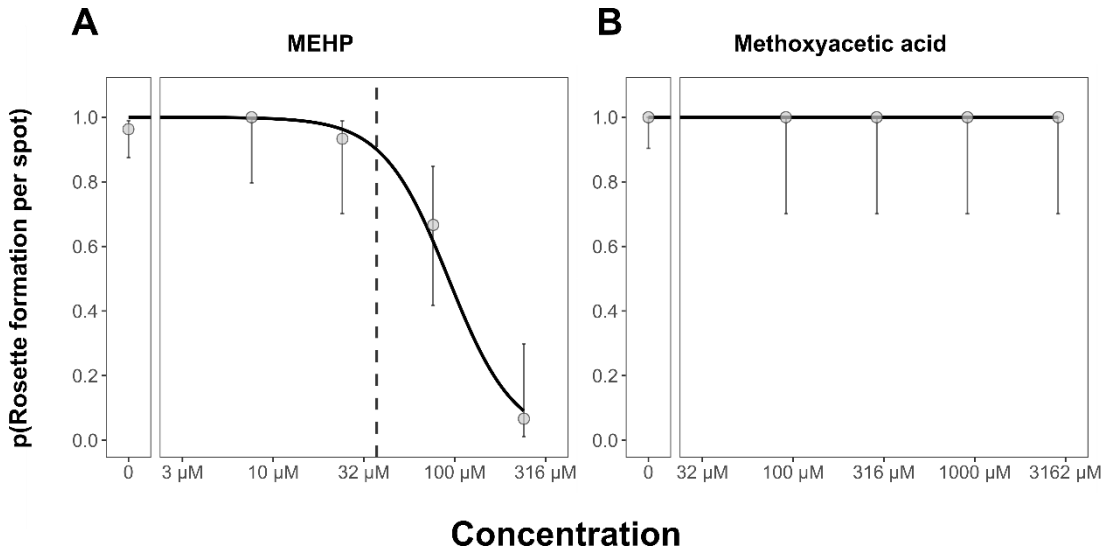


Figure 31: Disruption of neural rosette formation by exposure to MEHP and Methoxyacetic acid. *The probability of rosette formation within a confined spot as derived by binomial concentration-response modelling is indicated on the y-axis. Error bars represent 95% confidence intervals. The dashed line indicates the EC10.*

To systematically evaluate the predictive potential of rosette formation relative to cytotoxicity alone, EC_{10} values derived from rosette formation curves were compared to cytotoxicity EC_{10} values using ESS analysis (**Figure 32**). This revealed that 5 out of 9 developmental toxicants and 3 of 9 non-toxicants disrupted rosette formation at non-cytotoxic concentrations (positive ESS), whereas the remaining compounds did not induce observable effects on rosette formation.

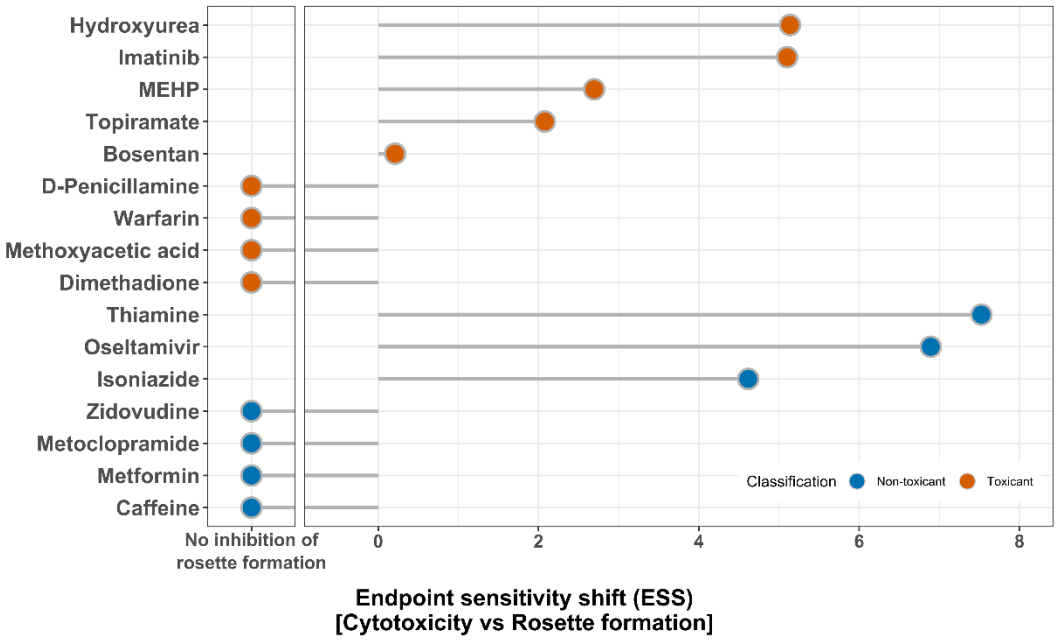


Figure 32: Endpoint sensitivity shift index (ESS) of late-stage cytotoxicity vs disruption of neural rosette formation. *The ESS was defined as the log2 ratio of median EC_{10} cytotoxicity values during late-stage neural induction and the EC_{10} value of disrupted neural rosette formation. Positive ESS values indicate disruption of neural rosette formation at non-cytotoxic concentrations. Orange data points indicate developmental toxicants, blue data points indicate non-toxic reference compounds.*

3.9 Integration of late-stage exposure enables identifying developmental toxicants falsely classified by the early-stage model

Given the pronounced compound effects on viability and self-organisation at the late stage, the next goal was to determine whether incorporating late-stage exposure data improves DevTox prediction. Employing the same LOOCV random forest workflow as described before, a late-stage classifier based on seven new IS (Table 12) achieved a higher AUC ($AUC_{late} = 0.840$) then the early-stage classifier ($AUC_{early} = 0.778$). Furthermore, binary classification based on a threshold to maximise classification accuracy of toxicants and non-toxicants yielded higher accuracy, sensitivity and specificity compared to the early-stage classifier (Table 13). Using the three-category framework defined above, 10 of 18 compounds were classified as toxic or non-toxic at their respective *in vivo* c_{max} (i.e., outside the uncertainty range), and all 10 classifications were correct (**Figure 33**).

Table 12: Interpretable summaries obtained from cytotoxicity and disrupted rosette formation after late-stage neural induction

Interpretable Summary	Explanation	Group	Type
lateStage_medianEC10	Concentration where a fitted concentration-viability curve attains a value of 0.9, 0.8 or 0.5.	Cytotoxicity	Benchmark concentration
lateStage_medianEC20			Benchmark concentration
lateStage_medianEC50			Benchmark concentration
RosetteFormation_EC10	Concentration where a fitted $LL.2_{bin}$ concentration-probability curve attains a value of 0.9, 0.8 or 0.5.	Rosette Formation	Benchmark concentration
RosetteFormation_EC20			Benchmark concentration
RosetteFormation_EC50			Benchmark concentration
Penalty.Rosette	Penalty if RosetteFormation_EC10, EC20 and EC50 values cannot be determined.		Binary penalty variable

Table 13: Prediction performance of early- and late-stage neural induction-based models.

Classifier	AUC	Brier-Score	Accuracy	Sensitivity	Specificity
Early-stage	0.778	0.200	72.2%	55.6%	88.9%
Late-stage	0.840	0.180	83.3%	66.7%	100%
Unified	0.840	0.201	77.8%	55.6%	100%

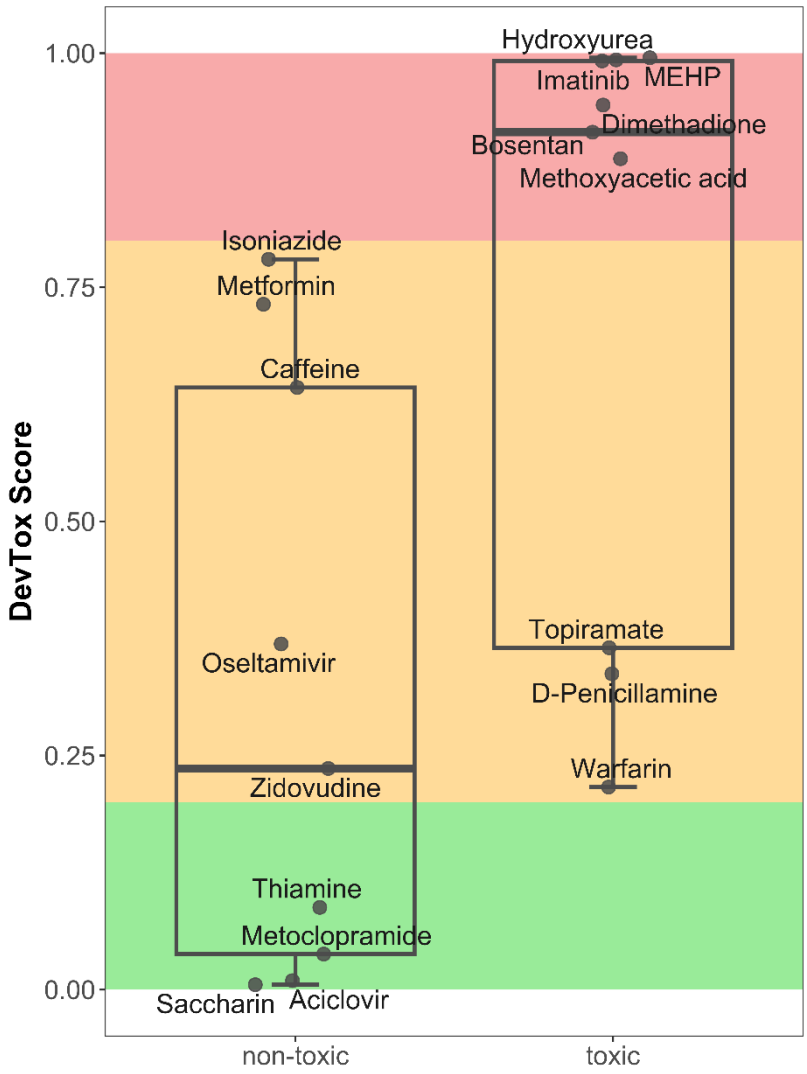


Figure 33: Prediction of developmental toxicity after late-stage neural induction. *DevTox scores for each compound were derived from Leave-One-Out cross validated random forest prediction. The colour scheme represents a three-category framework for compound classification, where data points in the red area represent compounds classified as toxic, data points in the yellow area are classified as uncertain and data points in the green area are classified as non-toxic.*

To illustrate how late-stage exposure alters concentration-dependent DevTox classification, Bosentan was examined in more detail, as it represents a compound that was not identified as toxic during early neural induction but was correctly classified as a developmental toxicant in the late-stage model. Although Bosentan is a known developmental toxicant at an *in vivo* c_{max} of 2.9 µM, the early-stage classifier identified it as toxic only at concentrations exceeding ~10 µM while yielding uncertain results below this concentration. In contrast, the late-stage model classifies Bosentan as toxic already at ~1 µM (**Figure 34**), thereby correctly capturing its developmental toxicity at *in vivo*-relevant concentrations. Importantly, this demonstrates that a subset of developmental toxicants cannot be identified during early neural induction but can be correctly detected when late-stage

exposure is incorporated, highlighting the added predictive value of stage-aware modelling. Concentration-dependent transition points between non-toxic, uncertain, and toxic classifications for all compounds are summarised in Supplementary table 3, and the corresponding concentration-prediction plots are provided in Supplementary figure 15 and Supplementary figure 16.

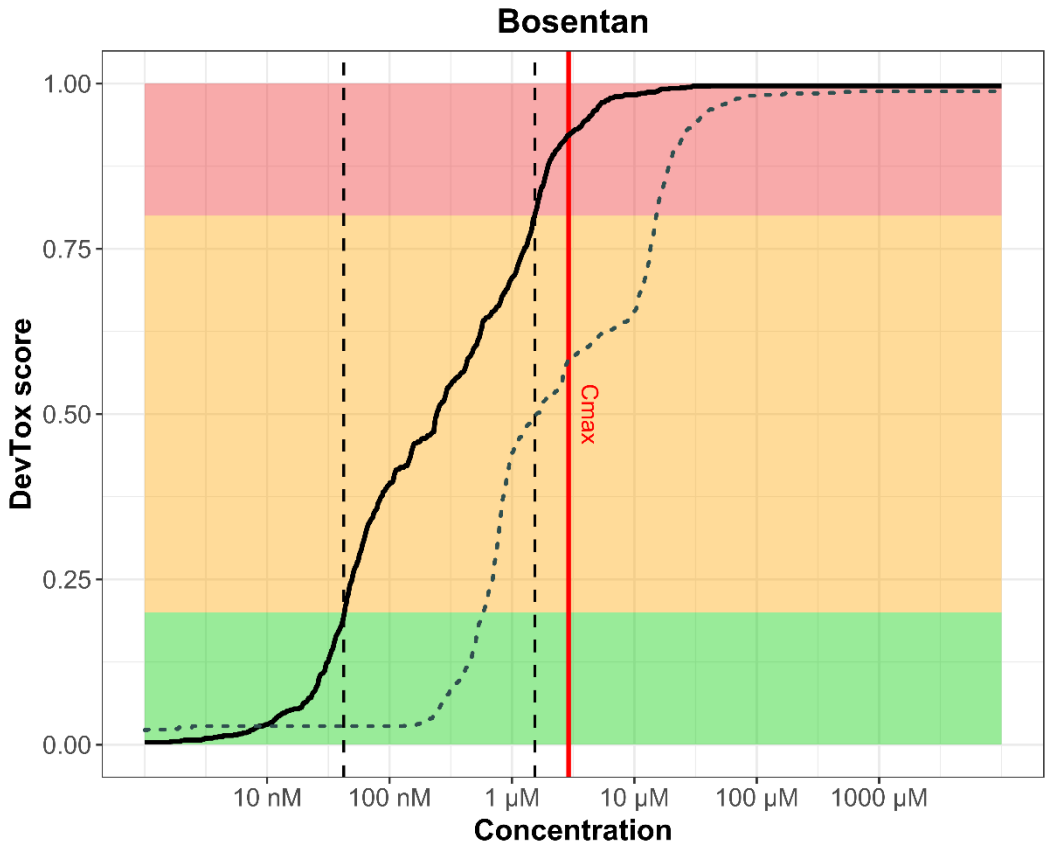


Figure 34: Concentration dependent developmental toxicity prediction of Bosentan after late-stage neural induction. *Red, yellow and green areas indicate whether a prediction leads to toxic, uncertain or non-toxic classification. Straight dashed lines indicate maximum safe concentration and the minimal toxic concentration. Red line indicates c_{max}. The dotted line indicates concentration-dependent prediction following early-stage neural induction.*

Notably, comparing the late-stage classifier with prediction based solely on cytotoxicity revealed that inclusion of rosette re-organisation did not increase predictive performance. Specifically, the AUC of the late-stage classifier trained on both cytotoxicity- and rosette-based IS was comparable to that of a classifier trained exclusively on viability-based late-stage IS (0.852). This indicated that, while assessment of rosette formation provided mechanistically interpretable information on how compound exposure perturbs cellular self-organisation, it did not improve overall DevTox prediction accuracy at the late stage.

Given the observed improvement of DevTox prediction achieved by incorporating late-stage exposure, the next question was whether a unified classifier integrating IS form both early- and late-stage exposure could further enhance prediction accuracy.

Intriguingly, the unified classifier did not improve classification of developmental toxicants and non-toxicants, as reflected by a similar AUC and a lower accuracy compared to the classifier based on the late stage exclusively. Furthermore, the unified classifier exhibited a higher Brier score than the late-stage classifier, indicating a less accurate alignment between predicted DevTox scores and true classifications. Notably, Bosentan was again not correctly identified in the unified model, remaining classified as false negative or uncertain at *in vivo* relevant concentrations. These findings indicate that integration of early- and late-stage IS can dilute stage-specific predictive signals and highlight the importance of stage-aware modelling strategies for DevTox prediction.

4 Discussion

Developmental toxicity testing is an essential pillar of chemical risk assessment. To date, the evaluation of developmental toxicity remains dependent on animal experimentation. However, developmental toxicity assessment *in vivo* is limited by low accuracy, questionable human relevance, and ethical considerations, raising an urgent need for alternative methods in this field. Here, pluripotent stem cell-based approaches that mimic early embryonic development are considered a promising perspective. In particular, alternative methods that employ directed neuroectodermal differentiation as a model for early embryonic neurodevelopment have been an important focus, such as the UKN1 assay, which combines neural induction of induced pluripotent stem cells (iPSC) with whole-transcriptome microarray analysis. Although this assay previously enabled highly accurate discrimination of developmental toxicants and non-toxicants at *in vivo*-relevant concentrations, key limitations constrain feasibility: The reliance on whole-transcriptome-based prediction models without functional anchoring may restrict mechanistic interpretability, while genome-wide expression profiling is associated with substantial costs and workload. Furthermore, focussing exclusively on early neural induction might limit detection of stage-dependent toxicants whose effects manifest only during later developmental windows. In this thesis, it was found that targeted RNA expression measurement (TREx) of a data-driven transcriptional biomarker panel enabled efficient identification of distinct compound-specific perturbation patterns during early neuroectodermal development, thereby supporting mechanistic interpretation of early neurodevelopmental disruption. Furthermore, integration of a later-stage assay capturing self-organisation of neuroepithelial progenitor-like cells into neural rosettes enabled detection of a developmental toxicant that could not be reliably identified during early differentiation alone. Importantly, although transcriptional profiling and assessment of disrupted self-organisation enabled mechanistic interpretation of compound-induced developmental perturbations, predictive performance was largely driven by cytotoxicity in a stage-dependent manner. Taken together, these observations emphasize the relevance of capturing discrete stages of neuroectodermal development to enhance both interpretability and predictive capacity.

