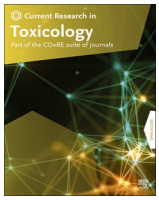
Current Research in Toxicology 8 (2025) 100223

Contents lists available at ScienceDirect

Current Research in Toxicology

journal homepage: www.journals.elsevier.com/current-research-in-toxicology



Research Paper

Profiling assay performance in the DevTox germ layer reporter platform☆

John T. Gamble [a,b], Chad Deisenroth [a,*]

[a] Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711 United States
[b] Oak Ridge Institute for Science and Education (ORISE), Oak Ridge, TN 37831, United States

ARTICLE INFO

Keywords:
High-throughput screening
Gastrulation
Developmental toxicity
Ectoderm
Mesoderm
Endoderm

ABSTRACT

The U.S. Environmental Protection Agency (U.S. EPA) is mandated to develop new approach methods (NAMs) to detect chemicals risks to susceptible populations, including effects on pregnant women and their offspring. With limited hazard information available for current and new chemicals, NAMs can provide greater relevance to human biology, mechanistic insight, and higher testing capacity than traditional animal models. The DevTox Germ Layer Reporter (GLR) model platform was recently established for high-throughput screening and prioritization of potential developmental hazards. The model platform utilizes the RUES2-GLR pluripotent stem cell reporter line that expresses fluorescent fusion protein biomarkers SOX17 (endoderm), Brachyury (mesoderm), and SOX2 (ectoderm and pluripotency); enabling a multi-lineage readout of gastrulation lineages. The DevTox GLR-Endo assay used the model platform to evaluate chemical effects on differentiating endoderm, yielding a balanced accuracy (BA) of 72% against a training set of 43 developmental toxicants and 23 non-developmental toxicants. To assess the predictivity of additional early embryonic lineages, assays for pluripotency (DevTox GLR-Pluri), ectoderm (DevTox GLR-Ecto), and mesoderm (DevTox GLR-Meso) were developed. Chemical reference set (12 developmental toxicants and 4 non-developmental toxicants) activity for each assay revealed BAs of 92% for DevTox GLR-Endo and DevTox GLR-Pluri, 71% for DevTox GLR-Ecto, and 58% for DevTox GLR-Meso. Expanded testing of the DevTox GLR-Endo and DevTox GLR-Pluri with 63 developmental and non-developmental toxicants yielded BAs of 75% and 68%, respectively. Amongst the four DevTox GLR platform assays, the DevTox GLR-Endo assay maintained the highest degree of efficacy and overall predictive accuracy for the compound set evaluated in this study.

Introduction

The U.S. EPA is committed to protecting human health and the environment. Humans are exposed to thousands of chemicals in consumer products and the environment that have little to no data regarding the impact on human health (Dix et al., 2006). Animal testing is currently used to determine adverse reproductive and developmental toxicity outcomes to make regulatory decisions in humans (Estevan et al., 2017). Recognizing the ethical concerns and scientific limitations associated with animal models, the U.S. EPA has increasingly embraced human cell-based methods as invaluable tools to enhance testing capabilities on chemicals with limited safety data (National Research Council, 2000, 2007). Shifting from traditional animal models to human cell-based models offers compelling advantages, including a more direct and relevant representation of human biology and the ability to study mechanistic responses with greater precision (Scialli et al., 2018). Additionally, the scalability and reproducibility of human cell-based models facilitate high-throughput screening, accelerating the pace of hazard identification. While the U.S. EPA utilizes numerous high-throughput screening approaches to identify environmental toxicants that disrupt key cellular functions, high-throughput approaches that measure perturbations to human embryonic development are less substantial (Dix et al., 2006; Judson et al., 2014; Kavlock et al., 2012; Richard et al., 2016). To address these challenges, human *in vitro* models that recapitulate early embryonic developmental pathways are needed (Scialli et al., 2018).

The utilization of human pluripotent stem cells (hPSC) provides a dynamic model system for studying early human development (Thomson et al., 1998). Given their ability to differentiate into each of

☆ This article is part of a special issue entitled: 'Stem Cells' published in Current Research in Toxicology.

* Corresponding author at: Advanced Experimental Toxicology Models Branch, Biomolecular and Computational Toxicology Division, Center for Computational Toxicology and Exposure, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, United States.
E-mail address: deisenroth.chad@epa.gov (C. Deisenroth).

https://doi.org/10.1016/j.crtox.2025.100223
Received 8 August 2024; Received in revised form 22 January 2025; Accepted 31 January 2025
Available online 8 February 2025

Nomenclature

Acronyms and Abbreviations

ACC	activity concentration at cutoff
AEID	assay endpoint identification
AIC	akaike information criterion
bmad	baseline median absolute deviation
BRA	Brachyury
CCTE	Center for Computational Toxicology and Exposure
CTOX	cytotoxicity
DevTox	developmental toxicity
DMSO	dimethyl sulfoxide
DPBS	Dulbecco's phosphate buffered saline
DT	developmental toxicity
Ecto	ectoderm
Endo	endoderm
EPA	Environmental Protection Agency
ESC	embryonic stem cell
FDA	Food and Drug Administration
FN	false negative
FP	false positive
GLR	germ layer reporter
hPSC	human pluripotent stem cell
Meso	mesoderm
NAM	new approach method
NDT	non-developmental toxicant
Pluri	pluripotent
rCV	robust coefficient of variation
RUES2-GLR	Rockefeller University Embryonic Stem cell line 2 – Germ Layer Reporter
rS/B	robust signal-to-background
rZ'	robust Z-prime factor
SOX17	SRY-box transcription factor 17
SOX2	SRY-box transcription factor 2
tcpl	ToxCast data pipeline
TN	true negative
TP	true positive
TSCA	Toxic Substances Control Act

the three germ layers (endoderm, mesoderm, and ectoderm) (Hong & Jeung, 2013; Siggia & Warmflash, 2018), insights may be gained regarding the potential role of environmental chemicals in disrupting critical developmental processes. The DevTox GLR model platform has the capacity to rapidly screen and identify potential chemical hazards for each germ layer lineage. The platform utilizes the RUES2-GLR pluripotent stem cell reporter line which expresses fluorescent fusion protein biomarkers for endoderm (SOX17-tdTomato), mesoderm (Brachyury (BRA)-mCerulean), and ectoderm or pluripotency (SOX2-mCitriline) to enable a multi-lineage high-throughput readout of gastrulation lineages (Martyn et al., 2018). Previously, the DevTox GLR-Endo assay, adapted from the Human Pluripotent Stem Cell Test (Kameoka et al., 2013), was evaluated against a broad range of pharmaceutical and environmental chemicals yielding a balanced accuracy of 72 % (Gamble et al., 2022). In this study, evaluation of the DevTox GLR model for the pluripotent state (DevTox GLR-Pluri assay) or differentiated toward the mesoderm lineage (DevTox GLR-Meso) or ectoderm lineage (DevTox GLR-Ecto), was conducted against developmental and non-developmental toxicants to determine the technical performance and predictivity across all assay modes in the platform.

Materials and methods

RUES2-GLR pluripotent stem cell culture

The RUES2-GLR (Rockefeller University Embryonic Stem cell line 2 – Germ Layer Reporter) cell line has been transgenically modified to express endogenous fluorescent reporter fusion protein biomarkers specific to each germ lineage (Martyn et al., 2018). Cell cultures were maintained with feeder-free complete Essential 8 Flex medium (Gibco, Waltham, MA) on vitronectin (Gibco, Waltham, MA) (0.5 μg/cm^2) as previous described (Gamble et al., 2022).

