CCTE_Deisenroth_DEVTOX-GLR_Pluri_CellCount

1. General Information

1.1 Assay Title: CCTE's DevTox GLR-Pluri Assay Evaluation of Total Cell Counts, Deisenroth Lab

1.2 Assay Summary: CCTE_Deisenroth_DEVTOX-GLR_Pluri assay is a cell-based, multiplex-readout assay that uses RUES2-GLR, a human pluripotent cell line, with measurements taken at 48 hours after chemical dosing in a 384-well microplate CCTE_Deisenroth_DEVTOX-GLR_Pluri_CellCount is 1 of 4 assay components calculated from the CCTE_Deisenroth_DEVTOX-GLR_Pluri assay and it measures cell count. Using fluorescent nuclei staining, gain-of-signal activity can be used to understand changes in cell viability with fluorescent microscopy. It is designed to make measurements of nuclei counts as a surrogate for cell number, as detected with optical fluorescence microscopy by Perkin Elmer Harmony nuclei detection algorithm. To generalize the intended target to other relatable targets, this assay component is annotated to the cell cycle intended target family, where the subfamily is cytotoxicity. Data from the assay component CCTE_Deisenroth_DEVTOX-GLR_Pluri_CellCount was analyzed as 1 assay endpoint. The endpoint the percent change in the cell count relative to the 0.2% DMSO control in the CCTE_Deisenroth_DEVTOX-GLR_Pluri assay. Furthermore, this assay endpoint can be referred to as a primary readout, because this assay has produced multiple assay endpoints where this one serves as a measure of cytotoxicity.

1.3 Date of Document Creation: September 05 2024

1.4 Authors and Contact Information:

US Environmental Protection Agency (EPA) Center for Computational Toxicology and Exposure (CCTE)
109 T.W. Alexander Drive (Mail Code D143-02)
Research Triangle Park, NC 27711

1.5 Assay Source: The EPA Center for Computational Toxicology and Exposure labs focus on developing and implementing in vitro methods to identify potential environmental toxicants. Principal investigators include Steve Simmons, Joshua Harrill, and Chad Deisenroth.

1.6 Date of Assay Development: For date of assay development, see *Section 6: Bibliography*.

1.7 References: For complete list of references, see *Section 6: Bibliography*.

1.8 Proprietary Elements: The human pluripotent stem cell test (hPST) is an in vitro test method for rapidly identifying potential human developmental toxicants that employs directed differentiation of embryonic stem cells to measure reductions in SOX17 biomarker expression and nuclear localization1. The multi-lineage DevTox GLR platform expands on the hPST principles by utilizing a transgenic pluripotent stem cell line expressing fluorescent reporter fusion protein biomarkers for SOX17 (endoderm marker), BRA (mesoderm marker), and SOX2 (ectoderm and pluripotency marker). Initial assay adaption to definitive endoderm (DevTox GLR-Endo) was performed to emulate the hPST SOX17 endpoint and enable comparative evaluation of concordant chemical effects. Assay duration was reduced to two days and screening throughput scaled to 384-well format for enhanced speed and efficiency. Assay performance for 66 chemicals derived from reference and training set data resulted in a balanced accuracy of 72% (79% sensitivity and 65% specificity). The DevTox GLR-Endo assay demonstrated successful adaptation of the hPST concept with increased throughput, shorter assay duration, and minimal endpoint processing (Gamble 2022). Method development and performance evaluation has also been completed for the DevTox GLR-Ecto (Ectoderm), DevTox GLR-Meso (Mesoderm), and DevTox GLR-Pluri (Pluripotent) assays.

1.9 Assay Throughput: 384-well plate.

1.10 Status: The assay is fully developed, and data are publicly available in ToxCast's invitroDB.

1.11 Abbreviations:

AIC: Akaike Information Criterion
AOP: Adverse Outcome Pathway
CV: Coefficient of Variation
DMSO: Dimethyl Sulfoxide
ToxCast: US EPA's Toxicity Forecaster Program
tcpl: ToxCast Data Analysis Pipeline R Package
SSMD: Strictly Standardized Mean Difference

2. Test Method Description

2.1 Purpose: CCTE_Deisenroth_DEVTOX-GLR_Pluri_CellCount was designed to measure changes in the number of cells where a decrease is indicative of cytotoxicity.

The DevTox GLR model 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 that may be indicative of developmental toxicity.

