STM_H9_OrnCyssISnorm_ratio

1. General Information

1.1 Assay Title: Stemina devTOX quickPredict H9 human embryonic stem cell (hESC)-based Assay for Ornithine (ORN):Cystine (CYSS) Ratio

1.2 Assay Summary: STM_H9_secretome utilizes undifferentiated H9 cells to measure relative changes in two metabolites, ornithine (ORN) and cystine (CYSS) in the secretome, targeting a critical drop in the ORN/CYSS ratio as a biomarker for developmental toxicity. STM_H9_OrnCyssISnorm_ratio is one of 10 components calculated in the STM_H9_secretome assay. It is the ratio value of OrnISnorm_perc to CyssISnorm_perc. Data from the assay component STM_H9_OrnCyssISnorm_ratio was analyzed at the assay endpoint STM_H9_OrnCyssISnorm_ratio in the positive analysis fitting direction relative to DMSO as the negative control and baseline of activity. Using a type of binding reporter, gain or loss-of-signal activity can be used to understand relative metabolite change.

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: Stemina Biomarker Discovery is a Contract Research Organization (CRO) that provides stem cell-based developmental toxicity screening for chemical compounds.

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: devTOX quickPredict platform is a technology trademarked and patented by Stemina Biomarker Discovery, Inc.

1.9 Assay Throughput: 96-well plate. All treatments were carried out in Matrigel-coated 96-well plates. H9 cells were plated with a seeding density of 100000 cells per well in mTeSR1 medium containing 10uM Y27632 Rho-associated kinase inhibitor (ATCC, Manassas, Virginia) to increase plating efficiency. Y27632 was removed prior to compound addition at 24h after plating.

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: Change in the ratio value of Orn release to Cyss utilization, indicative of the targeted biomarker of the assay.

The Stemina devTOX quickPredict platform is a human pluripotent stem cell-based assay that predicts the developmental toxicity potential based on changes in cellular metabolism following chemical exposure.

2.2 Scientific Principles: Profiling hESCs for their secreted metabolites has been proposed as an alternative testing platform for identifying compounds with potential developmental toxicity. Dynamic variations in metabolite abundance with functional changes in biochemical pathways and cellular metabolic response may be direct or secondary consequences of chemical exposure. the profile of intermediary metabolites and small molecules released by hESCs to their environment (secretome) could lead to identification of the extent of adverse outcome pathways in the developing embryo. The ToxCast STM platform described here provides a potency read-out of a chemical compound's exposure-based potential for developmental toxicity based on a critical imbalance in the targeted biomarker (decreased ORN/CYSS ratio detected in the H9 hESC conditioned medium).

2.3 Experimental System: monolayer hESC cell-based used. H9 cells (NIH code WA09, WiCell Research Institute, Inc, Madison, Wisconsin) were used as approved for federally funded research and selected because of their commercial availability, genetic stability (normal female karyotype), and scientific legacy (hundreds of publications). Derivation and characterization of the H9 cell line was originally reported by Thomson et al. (1998). Cells were handled as described (Palmer et al., 2013). Briefly, cells were maintained under feeder-free conditions with mTeSR1 media (StemCell Technologies, Vancouver, Canada) on Matrigel hESC-Qualified Matrix (Corning, Bedford, Massachusetts) coated 6-well plates. Cultures were incubated at 37C in a humidified atmosphere of <5% CO2. Differentiated colonies were removed daily through aspiration to maintain the undifferentiated stem cell population. Differentiation was based on visual inspection; there is typically < 5% differentiation in a culture, and only highly pure undifferentiated H9 cell populations were used for these experiments. Cultures were passaged using Versene (Life Technologies, Grand Island, New York) or ReLeSR (StemCell Technologies) at 85%–90% confluency, karyotyped approximately every 10 passages, and the absence of mycoplasma was routinely confirmed with the MycoAlert Mycoplasma Detection Kit (Lonza, Rockland, Maine). The passage number of H9 cells used over the course of this study ranged from 31 to 48; anything above passage 40 was karyotyped within 10 passages prior to use in the assay.

2.4 Metabolic Competence: hES cells are an innovative in vitro model system that is metabolically similar to embryonic epiblast cells at gastrulation. Perturbations in cellular metabolism may be indicative of teratogenicity.