Test chemical selection and preparation

Negative reference chemicals (acetaminophen, folic acid, saccharin, penicillin G) were selected based on FDA pregnancy categories A and B, and demonstrate minimal bioactivity in other cell-based developmental toxicity assays (Gamble et al., 2022; Kameoka et al., 2013; Zurlinden et al., 2020). Positive reference chemicals (13-cis retinoic acid; 5,5-diphenylhydantoin; 5-fluorouracil; all-trans retinoic acid; bisphenol A; busulfan; diethylstibestrol; methotrexate; pomalidomide; sunitinib; thalidomide; valproic acid) were selected to represent FDA pregnancy categories D, X or undetermined, and exhibit bioactivity in cell-based developmental toxicity models (Table 1). A larger training set of 47 chemicals comprised of 28 positive and 19 negative chemicals was used to evaluate the predictive accuracy of the DevTox GLR-Pluri assay relative to historical data generated in the DevTox GLR-Endo assay (Gamble et al., 2022) (Supplementary Table S1). Chemicals were solubilized in dimethyl sulfoxide (DMSO) at a concentration of 100 mM, or up to the limit of solubility, as previously described (Gamble et al., 2022).

TaqMan hPSC scorecard gene expression panel

The TaqMan hPSC Scorecard panel was initially used to evaluate the RUES2-GLR differentiation efficiency for the ectoderm and mesoderm lineages using the commercial STEMdiff Trilineage Kit differentiation protocols (StemCell Tech, Cambridge, MA). Cells were seeded into vitronectin-coated (0.5 μg/cm^2) 6-well plates with 1X RevitaCell at 2.7 × 10^5 cells for ectoderm induction and 5.4 x 10^5 cells for mesoderm induction. Pluripotent control wells were seeded at a density of 2.7 × 10^5 cells and maintained in complete Essential 8 Flex medium. Differentiation was induced according to manufacturer protocols for a period of two or four days. Cells were harvested and RNA extracted using a RNeasy Mini Purification Kit (Qiagen, Hilden, Germany). The purified mRNA was converted to cDNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Waltham, MA). Gene expression was analyzed as previously reported (Gamble et al., 2022). Z-test score values were calculated using the average difference in the cycle threshold (ΔCt) value of two biological replicates and compared to the ΔCt values from 10 pluripotent stem cell lines maintained at pluripotency (Tsankov et al., 2015).

DevTox GLR assays

The DevTox GLR platform is comprised of four assay modes: DevTox GLR-Endo (Endoderm), DevTox GLR-Ecto (Ectoderm), DevTox GLR-Meso (Mesoderm), and DevTox GLR-Pluri (Pluripotent) which assess perturbations to specific gastrulation-associated cellular states. All assay protocols are aligned to the DevTox GLR-Endo assay workflow (Gamble et al., 2022). Briefly, for the DevTox GLR-Endo assay, RUES2-GLR cells were cultured to 70–85 % confluence and dissociated with Accutase to a single cell suspension according to the manufacturer's protocols (Gibco, Waltham, MA). Cells were counted with a trypan blue exclusion method using a Countess II automated cytometer (Life Technologies, Carlsbad,

Table 1
Reference chemicals. Chemical name, CAS registry number (CASRN), classification (developmental toxicant (DT), non-developmental toxicant (NDT)), FDA pregnancy category, and expected assay activity.

Chemical	CASN	Classification	FDA Category	Expected Assay Activity
Bisphenol A	80–05-7	DT	Undetermined	Positive
Retinoic Acid	302–79-4	DT	X	Positive
Diethylstilbestrol	56–53-1	DT	X	Positive
Valproic Acid	99–66-1	DT	D	Positive
13-cis Retinoic acid	4759–48-2	DT	X	Positive
Methotrexate	133073–73-1	DT	X	Positive
Busulfan	55–98-1	DT	D	Positive
Sunitinib	341031–54-7	DT	D	Positive
5,5-Diphenylhydantoin	57–41-0	DT	D	Positive
Thalidomide	50–35-1	DT	X	Positive
Pomalidomide	19171–19-8	DT	X	Positive
5-Fluorouracil	51–21-8	DT	D	Positive
Acetaminophen	103–90-2	NDT	B	Negative
Folic Acid	59–30-3	NDT	A	Negative
Saccharin	81–07-2	NDT	A	Negative
Penicillin G	113–98-4	NDT	B	Negative

CA) and seeded in Essential 8 Flex medium (Gibco, Waltham, MA) into vitronectin (0.5 μg/cm^2) coated 384-well PhenoPlates (Perkin Elmer, Shelton, CT) at 3,000 cells per well using a Certus FLEX Micro Dispenser (Fritz Gyger, Gwatt, CHE) with a 0.45 μm valve. After 24 (±2) hours, cells were differentiated to definitive endoderm lineage using 40 μL of Medium A in the PSC Definitive Endoderm Induction kit (Gibco, Waltham, MA) while exposed to test chemical. Medium was changed at 24 (±2) hours post exposure to Medium B with continued chemical exposure. At the assay termination point, cells were fixed in 4 % paraformaldehyde for 20 min, nuclei stained with HCS NuclearMask Deep Red stain (Invitrogen, Waltham, MA) for 30 min, and maintained in 50 μL Dulbecco's phosphate buffered saline (DPBS) (Gibco, Waltham, MA) for imaging.

The DevTox GLR-Meso assay was conducted in the same manner as the DevTox GLR-Endo assay with the following exceptions. RUES2-GLR cells were seeded at 6,000 cells per well in Essential 8 Flex medium with 1X RevitaCell supplement. At 24 (±2) hours seeding, medium was replaced with 40 μL of STEMdiff Trilineage Mesoderm medium (StemCell Tech, Cambridge, MA) with chemical exposure. After an additional 24 (±2) hours, medium was replaced with fresh Mesoderm medium with chemical exposure (Fig. 1A).

The DevTox GLR-Ecto assay was conducted in the same manner as the DevTox GLR-Endo assay with the following exceptions. RUES2-GLR cells were seeded at 3,000 cells per well in STEMdiff Trilineage Ectoderm medium (StemCell Tech, Cambridge, MA) with 1X RevitaCell supplement. After 24 (±2) hours post seeding, medium was replaced with 40 μL of fresh Ectoderm medium with chemical exposure. After an additional 24 (±2) hours, medium was replaced with fresh Ectoderm medium with chemical exposure.

The DevTox GLR-Pluri assay was conducted in the same manner as the DevTox GLR-Endo assay with the following exceptions. RUES2-GLR cells were seeded at 3,000 cells per well in Essential 8 Flex medium with 1X RevitaCell supplement. At 24 (±2) hours post seeding, medium was replaced with 40 μL fresh Essential 8 Flex medium with chemical exposure. After an additional 24 (±2) hours, the medium was replaced with chemical exposure.

For reference chemical testing, each chemical was dispensed at ten different concentrations using a log scale ranging from 100 pM – 200 μM, with four technical replicates per plate (Supplementary Fig. S1A-C). Each assay included 16 technical replicates of positive, negative, and solvent (0.2 % DMSO) controls. Training set chemical testing in the DevTox GLR-Pluri assay utilized a ten-point log-based concentration series for each chemical ranging from 1 nM – 200 μM without technical replicates and exclusion of border wells (Supplementary Fig. S1D). Positive, negative, and solvent controls ranged from eight to ten technical replicates. Four experimental replicates were completed for both reference and training set chemical testing.