2.2 Scientific Principles: Birth defects impact approximately 3% of births in the United States annually and are a major contributor to infant morbidity and mortality (Yoon 1997, Hoyert 2006). The majority of developmental anomalies are of unknown etiology but there is increasing evidence that exposure to certain environmental chemicals is a contributing factor (Stillerman 2008). Identifying prenatal developmental toxicants is challenging since target effects can span critical stages of fetal development (e.g. conception through organogenesis), resulting in adverse outcomes including low birth weight, congenital defects, functional deficits, and pregnancy loss (Stillerman 2008). Further complicating hazard identification are fetal-maternal interactions where altered chemical toxicokinetic and toxicodynamic parameters can be ascribed to xenobiotic metabolism (Fantel 1982, Webster 2002), placental transport functions (Caserta 2013, Birks 2016, Grindler 2018), or general adverse effects on maternal physiology. The DevTox GLR model platform has the capacity to rapidly screen and identify potential chemical hazards during the gastrulation phase of early embryogenesis.

2.3 Experimental System: adherent RUES2-GLR cell-based used. The platform utilizes the RUES2-GLR (Rockefeller University Embryonic Stem cell line 2 – Germ Layer Reporter) 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 (Martyn 2018).

2.4 Metabolic Competence: Xenobiotic biotransformation potential has not been characterized.

2.5 Exposure Regime: All assay protocols are aligned to the DevTox GLR-Endo assay workflow2. Cells are seeded into 384-well microplates for 24 hours prior to initiating differentiation in defined media for a total duration of 48 hours. A total of two chemical exposures are administered in 24-hour intervals during the 48-hour differentiation period. At the assay termination point, cells are fixed and stained with a fluorescent dye demarcating the nucleus. High-content imaging is used to acquire brightfield and fluorescent channel images across 5 fields per well. Cell identification and mean biomarker fluorescence intensities are calculated with image analysis software to determine the total cell counts for viability assessment and the biomarker expression frequency for each lineage-dependent assay. A typical experimental design tests a 10-point concentration series (1 nM – 200 uM) for 28 compounds in single technical replicate. Assay plate controls include non-differentiated pluripotent cells (baseline control), lineage-specific differentiated cells (differentiation control), and DMSO solvent exposed differentiated cells (solvent control). Experiments typically comprise a minimum of four experimental replicates (n=4).

ASSAY DESIGN SUMMARY	
Nominal number of tested concentrations: 10	Target (nominal) number of replicates: 5
Standard minimum concentration tested: 0.000998 µM	Standard maximum concentration tested: 200 µM
Key positive control: NA	Neutral vehicle control: DMSO
Baseline median absolute deviation for the assay (bmad): 15.862	
Response cutoff threshold used to determine hit calls: 47.587	
Detection technology used: Perkin Elmer Harmony nuclei detection algorthm (microscopy)	

2.6 Response: The DevTox GLR-Pluri assay evaluates SOX2 protein expression as a biomarker of the pluripotent state that precedes germ layer development during the gastrulation phase of embryogenesis. Well-level biomarker expression frequency is measured as the 'percent responders'. Well-level 'cell counts', reflected by the sum of total counted nuclei from all imaged fields, are used to calculate the 'percent responders' for each respective biomarker and evaluate cytotoxicity.

2.7 Quality and Acceptance Criteria: Each assay may utilize different acceptance criteria and quality assurance methods as it pertains to the individual assay platform and implementation. Pre-processing transformations may indicate issues in plates or wells by setting well quality (wllq) values to 0. Analytical QC calls per sample and substance can also be considered to understand the applicability domain of the chemicals for screening.

2.8 Technical Limitations: ToxCast data can provide initial (screening) information about the capacity for a chemical to illicit a biological response; caution is advised with extrapolation of these results to organism-level responses. The potential for a chemical to elicit adverse health outcomes in living systems is a function of multiple factors, and this assay is not intended to provide predictive details regarding long term or indirect adverse effects in complex biological systems but can aid in the prioritization of compound selection for more resource intensive toxicity studies. See *Section 4.4*. for more information on the chemical applicability of the assay.

2.9 Related Assays: For related assays, consult the following assay lists or intended target families. This assay is present in the following assay lists:

Developmental Toxicity: Assays associated with developmental toxicity

Additionally, this assay was annotated to the intended target family of cell cycle.