2.5 Exposure Regime: H9 hESCs maintained in a pluripotent state were exposed to test compound for 72h with chemical replenishment with media replacement every 24h. Cell-conditioned media from the final 24-h treatment period was collected for analysis of the targeted biomarker, and cell viability of the corresponding cell layer was assessed. The chemical library was tested in blinded fashion. Each test plate included 1uM Methotrexate (MTX; Selleck Chemicals, Houston, Texas) as a positive reference, 5 nM MTX as a negative reference, 0.1% DMSO as the neutral (vehicle) control, and sample-level media blanks. H9 cell-conditioned media from the final 24-h treatment period was collected for analysis of the targeted biomarker and cell viability was measured from the corresponding cell layer. Cell viability was measured using the CellTiter-Fluor assay (Promega, Madison, Wisconsin) based on proteolytic cleavage of a substrate to fluorescent signal proportional to the number of living cells. The cell viability Relative Fluorescence Unit (RFU) was background corrected and normalized to mean RFU of the neutral control (0.1% DMSO). The collected H9-cell-conditioned media samples were processed for targeted biomarker analysis as described (Palmer et al., 2013). Briefly, spent media samples were deproteinized (40% acetonitrile) and processed for UPLC-HRMS. Data acquisition was performed using 4 separate UPLC-HRMS systems, consisting of an Agilent 1290 Infinity LC system (Agilent Technologies) interfaced with an Agilent high-resolution mass spectrometer (models G6520A, G6520B, G6530A, and G6224A). A Waters Acquity UPLC BEH Amide column (2.1mm × 50mm, 1.7lm particle size; Waters, Milford, Massachusetts) maintained at 40C was applied for separation of metabolites using a 6.5min solvent gradient with 0.1% formic acid in water and 0.1% formic acid in acetonitrile (1.0ml/min flow rate). Data were acquired using MassHunter Acquisition software (version B 04.00, Agilent Technologies). The extracted ion chromatogram (EIC) areas for ORN, CYSS, and their respective spike-in C13-standards were determined using the Agilent MassHunter Quantitative Analysis software, version B.05.00 or newer (Agilent Technologies), and data were normalized as described in Palmer et al. (2017).

ASSAY DESIGN SUMMARY	
Nominal number of tested concentrations: 8	Target (nominal) number of replicates: 3
Standard minimum concentration tested: 300 μM	Standard maximum concentration tested: 1e+06 μM
Key positive control: Methotrexate	Neutral vehicle control: DMSO
Baseline median absolute deviation for the assay (bmad): 0.136	

Response cutoff threshold used to determine hit calls: 0.409
Detection technology used: UPLC-HRMS (Mass spectrometry)

2.6 Response: This assay measures relative changes in 2 metabolites, ornithine (ORN) and cystine (CYSS), targeting the ORN/CYSS ratio as a biomarker for developmental toxicity. Ornithine is a nonproteogenic amino acid that functions in several biochemical pathways including ammonia detoxification in the urea cycle, pyrimidine synthesis via ornithine transcarbamy-lase, and polyamine synthesis via ornithine decarboxylase. Ornithine is initially absent from the medium but released from viable cells; as such, decreased cellular release reflects general metabolic states for these pathways. Cystine is initially present in the medium and used by cells in glutathione production; as such, the change connected to decreased CYSS uptake likely reflects a change in cellular glutathione synthesis and redox balance. Although minor effects on cell viability could account for changes in cellular ORN release and/or CYSS uptake measured in the ORN/CYSS ratio, altered cell growth and/or survival are potential modes of action in developmental toxicity. As such, compounds that impact the ORN/CYSS biomarker due to minimal effects on cell viability should not be discounted because of it. Although cell viability is measured, it is not included in the prediction by the assay, which is based solely on the ORN/CYSS ratio response. Cell viability is provided as an additional endpoint to aid in the interpretation of the data.

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 metabolite.

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: All raw data and metadata were loaded into a central database for ToxCast using standard nomenclature to pipeline the data into invitrodb under for the Stemina DevToxqP assay source identifier (asid) for the ToxCast platform, designated STM_H9 (asid 14); specifically, the 2 assay identifiers (aid) STM_H9_secretome (aid 428) and STM_H9_viability (aid 437)representing data measures for the conditioned media and cell monolayer, respectively. This included identifiers for tracking each chemical sample (spid), plate identifier (apid), well position (row, column), micromolar concentration tested (dose), well quality (0=fail, 1=pass), and well type (wllt). The well type identifiers included media blank "b" lacking H9 cells, neutral control "n" 0.1% DMSO, test compound "t," negative control "m" 0.005uM MTX, and positive control "p" 1.0uM MTX. Raw data values (rval) were stored under the following assay components (acid): peak area for C13-cystine (acid 1023) and C13-ornithine (acid 1024) tracers, peak area for measured cystine (acid 1025) and cystine standardized to the C13 tracer (acid 1026), cystine normalized to the plate median value of the neutral controls (acid 1027), peak area for measured ornithine (acid 1028) and ornithine standardized to the C13 tracer (acid