Basic characterisation of directed neuroectodermal differentiation revealed that dual-SMAD inhibition-based neural induction recapitulates key features of early neurodevelopment. Initial upregulation of neuroectoderm-associated transcription factors alongside downregulation of established pluripotency marker genes demonstrated transition from an undifferentiated stem cell-state towards a neuroectodermal progenitor-identity, consistent with transcriptional programs governing neuroectodermal development *in vivo*. Interestingly, differentiation was

accompanied by a moderate upregulation of MSX1. This transcription factor is associated with neural crest identity, a cell population that arises at the lateral border of the neuroectoderm *in vivo* (144). Although dual-SMAD inhibition is designed to promote neuroectodermal differentiation, induction of neural crest subpopulations has been reported, especially under suboptimal initial cell densities (72). In the present thesis, initial cell densities were tightly controlled, and the modest increase of MSX1 expression (log2FC ~ 1) suggests a minor activation of neural crest-related transcriptional programs rather than a major lineage deviation. This observation is consistent with an early differentiation stage permitting neural crest specification during initial neuroectodermal induction, a feature that may be particularly relevant for capturing toxicant-induced perturbation of lineage specification. Furthermore, the observed increase of CDH2 expression accompanied by downregulation of CDH1 reflects an additional hallmark of early neuroectodermal development. The cadherin switch from CDH1 to CDH2 is regarded as essential for structural reorganisation during neuroectodermal lineage specification (145). Importantly, regulation of CDH1 and CDH2 was only observed after 11 days of neural induction, but not at day 6, consistent with an early differentiation stage in which structural maturation has not yet been fully established. Notably, the CDH1/CDH2-switch was more pronounced in suspension culture compared to standard monolayer differentiation. Employing suspension culture for prolonged neural induction was originally motivated by substantial overgrowth during extended monolayer cultivation. However, the more pronounced cadherin-switch suggests that this format may also enhance neural induction efficiency, consistent with reports of increased $PAX6^+$ neuroectodermal progenitor cells yielded by suspension-based differentiation protocols compared to monolayer differentiation (146). Collectively, the findings demonstrate that dual-SMAD inhibition-based neural induction recapitulates key aspects of early neurodevelopment, thereby establishing a defined baseline for subsequent exposure experiments.

Systematic assessment of cytotoxicity at early and late stages of neuroectodermal differentiation revealed pronounced stage-dependent susceptibility patterns. Notably, certain compounds exhibited higher cytotoxicity in the early stage (e.g., Caffeine), whereas others showed increased sensitivity in the late stage (e.g., Hydroxyurea), while a third subset showed no difference (e.g., Warfarin). This stage-dependent sensitivity may, at least in part, be driven by differences in proliferative activity across the differentiation trajectory (147), given that proliferative status has been shown to influence cellular susceptibility to cytotoxic insults (148,149). Consistent with this, Hydroxyurea – a DNA synthesis inhibitor preferentially inducing cytotoxicity in proliferating cells (150) – exhibited pronounced stage-dependent sensitivity. Furthermore, studies linking the developmental toxicity of Hydroxyurea *in vivo* to its cytotoxic activity (151) are consistent with the present

observation that Hydroxyurea induces cytotoxicity *in vitro* at concentrations associated with developmental effects *in vivo*. In contrast, the developmental toxicant Warfarin required substantially higher concentrations to induce cytotoxicity than those reported to cause developmental toxicity *in vivo*. Warfarin-induced developmental toxicity has been attributed to fetal haemorrhage (152), as well as to disruption of calcification by inhibition of Matrix Gla protein (153). Given that these mechanisms do not primarily involve direct cytotoxicity, the absence of cell death at concentrations regarded as developmentally toxic *in vivo* is mechanistically plausible. Together, these observations indicate that stage-dependent cytotoxicity reflects biologically meaningful differences in cellular vulnerability and that cytotoxicity may serve as a predictive marker primarily for compounds whose developmental toxicity is mechanistically linked to direct cell death.

To capture perturbed differentiation beyond general cytotoxicity, TREx-based gene expression profiling enabled identification of transcriptional biomarker signatures indicative of disrupted neuroectodermal differentiation. Targeted RNA sequencing of a biomarker panel tailored to detect chemical-induced perturbations of neuroectodermal development represents a resource-efficient and scalable approach. However, in comparison to whole-transcriptome assessment, limiting analysis to a set of 190 biomarker genes may result in omission of relevant gene expression changes. Nevertheless, several findings from the present work support the capacity of the transcriptional biomarker panel to capture exposure-induced disruption of differentiation. Gene ontology (GO) enrichment analysis demonstrated that the biomarker panel is enriched for genes associated with embryonic development rather than general stress response markers. Moreover, assessing biomarker responses in a non-differentiating control model, more specifically, Caco2 cells cultured over a period of 3 weeks to ensure complete differentiation before initiating exposure, revealed substantially fewer differentially expressed genes, supporting the specificity of the panel for neuroectodermal differentiation. Crucially, integrative pattern analysis of principal component loadings in the context of supervised clustering unravelled distinct transcriptional perturbation signatures that reflect distinct modes of differentiation disruption rather than nonspecific shifts from the untreated control. These modes include impaired differentiation progression, as observed following exposure to the developmental toxicant Methoxyacetic acid, which inhibited the transition from a pluripotent state towards neuroectodermal identity. Methoxyacetic acid-induced developmental toxicity has been attributed to histone deacetylase (HDAC) inhibition (154). Moreover, prenatal exposure to this compound has been associated with neural tube defects (155). Notably, similar patterns of disturbed neuroectodermal development have been reported for other HDAC inhibitors (156), collectively supporting a coherent mechanistic link between HDAC inhibition, impaired acquisition of neuroectodermal

identity *in vitro* and disrupted neural tube development *in vivo*. Beyond impaired differentiation progression, activation of divergent transcriptional programs constituted an additional mode of differentiation disruption, exemplified by Dimethadione. The transcriptional response to Dimethadione exposure was predominantly marked by activation of neural crest-associated genes. Consistent with the observation that during early differentiation stages neural crest-related programs remain inducible, this finding illustrates the potential of neuroectodermal differentiation-based models to reveal perturbations of lineage specification. Notably, perturbation of neural crest-associated development is linked to diverse adverse outcomes *in vivo*, as neural crest cells contribute to the formation of multiple organ systems, including craniofacial structures and the heart (157). In line with this, prenatal exposure to Dimethadione has been reported to induce congenital malformations affecting both cardiac and craniofacial development (97). While this correspondence is conceptually consistent, it should be noted that many key regulators of lineage specification are multifunctional. For instance, WNT signalling contributes to diverse developmental processes, with target gene activation varying across tissues and developmental stages (158). Thus, although the consistency between neural crest-associated gene activation and reported malformations is noteworthy, it does not imply a singular causal pathway, as many key developmental regulators elicit context-dependent effects. Rather, in combination with consistent impairment of differentiation progression induced by HDAC inhibitors, identification of divergent transcriptional programs demonstrates the capacity to unravel distinct modes of disrupted differentiation. Together, these observations illustrate the potential of stem cell-based models to identify mechanistically informed perturbation profiles that may be essential for chemical classification.

Whereas early-stage neural induction revealed distinct transcriptional signatures of disrupted differentiation, assessing neural rosette formation enabled evaluation of impaired cellular self-organisation. Neural rosettes recapitulate key features of the developing neural tube *in vivo*, including coordinated morphogenesis around a central lumen, establishment of apical-basal polarity, and responsiveness to developmental patterning cues (159). Accordingly, assessing impaired neural rosette formation is likely to capture perturbation of key morphogenetic mechanisms that are critical to tissue-level self-organisation. Importantly, a subset of validation compounds disrupted neural rosette formation at non-cytotoxic concentrations, demonstrating that the impairment of self-organisation was not secondary to viability loss. Notably, some of the compounds which disrupted neural rosette formation had no influence on gene expression during the early stage of differentiation, as exemplified by MEHP, a ubiquitous plastic-contaminant associated with neural tube defects through redox-mediated mechanisms (160). In contrast, other compounds that disrupted early differentiation (such as

Methoxyacetic acid) did not affect cellular self-organisation at non-cytotoxic concentrations. This finding illustrates how distinct compounds can perturb development at different biological levels and underscores the importance of complementary endpoints to achieve mechanistically interpretable *in vitro* DevTox modelling (**Figure 35**).

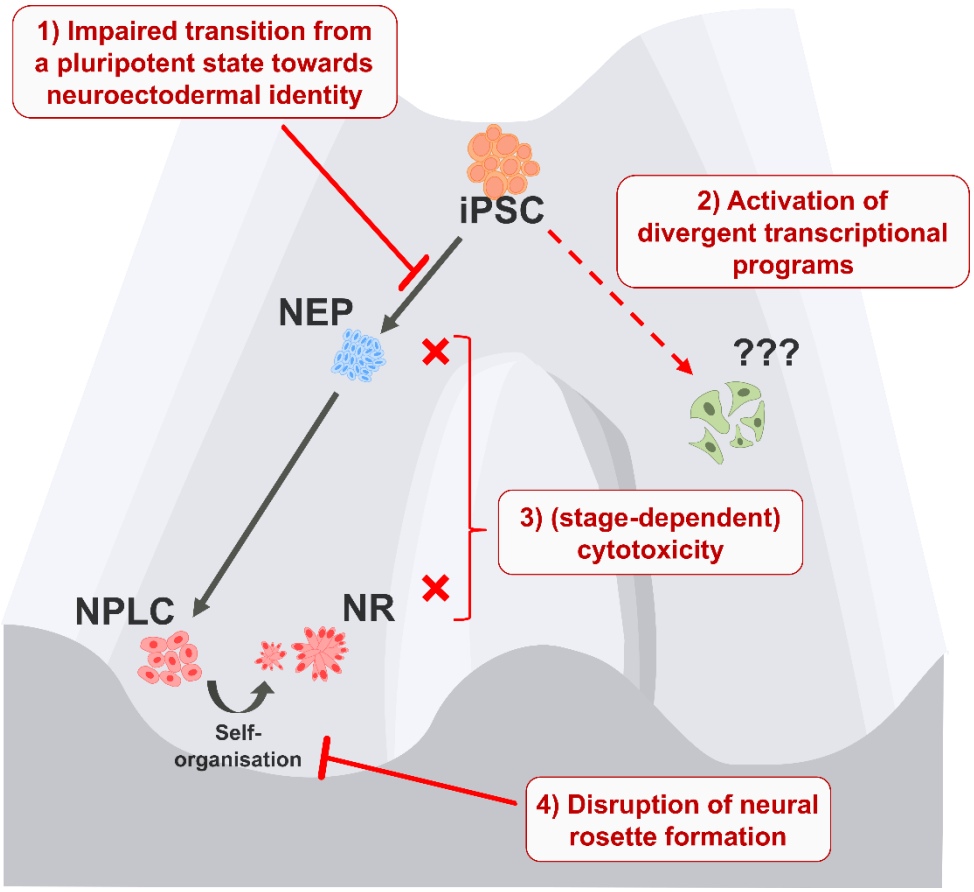


Figure 35: Distinct perturbation modes on different levels of neuroectodermal development. *iPSC: Induced pluripotent stem cells, NEP: Neuroectodermal progenitor cells, NPLC: Neuroepithelial progenitor-like cells, NR: Neural rosettes. Grey arrows indicate normal development. Red arrows indicate perturbation of development.*

Evaluation of predictive performance revealed that early-stage neural induction, implemented in form of the UKN1 assay, achieved only moderate classification accuracy. Notably, the observed accuracy of 72.2% was substantially lower than the 87-90% reported in earlier UKN1 studies (42). Although methodological differences between earlier studies and the present thesis limit direct comparability, the variation in reported accuracies highlights that classification performance is heavily dependent on the composition of compounds employed for validation. Importantly, addition of the late-stage model enhanced prediction accuracy. For instance, Bosentan was correctly classified as developmentally toxic only by the late-stage classifier. Bosentan, a dual endothelin A and B receptor antagonist, is clinically used for treatment of pulmonary hypertension (161). As endothelin-associated signalling has a central role for different morphogenetic processes, including the development

of neural crest-derived tissues (162), the craniofacial and cardiovascular malformations observed following prenatal Bosentan exposure are attributed to its receptor antagonism (163). Notably, a unified classifier integrating early- and late-stage data failed to correctly classify Bosentan at *in vivo*-relevant concentrations, indicating that stage-aware modelling may retain discriminatory information that are obscured by data aggregation in this context.

Comparison with other stem cell-based alternative methods reveals a similar range of predictive accuracies to that observed for early- and late-stage neuroectodermal differentiation (Table 14). Collectively, these observations suggest that predictive performance in developmental toxicity testing may not be maximized by a single high-performing assay, but rather through the complementary integration of multiple models. In this context, the enhanced predictive performance obtained by integration of multiple developmental stages should motivate assembling a stage-resolved battery of complementary test systems.

Table 14: Prediction accuracy of different stem cell-based alternative methods for developmental toxicity testing

Alternative Assay	Accuracy	Reference
TeraTox	69%	(40)
devTOX qp	77%	(39)
mEST	78%	(33)
DevTox GLR	72%	(164)

To strengthen confidence in alternative models for toxicity prediction, assessing reliable ranges and uncertainty is essential. In the present thesis, a three-category framework was implemented to define concentration ranges yielding highly reliable predictions alongside an intermediate range in which a compound is classified as uncertain. This framework requires monotonic concentration-response relationships, as non-monotonic patterns would prevent definition of meaningful concentration-ranges. Although random forest-based prediction models are widely regarded as robust tools for *in vitro* toxicity assessment, their flexibility allows non-monotonic concentration-response patterns to emerge when such patterns provide a locally optimal explanation of noisy data. As such patterns reflect global model optimisation to training set variation rather than compound-specific observations, they would be applied to all predictions based on this model, undermining biological plausibility, complicating interpretability and signalling overfitting to the training set. Accordingly, prediction models in the present work were trained with monotonic constraints to enforce biologically plausible monotonic concentration-response relationships and ensure applicability of the three-category framework. Importantly, two compounds – Warfarin and Caffeine – remained misclassified by the early-stage model even in the three-category framework, being assigned to the opposite class rather than the uncertain range at *in vivo* relevant concentrations. A

plausible explanation is statistical bias introduced by cross validation: While improving generalisability, leave-one-out cross validation (LOOCV) may bias prediction for compounds at the margins of the training set distribution. For instance, in the early stage classifier, Caffeine exhibited the strongest cytotoxicity relative to its *in vivo*-relevant concentration among all other non-toxic reference compounds, whereas Warfarin displayed the lowest relative cytotoxicity among the toxicants. When employing LOOCV, the prediction model for a certain compound is constructed based exclusively on the data of the remaining compounds. In the case of e.g. Caffeine, this leads to a training set in which all non-toxic reference compounds exhibit lower relative cytotoxicity, altering the relative feature distribution and overestimating the developmental toxicity of Caffeine. Together with the observation that predictive accuracy depends strongly on the composition of test compounds, this underscores the need for validation across larger and more diverse sets of chemicals. A central challenge in this context will be the availability of relevant *in vivo*-reference concentrations for an expanded set of compounds. Whereas in the present work, maternal maximum plasma concentrations were used as the best available surrogate of embryonic exposure, direct measurement of compound concentrations in embryonic tissue would represent the best analogue of *in vitro* exposure levels. However, such data is rarely available. Furthermore, while physiologically based pharmacokinetic (PBPK) modelling permits simulation of compound concentrations across various compartments of an organism, no published framework currently enables reliable modelling of exposure in early embryonic tissues (165). Thus, obtaining more relevant *in vivo* concentrations will proof to be an essential step towards more accurate DevTox prediction.

Comparison with models trained solely on viability-based featured revealed that prediction performance was mainly driven by cytotoxicity, whereas inclusion of functional endpoints such as gene expression and self-organisation did not improve classification. Notably, similar patterns have been reported for several other assays (34), and the close correlation between functional impairment and cytotoxicity is well recognized in NAM development (166). Although assessment of cytotoxicity is regarded an essential baseline characterisation for targeted toxicity evaluation (167), it is often considered to reflect non-specific cellular stress responses that provide limited mechanistic insight (168). In this context, assessment of functional endpoints holds a distinct epistemic value by providing mechanistic interpretation of the relevance of *in vitro* effects for *in vivo* outcomes. Building confidence in NAM-based predictions without mechanistic interpretability is considered challenging, as absence of understanding of the biological relevance may constrain applicability to boundaries tightly defined by the data used to generate and validate the prediction model (169). Here, targeted gene expression profiling and assessment of neural rosette formation allow assessment of biological relevancy that can enhance

confidence in toxicity assessment based on directed neuroectodermal development in stem cell-based assays.