Image acquisition and analysis

Fixed cells were imaged using an Opera Phenix High-Content Screening System with a plate::Handler robotic arm (Perkin Elmer, Waltham, MA). Fluorescent images for biomarkers SOX17 (tdTomato; ex 554/em 581), BRA (mCerulean; ex 433/em 475), SOX2 (mCitrine; ex 516/em 529), and nuclear stain (HCS NuclearMask Deep Red Stain; ex 636/em 685) were acquired using 2 x 2-pixel binning from five fields per well with a 20X water immersion objective. Using Harmony image analysis software (Perkin Elmer, Waltham, MA), cell nuclei were identified using algorithm C for the DevTox GLR-Endo assay, and algorithm M for the DevTox GLR-Meso, −Ecto and −Pluri assays. Border objects were excluded from analysis. For each cell, median cell biomarker intensity statistics were gathered and exported for further analysis using Python version 3.7.4 (Van Rossum and Fred, 2009). Plate-level negative and positive control data were used to establish a baseline intensity value cutoff threshold for determining primary biomarker positive cells. For each biomarker, median cell-level intensities for both the positive and negative control wells were averaged. In test wells, cells with median cell intensities above this average were considered biomarker positive. Cell counts reflect the sum of the total counted nuclei count in all five fields per well and were used to normalize the percent biomarker positive cells and evaluate cytotoxicity. Well-level results were used for data modeling in the ToxCast data pipeline (tcpl 3.1.0). Matplotlib version 3.1.1 (Hunter, 2007) in Python was used for scatter plot, rank order and box graphs. GraphPad Prism 9.5.0 (GraphPad Software, Boston, MA) was used for the reference chemical plots.

ToxCast data pipeline analysis

The ToxCast Data Pipeline R package (v3.1.0) was used to normalize, fit, and qualify concentration–response screening data (Feshuk et al., 2023). The raw data was loaded into the ToxCast database, invitrodb (to be released in version 4.2, expected Fall 2024) as the following assay endpoints under the Center for Computational Toxicology and Exposure (CCTE) assay source identifiers:

Assay endpoint ID (AEID)	Assay endpoint name
3223	CCTE_Deisenroth_DEVTOX-GLR_Endo_Sox17
3226	CCTE_Deisenroth_DEVTOX-GLR_Endo_CellCount
3229	CCTE_Deisenroth_DEVTOX-GLR_Meso_Bra
3230	CCTE_Deisenroth_DEVTOX-GLR_Meso_CellCount
3232	CCTE_Deisenroth_DEVTOX-GLR_Ecto_Sox2
3234	CCTE_Deisenroth_DEVTOX-GLR_Ecto_CellCount
3236	CCTE_Deisenroth_DEVTOX-GLR_Pluri_Sox2
3238	CCTE_Deisenroth_DEVTOX-GLR_Pluri_CellCount

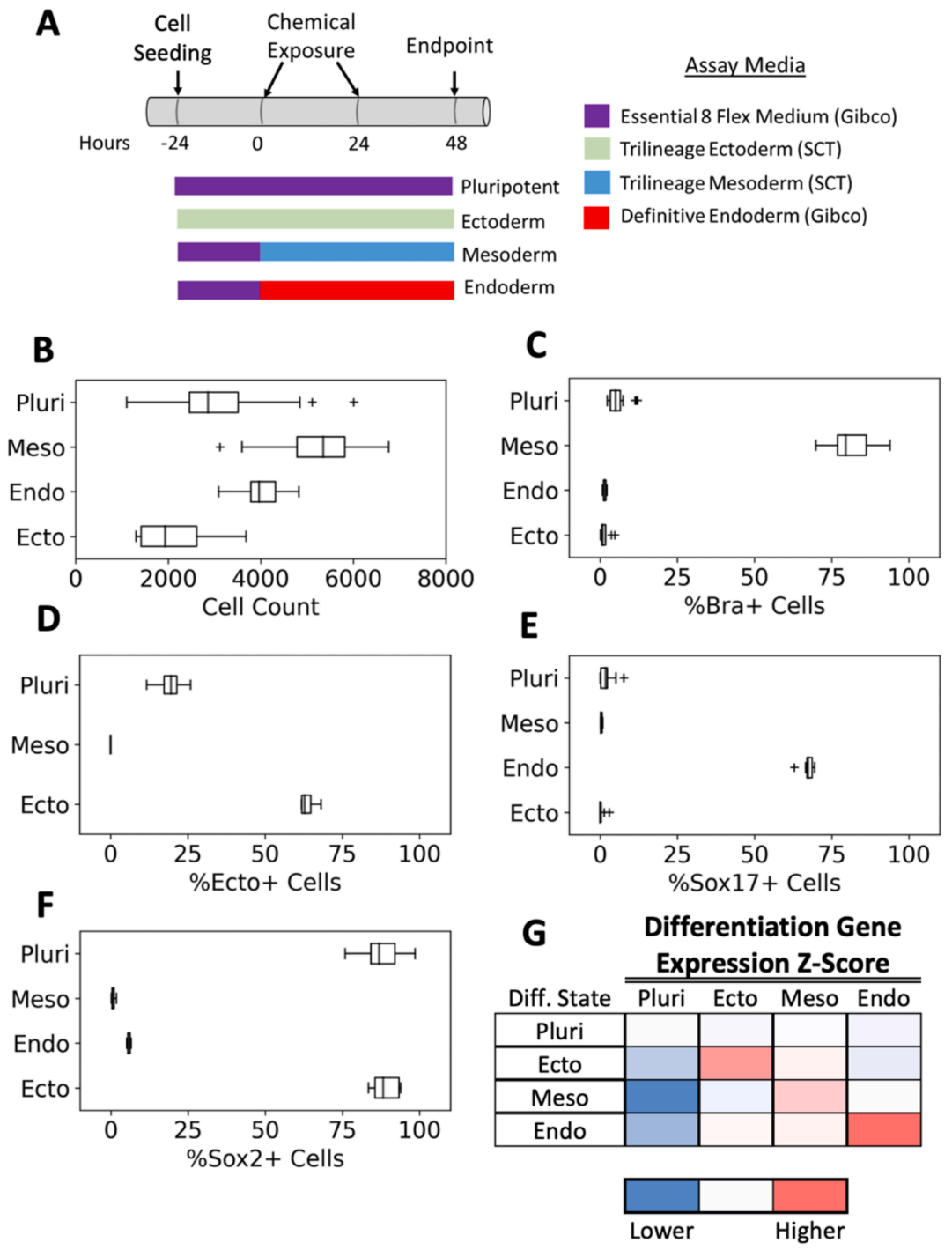


Fig. 1. Directed differentiation efficiency of RUES2-GLR to germ layer lineages. A) DevTox GLR Platform schematic, B) cell count data, C-F) biomarker expression profiles across each assay condition. Box and whisker plots: center line – median, box – first and third quartiles, whiskers – 1.5*inter-quartile range, + outliers; $n = 3$. G) Heatmap of Z-test scores using multiple human pluripotent stem cell lines at pluripotency for comparison. Gene expression data was gathered using the Taqman hPSC Scorecard array for RUES2-GLR cells in the pluripotent (Pluri) state or following 48 h of directed differentiation to ectoderm (Ecto), mesoderm (Meso), and endoderm (Endo) lineages; $n = 2$.