1029), ornithine normalized to the plate median value of the neutral controls (acid 1030), the ORN:CYSS ratio calculated from DMSO-normalized values (acid 1031), the targeted biomarker prediction (acid 1032, which is an empty placeholder), background-corrected RFU from the CellTiter-Fluor assay (acid 1113), and cell viability normalized to mean RFU of the DMSO control (acid 1114). The corresponding assay endpoint identifiers (aeids) analyzed for the predictive model were: the decrease in media ornithine reflecting reduced cellular release (STM_H9_ornithineISnorm_perc, aeid 1688), the increase in media cystine reflecting less utilization (STM_H9_cystineISnorm_perc, aeid 1682), the ORN/CYSS ratio reflecting a decrease in the ORN:CYSS ratio as the primary biomarker (STM_H9_OrnCyssISnorm_ratio, aeid 1690), and normalized cell viability (STM_H9_NormalizedViability, aeid 1858).

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:

1: none (Use corrected response value (cval) as is; cval = cval. No additional mc2 methods needed for component-specific corrections.), 2: log2 (Transform the corrected response value (cval) to log-scale (base 2).), 3: rmneg (Exclude wells with negative corrected response values (cval) and downgrading their well quality (wllq); if cval < 0, wllq = 0.), 4: rmzero (Exclude wells with corrected response values (cval) equal to zero and downgrading their well quality (wllq); if cval = 0, wllq = 0.)

Level 3: Endpoint-specific normalization include:

1: none (Set the corrected response value (cval) as the normalized response value (resp); cval = resp. No additional mc3 methods needed for endpoint-specific normalization.)

Level 4: Baseline and required tcplFit2 parameters defined by:

1: bmad.aeid.lowconc.twells (Calculate the baseline median absolute value (bmad) as the median absolute deviation of normalized response values (rep) for test compound wells (wllt = t) with concentration index (cndx) equal to 1 or 2. Calculate one standard deviation of the normalized response for test compound wells (wllt = t) with a concentration index (cndx) of 1 or 2; 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.), 27: ow_bidirectional_loss (Multiply winning model hitcall (hitc) by -1 for models fit in the positive analysis direction. Typically used for endpoints where only negative responses are biologically relevant.)

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.)

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: 428		Number of chemicals tested: 428
ACTIVITY HIT CALLS		
Active hit count: hitc≥0.9 210	Inactive hit count: 0≤hitc<0.9 116	NA hit count: *hitc<0* *102*
WINING MODEL SELECTION		
Number of sample-assay endpoints with winning *hill* model:	33	
gain-loss (gnls) model:	18	
power(pow) model:	58	
linear-polynomial (poly1) model:	86	
quadratic-polynomial(poly2) model:	73	
exponential-2 (exp2) model:	24	
exponential-3 (exp3) model:	12	
exponential-4 (exp4) model:	59	
exponential-5 (exp5) model:	65	

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*	1
Neutral control median absolute deviation, by plate: *nmad*	0.06
Coefficient of variation (CV%) in neutral control wells: *(nmad/nmed)*100*	5.96%
POSITIVE CONTROL (well type = "p")	
Positive control well median response value, by plate: *pmed*	0.108
Positive control well median absolute deviation, by plate: *pmad*	0.004
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($pmad^2$ + $nmad^2$)*	-14.139
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*	1.011
Negative control well median absolute deviation value, by plate: *mmad*	0.038
Z Prime Factor for median negative and neutral control across all plates:	NA

*(1 - ((3 * (mmad + nmad)) / abs(mmed - nmed))*	
Strictly standardized mean difference (SSMD) for negative compared to neutral control wells: *((mmed - nmed) / sqrt(mmad^2 + nmad^2)*	0.124
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: 65.

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:** Zurlinden TJ, Saili KS, Rush N, Kothiya P, Judson RS, Houck KA, Hunter ES, Baker NC, Palmer JA, Thomas RS, Knudsen TB. Profiling the ToxCast Library With a Pluripotent Human (H9) Stem Cell Line-Based Biomarker Assay for Developmental Toxicity. Toxicol Sci. 2020 Apr 1;174(2):189-209. doi: 10.1093/toxsci/kfaa014. PMID: 32073639., 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 research. Part B, Developmental and reproductive toxicology, 98(4), 343–363. https://doi.org/10.1002/bdrb.21078, Palmer, J. A., Smith, A. M., Egnash, L. A., Colwell, M. R., Donley, E. L. R., Kirchner, F. R., & Burrier, R. E. (2017). A human induced pluripotent stem cell-based in vitro assay predicts developmental toxicity through a retinoic acid receptor-mediated pathway for a series of related retinoid analogues. Reproductive toxicology (Elmsford, N.Y.), 73, 350–361. https://doi.org/10.1016/j.reprotox.2017.07.011

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.