In conclusion, modelling early embryonic neurodevelopment by directed neuroectodermal differentiation of pluripotent stem cells was shown to be a promising tool for animal-free developmental toxicity testing. While targeted gene expression profiling and assessment of disrupted self-organisation provided mechanistically informed insights into compound-induced developmental perturbations, extending exposure windows to capture later developmental stages was demonstrated to improve predictive performance.

5 References

1. OECD. Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment. Paris; 2008.
2. Drinking Water and Health, Volume 6. Washington: National Academies Press; 1986. 11 p.
3. OECD. Test No. 414: Prenatal Developmental Toxicity Study [Internet]. OECD; 2018 [cited 2026 Mar 17]. Available from: https://www.oecd.org/en/publications/test-no-414-prenatal-development-toxicity-study_9789264070820-en.html doi:10.1787/9789264070820-en
4. Giordano G, Costa LG. Developmental Neurotoxicity: Some Old and New Issues. ISRN Toxicol. 2012 Jun 24;2012:1–12. doi:10.5402/2012/814795
5. Holmes LB, Harvey EA, Coull BA, Huntington KB, Khoshbin S, Hayes AM, et al. The Teratogenicity of Anticonvulsant Drugs. N Engl J Med. 2001;(344):1132–8.
6. Greene NDE, Copp AJ. Neural Tube Defects. Annu Rev Neurosci. 2015 Jul 1;37:221–42. doi:10.1146/annurev-neuro-062012-170354
7. Copp AJ, Adzick NS, Chitty LS, Fletcher JM, Holmbeck GN, Shaw GM. Spina bifida. Nat Rev Dis Primer. 2016 Jun 8;(1):15007. doi:10.1038/nrdp.2015.7
8. Finnell RH, Caiaffa CD, Kim SE, Lei Y, Steele J, Cao X, et al. Gene Environment Interactions in the Etiology of Neural Tube Defects. Front Genet. 2021 May 10;12:659612. doi:10.3389/fgene.2021.659612
9. Friedman JM. The principles of teratology: Are they still true? Birt Defects Res A Clin Mol Teratol. 2010 May 12;(88):766–8. doi:10.1002/bdra.20697
10. Bailey J, Knight A, Balcombe J. The future of teratology research is in vitro. Biog Amines. 2005 May 1;19:97–145. doi:10.1163/1569391053722755
11. ICH S5 (R3) Guideline on detection of reproductive and developmental toxicity for human pharmaceuticals. 2020.
12. OECD. Test No. 443: Extended One-Generation Reproductive Toxicity Study [Internet]. OECD; 2025 [cited 2026 Mar 17]. Available from: https://www.oecd.org/en/publications/test-no-443-extended-one-generation-reproductive-toxicity-study_9789264185371-en.html doi:10.1787/9789264185371-en
13. Commission Staff Working Document, European Commission. Summary Report on the statistics on the use of animals for scientific purposes in the Member States of the European Union and Norway in 2022. Brussels; 2024.
14. Blau HM, Brazelton TR, Weimann JM. The Evolving Concept of a Stem Cell: Entity or Function? Cell. 2001 Jun 29;105:829–41.
15. Thomson J, Itskovitz-Eldor J, Shapiro S, Waknitz M, Swiergiel J, Marshall V, et al. Embryonic stem cell lines derived from human blastocysts. Science. 1998 Nov 6;282(5391):1145. doi:10.1126/science.282.5391.1145
16. Smith A. A glossary for stem-cell biology. Nature. 2006 Jun 29;441(7097):1060–1060. doi:10.1038/nature04954
17. Keller GM. In vitro differentiation of embryonic stem cells. Curr Opin Cell Biol. 1995 Jan;7(6):862–9. doi:10.1016/0955-0674(95)80071-9
18. Murry CE, Keller G. Differentiation of Embryonic Stem Cells to Clinically Relevant Populations: Lessons from Embryonic Development. Cell. 2008 Feb;132(4):661–80. doi:10.1016/j.cell.2008.02.008
19. Itoh F, Watabe T, Miyazono K. Roles of TGF-β family signals in the fate determination of pluripotent stem cells. Semin Cell Dev Biol. 2014 Aug;32:98–106. doi:10.1016/j.semcdb.2014.05.017

20. Niwa H, Miyazaki J ichi, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet. 2000 Apr;24(4):372–6. doi:10.1038/74199

21. Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, et al. Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. Cell. 2003 May;113(5):643–55. doi:10.1016/S0092-8674(03)00392-1

22. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, et al. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. Cell. 2005 Sep;122(6):947–56. doi:10.1016/j.cell.2005.08.020

23. Bundesministerium der Justiz und für Verbraucherschutz, Bundesamt für Justiz. Gesetz zum Schutz von Embryonen (Embryonenschutzgesetz - ESchG). 1991 Jan 1.

24. Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell. 2006 Aug;126(4):663–76. doi:10.1016/j.cell.2006.07.024

25. OECD. Test No. 432: In Vitro 3T3 NRU Phototoxicity Test [Internet]. OECD; 2019 [cited 2026 Mar 17]. Available from: https://www.oecd.org/en/publications/test-no-432-in-vitro-3t3-nru-phototoxicity-test_9789264071162-en.html doi:10.1787/9789264071162-en

26. Montesinos López OA, Montesinos López A, Crossa J. Overfitting, Model Tuning, and Evaluation of Prediction Performance. In: Multivariate Statistical Machine Learning Methods for Genomic Prediction [Internet]. Cham: Springer International Publishing; 2022 [cited 2026 Mar 17]. p. 109–39. Available from: https://link.springer.com/10.1007/978-3-030-89010-0_4 doi:10.1007/978-3-030-89010-0_4

27. Albrecht W, Kappenberg F, Brecklinghaus T, Stoeber R, Marchan R, Zhang M, et al. Prediction of human drug-induced liver injury (DILI) in relation to oral doses and blood concentrations. Arch Toxicol. 2019 Jun;93(6):1609–37. doi:10.1007/s00204-019-02492-9

28. Worth AP, Cronin MTD. Prediction Models for Eye Irritation Potential Based on Endpoints of the HETCAM and Neutral Red Uptake Tests. In Vitr Mol Toxicol. 2001 Sep;14(3):143–56. doi:10.1089/109793301753407911

29. Stolte M, Albrecht W, Brecklinghaus T, Gründler L, Chen P, Hengstler JG, et al. Classification of hepatotoxicity of compounds based on cytotoxicity assays is improved by additional interpretable summaries of high-dimensional gene expression data. Comput Toxicol. 2023 Nov;28:100288. doi:10.1016/j.comtox.2023.100288

30. Bates S, Hastie T, Tibshirani R. Cross-Validation: What Does It Estimate and How Well Does It Do It? J Am Stat Assoc. 2024 Apr 2;119(546):1434–45. doi:10.1080/01621459.2023.2197686

31. Bal-Price A, Hogberg HT, Crofton KM, Daneshian M, FitzGerald RE, Fritsche E, et al. Recommendation on test readiness criteria for new approach methods in toxicology: Exemplified for developmental neurotoxicity. ALTEX. 2018;35:306–52. doi:10.14573/altex.1712081

32. Spielmann H, Pohl I, Doring B, Liebsch M, Moldenhauer F. The embryonic stem cell test, an in vitro embryotoxicity test using two permanent mouse cell lines: 3T3 fibroblasts and embronic stem cells. J Mol Cell Toxicol. 1997;10(1).

33. Genschow E, Spielmann H, Scholz G, Pohl I, Seiler A, Clemann N, et al. Validation of the Embryonic Stem Cell Test in the International ECVAM Validation Study on Three *In Vitro* Embryotoxicity Tests. Altern Lab Anim. 2004 Sep;32(3):209–44. doi:10.1177/026119290403200305

34. Marx-Stoelting P, Adriaens E, Ahr HJ, Bremer S, Garthoff B, Gelbke HP, et al. A Review of the Implementation of the Embryonic Stem Cell Test (EST): The Report and Recommendations of an ECVAM/ReProTect Workshop <sup/>. Altern Lab Anim. 2009 Jul;37(3):313–28. doi:10.1177/026119290903700314

35. Buesen R, Genschow E, Slawik B, Visan A, Spielmann H, Luch A, et al. Embryonic Stem Cell Test Remastered: Comparison between the Validated EST and the New Molecular FACS-EST for Assessing Developmental Toxicity In Vitro. Toxicol Sci. 2009 Apr;108(2):389–400. doi:10.1093/toxsci/kfp012

36. Suzuki N, Ando S, Yamashita N, Horie N, Saito K. Evaluation of Novel High-Throughput Embryonic Stem Cell Tests with New Molecular Markers for Screening Embryotoxic Chemicals In Vitro. Toxicol Sci. 2011 Dec;124(2):460–71. doi:10.1093/toxsci/kfr250

37. Simms L, Rudd K, Palmer J, Czekala L, Yu F, Chapman F, et al. The use of human induced pluripotent stem cells to screen for developmental toxicity potential indicates reduced potential for non-combusted products, when compared to cigarettes. Curr Res Toxicol. 2020 Jun;1:161–73. doi:10.1016/j.crtox.2020.11.001

38. West PR, Weir AM, Smith AM, Donley ELR, Cezar GG. Predicting human developmental toxicity of pharmaceuticals using human embryonic stem cells and metabolomics. Toxicol Appl Pharmacol. 2010 Aug;247(1):18–27. doi:10.1016/j.taap.2010.05.007

39. Palmer JA, Smith AM, Egnash LA, Conard KR, West PR, Burrier RE, et al. Establishment and Assessment of a New Human Embryonic Stem Cell-Based Biomarker Assay for Developmental Toxicity Screening. Birth Defects Res B Dev Reprod Toxicol. 2013 Aug;98(4):343–63. doi:10.1002/bdrb.21078

40. Jaklin M, Zhang JD, Schäfer N, Clemann N, Barrow P, Küng E, et al. Optimization of the TeraTox Assay for Preclinical Teratogenicity Assessment. Toxicol Sci. 2022 Jun 28;188(1):17–33. doi:10.1093/toxsci/kfac046

41. Krug AK, Kolde R, Gaspar JA, Rempel E, Balmer NV, Meganathan K, et al. Human embryonic stem cell-derived test systems for developmental neurotoxicity: a transcriptomics approach. Arch Toxicol. 2013 Jan;87(1):123–43. doi:10.1007/s00204-012-0967-3

42. Seidel F, Cherianidou A, Kappenberg F, Marta M, Dreser N, Blum J, et al. High Accuracy Classification of Developmental Toxicants by In Vitro Tests of Human Neuroepithelial and Cardiomyoblast Differentiation. Cells. 2022 Oct 27;11(21):3404. doi:10.3390/cells11213404

43. Jamalpoor A, Hartvelt S, Dimopoulou M, Zwetsloot T, Brandsma I, Racz PI, et al. A novel human stem cell-based biomarker assay for in vitro assessment of developmental toxicity. Birth Defects Res. 2022 Nov 15;114(19):1210–28. doi:10.1002/bdr2.2001

44. Cherianidou A, Seidel F, Kappenberg F, Dreser N, Blum J, Waldmann T, et al. Classification of Developmental Toxicants in a Human iPSC Transcriptomics-Based Test. Chem Res Toxicol. 2022 May 16;35(5):760–73. doi:10.1021/acs.chemrestox.1c00392

45. Tyser RCV, Srinivas S. Recent advances in understanding cell types during human gastrulation. Semin Cell Dev Biol. 2022 Nov;131:35–43. doi:10.1016/j.semcdb.2022.05.004

46. Wu MY, Hill CS. TGF-β Superfamily Signaling in Embryonic Development and Homeostasis. Dev Cell. 2009 Mar;16(3):329–43. doi:10.1016/j.devcel.2009.02.012

47. Ben-Haim N, Lu C, Guzman-Ayala M, Pescatore L, Mesnard D, Bischofberger M, et al. The Nodal Precursor Acting via Activin Receptors Induces Mesoderm by Maintaining a Source of Its Convertases and BMP4. Dev Cell. 2006 Sep;11(3):313–23. doi:10.1016/j.devcel.2006.07.005

48. Meyers EA, Kessler JA. TGF-β Family Signaling in Neural and Neuronal Differentiation, Development, and Function. Cold Spring Harb Perspect Biol. 2017 Aug;9(8):a022244. doi:10.1101/cshperspect.a022244

49. Muñoz-Sanjuán I, Brivanlou AH. Neural induction, the default model and embryonic stem cells. Nat Rev Neurosci. 2002 Apr;3(4):271–80. doi:10.1038/nrn786

50. Ybot-Gonzalez P, Gaston-Massuet C, Girdler G, Klingensmith J, Arkell R, Greene NDE, et al. Neural plate morphogenesis during mouse neurulation is regulated by antagonism of Bmp signalling. Development. 2007 Sep 1;134(17):3203–11. doi:10.1242/dev.008177

51. Trevers KE, Lu HC, Yang Y, Thiery AP, Strobl AC, Anderson C, et al. A gene regulatory network for neural induction. eLife. 2023 Mar 3;12:e73189. doi:10.7554/eLife.73189

52. Takemoto T. Mechanism of cell fate choice between neural and mesodermal development during early embryogenesis. Congenit Anom. 2013 Jun;53(2):61–6. doi:10.1111/cga.12017

53. Panaliappan TK, Wittmann W, Jidigam VK, Mercurio S, Bertolini JA, Sghari S, et al. Sox2 is required for olfactory pit formation and olfactory neurogenesis through BMP restriction and Hes5 upregulation. Development. 2018 Jan 15;145(2):dev153791. doi:10.1242/dev.153791

54. Sagner A, Briscoe J. Morphogen interpretation: concentration, time, competence, and signaling dynamics. WIREs Dev Biol. 2017 Jul;6(4):e271. doi:10.1002/wdev.271

55. Semprich CI, Davidson L, Amorim Torres A, Patel H, Briscoe J, Metzis V, et al. ERK1/2 signalling dynamics promote neural differentiation by regulating chromatin accessibility and the polycomb repressive complex. Bronner ME, editor. PLOS Biol. 2022 Dec 1;20(12):e3000221. doi:10.1371/journal.pbio.3000221

56. Osumi N, Shinohara H, Numayama-Tsuruta K, Maekawa M. Concise Review: Pax6 Transcription Factor Contributes to both Embryonic and Adult Neurogenesis as a Multifunctional Regulator. Stem Cells. 2008 Jul 1;26(7):1663–72. doi:10.1634/stemcells.2007-0884

57. Heusinkveld HJ, Staal YCM, Baker NC, Daston G, Knudsen TB, Piersma A. An ontology for developmental processes and toxicities of neural tube closure. Reprod Toxicol. 2021 Jan;99:160–7. doi:10.1016/j.reprotox.2020.09.002

58. Aruga J. The role of Zic genes in neural development. Mol Cell Neurosci. 2004 Jun;26(2):205–21. doi:10.1016/j.mcn.2004.01.004

59. Cunningham TJ, Brade T, Sandell LL, Lewandoski M, Trainor PA, Colas A, et al. Retinoic Acid Activity in Undifferentiated Neural Progenitors Is Sufficient to Fulfill Its Role in Restricting Fgf8 Expression for Somitogenesis. Mallo M, editor. PLOS ONE. 2015 Sep 14;10(9):e0137894. doi:10.1371/journal.pone.0137894

60. Eom DS, Amarnath S, Fogel JL, Agarwala S. Bone morphogenetic proteins regulate hinge point formation during neural tube closure by dynamic modulation of apicobasal polarity. Birt Defects Res A Clin Mol Teratol. 2012 Oct;94(10):804–16. doi:10.1002/bdra.23052

61. Zhang J, Raghunathan R, Rippy J, Wu C, Finnell RH, Larin KV, et al. Tissue biomechanics during cranial neural tube closure measured by Brillouin microscopy and optical coherence tomography. Birth Defects Res. 2019 Aug 15;111(14):991–8. doi:10.1002/bdr2.1389

62. Kalcheim C. Epithelial–Mesenchymal Transitions during Neural Crest and Somite Development. J Clin Med. 2016;5(1):1. doi:10.3390/jcm5010001

63. Taneyhill LA. To adhere or not to adhere: The role of Cadherins in neural crest cell development. Cell Adhes Migr. 2008 Oct;2(4):223–30. doi:10.4161/cam.2.4.6835

64. Nieto MA. The early steps of neural crest development. Mech Dev. 2001 Jul;105(1–2):27–35. doi:10.1016/S0925-4773(01)00394-X

65. Meulemans D, Bronner-Fraser M. Gene-Regulatory Interactions in Neural Crest Evolution and Development. Dev Cell. 2004;7:291–9.