Well-level raw data response values (*rval*) were obtained as either the percentage of SOX17, SOX2, or BRA positive cells (developmental toxicity) or total cell counts (cytotoxicity). Technical replicates were aggregated by taking the plate-wise median response by sample and concentration index. The *rval* were normalized as response percent control (pc) relative to DMSO solvent control using the following equations,

$$resp.pc = -1^{*}\left(\frac{rval - bval}{pval - bval}\right)^{*}100 \quad (1)$$

where *resp.pc* is the normalized response to be fit, *pval* is set as 0 to establish uniform scaling across plates, and *bval* is the median baseline value of DMSO control wells on the same plate. The *resp.pc* values for biomarker positive (AEID 3223, 3229, 3232, 3236) and cell count positive (AEIDs 3226, 3230, 3234, 3238) endpoints were zero-centered,

sign inverted (to align with change in biomarker population), and fit to ten models (constant, linear, quadratic, power, hill, gain-loss, exponential 2, exponential 3, exponential 4, exponential 5), with the winning model determined through the minimum of the Akaike Information Criteria (AIC). Efficacy cutoffs (coff) to define thresholds for bioactivity hit calls were defined as three times baseline median absolute deviation (3*bmad), where the baseline median absolute deviation (bmad) was calculated using the normalized responses of the DMSO solvent control wells across all plates in the screen. In this analysis, classification criteria for continuous activity hit calls (hitc) were defined as: hitc < 0 is not applicable, $0 \leq$ hitc < 0.9 is negative/equivocal, hitc ≥ 0.9 is positive (Supplementary TCPL data.zip)

Assay performance evaluation

Robust statistics were used to evaluate biomarker fluorescence intensity dynamic range (signal-to-background; rS/B), inter-plate precision (coefficient of variation; rCV), and screening quality (Z'-factor; rZ') as follows (Zhang et al., 1999; Zhang, 2011):

$$rS/B = \frac{\widetilde{\chi}_{Pos}}{\widetilde{\chi}_{Neg}} \tag{2}$$

$$rCV = \frac{\widetilde{\chi}_{Pos}}{mad_{Pos}} * 100 \tag{3}$$

$$rZ' = 1 - \frac{3(mad_{Pos} + mad_{Neg})}{\left|\widetilde{\chi}_{Pos} - \widetilde{\chi}_{Neg}\right|} \tag{4}$$

where $\widetilde{\chi}$ represents the median percent biomarker cell value of the control wells, *mad* is the median absolute deviation of the percent biomarker cell value control wells, *Pos* is the positive control wells under differentiation/pluripotency conditions and *Neg* is the negative control wells under pluripotent conditions for assays DevTox GLR-Endo, −Meso and −Ecto assays. The *Neg* control well for the DevTox GLR-Pluri assay was mesoderm induction wells corresponding to a loss of SOX2 expression. The inter-plate rCV calculation is the median of all plate-based rCV values.

Confusion matrix statistics, sensitivity (true positive rate), specificity (true negative rate), and balanced accuracy were used for determining assay performance as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN} * 100 \tag{5}$$

$$\text{Specificity} = \frac{TN}{TN + FP} * 100 \tag{6}$$

$$\text{Balanced Accuracy} = \frac{Sensitivity + Specificity}{2} * 100 \tag{7}$$

where *TP* is true positive, *TN* is true negative, *FP* is false positive, and *FN* is false negative.

Results

DevTox germ layer reporter platform

The DevTox GLR-Endo assay (Gamble et al., 2022) was previously adapted from the Human Pluripotent Stem Cell assay (Kameoka et al., 2013) and the first lineage-specific assay to be developed on the DevTox-GLR platform. Recognizing the parameters for the DevTox GLR-Endo assay may be suboptimal for the ectoderm, mesoderm, and pluripotent assays, seeding densities and total differentiation time for the new assay methods were evaluated in a 384-well format to assess the magnitude of biomarker expression. Cells were seeded between 1,000 and 6,000 cells per well and differentiated for up to six (mesoderm) or eight (ectoderm) days with fluorescent biomarker intensity and cell counts collected daily (Supplementary Figs. S2- S4). Cell seeding density of 3,000 cells per well was adequate to produce exclusive expression of SOX2 for the ectoderm and pluripotent conditions. A higher density of 6,000 cells per well for the mesoderm assay was necessary to obtain optimal expression of BRA without significant expression of the other biomarkers. In addition, 48 h was found to be sufficient for all assays to achieve exclusive primary biomarker expression; in alignment with the differentiation period for the endoderm assay (Fig. 1A).

Following a 48-hour induction period, median cell counts ranged from 2,000 to 6,000 cells per well depending on lineage type (Fig. 1B). The percentage of lineage-specific biomarker expression was evident in the majority of the culture population (Endoderm – 67.3 % SOX17+, Mesoderm – 80 % BRA+, Ectoderm – 65 % Ecto+, Pluripotent – 90 % SOX2 +) (Fig. 1C-F). While ectoderm and pluripotency lineages share the same biomarker (SOX2), ectoderm induction increased SOX2 biomarker intensity more than two-fold relative to the pluripotent conditions, making it possible to distinguish between the two cell states (Supplementary Fig. S5C). Taqman hPSC Scorecard arrays were used to obtain a broader indication of lineage differentiation efficiency beyond the single biomarkers. The qPCR Scorecard arrays are comprised of a curated set of genes that are specific to one of the four gastrulation related states (pluripotent, endoderm, mesoderm and ectoderm) (Tsankov et al., 2015). Following directed differentiation, Z-scores revealed successful differentiation for each method where higher scores indicate a larger difference in expression of lineage specific genes compared to pluripotent cell lines (Fig. 1G).

Reference chemical evaluation

A chemical reference set comprised of 12 chemicals known to elicit developmental toxicity effects within *in vitro* cell-based assays and four known non-developmental toxicants previously used to profile the predictivity of the DevTox GLR-Endo assay (Gamble et al., 2022). To evaluate against this benchmark set, all 16 chemicals were screened in the DevTox GLR-Meso, −Ecto and −Pluri assays (Table 1). For each assay, a 10-point concentration series was tested ranging from 100 pM to 200 μM. One exception was sunitinib which precipitated out of solution above 2 mM and hence had an adjusted concentration range of (100 pM to 1.6 μM). Results were compared to the DevTox GLR-Endo assay benchmark dataset which was reanalyzed using improved methods, specifically using both positive and negative controls to establish gating criteria for biomarker expression. No effects were observed for the non-

Table 2
Reference chemical performance in the DevTox GLR platform assays. Reference chemicals were classified across each assay (TP – True Positive; FP – False Positive; TN – True Negative; FN – False Negative). Pomalidomide activity could not be assessed in the DevTox GLR-Meso assay due to autofluorescence signal interference.

	DevTox GLR Platform Assays			
Chemicals	Endo	Ecto	Meso	Pluri
13-cis Retinoic acid	TP	TP	FN	TP
5,5-Diphenylhydantoin	FN	FN	FN	TP
5-Fluouracil	TP	TP	FN	TP
Acetaminophen	TN	TN	TN	TN
all-trans Retinoic Acid	TP	TP	FN	TP
Bisphenol A	TP	FN	FN	TP
Busulfan	TP	TP	FN	TP
Diethylstilbestrol	TP	FN	TP	TP
Folic Acid	TN	TN	TN	TN
Methotrexate	TP	TP	FN	TP
Penicillin G	TN	TN	TN	TN
Pomalidomide	TP	FN	–	TP
Saccharin	TN	TN	TN	TN
Sunitinib	TP	FN	FN	TP
Thalidomide	TP	FN	FN	FN
Valproic Acid	FN	FN	FN	FN

developmental toxicants in any of the four assays (Table 2, Fig. 2). All assays provided acceptable robust Z'-factors greater than or equal to 0.6 (Table 3). Precision of measurements across experiments indicated the inter-experimental rCV was less than 2 %. The DevTox GLR-Meso assay was enriched with false negative observations, resulting in low sensitivity (9.1 %) and modest balanced accuracy (54.5 %). Pomalidomide results were excluded for the DevTox GLR-Meso assay due to autofluorescence signal interference. The DevTox GLR-Ecto assay performed with a sensitivity of 58.3 % and balanced accuracy of 79.2 %. Both the DevTox GLR-Pluri and −Endo assays performed equally well with 83.3 % sensitivity and balanced accuracies of 91.7 %. Both 5,5-diphenylhydantoin and valproic acid were identified as false negatives in the DevTox GLR-Endo assay. Thalidomide and valproic acid were false negatives in the DevTox GLR-Pluri assay.