3. Data Interpretation

The ToxCast Data Analysis Pipeline (tcpl) R package includes processing functionality for two screening paradigms: (1) single-concentration ("SC") and (2) multiple-concentration ("MC") screening. SC screening consists of testing chemicals at one concentration, often for the purpose of identifying potentially active chemicals to test in the multiple-concentration format. MC screening consists of testing chemicals across a concentration range, such that the modeled activity can give an estimate of potency, efficacy, etc. MC data is the focus of this documentation, with SC data processing metrics to be incorporated in the future.

3.1 Responses captured in prediction model: See *Section 2.6* for additional information on responses measured.

3.2 Data Analysis: Concentration-response modeling of the 'percent responders' and 'cell counts' endpoints is used to derive quantitative potency and efficacy values. Normalized values reflect the percent response of the lineage-specific differentiated cells (differentiation control) set at 100 percent.

Prior to the data processing, all the data must go through pre-processing to transform the heterogeneous data into a uniform format before it can be loaded into a database. Level 0 pre-processing is done in R by vendor/dataset-specific scripts with all manual transformations to the data documented with justification. Common examples of manual transformations include fixing a sample ID typo or changing well quality value(wllq) to 0 after identifying problems such a plate row/column missing an assay reagent.

Once data is loaded into the database, tcpl utilizes generalized processing functions provided to process, normalize, model, qualify, and visualize the data. To promote reproducibility, all method assignments must occur through the database and should come from the available list of methods for each processing level. Assigned multiple concentration processing methods include:

Level 2: Component-specific corrections include:

21: agg.median.rep.apid (Aggregate technical replicates by taking the plate-wise median per sample id and concentration index.)

Level 3: Endpoint-specific normalization include:

5: resp.pc (Calculate the normalized response (resp) as a percent of control, i.e. the ratio of the difference between the corrected (cval) and baseline (bval) values divided the difference between the positive control (pval) and baseline (bval) values multiplied by 100; resp = (cval-bval)/(pval-bval)*100.), 6: resp.multneg1 (Multiply the normalized response value (resp) by -1; -1*resp.), 11: bval.apid.nwlls.med

(Calculate the baseline value (bval) as the plate-wise median, by assay plate ID (apid), of the corrected values (cval) for neutral control wells (wllt = n).), 32: pval.zero (Set the positive control value (pval) to 0; pval = 0.)

Level 4: Baseline and required tcplFit2 parameters defined by:

2: bmad.aeid.lowconc.nwells (Calculate the baseline median absolute value (bmad) as the median absolute deviation of normalized response values (resp) for neutral control wells (wllt = n). Calculate one standard deviation of the normalized response for neutral control wells (wllt = n); onesd = sqrt(sum((resp - mean resp)^2)/sample size - 1). Onesd is used to establish BMR and therefore required for tcplfit2 processing.)

Level 5: Possible cutoff thresholds, where higher value for endpoint is selected, include:

1: bmad3 (Add a cutoff value of 3 multiplied by the baseline median absolute deviation (bmad) as defined at Level 4.)

Level 6: Cautionary flagging include:

5: modl.directionality.fail (Flag series if model directionality is questionable, i.e. if the winning model direction was opposite, more responses (resp) would have exceeded the cutoff (coff). If loss was winning directionality (top < 0), flag if count(resp < -1*coff) < 2*count(resp > coff). If gain was winning directionality (top > 0), flag if count(resp > coff) < 2*count(resp < -1*coff).), 6: singlept.hit.high (Flag single-point hit that's only at the highest conc tested, where series is an active hit call (hitc >= 0.9) with the median response observed above baseline occurring only at the highest tested concentration tested.), 7: singlept.hit.mid (Flag single-point hit that's not at the highest conc tested, where series is an active hit call (hitc >= 0.9) with the median response observed above baseline occurring only at one concentration and not the highest concentration tested.), 8: multipoint.neg (Flag multi-point miss, where series is an inactive hit call (hitc < 0.9) with multiple median responses observed above baseline.), 9: bmd.high (Flag series if modeled benchmark dose (BMD) is greater than AC50 (concentration at 50 percent maximal response). This is indicates high variability in baseline response in excess of more than half of the maximal response.), 10: noise (Flag series as noisy if the quality of fit as calculated by the root mean square error (rmse) for the series is greater than the cutoff (coff); rmse > coff.), 11: border (Flag series if borderline activity is suspected based on modeled top parameter (top) relative to cutoff (coff); |top| <= 1.2(coff) or |top| >= 0.8(coff).), 13: low.nrep (Flag series if the average number of replicates per concentration is less than 2; nrep < 2.), 14: low.nconc (Flag series if 4 concentrations or less were tested; nconc <= 4.), 15: gnls.lowconc (Flag series where winning model is gain-loss (gnls) and the gain AC50 is less than the minimum tested concentration, and the loss AC50 is less than the mean tested concentration.), 17: efficacy.50 (Flag low efficacy hits if series has an active hit call (hitc >= 0.9) and efficacy values (e.g. top and maximum median response) less than 50 percent; intended for biochemical assays. If hitc >= 0.9 and coff >= 5, then flag when top < 50 or max_med < 50. If hitc >= 0.9 and coff < 5, then flag when top < log2(1.5) or max_med < log2(1.5).), 18: ac50.lowconc (Flag series with an active hit call (hitc >= 0.9) if AC50 (concentration at 50 percent maximal response) is less than the lowest concentration tested;if hitc >= 0.9 and ac50 < 10^logc_min, then flag.), 19: viability.gnls (Flag series with an active hit call (hitc >= 0.9) if denoted as cell viability assay with winning model is gain-loss (gnls); if hitc >= 0.9, modl=="gnls" and cell_viability_assay == 1, then flag.), 20: no.med.gt.3bmad (Flag series where no median response values are greater than baseline as defined by 3 times the baseline median absolute deviation (bmad); nmed_gtbl_pos and nmed_gtbl_neg both = 0, where nmed_gtbl_pos/_neg is the number of medians greater than 3*bmad/less than -3*bmad.)

The following is an aggregate endpoint summary of the number of samples and chemicals tested, as well as active or inactive hit calls (hitc) and predicted winning models for all samples tested in this endpoint.

SAMPLE AND CHEMICAL COVERAGE

Number of samples tested: 65 Number of chemicals tested: 65

ACTIVITY HIT CALLS

Active hit count: hitc≥0.9	Inactive hit count: 0≤hitc<0.9	NA hit count: *hitc<0*
26	39	*0*

WINING MODEL SELECTION	
Number of sample-assay endpoints with winning *hill* model:	4
gain-loss (gnls) model:	4
power(pow) model:	6
linear-polynomial (poly1) model:	15
quadratic-polynomial(poly2) model:	9
exponential-2 (exp2) model:	1
exponential-3 (exp3) model:	0
exponential-4 (exp4) model:	17
exponential-5 (exp5) model:	9

For each concentration series, several point-of-departure (POD) estimates are calculated for the winning model. The major estimates include: (1) the activity concentration at the specified benchmark response (BMR) (bmd), (2) the activity concentration at 50% of the maximal response (ac50), (3) the activity concentration at the efficacy cutoff (acc), (4) the activity concentration at 10% of the maximal response, and (5) the concentration at 5% of the maximal response.

3.3 Prediction Model: All statistical analyses were conducted using R programming language, employing the tcpl package to generate model parameters and confidence intervals. Each chemical concentration response series is fit to ten predictive models, encoded by the dependency package tcplfit2. The models include the constant, Hill, gain-loss, two polynomials (i.e. linear and quadratic), power, and four exponential variants. The polynomials, power, and exponential models are all based on BMDExpress2. The winning model (modl) is selected based on the lowest AIC value and is used to determine the activity (or hit call) for the concentration series. If two models have equal AIC values, then the simpler model (i.e. model with fewer parameters) wins. In invitrodb, levels 4 and 5 capture model fit information. mc4 captures summary values calculated for each concentration series, whereas mc4_param stores the estimated model parameters for all models fit to data in long format. mc5 captures the winning model selected and the activity hit call, whereas mc5_param stores the estimated model parameters for the selected winning model in long format. Activity for each concentration-response series is determined by calculating a continuous hit-call for the winning model, which is the product of three proportional weights. The first weight reflects whether there is at least one median response outside the efficacy cutoff band. Second, the top (or maximal change in the predicted response) is larger than the cutoff. The last weight reflects whether the AIC of the winning model is less than the constant model, i.e. the winning model is better fit than a flat line.