66. Fuller LC, Cornelius SK, Murphy CW, Wiens DJ. Neural crest cell motility in valproic acid. Reprod Toxicol. 2002 Nov;16(6):825–39. doi:10.1016/S0890-6238(02)00059-X

67. Ying QL, Stavridis M, Griffiths D, Li M, Smith A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol. 2003 Feb;21(2):183–6. doi:10.1038/nbt780

68. Itoh F, Watabe T, Miyazono K. Roles of TGF-β family signals in the fate determination of pluripotent stem cells. Semin Cell Dev Biol. 2014 Aug;32:98–106. doi:10.1016/j.semcdb.2014.05.017

69. Groppe J, Greenwald J, Wiater E, Rodriguez-Leon J, Economides A, Kwiatkowski W, et al. Structural Basis of BMP Signaling Inhibition by Noggin, a Novel Twelve-Membered Cystine Knot Protein. J Bone Jt Surg. 2003;85:52–8.

70. Smith JR, Vallier L, Lupo G, Alexander M, Harris WA, Pedersen RA. Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. Dev Biol. 2008 Jan;313(1):107–17. doi:10.1016/j.ydbio.2007.10.003

71. Inman GJ, Nicolás FJ, Callahan JF, Harling JD, Gaster LM, Reith AD, et al. SB-431542 Is a Potent and Specific Inhibitor of Transforming Growth Factor-β Superfamily Type I Activin Receptor-Like Kinase (ALK) Receptors ALK4, ALK5, and ALK7. Mol Pharmacol. 2002 Jul;62(1):65–74. doi:10.1124/mol.62.1.65

72. Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, Studer L. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol. 2009 Mar;27(3):275–80. doi:10.1038/nbt.1529

73. Elkabetz Y, Panagiotakos G, Al Shamy G, Socci ND, Tabar V, Studer L. Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. Genes Dev. 2008 Jan 15;22(2):152–65. doi:10.1101/gad.1616208

74. Elkabetz Y, Studer L. Human ESC-derived Neural Rosettes and Neural Stem Cell Progression. Cold Spring Harb Symp Quant Biol. 2008 Jan 1;73(0):377–87. doi:10.1101/sqb.2008.73.052

75. Dreser N, Madjar K, Holzer AK, Kapitza M, Scholz C, Kranaster P, et al. Development of a neural rosette formation assay (RoFA) to identify neurodevelopmental toxicants and to characterize their transcriptome disturbances. Arch Toxicol. 2019 Nov 11;94(1):151–71. doi:10.1007/s00204-019-02612-5

76. Morrison M, Klein C, Clemann N, Collier DA, Hardy J, Heiβerer B, et al. StemBANCC: Governing Access to Material and Data in a Large Stem Cell Research Consortium. Stem Cell Rev Rep. 2015 Oct;11(5):681–7. doi:10.1007/s12015-015-9599-3

77. Kim J, Jeon J, Song B, Lee N, Ko S, Cha Y, et al. Spotting-based differentiation of functional dopaminergic progenitors from human pluripotent stem cells. Nat Protoc. 2022 Mar;17(3):890–909. doi:10.1038/s41596-021-00673-4

78. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods. 2001 Dec;25(4):402–8. doi:10.1006/meth.2001.1262

79. Adrian Alexa JR. topGO [Internet]. Bioconductor; 2025 [cited 2026 Mar 18]. Available from: https://bioconductor.org/packages/topGO doi:10.18129/B9.BIOC.TOPGO

80. Chen J, Bao Z, Huang Y, Wang Z, Zhao Y. Selection of Suitable Reference Genes for qPCR Gene Expression Analysis of HepG2 and L02 in Four Different Liver Cell Injured Models. Ergün S, editor. BioMed Res Int. 2020 Jan;2020(1):8926120. doi:10.1155/2020/8926120

81. Riedel G, Rüdrich U, Fekete-Drimusz N, Manns MP, Vondran FWR, Bock M. An Extended ΔCT-Method Facilitating Normalisation with Multiple Reference Genes Suited for Quantitative RT-PCR Analyses of Human Hepatocyte-Like Cells. Bartosch B, editor. PLoS ONE. 2014 Mar 21;9(3):e93031. doi:10.1371/journal.pone.0093031

82. Sharan RN, Vaiphei ST, Nongrum S, Keppen J, Ksoo M. Consensus reference gene(s) for gene expression studies in human cancers: end of the tunnel visible? Cell Oncol. 2015 Dec;38(6):419–31. doi:10.1007/s13402-015-0244-6

83. Augustyniak J, Lenart J, Lipka G, Stepien PP, Buzanska L. Reference Gene Validation via RT–qPCR for Human iPSC-Derived Neural Stem Cells and Neural Progenitors. Mol Neurobiol. 2019 Oct;56(10):6820–32. doi:10.1007/s12035-019-1538-x

84. Panina Y, Germond A, Masui S, Watanabe TM. Validation of Common Housekeeping Genes as Reference for qPCR Gene Expression Analysis During iPS Reprogramming Process. Sci Rep. 2018 Jun 7;8(1):8716. doi:10.1038/s41598-018-26707-8

85. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. NAR Genomics Bioinforma. 2020 Sep 1;2(3):lqaa078. doi:10.1093/nargab/lqaa078

86. Michael Love SA. DESeq2 [Internet]. Bioconductor; 2017 [cited 2026 Mar 18]. Available from: https://bioconductor.org/packages/DESeq2 doi:10.18129/B9.BIOC.DESEQ2

87. Nell P, Kattler K, Feuerborn D, Hellwig B, Rieck A, Salhab A, et al. Identification of an FXR-modulated liver-intestine hybrid state in iPSC-derived hepatocyte-like cells. J Hepatol. 2022 Nov;77(5):1386–98. doi:10.1016/j.jhep.2022.07.009

88. Andrzej Oleś GP. EBImage [Internet]. Bioconductor; 2017 [cited 2026 Mar 18]. Available from: https://bioconductor.org/packages/EBImage doi:10.18129/B9.BIOC.EBIMAGE

89. Breiman L, Cutler A, Liaw A, Wiener M. randomForest: Breiman and Cutlers Random Forests for Classification and Regression [Internet]. 2002 [cited 2026 Mar 19]. p. 4.7-1.2. Available from: https://CRAN.R-project.org/package=randomForest doi:10.32614/CRAN.package.randomForest

90. Ritz C, Baty F, Streibig JC, Gerhard D. Dose-Response Analysis Using R. Xia Y, editor. PLOS ONE. 2015 Dec 30;10(12):e0146021. doi:10.1371/journal.pone.0146021

91. Akaike H. A New Look at the Statistical Model Identification. IEEE Trans Autom Control. 1974;19(6):716–23.

92. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011 Dec;12(1):77. doi:10.1186/1471-2105-12-77

93. Künzel S, Saarinen T, Walter S, Antonyan S, Liu E, Tang A, et al. Rforestry: Random Forests, Linear Trees, and Gradient Boosting for Inference and Interpretability [Internet]. 2021 [cited 2026 Mar 19]. p. 0.11.1.0. Available from: https://CRAN.R-project.org/package=Rforestry doi:10.32614/CRAN.package.Rforestry

94. European Medicines Agency. Assessment Report for Tracleer. London; 2009.

95. De Raaf MA, Beekhuijzen M, Guignabert C, Vonk Noordegraaf A, Bogaard HJ. Endothelin-1 receptor antagonists in fetal development and pulmonary arterial hypertension. Reprod Toxicol. 2015 Aug;56:45–51. doi:10.1016/j.reprotox.2015.06.048

96. Dingemanse J, Bodin F, Weidekamm E, Kutz K, Van Giersbergen P. Influence of Food Intake and Formulation on the Pharmacokinetics and Metabolism of Bosentan, a Dual Endothelin Receptor Antagonist. J Clin Pharmacol. 2002 Mar;42(3):283–9. doi:10.1177/00912700222011300

97. German J, Kowal A, Ehlers KH. Trimethadione and human teratogenesis. Teratology. 1970 Nov;3(4):349–61. doi:10.1002/tera.1420030412

98. Kelly TE. Teratogenicity of anticonvulsant drugs. I: Review of the literature. Am J Med Genet. 1984 Nov;19(3):413–34. doi:10.1002/ajmg.1320190302

99. Zackai EH, Mellman WJ, Neiderer B, Hanson JW. The fetal trimethadione syndrome. J Pediatr. 1975;87(2):280–4.

100. Bruni J, Albright PS. The Clinical Pharmacology of Antiepileptic Drugs: Clin Neuropharmacol. 1984 Mar;7(1):1–34. doi:10.1097/00002826-198403000-00001

101. Endres W. D-penicillamine in pregnancy — to ban or not to ban? Klin Wochenschr. 1981 Jun;59(11):535–7. doi:10.1007/BF01716453

102. Ostensen M, Ramsey-Goldman R. Treatment of Inflammatory Rheumatic Disorders in Pregnancy: What are the Safest Treatment Options? Drug Saf. 1998;19(5):389–410. doi:10.2165/00002018-199819050-00006

103. Brooks PM, editor. Slow-acting antirheumatic drugs and immunosuppressives. London Philadelphia: Baillière Tindall; 1990. 553–574 p. (Baillière's clinical rheumatology; volume 4, number 3).

104. Dong M, Ware RE, Dallmann A, Vinks AA. Hydroxyurea treatment for sickle cell anemia during pregnancy and lactation: Current evidence and knowledge gaps. Pharmacother J Hum Pharmacol Drug Ther. 2023 May;43(5):419–29. doi:10.1002/phar.2793

105. National Toxicology Program. NTP-CERHR monograph on the potential human reproductive and developmental effects of hydroxyurea. Ntp Cerhr Mon. 2008 Oct;(21):vii–viii, v, ix-III1. PubMed PMID: 19407858.

106. Apperley J. Issues of Imatinib and Pregnancy Outcome. J Natl Compr Canc Netw. 2009 Nov;7(10):1050–8. doi:10.6004/jnccn.2009.0069

107. Berman E. Family Planning and Pregnancy in Patients with Chronic Myeloid Leukemia. Curr Hematol Malig Rep. 2023 Apr;18(2):33–9. doi:10.1007/s11899-023-00689-5

108. Peng B, Lloyd P, Schran H. Clinical Pharmacokinetics of Imatinib: Clin Pharmacokinet. 2005;44(9):879–94. doi:10.2165/00003088-200544090-00001

109. Park S, Zimmerman E, Huerta-Montañez G, Rosario-Pabón Z, Vélez-Vega CM, Cordero JF, et al. Gestational Exposure to Phthalates and Phthalate Replacements in Relation to Neurodevelopmental Delays in Early Childhood. Toxics. 2023 Jan 11;11(1):65. doi:10.3390/toxics11010065

110. Svoboda LK, Ishikawa T, Dolinoy DC. Developmental toxicant exposures and sex-specific effects on epigenetic programming and cardiovascular health across generations. Environ Epigenetics. 2022 Oct 26;8(1):dvac017. doi:10.1093/eep/dvac017

111. Daston GP, Beyer BK, Carney EW, Chapin RE, Friedman JM, Piersma AH, et al. Exposure-Based Validation List for Developmental Toxicity Screening Assays. Birth Defects Res B Dev Reprod Toxicol. 2014 Dec;101(6):423–8. doi:10.1002/bdrb.21132

112. Johanson G. Toxicity Review of Ethylene Glycol Monomethyl Ether and its Acetate Ester. Crit Rev Toxicol. 2000 Jan;30(3):307–45. doi:10.1080/10408440091159220

113. Welsch F. The mechanism of ethylene glycol ether reproductive and developmental toxicity and evidence for adverse effects in humans. Toxicol Lett. 2005 Mar;156(1):13–28. doi:10.1016/j.toxlet.2003.08.010

114. Gargas ML, Tyler TR, Sweeney LM, Corley RA, Weitz KK, Mast TJ, et al. A Toxicokinetic Study of Inhaled Ethylene Glycol Monomethyl Ether (2-ME) and Validation of a Physiologically Based Pharmacokinetic Model for the Pregnant Rat and Human. Toxicol Appl Pharmacol. 2000 May;165(1):53–62. doi:10.1006/taap.2000.8928

115. Nucera B, Brigo F, Trinka E, Kalss G. Treatment and care of women with epilepsy before, during, and after pregnancy: a practical guide. Ther Adv Neurol Disord. 2022 Jan;15:1–31. doi:10.1177/17562864221101687

116. Shank RP, Maryanoff BE. Molecular Pharmacodynamics, Clinical Therapeutics, and Pharmacokinetics of Topiramate. CNS Neurosci Ther. 2008 Jun;14(2):120–42. doi:10.1111/j.1527-3458.2008.00041.x

117. Johannessen SI. Pharmacokinetics and Interaction Profile of Topiramate: Review and Comparison with Other Newer Antiepileptic Drugs. Epilepsia. 1997 Jan;38(s1). doi:10.1111/j.1528-1157.1997.tb04512.x

118. Warkany J. Warfarin embryopathy. Teratology. 1976 Oct;14(2):205–9. doi:10.1002/tera.1420140211

119. Wainwright H, Beighton P. Warfarin embryopathy: fetal manifestations. Virchows Arch. 2010 Dec;457(6):735–9. doi:10.1007/s00428-010-0982-9

120. Routledge P, Chapman P, Davies D, Rawlins M. Pharmacokinetics and pharmacodynamics of warfarin at steady state. Br J Clin Pharmacol. 1979 Sep;8(3):243–7. doi:10.1111/j.1365-2125.1979.tb01009.x

121. Stone KM, Reiff-Eldridge R, White AD, Cordero JF, Brown Z, Alexander ER, et al. Pregnancy outcomes following systemic prenatal acyclovir exposure: Conclusions from the international acyclovir pregnancy registry, 1984–1999. Birt Defects Res A Clin Mol Teratol. 2004 Apr;70(4):201–7. doi:10.1002/bdra.20013

122. Palma-Aguirre JA, Absalón-Reyes JA, Novoa-Heckel G, De Lago A, Oliva I, Rodríguez Z, et al. Bioavailability of two oral suspension and two oral tablet formulations of acyclovir 400 mg: Two single-dose, open-label, randomized, two-period crossover comparisons in healthy Mexican adult subjects. Clin Ther. 2007 Jun;29(6):1146–52. doi:10.1016/j.clinthera.2007.06.007

123. Brent RL, Christian MS, Diener RM. Evaluation of the reproductive and developmental risks of caffeine. Birth Defects Res B Dev Reprod Toxicol. 2011 Apr;92(2):152–87. doi:10.1002/bdrb.20288

124. Doepker C, Franke K, Myers E, Goldberger JJ, Lieberman HR, O'Brien C, et al. Key Findings and Implications of a Recent Systematic Review of the Potential Adverse Effects of Caffeine Consumption in Healthy Adults, Pregnant Women, Adolescents, and Children. Nutrients. 2018 Oct 18;10(10):1536. doi:10.3390/nu10101536

125. White JR, Padowski JM, Zhong Y, Chen G, Luo S, Lazarus P, et al. Pharmacokinetic analysis and comparison of caffeine administered rapidly or slowly in coffee chilled or hot versus chilled energy drink in healthy young adults. Clin Toxicol. 2016 Apr 20;54(4):308–12. doi:10.3109/15563650.2016.1146740

126. Bothamley G. Drug Treatment for Tuberculosis during Pregnancy: Safety Considerations. Drug Saf. 2001;24(7):553–65. doi:10.2165/00002018-200124070-00006

127. Surarak T, Chumnumwat S, Nosoongnoen W, Tragulpiankit P. Efficacy, safety, and pharmacokinetics of isoniazid affected by *NAT2* polymorphisms in patients with tuberculosis: A systematic review. Clin Transl Sci. 2024 Apr;17(4):e13795. doi:10.1111/cts.13795

128. Brand KMG, Saarelainen L, Sonajalg J, Boutmy E, Foch C, Vääräsmäki M, et al. Metformin in pregnancy and risk of adverse long-term outcomes: a register-based cohort study. BMJ Open Diabetes Res Care. 2022 Jan;10(1):e002363. doi:10.1136/bmjdrc-2021-002363

129. Gilbert C, Valois M, Koren G. Pregnancy outcome after first-trimester exposure to metformin: a meta-analysis. Fertil Steril. 2006 Sep;86(3):658–63. doi:10.1016/j.fertnstert.2006.02.098

130. Scheen AJ. Clinical Pharmacokinetics of Metformin: Clin Pharmacokinet. 1996 May;30(5):359–71. doi:10.2165/00003088-199630050-00003

131. Koren G. Safety considerations surrounding use of treatment options for nausea and vomiting in pregnancy. Expert Opin Drug Saf. 2017 Nov 2;16(11):1227–34. doi:10.1080/14740338.2017.1361403

132. Graffner C, Lagerstrom P, Lundborg P, Ronn O. Pharmacokinetics of metoclopramide intravenously and orally determined by liquid chromatography. Br J Clin Pharmacol. 1979 Nov;8(5):469–74. doi:10.1111/j.1365-2125.1979.tb01028.x

133. Beau A, Hurault-Delarue C, Vial T, Montastruc J, Damase-Michel C, Lacroix I. Safety of oseltamivir during pregnancy: a comparative study using the EFEMERIS database. BJOG Int J Obstet Gynaecol. 2014 Jun;121(7):895–900. doi:10.1111/1471-0528.12617

134. Chambers CD, Johnson D, Xu R, Luo Y, Jones KL, the OTIS Collaborative Research Group. Oseltamivir use in pregnancy: Risk of birth defects, preterm delivery, and small for gestational age infants. Birth Defects Res. 2019 Nov 15;111(19):1487–93. doi:10.1002/bdr2.1566

135. Ehrenstein V, Kristensen NR, Monz BU, Clinch B, Kenwright A, Sørensen HT. Oseltamivir in pregnancy and birth outcomes. BMC Infect Dis. 2018 Dec;18(1):519. doi:10.1186/s12879-018-3423-z

136. Chairat K, Jittamala P, Hanpithakpong W, Day NPJ, White NJ, Pukrittayakamee S, et al. Population pharmacokinetics of oseltamivir and oseltamivir carboxylate in obese and non-obese volunteers. Br J Clin Pharmacol. 2016 Jun;81(6):1103–12. doi:10.1111/bcp.12892

137. Pope E, Koren G, Bozzo P. Sugar substitutes during pregnancy. Can Fam Physician Med Fam Can. 2014 Nov;60(11):1003–5. PubMed PMID: 25392440; PubMed Central PMCID: PMC4229159.