Table 3
DevTox GLR Platform assay metrics. Assay performance metrics for screening quality (rZ'-Factor), dynamic range (rS/B), and precision (rCV) are noted. Assay sensitivity (true positive rate), specificity (true negative rate), and balanced accuracy are noted.

	DevTox GLR Platform Assays			
	Endo	Ecto	Meso	Pluri
rZ'-Factor	0.8	0.6	0.8	0.9
rS/B	112.6	3.2	27	315.1
rCV	1.1 %	1.5 %	1.9 %	1.4 %
Sensitivity	83.3 %	41.7 %	9.1 %	83.3 %
Specificity	100 %	100 %	100 %	100 %
Balanced Accuracy	91.7 %	70.8 %	54.5 %	91.7 %

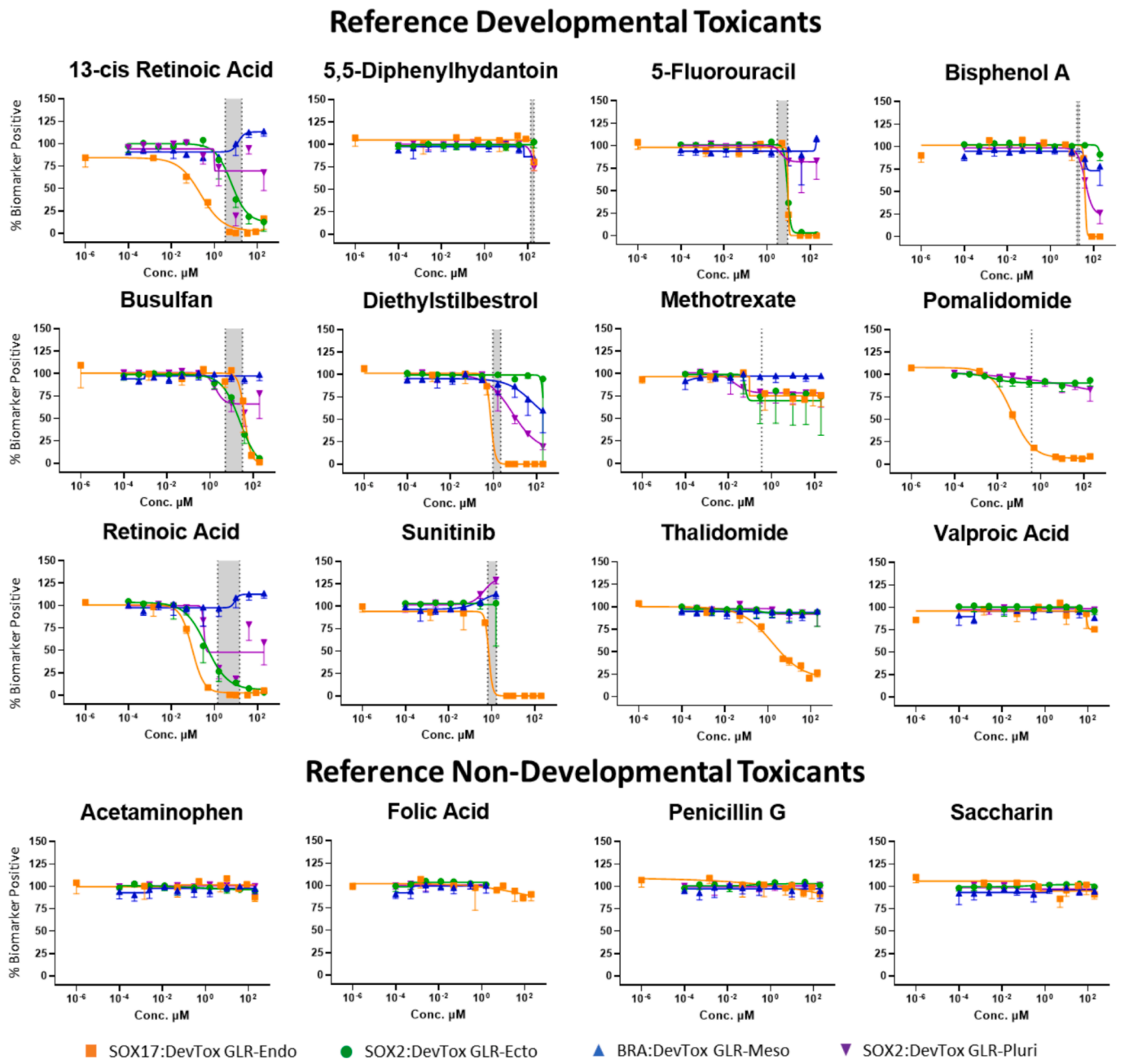


Fig. 2. Reference chemical testing in the DevTox GLR Platform Assays. Chemical exposures for 48 h in the DevTox GLR-Endo (Orange), DevTox GLR-Ecto (Green), DevTox GLR-Meso (Blue), and DevTox GLR-Pluri (Purple) assays. Dotted lines with gray shaded rectangle indicate the cytotoxic threshold range across all assays. Data points are the median value of all experimental replicates (n = 4). Error bars are the 95 % confidence interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Comparative analysis of reference chemical potency and efficacy was conducted across all four assays (Fig. 3). Overall, the DevTox GLR-Endo assay exhibited the highest efficacy for chemicals classified as developmental toxicants with 9/12 inducing > 75 % maximum inhibition. The assay also yielded the lowest activity concentration at the assay cutoff (ACC) value, an assay-driven point of departure, for 5/12 developmental toxicants, in contrast to the DevTox GLR-Pluri (4/12), DevTox GLR-Ecto (1/12), and DevTox GLR-Meso (0/12) assays.

Chemical training set evaluation

With the DevTox GLR-Pluri assay providing similar results to those observed in the DevTox GLR-Endo assay, a larger training set of chemicals (28 developmental and 19 non-developmental toxicants) were screened in the DevTox GLR-Pluri assay (Fig. 4). Training set chemicals were previously evaluated in the DevTox GLR-Endo assay and establish a comparative benchmark between the two assays (Gamble et al., 2022). Training set performance for the DevTox GLR-Pluri assay displayed a specificity of 57.9 %, sensitivity of 67.8 %, and balanced accuracy of 62.9 %. This compared to 68.4 % specificity, 71.4 % sensitivity, and 70.0 % balanced accuracy for the same chemicals in the DevTox GLR-Endo assay. When comparing all chemicals tested, reference and training set, the DevTox GLR-Pluri assay had a specificity of 65.2 %, sensitivity of 72.5 %, and balanced accuracy of 68.9 % which was less than DevTox GLR-Endo performance of 73.9 %, 75.0 %, and 74.5 % respectively (Table S1). Both assays exhibited false positive activity for ketanserin, tegaserod, retinol, esomeprazole, butylparaben, and catechin despite being categorized as non-developmental toxicants. The non-developmental toxicants methyldopa and sulfasalazine were false positives in the DevTox GLR-Pluri but not the DevTox GLR-Endo assay.