The continuous hit call value (hitc), fit category (fitc), and cautionary flags (mc6) can be used to understand the goodness-of-fit, enabling the user to decide the stringency with which to filter and interpret results. Hitc may be further binarized into active or inactive, depending on the level of stringency required by the user; herein, hitc greater than or equal to 0.90 is active, hitc between 0 and 0.90 is inactive, and hitc less than 0 is not applicable, but different thresholds may be used. Fitc was summarize curve behavior relative activity, efficacy, and potency comparisons between the AC50 and the concentration range screened. Cautionary flags on fitting were developed in previous versions of tcpl and have been stored at level 6. These flags are programmatically generated and indicate characteristics of a curve that need extra attention or potential anomalies in the curve or data. Users may review these filtered groupings to understand high-confidence curves.

3.4 Software: The ToxCast Data Analysis Pipeline (tcpl) is an R package that manages, curve-fits, plots, and stores ToxCast data to populate its linked MySQL database, invitrodb. Data for invitrodb v4.2 was processed using the tcpl v3.2. See *Section 7: Supporting Information* on the ToxCast program and tcpl R package.

4. Test Method Performance

4.1 Robustness: The following assay performance metrics surmise the robustness of the method i.e. the reliability of the experimental results and the prediction capability of the model used.

NEUTRAL CONTROL (well type = "n")	
Neutral control well median response value, by plate: *nmed*	2634.5
Neutral control median absolute deviation, by plate: *nmad*	365.09
Coefficient of variation (CV%) in neutral control wells: *(nmad/nmed)*100*	17.67%
POSITIVE CONTROL (well type = "p")	
Positive control well median response value, by plate: *pmed*	3153.75
Positive control well median absolute deviation, by plate: *pmad*	431.066
Z Prime Factor for median positive and neutral control across all plates: *(1 - ((3 * (pmad + nmad)) / abs(pmed - nmed))*	NA
Strictly standardized mean difference (SSMD) for positive compared to neutral control wells: *((pmed - nmed) / sqrt($pmad^2$ + $nmad^2$)*	0.18
Positive control signal-to-noise: *((pmed-nmed)/nmad)*	NA
Positive control signal-to-background: *(pmed/nmed)*	NA
NEGATIVE CONTROL (well type = "m")	
Negative control well median, by plate: *mmed*	NA
Negative control well median absolute deviation value, by plate: *mmad*	NA
Z Prime Factor for median negative and neutral control across all plates: *(1 - ((3 * (mmad + nmad)) / abs(mmed - nmed))*	NA
Strictly standardized mean difference (SSMD) for negative compared to neutral control wells: *((mmed - nmed) / sqrt($mmad^2$ + $nmad^2$)*	NA
Signal-to-noise (median across all plates, using negative control wells): *((mmed-nmed)/nmad)*	NA
Signal-to-background (median across all plates, using negative control wells): *(mmed/nmed)*	NA

4.2 Reference Chemical Information: Reference chemical curation is ongoing, and this section will be updated as more information becomes available.

4.3 Performance Measures and Predictive Capacity: The performance and predictivity for a given assay may be evaluated with a variety of performance statistics but is dependent upon available data. Predictive capacity (i.e. false negative, false positive rates) will be assessed when reference chemical information is available. Ideally, assays will have sufficient data on reference chemicals (i.e. positive and negative controls) to enable estimation of accuracy statistics, such as sensitivity and specificity.

ToxCast targets may align to a range of event types in the Adverse Outcome Pathway (AOP) framework, however each assay technology may have specific limitations, which may require user discretion for more complex interpretations of the data.

The median root mean squared error (RMSE) across all winning models for active hits was calculated as: 9.

4.4 Chemical Library Scope and Limitations: The ToxCast Chemical Library was designed to capture a large spectrum of structurally and physicochemically diverse compounds. This chemical inventory incorporates toxicity data-rich chemicals, chemicals spanning major use-categories, and chemicals with exposure potential, including but

not limited to pesticides, antimicrobials, fragrances, green chemistry alternatives, food additives, toxicity reference compounds and failed pharmaceuticals. In addition to environmental or exposure concerns, chemical selection criteria also consider practical constraints, such as commercial availability, dimethyl sulfoxide (DMSO) solubility and stability, and suitability for testing in automated or semi-automated systems (e.g., low volatility and moderate LogP values). Under these constraints, there were three major, interrelated drivers for chemical selection: availability of animal toxicity data or mechanistic knowledge, exposure potential, and EPA regulatory interest. The first driver would provide the necessary *in vivo* and mechanistic data to anchor and validate subsequent prediction modeling efforts, whereas the latter two were intended to provide coverage of the chemical landscape to which humans and ecosystems are potentially exposed and for which toxicity data are mostly lacking. Analytical QC calls per sample and substance should be considered to understand the applicability domain of the chemicals for screening.