138. Colburn WA, Bekersky I, Blumenthal HP. Dietary saccharin kinetics. Clin Pharmacol Ther. 1981 Oct;30(4):558–63. doi:10.1038/clpt.1981.203

139. Heller S, Salkeld RM, Körner WF. Vitamin B1 status in pregnancy. Am J Clin Nutr. 1974 Nov;27(11):1221–4. doi:10.1093/ajcn/27.11.1221

140. Nelson-Percy C. Treatment of Nausea and Vomiting in Pregnancy: When Should it be Treated and What can be Safely Taken? Drug Saf. 1998;19(2):155–64. doi:10.2165/00002018-199819020-00006

141. Tallaksen CME, Sande A, B&#xfffd;hmer T, Bell H, Karlsen J. Kinetics of thiamin and thiamin phosphate esters in human blood, plasma and urine after 50 mg intravenously or orally. Eur J Clin Pharmacol. 1993 Jan;44(1):73–8. doi:10.1007/BF00315284

142. Centers for Disease Control and Prevention, HIV Medicine Association of the Infectious Diseases Society of America, Pediatric Infectious Diseases Society, HHS Panel on Treatment of HIV During Pregnancy and Prevention of Perinatal Transmission— A Working Group of the Office of AIDS Research Advisory Council (OARAC). Recommendations for the Use of Antiretroviral Drugs During Pregnancy and Interventions to Reduce Perinatal HIV Transmission in the United States [Internet]. 2025 [cited 2023 Jun 30]. Available from: https://clinicalinfo.hiv.gov/en/guidelines/perinatal

143. Acosta EP, Page LM, Fletcher CV. Clinical Pharmacokinetics of Zidovudine: An Update. Clin Pharmacokinet. 1996 Apr;30(4):251–62. doi:10.2165/00003088-199630040-00001

144. Simões-Costa M, Bronner ME. Establishing neural crest identity: a gene regulatory recipe. Development. 2015 Jan 15;142(2):242–57. doi:10.1242/dev.105445

145. Punovuori K, Malaguti M, Lowell S. Cadherins in early neural development. Cell Mol Life Sci. 2021 May;78(9):4435–50. doi:10.1007/s00018-021-03815-9

146. Chandrasekaran A, Avci HX, Ochalek A, Rösingh LN, Molnár K, László L, et al. Comparison of 2D and 3D neural induction methods for the generation of neural progenitor cells from human induced pluripotent stem cells. Stem Cell Res. 2017 Dec;25:139–51. doi:10.1016/j.scr.2017.10.010

147. McShane SG, Molè MA, Savery D, Greene NDE, Tam PPL, Copp AJ. Cellular basis of neuroepithelial bending during mouse spinal neural tube closure. Dev Biol. 2015 Aug;404(2):113–24. doi:10.1016/j.ydbio.2015.06.003

148. Siu WY, Arooz T, Poon RYC. Differential Responses of Proliferating versus Quiescent Cells to Adriamycin. Exp Cell Res. 1999 Jul;250(1):131–41. doi:10.1006/excr.1999.4551

149. Johnson TI, Minteer CJ, Kottmann D, Dunlop CR, Fernández SBDQ, Carnevalli LS, et al. Quantifying cell cycle-dependent drug sensitivities in cancer using a high throughput synchronisation and screening approach. EBioMedicine. 2021 Jun;68:103396. doi:10.1016/j.ebiom.2021.103396

150. Singh A, Xu YJ. The Cell Killing Mechanisms of Hydroxyurea. Genes. 2016 Nov 17;7(11):99. doi:10.3390/genes7110099

151. Scott WJ, Ritter EJ, Wilson JG. DNA synthesis inhibition and cell death associated with hydroxyurea teratogenesis in rat embryos. Dev Biol. 1971 Oct;26(2):306–15. doi:10.1016/0012-1606(71)90129-1

152. Hirsh J, Cade JF, Gallus AS. Fetal Effects of Coumadin Administered During Pregnancy. Blood. 1970 Nov 1;36(5):623–7. doi:10.1182/blood.V36.5.623.623

153. Donaldson C, Harrington D. Therapeutic warfarin use and the extrahepatic functions of vitamin K-dependent proteins. Br J Biomed Sci. 2017 Oct 2;74(4):163–9. doi:10.1080/09674845.2017.1336854

154. Clarke D, Hungate L. Addition of serine to rodent embryo cultures does not protect against 2-methoxyacetic acid-induced neural tube defects. Teratology. 1993;47(5):402.

155. Li ASW, Marikawa Y. Methoxyacetic acid inhibits histone deacetylase and impairs axial elongation morphogenesis of mouse gastruloids in a retinoic acid signaling-dependent manner. Birth Defects Res. 2020 Aug;112(14):1043–56. doi:10.1002/bdr2.1712

156. Balmer NV, Weng MK, Zimmer B, Ivanova VN, Chambers SM, Nikolaeva E, et al. Epigenetic changes and disturbed neural development in a human embryonic stem cell-based model relating to the fetal valproate syndrome. Hum Mol Genet. 2012 Sep 15;21(18):4104–14. doi:10.1093/hmg/dds239

157. Okuno H, Okano H. Modeling human congenital disorders with neural crest developmental defects using patient-derived induced pluripotent stem cells. Regen Ther. 2021 Dec;18:275–80. doi:10.1016/j.reth.2021.08.001

158. Clevers H, Nusse R. Wnt/β-Catenin Signaling and Disease. Cell. 2012 Jun;149(6):1192–205. doi:10.1016/j.cell.2012.05.012

159. Conti L, Cattaneo E. Neural stem cell systems: physiological players or in vitro entities? Nat Rev Neurosci. 2010 Mar;11(3):176–87. doi:10.1038/nrn2761

160. Sant KE, Dolinoy DC, Jilek JL, Sartor MA, Harris C. Mono-2-ethylhexyl phthalate disrupts neurulation and modifies the embryonic redox environment and gene expression. Reprod Toxicol. 2016 Aug;63:32–48. doi:10.1016/j.reprotox.2016.03.042

161. Roux S, Rubin LJ. Bosentan: a dual endothelin receptor antagonist. Expert Opin Investig Drugs. 2002 Jul;11(7):991–1002. doi:10.1517/13543784.11.7.991

162. Kanai SM, Clouthier DE. Endothelin signaling in development. Development. 2023 Dec 15;150(24):dev201786. doi:10.1242/dev.201786

163. Spence S. Teratogenic effects of the endothelin receptor antagonist L-753,037 in the rat. Reprod Toxicol. 1999 Jan 2;13(1):15–29. doi:10.1016/S0890-6238(98)00064-1

164. Gamble JT, Hopperstad K, Deisenroth C. The DevTox Germ Layer Reporter Platform: An Assay Adaptation of the Human Pluripotent Stem Cell Test. Toxics. 2022 Jul 13;10(7):392. doi:10.3390/toxics10070392

165. Zhang Z, Imperial MZ, Patilea-Vrana GI, Wedagedera J, Gaohua L, Unadkat JD. Development of a Novel Maternal-Fetal Physiologically Based Pharmacokinetic Model I: Insights into Factors that Determine Fetal Drug Exposure through Simulations and Sensitivity Analyses. Drug Metab Dispos. 2017 Aug;45(8):920–38. doi:10.1124/dmd.117.075192

166. Fay KA, Villeneuve DL, Swintek J, Edwards SW, Nelms MD, Blackwell BR, et al. Differentiating Pathway-Specific From Nonspecific Effects in High-Throughput Toxicity Data: A Foundation for Prioritizing Adverse Outcome Pathway Development. Toxicol Sci. 2018 Jun 1;163(2):500–15. doi:10.1093/toxsci/kfy049

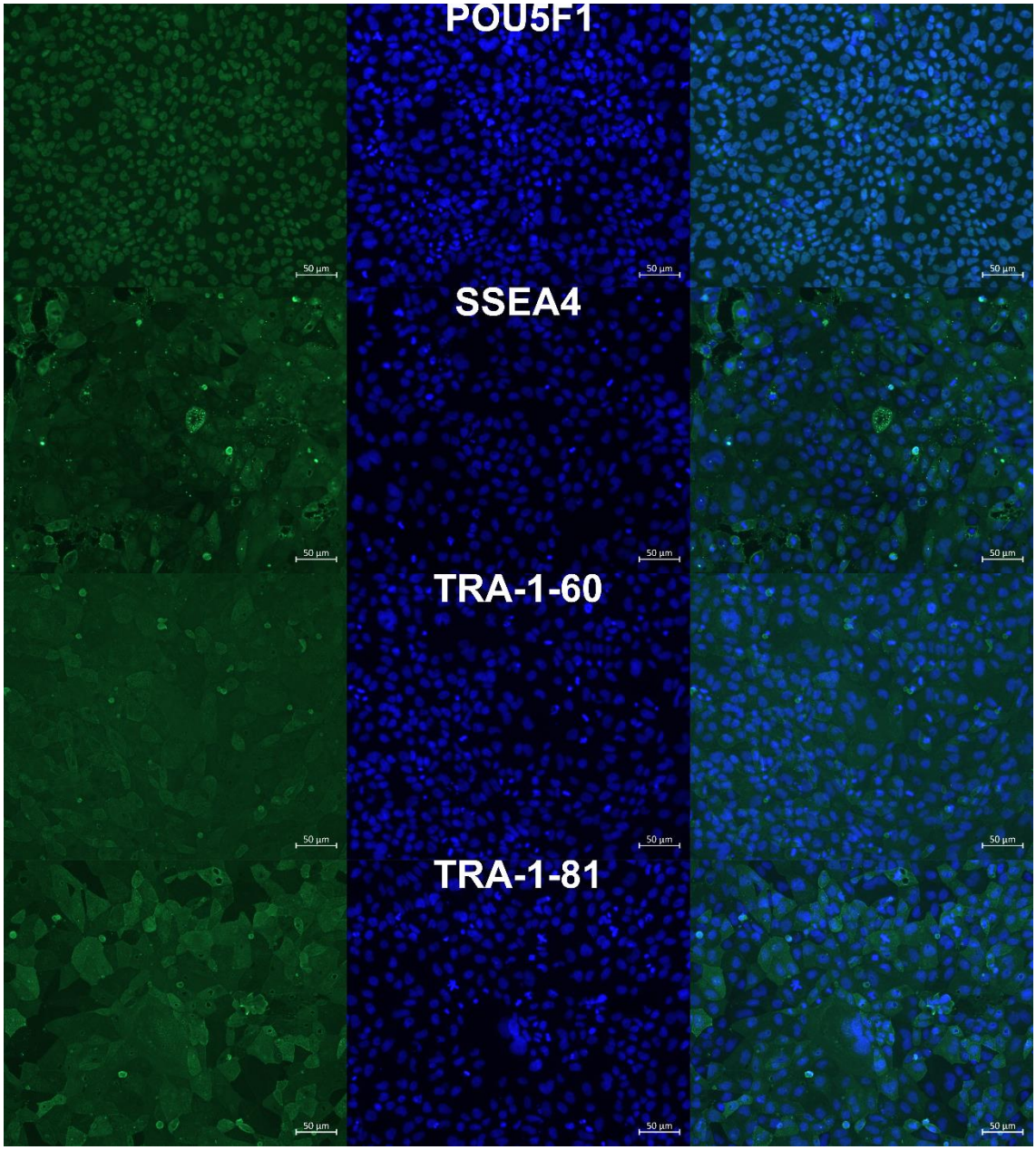
167. Vinken M, Blaauboer BJ. In vitro testing of basal cytotoxicity: Establishment of an adverse outcome pathway from chemical insult to cell death. Toxicol In Vitro. 2017 Mar;39:104–10. doi:10.1016/j.tiv.2016.12.004

168. Escher BI, Henneberger L, König M, Schlichting R, Fischer FC. Cytotoxicity Burst? Differentiating Specific from Nonspecific Effects in Tox21 in Vitro Reporter Gene Assays. Environ Health Perspect. 2020 Jul;128(7):077007. doi:10.1289/EHP6664

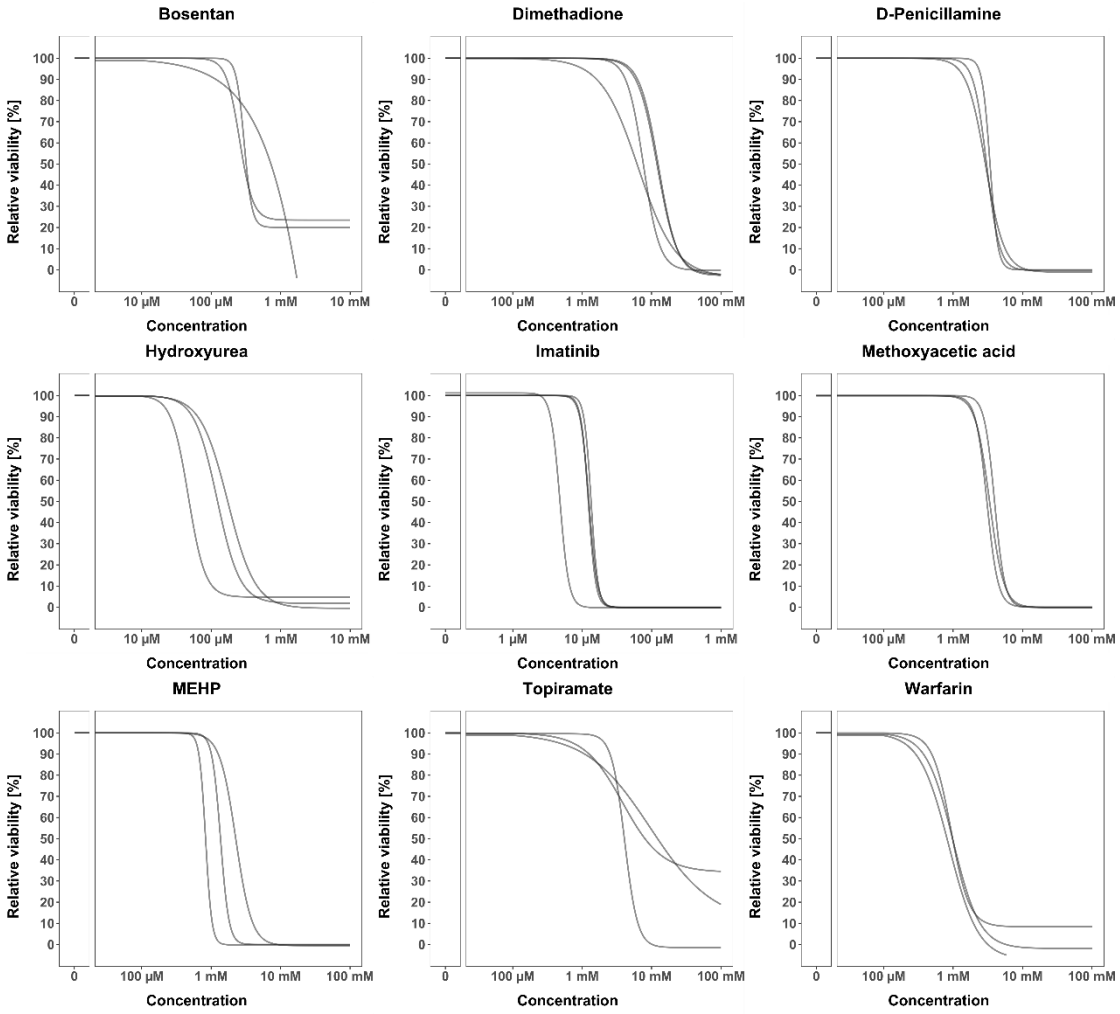
169. Iccvam. Validation, Qualification, and Regulatory Acceptance of New Approach Methodologies [Internet]. 2024 Mar [cited 2026 Mar 20]. p. 2. Report no.: 2. Available from: https://ntp.niehs.nih.gov/go/ICCVAM-submit doi:10.22427/NICEATM-2