Discussion

Maternal chemical exposures are a concern for early embryonic and fetal development. The causative factors are largely unknown in the majority of cases, but environmental chemical exposures are deemed to be a significant contributing component. With over 33,000 non-confidential chemicals designated as “active” on the TSCA Chemical Substance Inventory in the United States (https://comptox.epa.gov/dashboard/chemical-lists/TSCA_ACTIVE_NCTI_0320. Accessed May 3, 2024), rapid identification of human-relevant developmental toxicants is a critical concern for chemical risk assessment. The DevTox GLR model platform has the capacity to evaluate chemical effects on stem cells maintained in a pluripotent state or differentiated toward the three germ layer lineages. Despite these conceptual capabilities, it is unclear if one or more assay endpoints is necessary and sufficient to screen chemicals for developmental hazard identification.

Assessment of all four assays within the DevTox GLR platform indicated the DevTox GLR-Endo assay demonstrated robust and reliable technical performance, frequently exhibited the highest efficacy amongst the chemicals tested in this study, and had the best overall predictive accuracy for compound screening among the DevTox GLR assays. Improvements to the technical image analysis required reexamination of the previously published assay dataset. Consistent with prior observations, the assay maintained suitable screening quality (rZ’ = 0.80), but had higher inter-experimental precision (rCV = 1.1 %) and increased dynamic range (rS/B = 101.9). The balanced accuracy increased for the reference set (91.7 %) (Table 3) and the expanded training set (74.5 %) (Supplementary Table S1) maintaining suitability for primary screening of data-poor chemical sets.

The DevTox GLR-Pluri assay matched the DevTox GLR-Endo assay in predictivity with the reference chemical set but did not correctly identify select developmental toxicants (rifampicin and warfarin) and non-developmental toxicants (sulfasalazine and methyldopa) in the training set, leading to lower sensitivity and specificity in the assay (Fig. 4, Supplementary Table S1). For chemicals that were correctly identified as developmental toxicants, the two assays differed in potency with concentration differences greater than 1.5 $\log_2$ fold-change for 14

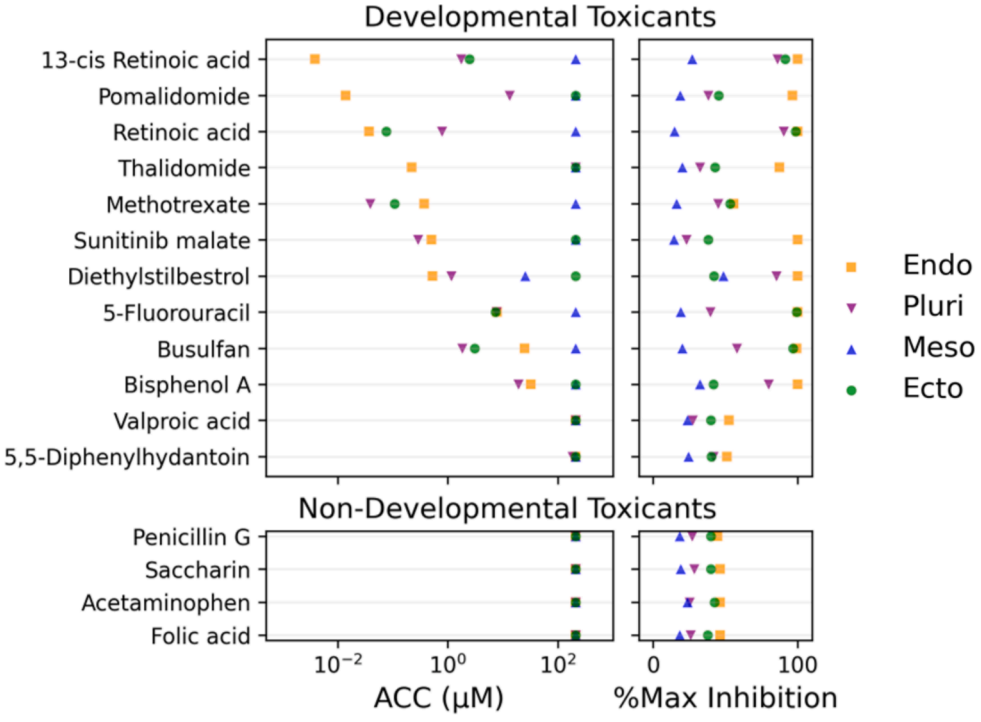


Fig. 3. Rank order plot of potency and maximum inhibition for reference chemicals across all the DevTox GLR Platform assays. Activity concentration at cutoff (ACC), left panel, and corresponding maximum effect size (% Max Inhibition), right panel. DevTox GLR-Endo (orange square), DevTox GLR-Pluri (purple point-side down triangle), DevTox GLR-Meso (blue point-side up triangle), DevTox GLR-Ecto (green circle). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

