CLD_ABCG2_48hr

1. General Information

1.1 Assay Title: CellzDirect Gene Expression Assay for ABC Transporter (ABCG2) at 48 hours

1.2 Assay Summary: CLD_48hr is a cell-based, multiplexed-readout assay that uses hepatocyte, a human liver primary cell, with measurements taken at 48 hours after chemical dosing in a 96-well plate. CLD_ABCG2_48hr is one of 16 assay component(s) measured or calculated from the CLD_48hr assay. It is designed to make measurements of mRNA induction, a form of inducible reporter, as detected with chemiluminescence signals by Quantitative Nuclease Protection Assay (qNPA) technology. Data from the assay component CLD_ABCG2_48hr was analyzed at the endpoint, CLD_ABCG2_48hr, in the positive analysis fitting direction relative to DMSO as the negative control and baseline of activity. Using a type of inducible reporter, gain-of-signal activity can be used to understand changes in the reporter gene as they relate to the gene ABCG2. Furthermore, this assay endpoint can be referred to as a primary readout. To generalize the intended target to other relatable targets, this assay endpoint is annotated to the transporter intended target family, where subfamily is ABC transporter.

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: Formerly CellzDirect, this Contract Research Organization (CRO) is now part of the Invitrogen brand of Thermo Fisher providing cell-based in vitro assay screening services using primary hepatocytes.

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: Human hepatocyte culture and qNPA assays are developed and performed for commercial purposes by CellzDirect/Invitrogen corporation (a part of Life Technologies).

1.9 Assay Throughput: 96-well plate. Cultures of primary human hepatocytes from 20,928 wells (96-well plate format) were prepared, cultured, and harvested across 4 time points (0, 6, 24, or 48 h) for a minimum of 5 concentrations of each chemical or positive control.

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:

The CellzDirect assays used metabolically competent primary cultures of human hepatocytes to monitor induction or suppression of xenobiotic metabolizing enzyme and transporter gene expression multiple liver-relevant pathways.

2.2 Scientific Principles: This model system was used to characterize the concentration- and time-response of chemicals for changes in expression of genes regulated by nuclear receptors. Nuclear receptor-mediated regulation of gene expression represents an important hepatic response to exposure to both endogenous and exogenous substrates (Nakata et al., 2006). These receptors regulate multiple gene targets involved in absorption, metabolism, disposition, and excretion of endogenous and foreign chemicals (and metabolites). Fourteen gene targets were monitored in quantitative nuclease protection assays: six representative cytochromes P-450, four hepatic transporters, three Phase II conjugating enzymes, and one endogenous

metabolism gene involved in cholesterol synthesis. These gene targets are sentinels of five major signaling pathways: AhR, CAR, PXR, FXR, and PPARa. Besides gene expression, the relative potency and efficacy for these chemicals to modulate cellular health and enzymatic activity were assessed.

2.3 Experimental System: adherent hepatocyte primary cell used. Primary cultures of human hepatocytes were prepared from human liver tissue derived from two separate male donors (Hu776 and Hu778) and had initial viabilities of 91 and 95%, respectively, at the time of plating. Donor Hu776 was a male Caucasian, 41 yr of age, weighing 180 lb, standing 5'10" tall, who consumed 4–6 alcoholic beverages per week and occasionally chewed tobacco. Donor Hu778 was a male Caucasian, 55 yr of age, weighing 162 lb, standing 5'6" tall, with no history of alcohol or tobacco consumption. Tissue specimens used for these studies were derived from the normal margins of resected liver tissue that was resected due to the presence of metastatic colon tumors. The research was carried out in accordance with the principles of the current version of the Helsinki Declaration. Each patient whose tissue was used in this study was fully consented under an institutional review board (IRB) application approved by the individual institutions from patients undergoing liver resection surgery. Eligible patients were between the ages of 18 and 75 yr and were not restricted to any gender or ethnic grouping. All samples were collected and preserved at the participating institution and shipped to CellzDirect's facility in Durham, NC, for processing under protocols approved during the IRB application process. Hepatocytes were isolated by a modification of the two-step collagenase perfusion method described previously (LeCluyse et al., 2005). Final cell viability, prior to plating, was determined by the trypan blue exclusion test and was ≥90% in both preparations. Following isolation, hepatocytes were resuspended in Dulbecco's modified Eagle's medium (DMEM) containing 5% fetal calf serum, insulin (4 ug/ml), and DEX (1 uM) and added to 96-well plates (BioCoat, BD Biosciences, San Jose, CA) coated with a simple collagen, type I, substratum. Hepatocytes were allowed to attach for 4–6 h at 37C in a humidified culture chamber with 95% relative humidity/5% air/CO2. After attachment, culture vessels were swirled and medium containing debris and unattached cells was aspirated. Fresh ice-cold serum-free DMEM/Ham's F12 containing 50 nM DEX, 6.25 ug/ml insulin, 6.25 ug/ml transferrin, 6.25 ng/ml selenium (ITS+), and 0.25 mg/ml ECM was added to the culture vessels and immediately returned to the culture chamber. Medium was changed on a daily basis thereafter. Cultures of hepatocytes were maintained for 24–48 h prior to initiating experiments with the chemicals.