6 Appendix

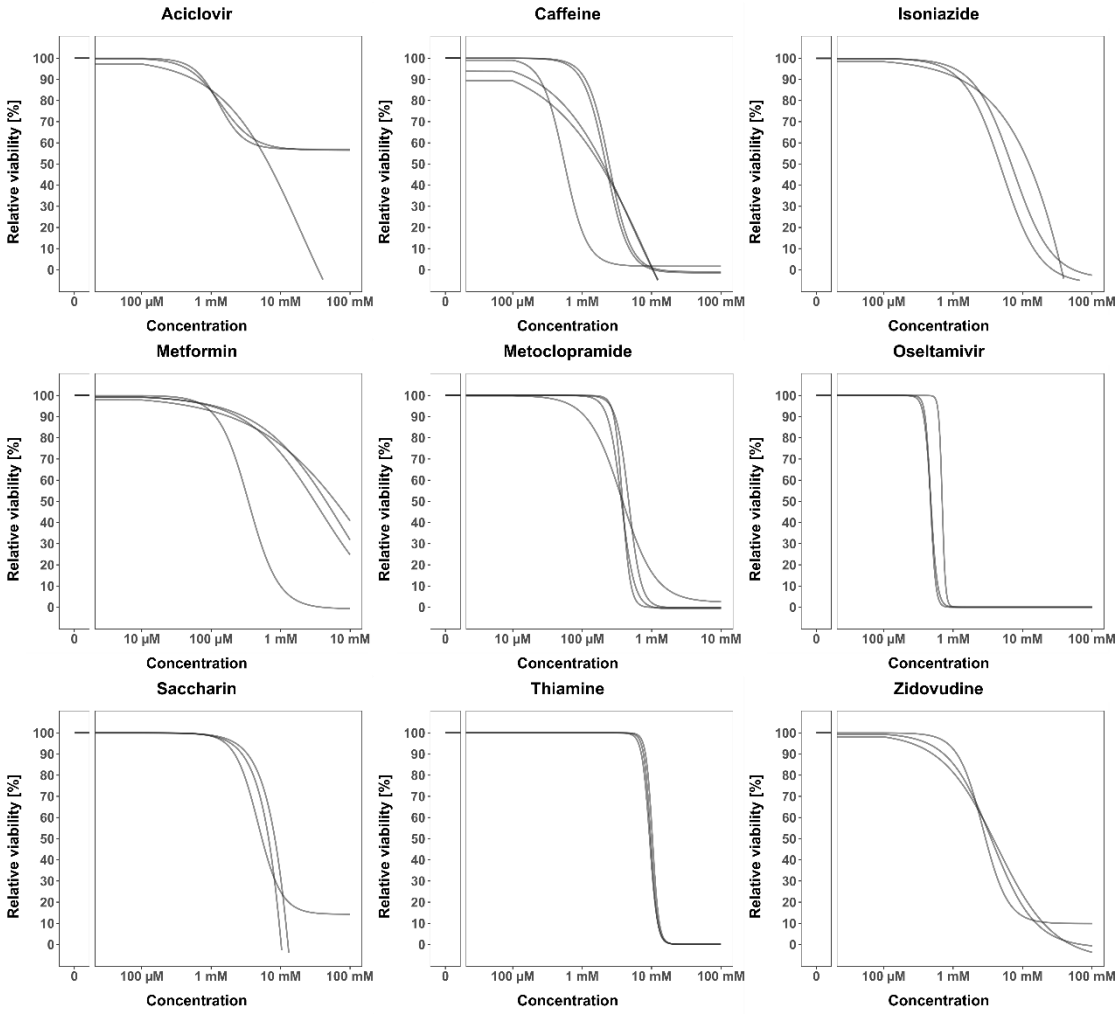
6.1 Supplementary Figures



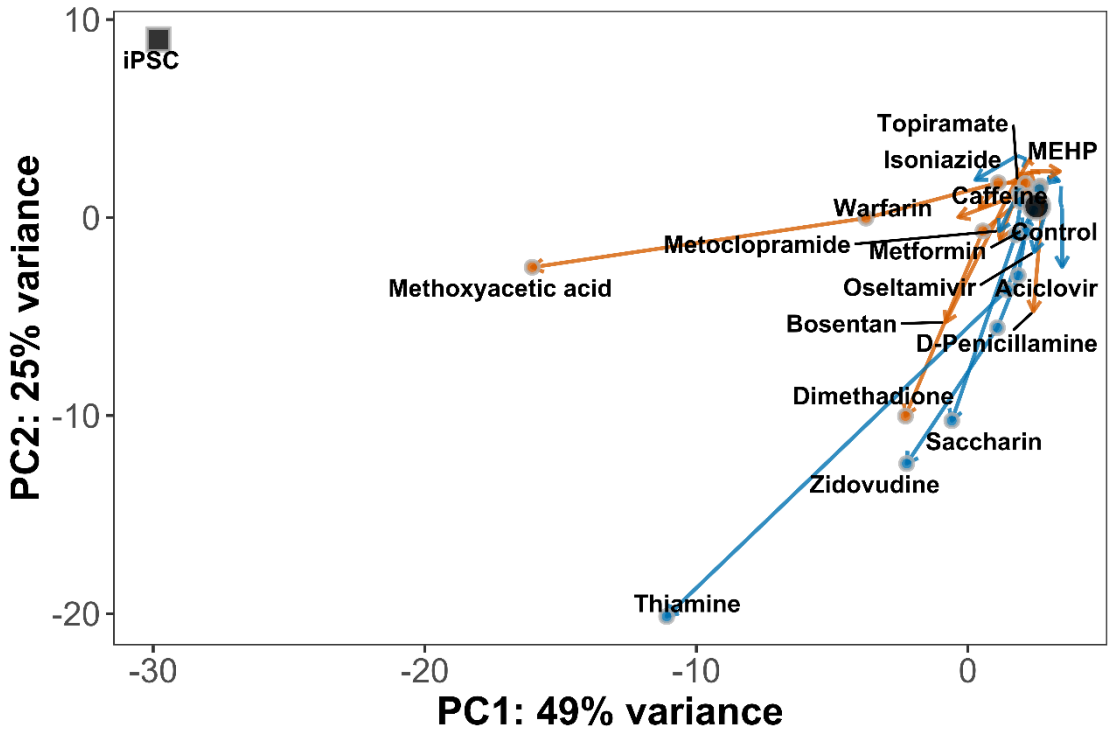
Supplementary figure 1: Immunofluorescence staining of established pluripotency markers. Undifferentiated cells were stained for POU5F1, SSEA4, TRA-1-60, TRA-1-81 (green), and nuclei were stained with DAPI (blue). Scale bar represents 50 µM.



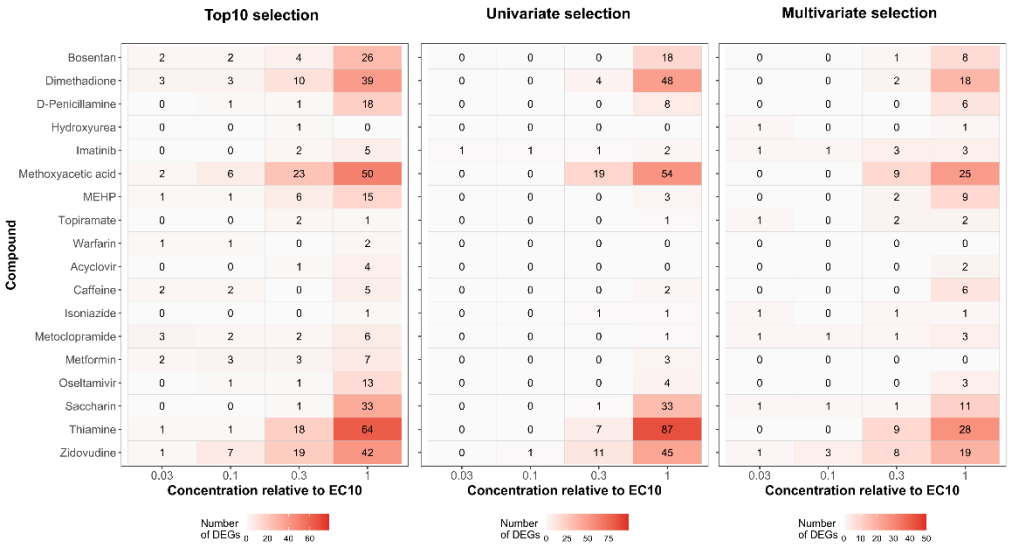
Supplementary figure 2: Cytotoxicity induced by exposure to developmental toxicants during early-stage neural induction. The different lines indicate concentration-viability curves from 3-5 independent experiments.



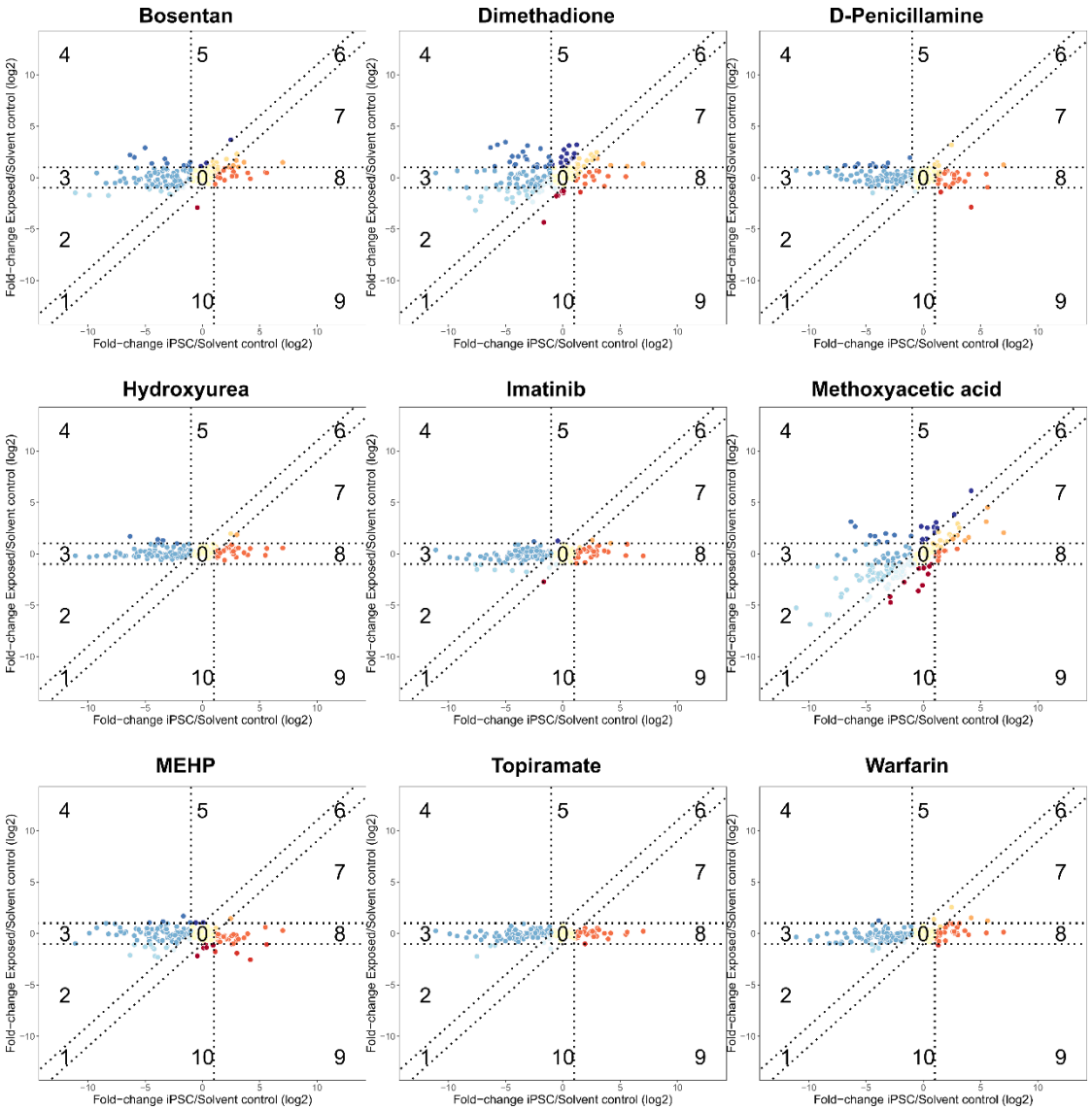
Supplementary figure 3: Cytotoxicity induced by exposure to non-toxic reference compounds during early-stage neural induction. The different lines indicate concentration-viability curves from 3-5 independent experiments.



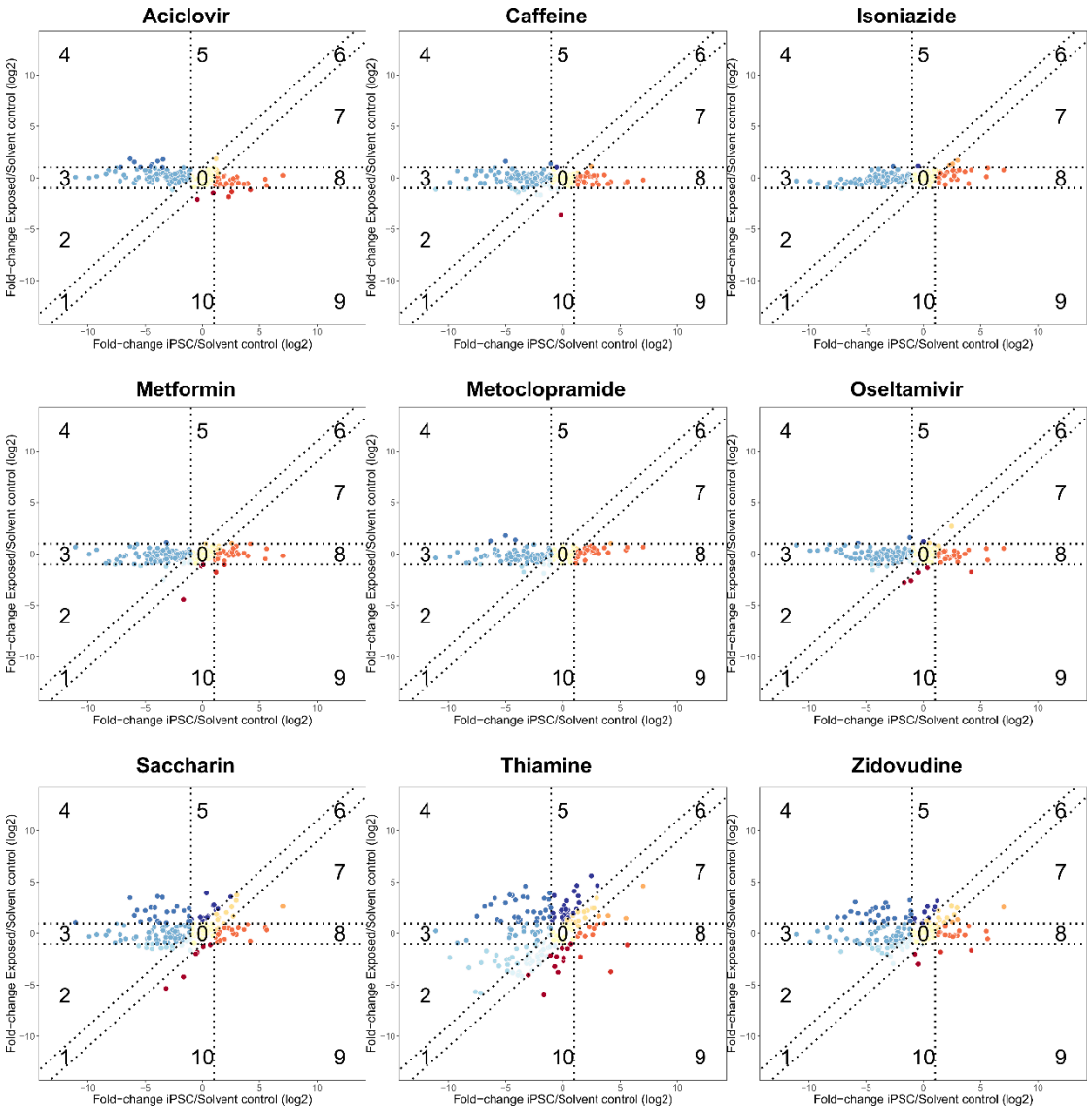
Supplementary figure 4: PCA of biomarkers after early neural induction. Biomarker expression was assessed by Targeted RNA Expression Measurement. Dots indicate neuroectodermal progenitor cells, square indicates undifferentiated iPSC. Arrows indicate increasing concentration. Black data points indicate controls, orange data points and arrows indicate toxicants, blue data points and arrows indicate non-toxicants.