5. Potential Regulatory Applications

5.1 Context of Use: Examples of end use scenarios could include, but are not limited to:

- *Support Category Formation and Read-Across:* The outcomes from the assay could be used to substantiate a hypothesis for grouping substances together for the purposes of read-across,
- *Priority Setting:* The assay might help prioritize substances within an inventory for more detailed evaluation,
- *Screening Level Assessment of a Biomarker or Mechanistic Activity or Response:* The screening level assessment may be sufficient to identify a hazard and provide a gauge of potency; or
- *Integrated approaches to testing and assessment (IATA):* The assay may form one component of an IATA.

6. **Bibliography:** 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), 392. https://doi.org/10.3390/toxics10070392, Kameoka, S., Babiarz, J., Kolaja, K., & Chiao, E. (2014). A high-throughput screen for teratogens using human pluripotent stem cells. Toxicological sciences : an official journal of the Society of Toxicology, 137(1), 76–90. https://doi.org/10.1093/toxsci/kft239, Yoon, P. W., Olney, R. S., Khoury, M. J., Sappenfield, W. M., Chavez, G. F., & Taylor, D. (1997). Contribution of birth defects and genetic diseases to pediatric hospitalizations. A population-based study. Archives of pediatrics & adolescent medicine, 151(11), 1096–1103., Hoyert, D. L., Mathews, T. J., Menacker, F., Strobino, D. M., & Guyer, B. (2006). Annual summary of vital statistics: 2004. Pediatrics, 117(1), 168–183. https://doi.org/10.1542/peds.2005-2587, Stillerman, K. P., Mattison, D. R., Giudice, L. C., & Woodruff, T. J. (2008). Environmental exposures and adverse pregnancy outcomes: a review of the science. Reproductive sciences (Thousand Oaks, Calif.), 15(7), 631–650. https://doi.org/10.1177/1933719108322436, Fantel A. G. (1982). Culture of whole rodent embryos in teratogen screening. Teratogenesis, carcinogenesis, and mutagenesis, 2(3-4), 231–242. https://doi.org/10.1002/1520-6866(1990)2:3/4<231::aid-tcm1770020305>3.0.co;2-1, Webster, W. S., Brown-Woodman, P. D., & Ritchie, H. E. (1997). A review of the contribution of whole embryo culture to the determination of hazard and risk in teratogenicity testing. The International journal of developmental biology, 41(2), 329–335., Caserta, D., Graziano, A., Lo Monte, G., Bordi, G., & Moscarini, M. (2013). Heavy metals and placental fetal-maternal barrier: a mini-review on the major concerns. European review for medical and pharmacological sciences, 17(16), 2198–2206., Birks, L., Casas, M., Garcia, A. M., Alexander, J., Barros, H., Bergström, A., Bonde, J. P., Burdorf, A., Costet, N., Danileviciute, A., Eggesbø, M., Fernández, M. F., González-Galarzo, M. C., Regina Gražulevičienė, Hanke, W., Jaddoe, V., Kogevinas, M., Kull, I., Lertxundi, A., Melaki, V., … Vrijheid, M. (2016). Occupational Exposure to Endocrine-Disrupting Chemicals and Birth Weight and Length of Gestation: A European Meta-Analysis. Environmental health perspectives, 124(11), 1785–1793. https://doi.org/10.1289/EHP208, Grindler, N. M., Vanderlinden, L., Karthikraj, R., Kannan, K., Teal, S., Polotsky, A. J., Powell, T. L., Yang, I. V., & Jansson, T. (2018). Exposure to Phthalate, an Endocrine Disrupting Chemical, Alters the First Trimester Placental Methylome and Transcriptome in Women. Scientific reports, 8(1), 6086. https://doi.org/10.1038/s41598-018-24505-w, 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

7. **Supporting Information:**
More information on the ToxCast program can be found at: https://www.epa.gov/chemical-research/toxicity-forecasting. The most recent version of downloadable data can be found at: https://www.epa.gov/chemical-research/exploring-toxcast-data-downloadable-data. The ToxCast Data Analysis Pipeline (tcpl) R package is available on CRAN or GitHub. Check out tcpl's vignette for comprehensive documentation describing ToxCast data processing, retrieval, and interpretation.