2.4 Metabolic Competence: This assay utilizes a metabolically competent, in vitro hepatocyte culture system consisting of primary cultures of human hepatocytes prepared from human liver tissue derived from two separate male donors. Primary human hepatocyte cultures are useful in vitro model systems of human liver because when cultured under appropriate conditions the hepatocytes retain liver-like functionality such as metabolism, transport, and cell signaling.

2.5 Exposure Regime: Human hepatocyte cultures were treated daily for two consecutive days with fresh dosing solutions containing appropriate concentrations of the 320 ToxCast chemicals (chemical identities were blinded to CellzDirect and other ToxCast collaborators), vehicle control (0.2% DMSO), and positive control inducers as summarized in Table 1. Due to the large number of treatment groups and chemicals examined, the ToxCast chemicals were divided into two groups (spanning two independent cultures of hepatocytes from separate donor preparations (Hu778: ToxCast plates 1 and 2; Hu776: ToxCast plates 3 and 4). Each preparation of human hepatocyte cultures was treated with respective vehicle control, medium only control, set of six positive controls/reference chemicals for the five hepatic receptor pathways, and a subset of eight of the ToxCast chemicals (indoxacarb, pyrithiobac-sodium, norflurazon, cyhalofop-butyl, methomyl, thiram, acetochlor, and propiconazole). These replicate data, coupled with internal replicates designed within the blinded ToxCast chemical library, provided additional data to evaluate interindividual differences between hepatocyte preparations and their potential impact on chemical profiles. Cell morphology and integrity were evaluated using phase-contrast microscopy as an indicator of hepatocyte cell health (Tyson & Green, 1987). Cultures for each treatment group (i.e., media, vehicle [0.2% DMSO], positive control inducers [multiple concentrations], and the ToxCast chemicals [multiple concentrations]) were observed and cell morphology was assessed relative to vehicle control cultures at each harvest time point (0, 6, 24, or 48 h). Any discernable morphological alterations such as changes in cell shape, nucleus size/shape, cytoplasmic alterations, and accumulation of vacuoles suggestive of dilated organelles and lipid droplets (Guillouzo et al., 1997) that were observed as a

consequence of chemical exposure were recorded in images captured using a Zeiss Axiovert inverted research microscope equipped with phase-contrast optics, a 3 CCD camera, and a computer with image capture and analysis software. Images were evaluated at the conclusion of the study to assess changes in cell integrity and account for the effects of chemical exposure on uncharacteristic changes in concentration-dependent gene expression profiles. The results from these determinations of apparent cytotoxicity were annotated (yes/no) alongside the corresponding qNPA data for reference. At the conclusion of each treatment period, hepatocyte cultures (96-well) were washed with 1 volume of HBSS, lysed by addition of 25 ul ArrayPlate lysis buffer (HTG, Tucson, AZ), and 70 ul/well of Denaturation Oil, denatured by incubation at 95C for 10 min, and frozen at approximately –70C until analysis by nuclease protection assay (qNPA) (Roberts et al., 2007). For qNPA analysis, cell lysates were thawed at 50C for approximately 30 min, qNPA probes were added, and samples were incubated at 95C for 10 min to begin the detection process by denaturing the target RNA, dissociating the duplexes and secondary structure hybridization. At the conclusion of the hybridization period, S1 nuclease reagent was added to each sample to digest all nonprotected nucleotides at 50C for 60–90 min. At the conclusion of the S1 nuclease digestions, all reactions were stopped by transfer of all the samples to fresh plates containing stop solution and incubated at 95C for 15 min to deactivate the enzyme, dissociate the mRNA/ DNA probe heterodimers, and hydrolyze the resulting single stranded mRNA, leaving a stoichiometric amount of single-stranded DNA nuclease protection probe, unmodified in sequence, as the only intact oligonucleotide left in the sample. Neutralization solution was subsequently applied to cooled (room temperature) plates, and samples were transferred to ArrayPlates for overnight incubations at 50C to allow probes to be captured onto programmed locations on the ArrayPlates. Half of the nucleotides comprising each nuclease protection probe are utilized for capture hybridization to the array. At the completion of the array capture of probes, plates were washed, detection linkers were hybridized to the other half of each nuclease protection probe, plates were washed again, and detection enzymes were applied. The final step in the process was the imaging of the plates with the OMIX Imaging System (HTG, Tucson AZ). The quantity of protected nuclease protection probe, and hence target mRNA in each well, was proportional to the luminescence intensity of the labeled detection oligonucleotides that bind each of the 16 spots within each well of a 96-well plate. Luminescence data were generated using the OMIX Imaging System software to generate endogenous control normalized data. These data were exported for bioinformatic analyses.