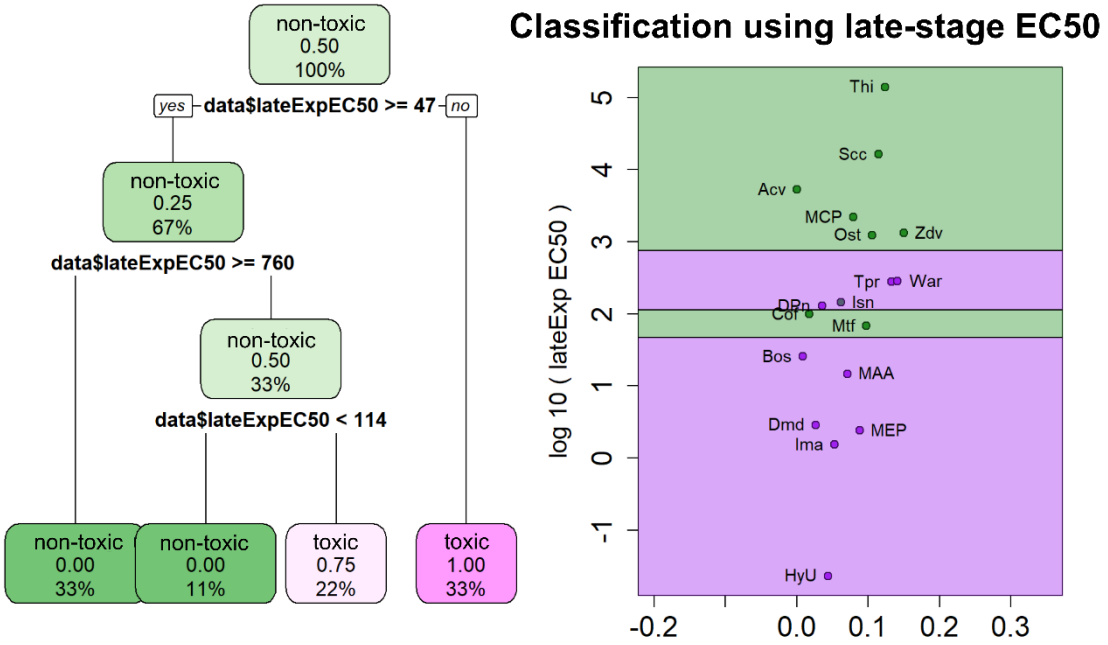
Supplementary figure 5: Heat map of differentially expressed biomarker genes obtained by three different selection strategies. Concentrations are given relative to the EC10 derived from cytotoxicity analysis during early neural induction.



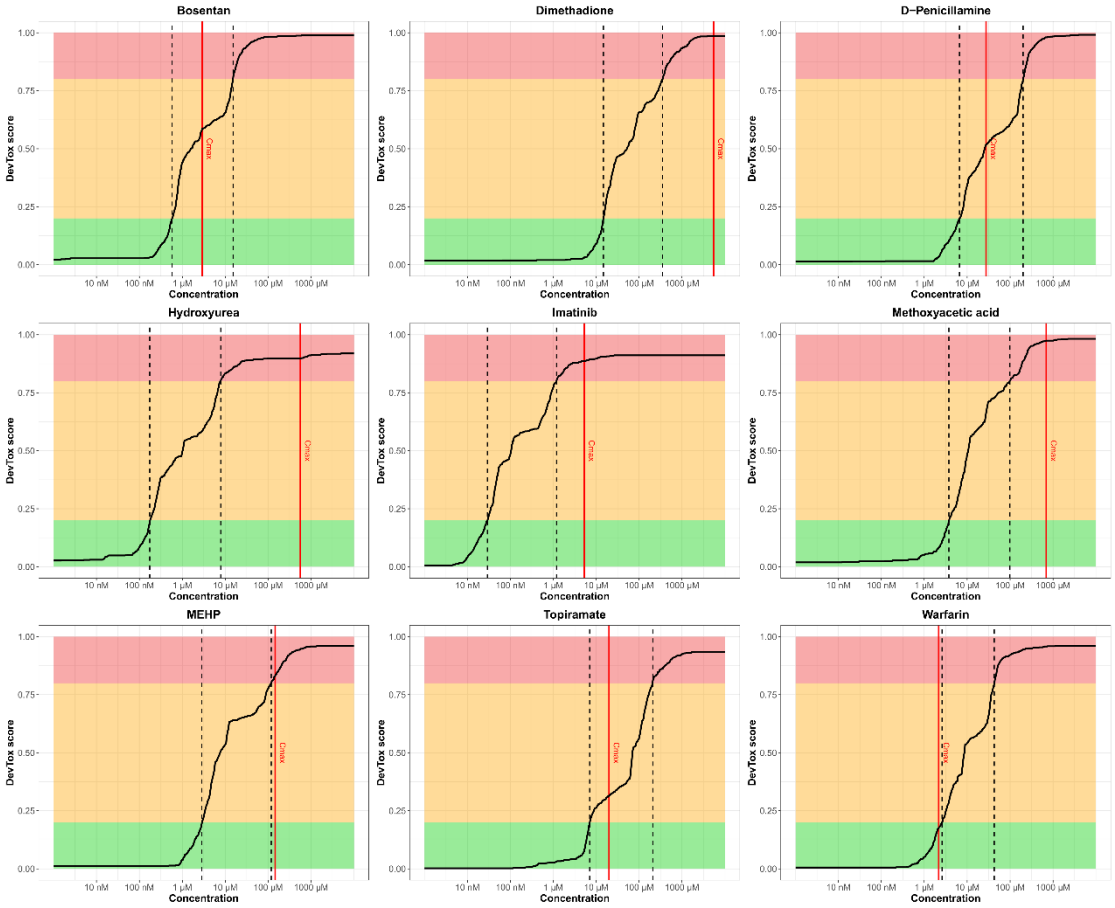
Supplementary figure 6: Differentiation Pattern Clustering after exposure to developmental toxicants. DiPaC plot showing gene expression changes induced by the highest tested concentration of 9 different developmental toxicants against gene expression changes during undisturbed differentiation. Dotted lines represent cut off values that define DPGs.



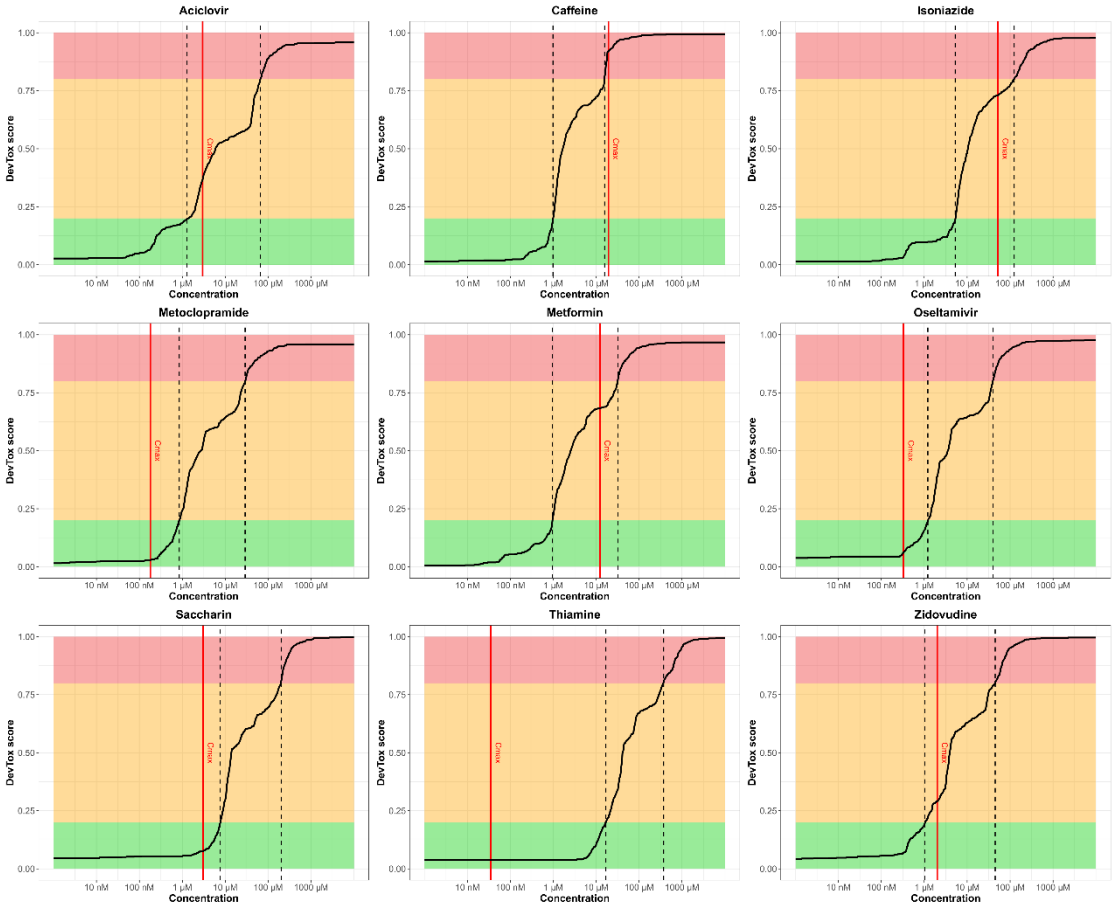
Supplementary figure 7: Differentiation Pattern Clustering after exposure to non-toxic reference compounds. DiPaC plot showing gene expression changes induced by the highest tested concentration of 9 different non-toxicants against gene expression changes during undisturbed differentiation. Dotted lines represent cut off values that define DPGs.



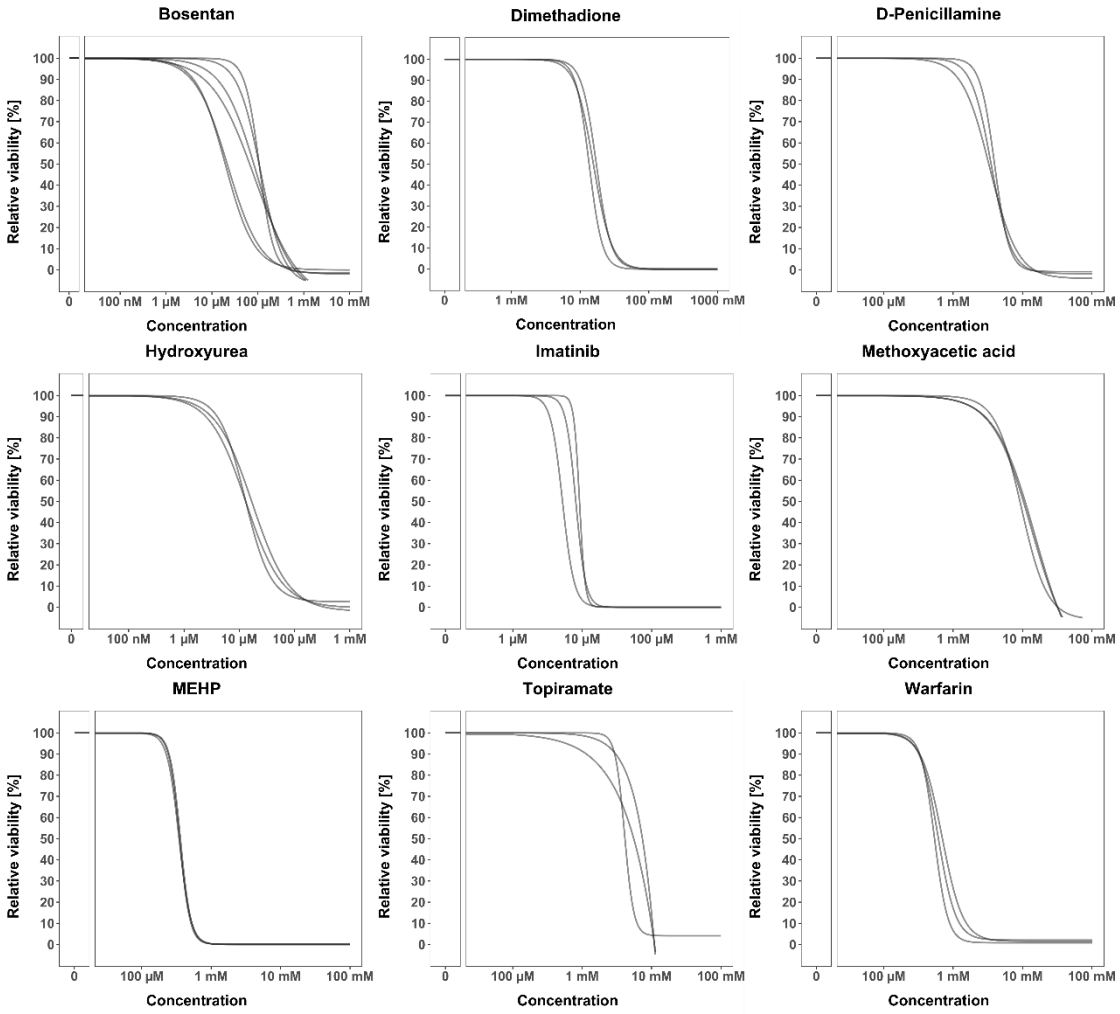
Supplementary figure 8: Random forest-based prediction of developmental toxicity without monotonic constraints leads to emergence of non-monotonic concentration-response patterns (right), exemplified by a single decision tree based on EC50 values of late-stage cytotoxicity (left).



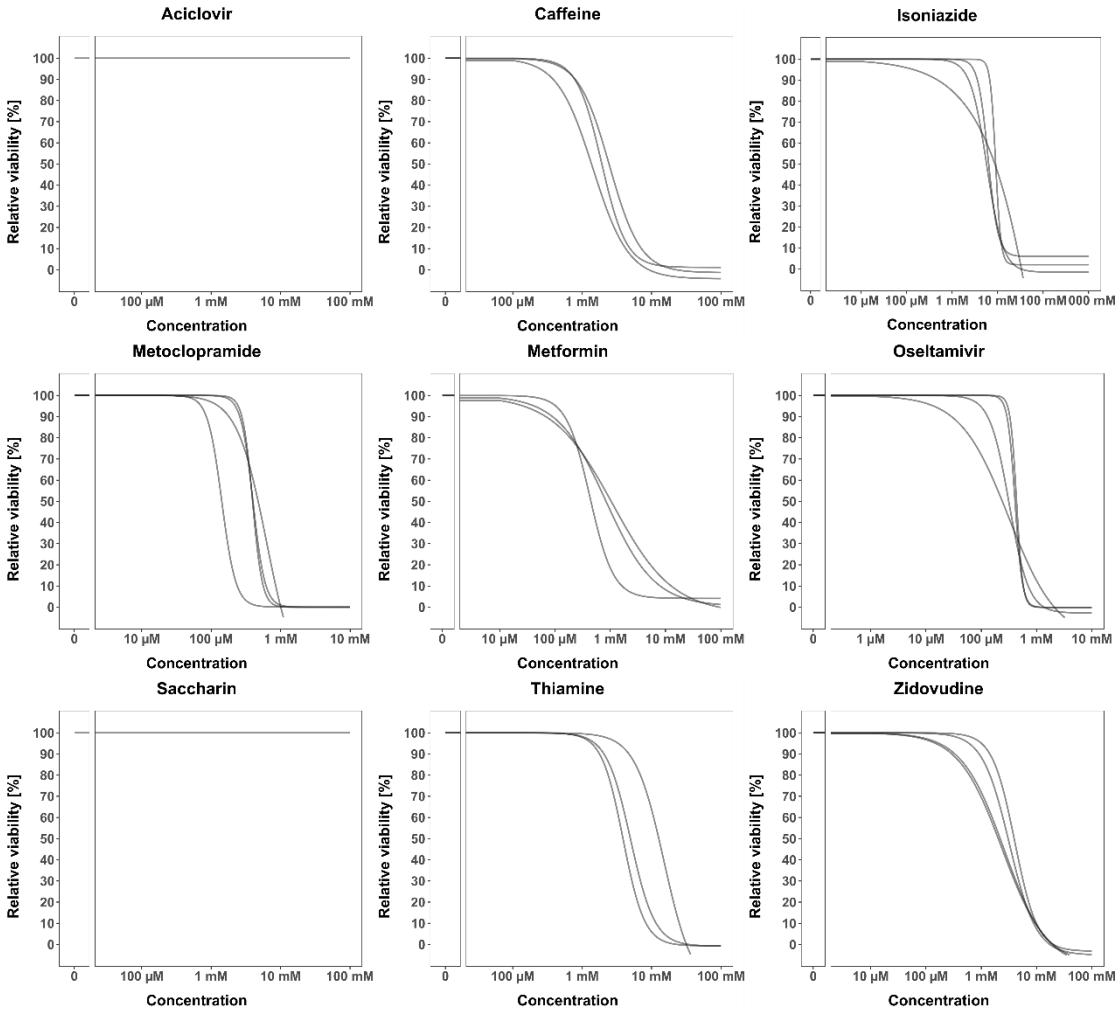
Supplementary figure 9: Concentration dependent developmental toxicity prediction of known toxicants after early-stage neural induction. Red, yellow and green areas indicate whether a prediction leads to toxic, uncertain or non-toxic classification. Straight dashed lines indicate maximum safe concentration and the minimal toxic concentration. Red line indicates c_{max}.