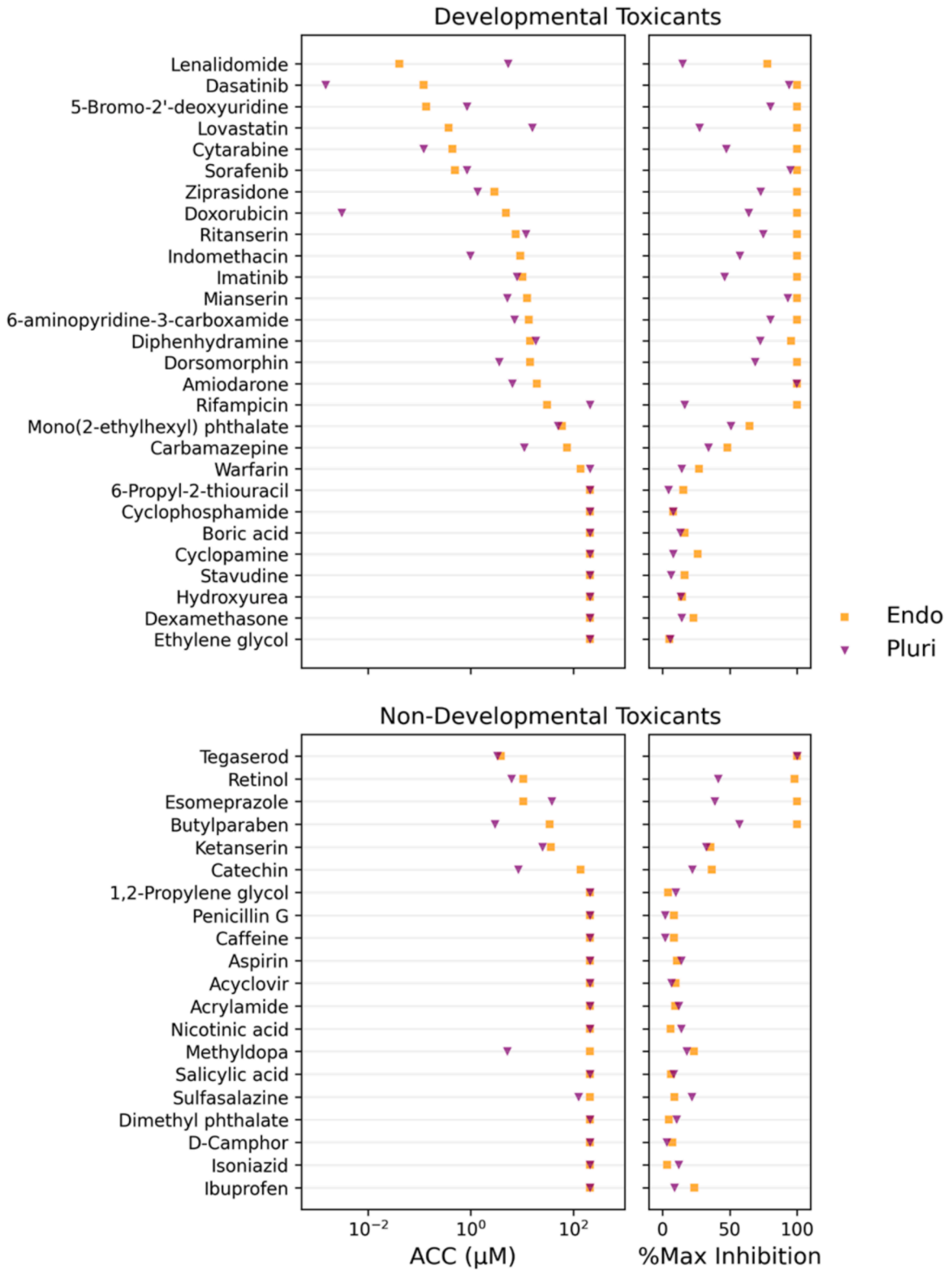


Fig. 4. Comparison of chemical responses between the DevTox GLR-Pluri and DevTox GLR-Endo assays for the chemical training set. Developmental toxicants and non-developmental toxicants are ranked by DevTox GLR-Endo potency (left panels; activity concentration at cutoff (ACC)) with corresponding maximum effect size (right panels; % Max Inhibition) for percent SOX17 positive cells (DevTox – orange square) and percent SOX2 positive cells (Pluri – purple triangle). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

chemicals. The DevTox GLR-Endo assay exhibited greater potency for six developmental toxicants including retinoids (13-cis retinoic and all-trans retinoic acid) and chemicals reported to inhibit endoderm differentiation (thalidomide, lenalidomide, and pomalidomide) (Belair et al., 2020). Increased potency in the DevTox GLR-Pluri assay was noted for chemotherapeutic drugs (doxorubicin, methotrexate, and cytarabine) and seizure medications (carbamazepine, and 5,5-diphenylhydantoin). When discriminating developmental toxicity bioactivity from cytotoxic

effects, the DevTox GLR-Pluri SOX2 biomarker was much more sensitive to chemical exposure with 19/28 true positive chemicals demonstrating bioactivity below the cytotoxicity point of departure. This, in part, was due to considerably higher variability in cell counting relative to the DevTox GLR-Endo assay. Pluripotent stem cells form dense colonies in culture, making it more difficult to identify single cells, leading to greater variability in overall counts and higher thresholds for activity determination. Acquiring more imaging fields per well could improve cell counting variability. The DevTox GLR-Pluri assay performed well with excellent reproducibility. However, most chemicals were not as effective at suppressing the SOX2 population compared to SOX17 inhibition in the DevTox GLR-Endo assay. Only five chemicals induced 90 % or greater inhibition in the DevTox GLR-Pluri assay relative to 28 chemicals inhibiting SOX17 to the same magnitude in the DevTox GLR-Endo assay.

The balanced accuracy in the devTOX *quick*Predict (83 %) assay was previously reported to be similar to the DevTox GLR-Endo (76 %) assay for a common benchmark set of 40 chemicals (Gamble et al., 2022; Zurlinden et al., 2020). The DevTOX *quick*Predict assay measures changes in the ornithine/cystine (o/c) ratio in the medium after 72 h of chemical exposure in human pluripotent stem cells to predict potential developmental toxicants (Palmer et al., 2013; Simms et al., 2020). By evaluating chemical effects on the pluripotent state over a 48-hour exposure duration, the DevTox GLR-Pluri assay provides a similar model for comparison. Considering the responses among a common subset of chemicals (25 developmental toxicants, 14 non-developmental toxicants), the DevTOX *quick*Predict assay outperforms the DevTox GLR-Pluri in sensitivity (64 % to 52 %), specificity (100 % to 79 %), and balanced accuracy (82 % to 65 %) (Table S2). Contrary to the DevTox GLR-Pluri assay, the DevTOX *quick*Predict assay correctly identified thalidomide, rifampicin, stavudine, dexamethasone, hydroxyurea, and valproic acid as developmental toxicants. Likewise, sulfasalazine, retinol, and butylparaben were identified as non-developmental toxicants. Interestingly, the DevTox GLR-Pluri assay correctly identified diethylstibestrol, 5,5-diphenylhydantoin, and bisphenol A as developmental toxicants whereas the DevTOX *quick*Predict did not. Possible explanations for the discordant observations include differences in compound exposure frequency and duration, biological variation between pluripotent stem cell lines, mechanistic differences in the measured biomarkers, as well as data modeling and analysis.

The DevTox GLR-Ecto and −Pluri assays utilized the same biomarker (SOX2) to measure perturbations. However, SOX2 protein levels were approximately two-fold higher in the ectoderm lineage at 48 h post-differentiation relative to the pluripotent state. Expression of SOX2 increases in early ectoderm differentiation and thus enables discrimination of early ectoderm cells from pluripotent cells (Bergsland et al., 2011; Collignon et al., 1996; Ferri et al., 2004) (Fig. 1). There were some unique observations in the DevTox GLR-Ecto assay that were not present across the other assays. For example, the retinoids (all-trans and 13-cis retinoic acids) induced proliferation at sub-cytotoxic concentrations. The synthetic non-steroidal estrogen, diethylstilbestrol, was another example where cytotoxicity was observed in all lineages except for ectoderm. These results reflect other *in vitro* cell-based models (Kim et al., 2013) where mesodermal and endodermal lineages were susceptible to estrogenic effects with little impact on ectoderm, indicating the estrogen receptor pathway may not be a relevant target in the DevTox GLR-Ecto assay.

The DevTox GLR-Meso assay performed the poorest with the lowest sensitivity. It only correctly identified one developmental toxicant, diethylstilbestrol, in the reference chemical set. However, eight of the 12 developmental toxicants caused cytotoxic effects in the differentiating mesoderm cells, suggesting BRA expression was not responsive to the chemicals tested (Supplementary TCPL data.zip). Additionally, although BRA expression was significantly increased from pluripotency at 48 h, expression continued to increase with more prolonged differentiation periods (Fig. S3). A longer differentiation period may be necessary to facilitate more mature mesoderm development and accurately identify chemical-mediated perturbations. Ultimately, these results suggest that BRA may not be an effective biomarker at the 48-hour time point to measure developmental perturbations. Likewise, lifestage susceptibility to the selected chemicals in this study, such as those impacting fetal development at later stages, may not adequately be captured by the relatively simple gastrulation simulations in the DevTox GLR assays.

In conclusion, the DevTox GLR model platform was designed as a human-relevant model for high-throughput screening of potential developmental toxicants. By profiling the performance of each lineage-specific assay (DevTox GLR-Endo, −Meso, −Ecto, −Pluri), an optimal default assay for routine screening could be identified. The DevTox GLR-Meso assay was incapable of adequately identifying developmental toxicants. The DevTox GLR-Ecto and −Pluri assays provided additional mechanistic coverage, but fell short in terms of overall predictivity. The DevTox GLR-Endo assay maintained robust and reliable technical performance, exhibited the highest efficacy amongst most test chemicals, and had the best overall predictive accuracy for compound screening which positions it as the default assay for future screening efforts.

Disclaimer

The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Funding

U.S. Environmental Protection Agency.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank Dr. Sreenivasa Ramaiahgari and Dr. Drake Phelps for technical review and critical reading of the manuscript, Madison Feshuk for computational support with the ToxCast data pipeline analysis and annotation, and Dr. Ali Brivanlou from The Rockefeller University for generously providing the RUES2-GLR cell line.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crtox.2025.100223.