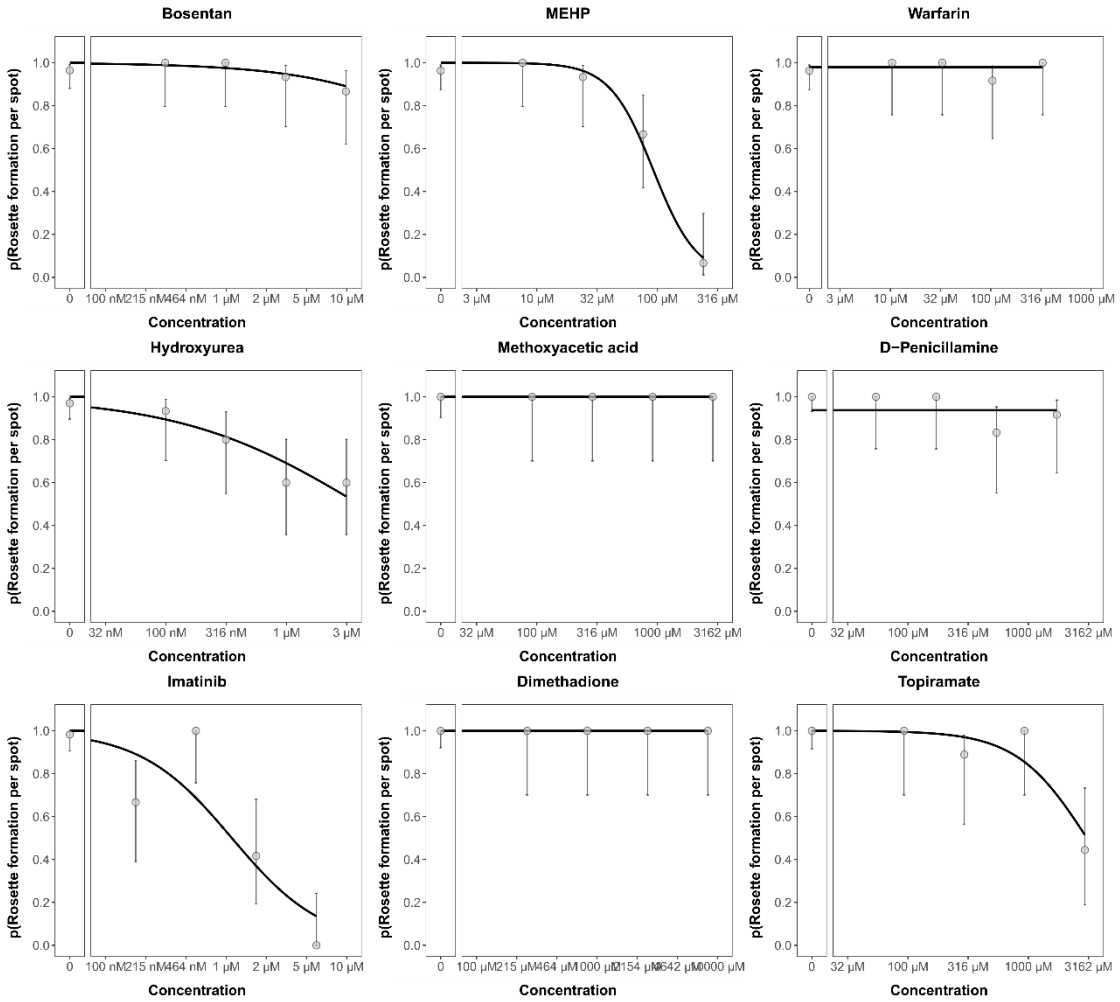
Supplementary figure 10: Concentration dependent developmental toxicity prediction of non-toxic reference compounds after early-stage neural induction. Red, yellow and green areas indicate whether a prediction leads to toxic, uncertain or non-toxic classification. Straight dashed lines indicate maximum safe concentration and the minimal toxic concentration. Red line indicates c_{max}.



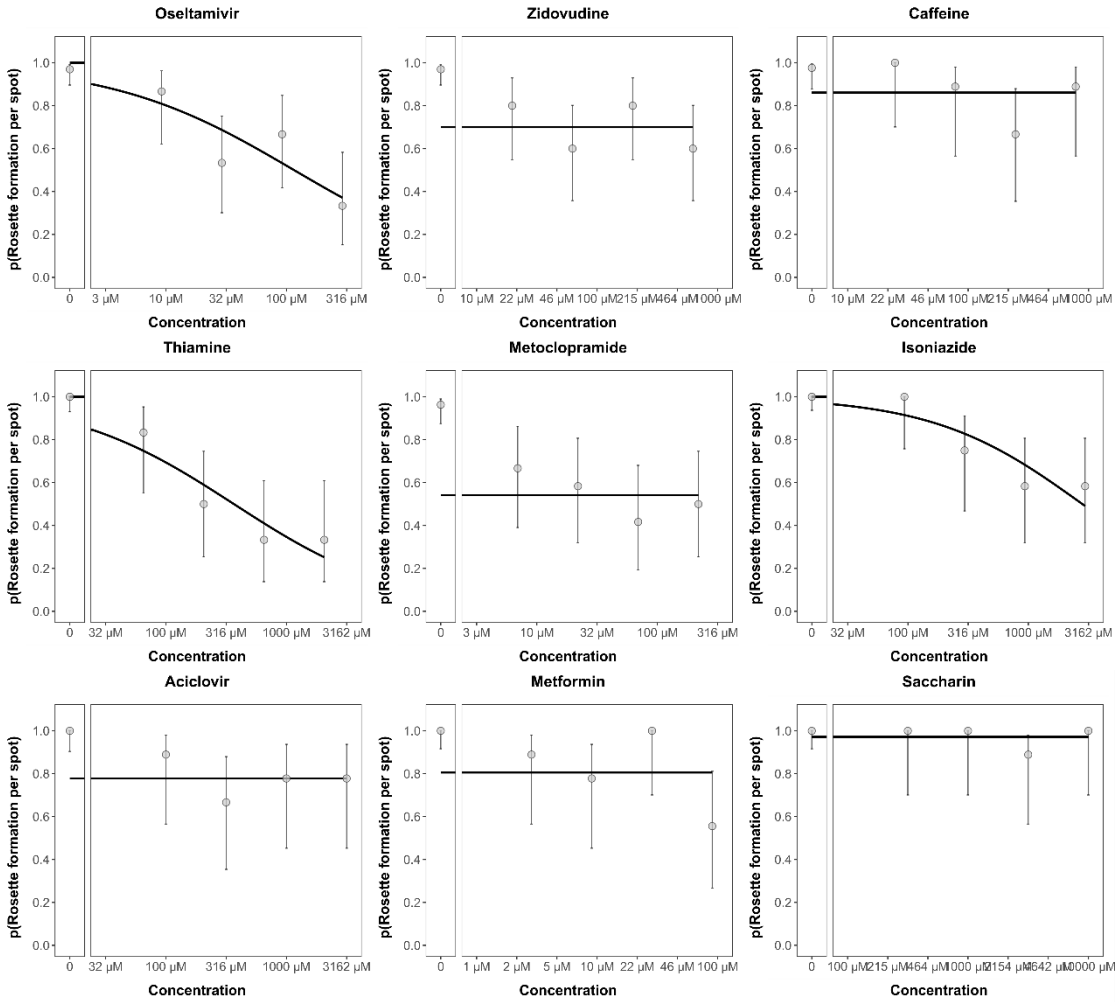
Supplementary figure 11: Cytotoxicity induced by exposure to developmental toxicants during late-stage neural induction. The different lines indicate concentration-viability curves from 3-5 independent experiments.



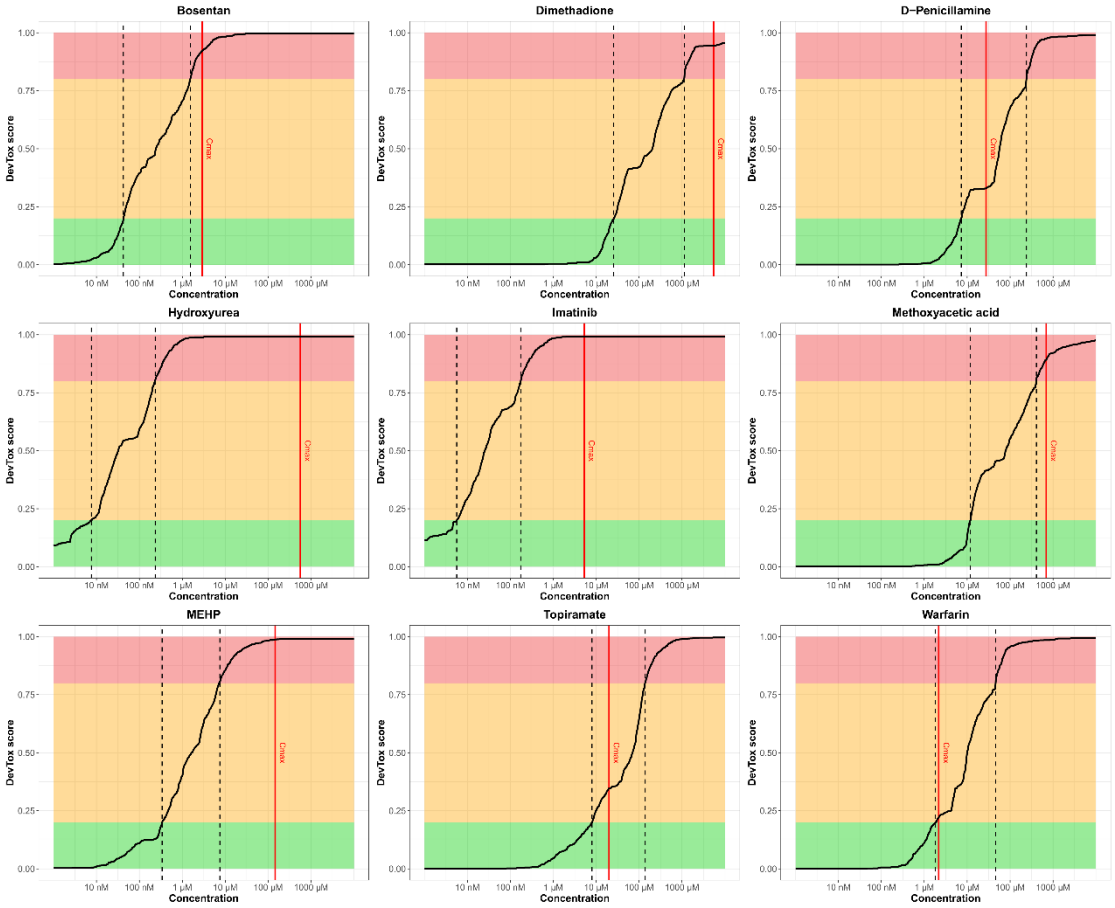
Supplementary figure 12: Cytotoxicity induced by exposure to non-toxic reference compounds during late-stage neural induction. The different lines indicate concentration-viability curves from 3-5 independent experiments.



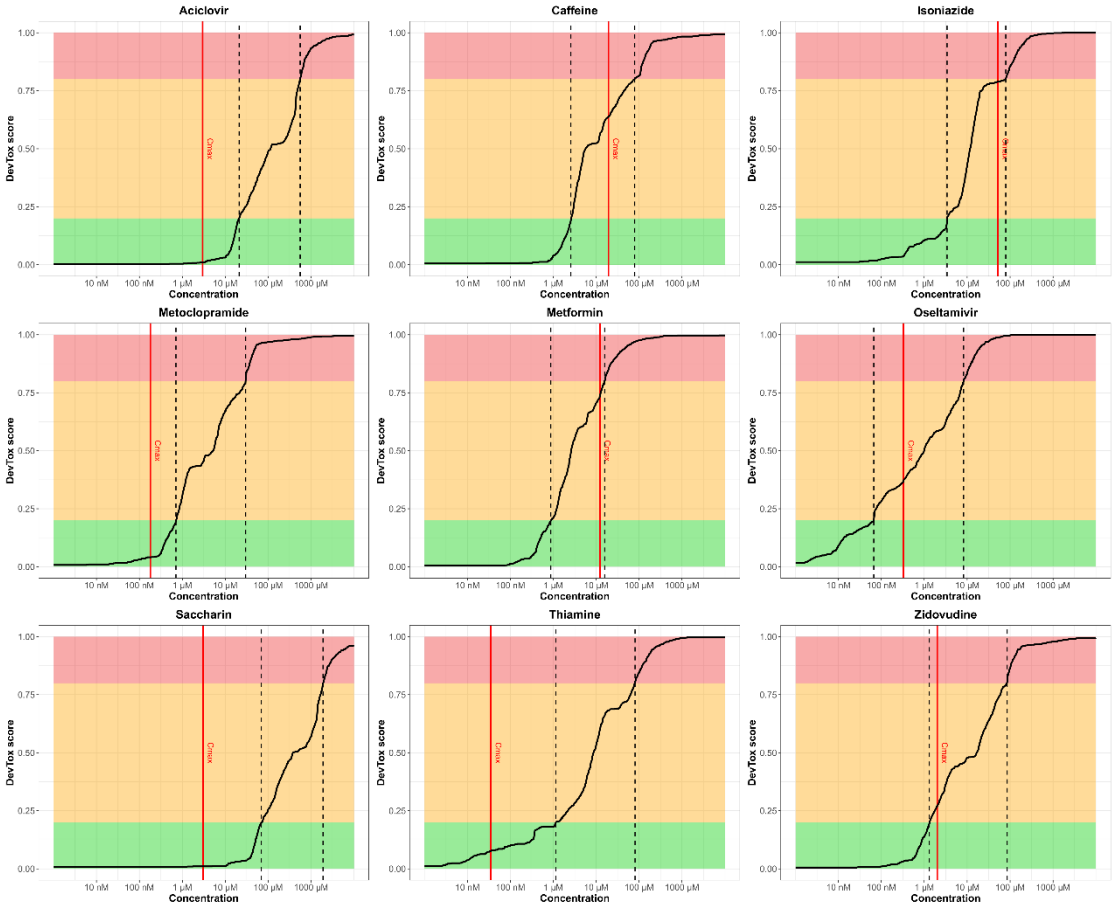
Supplementary figure 13: Disruption of neural rosette formation by exposure to developmental toxicants. The probability of rosette formation within a confined spot as derived by binomial concentration-response modelling is indicated on the y-axis. Error bars represent 95% confidence intervals.



Supplementary figure 14: Disruption of neural rosette formation by exposure to non-toxic reference compounds. The probability of rosette formation within a confined spot as derived by binomial concentration-response modelling is indicated on the y-axis. Error bars represent 95% confidence intervals.



Supplementary figure 15: Concentration dependent developmental toxicity prediction of known toxicants after late-stage neural induction. Red, yellow and green areas indicate whether a prediction leads to toxic, uncertain or non-toxic classification. Straight dashed lines indicate maximum safe concentration and the minimal toxic concentration. Red line indicates c_{max}.



Supplementary figure 16: Concentration dependent developmental toxicity prediction of non-toxic reference compounds after late-stage neural induction. Red, yellow and green areas indicate whether a prediction leads to toxic, uncertain or non-toxic classification. Straight dashed lines indicate maximum safe concentration and the minimal toxic concentration. Red line indicates c_{max}.

6.2 Supplementary tables

All supplementary tables are provided in the digital version of this thesis.

Supplementary table 1: List of transcriptional biomarker genes used for Targeted RNA Expression Measurement-based assessment of disturbed early neural induction as well as reference genes used for normalisation.

Supplementary table 2: Interpretable summary values for all compounds used for developmental toxicity prediction after early and late-stage neural induction.

Supplementary table 3: Classification of validation compounds based on early and late-stage prediction models, including maximal non-toxic concentration and minimum toxic concentration predicted in the three-category framework.

6.3 List of figures

Figure 1: Numbers of animals used for different types of safety assessment in the EU and Norway in 2022. Assessment of developmental toxicity (indicated in red) has

6.4 List of tables

7 Acknowledgements

First of all, I would like to express my thanks to Patrick Nell, who provided great support during the last years. I am very grateful for welcoming me into the working group, the scientific guidance and the on-going willingness to support me in all matters.

I would like to thank Prof. Dr. Jan G. Hengstler for giving me the opportunity to obtain my PhD at the Leibniz Research Centre for Working Environment and Human Factors, as well as for his continuous supervision and support of my development as a scientist.

Moreover, I would like to thank my fellow StemNet crew members Antonia Thomitzek, Sarah Seidel and David Feuerborn for their assistance during the project as well as the great office atmosphere. All the very valuable discussions during lunch break not only helped me in my project but also made work a lot more enjoyable.

Many thanks go to Prof. Dr. Jörg Rahnenführer for being second reviewer of this thesis and for the productive collaboration together with Annalena Weissert and Franziska Kappenberg, who greatly contributed to this work with invaluable statistical insights.

For their foundational contribution to the RNA sequencing experiments, I would like to thank Karolina Edlund and Katharina Derksen, who kept supporting me even in the face of hundreds of library preparations and weeks of pipetting. I truly appreciate it!

I would like to thank Florian Seidel for introducing me into the project when I started at IfADo, and I want to thank Daniel Hösle for supporting the project during his master's thesis. Moreover, many thanks go to all colleagues of the Toxicology department, for creating an incredibly welcoming, humorous, supportive, and collaborative atmosphere.

Many thanks go to my family and friends, especially to my parents Anke and Helmut Scholtz, as well as to Betti, Arne and Jule Strauß. Also, I want to thank my grandmother Helga Weiß, who somehow knew that I was going to get my PhD years before I realised it myself.

Finally, I want to thank my wife Vivi Illigens for the incredible and continuous support throughout the last years. Throughout long days and quite some weekend work, it has not always been easy, and I do not take it for granted – so I am as grateful as can be!