Data availability

Data will be accessible on the US EPA CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard/)

References

Belair, D.G., Lu, G., Waller, L.E., Gustin, J.A., Collins, N.D., Kolaja, K.L., 2020. Thalidomide Inhibits Human iPSC Mesendoderm Differentiation by Modulating CRBN-dependent Degradation of SALL4. *Sci. Rep.* 10 (1), 2864. https://doi.org/10.1038/s41598-020-59542-x.

Bergsland, M., Ramsköld, D., Zaouter, C., Klum, S., Sandberg, R., Muhr, J., 2011. Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev.* 25 (23), 2453–2464. https://doi.org/10.1101/gad.176008.111.

Collignon, J., Sockanathan, S., Hacker, A., Cohen-Tannoudji, M., Norris, D., Rastan, S., Stevanovic, M., Goodfellow, P.N., Lovell-Badge, R., 1996. A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. *Development* 122 (2), 509–520. https://doi.org/10.1242/dev.122.2.509.

Dix, D.J., Houck, K.A., Martin, M.T., Richard, A.M., Setzer, R.W., Kavlock, R.J., 2006. The toxcast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci.* 95 (1), 5–12. https://doi.org/10.1093/toxsci/kfl103.

Estevan, C., Pamies, D., Vilanova, E., Sogorb, M.A., 2017. Chapter 9 - OECD Guidelines for In Vivo Testing of Reproductive Toxicity. In: Gupta, R.C. (Ed.), *Reproductive and Developmental Toxicology*, Second Edition. Academic Press, pp. 163–178. https://doi.org/10.1016/B978-0-12-804239-7.00009-3.

Ferri, A.L.M., Cavallaro, M., Braida, D., Di Cristofano, A., Canta, A., Vezzani, A., Ottolenghi, S., Pandolfi, P.P., Sala, M., DeBiasi, S., Nicolis, S.K., 2004. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development* 131 (15), 3805–3819. https://doi.org/10.1242/dev.01204.

Feshuk, M., Kolaczkowski, L., Dunham, K., Davidson-Fritz, S.E., Carstens, K.E., Brown, J., Judson, R.S., Paul Friedman, K., 2023. The ToxCast pipeline: updates to curve-fitting approaches and database structure [Original Research]. *Front. Toxicol.* 5. https://doi.org/10.3389/ftox.2023.1275980.

Gamble, J.T., Hopperstad, K., Deisenroth, C., 2022. The DevTox germ layer reporter platform: an assay adaptation of the human pluripotent stem cell test. *Toxics* 10 (7). https://doi.org/10.3390/toxics10070392.

Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9 (3), 90–95. https://doi.org/10.1109/MCSE.2007.55.

Judson, R., Houck, K., Martin, M., Knudsen, T., Thomas, R.S., Sipes, N., Shah, I., Wambaugh, J., Crofton, K., 2014. In vitro and modelling approaches to risk assessment from the U.S. Environmental Protection Agency ToxCast programme. *Basic Clin. Paharmacol. Toxicol.* 115 (1), 69–76. https://doi.org/10.1111/bcpt.12239.

Kameoka, S., Babiarz, J., Kolaja, K., Chiao, E., 2013. A high-throughput screen for teratogens using human pluripotent stem cells. *Toxicol. Sci.* 137 (1), 76–90. https://doi.org/10.1093/toxsci/kft239.

Kavlock, R., Chandler, K., Houck, K., Hunter, S., Judson, R., Kleinstreuer, N., Knudsen, T., Martin, M., Padilla, S., Reif, D., Richard, A., Rotroff, D., Sipes, N., Dix, D., 2012. Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. *Chem. Res. Toxicol.* 25 (7), 1287–1302. https://doi.org/10.1021/tx3000939.

Kim, H., Kim, Y.Y., Ku, S.Y., Kim, S.H., Choi, Y.M., Moon, S.Y., 2013. The effect of estrogen compounds on human embryoid bodies. *Reprod. Sci.* 20 (6), 661–669. https://doi.org/10.1177/1933719112462630.

Martyn, I., Kanno, T.Y., Ruzo, A., Siggia, E.D., Brivanlou, A.H., 2018. Self-organization of a human organizer by combined Wnt and Nodal signalling. *Nature* 558 (7708), 132–135. https://doi.org/10.1038/s41586-018-0150-y.

National Research Council. (2000). *Scientific Frontiers in Developmental Toxicology and Risk Assessment*. The National Academies Press. Doi: 10.17226/9871.

National Research Council. (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. The National Academies Press. Doi: 10.17226/11970.

Palmer, J.A., Smith, A.M., Egnash, L.A., Conard, K.R., West, P.R., Burrier, R.E., Donley, E. L., Kirchner, F.R., 2013. Establishment and assessment of a new human embryonic stem cell-based biomarker assay for developmental toxicity screening. *Birth Defects Res. B* 98 (4), 343–363. https://doi.org/10.1002/bdrb.21078.

Richard, A.M., Judson, R.S., Houck, K.A., Grulke, C.M., Volarath, P., Thillainadarajah, I., Yang, C., Rathman, J., Martin, M.T., Wambaugh, J.F., Knudsen, T.B., Kancherla, J., Mansouri, K., Patlewicz, G., Williams, A.J., Little, S.B., Crofton, K.M., Thomas, R.S., 2016. ToxCast chemical landscape: paving the road to 21st century toxicology. *Chem. Res. Toxicol.* 29 (8), 1225–1251. https://doi.org/10.1021/acs.chemrestox.6b00135.

Scialli, A.R., Daston, G., Chen, C., Coder, P.S., Euling, S.Y., Foreman, J., Hoberman, A.M., Hui, J., Knudsen, T., Makris, S.L., Morford, L., Piersma, A.H., Stanislaus, D., Thompson, K.E., 2018. Rethinking developmental toxicity testing: Evolution or revolution? *Birth Defects Res.* 110 (10), 840–850. https://doi.org/10.1002/bdr2.1212.

Simms, L., Rudd, K., Palmer, J., Czekala, L., Yu, F., Chapman, F., Trelles Sticken, E., Wieczorek, R., Bode, L.M., Stevenson, M., Walele, T., 2020. 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.* 1, 161–173. https://doi.org/10.1016/j.crtox.2020.11.001.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., Jones, J.M., 1998. Embryonic stem cell lines derived from human blastocysts. *Science* 282 (5391), 1145–1147. https://doi.org/10.1126/science.282.5391.1145.

Tsankov, A.M., Akopian, V., Pop, R., Chetty, S., Gifford, C.A., Daheron, L., Tsankova, N. M., Meissner, A., 2015. A qPCR ScoreCard quantifies the differentiation potential of human pluripotent stem cells. *Nat. Biotechnol.* 33 (11), 1182–1192. https://doi.org/10.1038/nbt.3387.

Van Rossum, G.a.D., Fred, L., 2009. Python 3 reference manual. *CreateSpace*.

Zhang, J.H., Chung, T.D., Oldenburg, K.R., 1999. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen.* 4 (2), 67–73. https://doi.org/10.1177/108705719900400206.

Zhang, X.D., 2011. Illustration of SSMD, z Score, SSMD*, z* Score, and t statistic for hit selection in RNAi high-throughput screens. *J. Biomol. Screen.* 16 (7), 775–785. https://doi.org/10.1177/1087057111405851.

Zurlinden, T.J., Saili, K.S., Rush, N., Kothiya, P., Judson, R.S., Houck, K.A., Hunter, E.S., Baker, N.C., Palmer, J.A., Thomas, R.S., Knudsen, T.B., 2020. Profiling the ToxCast library with a pluripotent human (H9) stem cell line-based biomarker assay for developmental toxicity. *Toxicol. Sci.* 174 (2), 189–209. https://doi.org/10.1093/toxsci/kfaa014.