ASSAY DESIGN SUMMARY	
Nominal number of tested concentrations: 5	Target (nominal) number of replicates: 4
Standard minimum concentration tested: 0.004 µM	Standard maximum concentration tested: 40 µM
Key positive control:	Neutral vehicle control: DMSO
Baseline median absolute deviation for the assay (bmad): 0.202	
Response cutoff threshold used to determine hit calls: 1	
Detection technology used: Quantitative Nuclease Protection Assay (qNPA) (Luminescence)	

2.6 Response: This platform used quantitative Nuclease Protection Assays (qNPA) to simultaneously monitor an array of liver-relevant gene targets, primarily those regulated by nuclear receptors which serve as sentinels for key toxicant response mediated pathways; including gene products involved in absorption, metabolism, disposition and excretion of endogenous and foreign chemicals. The mRNA levels of 14 target and 2 control genes were measured: ABCB1, ABCB11, ABCG2, SLCO1B1, CYP1A1, CYP1A2, CYP2B6, CYP2C9, CYP2C19, CYP3A4, UGT1A1, GSTA2, SULT2A1, HMGCS2, and control genes ACTB, GAPDH. These genes represent 5 nuclear receptor signaling pathways: AHR, CAR, PXR, PPARa and FXR. In addition to measuring gene expression levels, visual observation of changes to cellular morphology (relative to vehicle control) are assessed at multiple time points (0, 6, 24, and 48 h) to quantify morphological alterations such as changes in cell shape, nucleus size/shape, cytoplasmic alterations, accumulation of vacuoles etc. resulting from xenobiotic exposures.

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:

NA

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

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: Data from an OMIX imager was received in comma-separated values (.csv) format with plate and well identifiers. These data were annotated with matching chemical and dosage information and compiled in a database. Foldover-control values for each respective time point were calculated for each treatment group. mRNA induction data is transformed to log2 fold induction over DMSO (vehicle control) signal and 5-point concentration response curves were generated following 48 hours of continuous exposure to test chemicals. Performance was compared to known inducers of xenobiotic response (e.g., phenobarbital and rifampicin) and response data were generated at 4 different time points during the test duration (0, 6, 24 and 48 hours).

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:

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:

35: resp.logfc (Calculate the normalized response (resp) as the fold change of logged, i.e. the difference between corrected (cval) and baseline (bval) log-scale values.), 38: bval.apid.nwllstcwllslowconc.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) or wells with a concentration index (cndx) of 1 or 2 and well type of test compound (wllt = t) or gain-of-signal control in multiple concentrations (wllt = c).)

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:

11: log2_2 (Add a cutoff value log2(2). Typically for fold change data.), 28: ow_bidirectional_gain (Multiply winning model hitcall (hitc) by -1 for models fit in the negative analysis direction. Typically used for endpoints where only positive 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.), 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: 319 Number of chemicals tested: 309

ACTIVITY HIT CALLS

Inactive hit count: 0≤hitc<0.9 NA hit count: *hitc<0*

Active hit count: hitc≥0.918	181	*120*

WINING MODEL SELECTION	
Number of sample-assay endpoints with winning *hill* model:	1
gain-loss (gnls) model:	24
power(pow) model:	14
linear-polynomial (poly1) model:	135
quadratic-polynomial(poly2) model:	67
exponential-2 (exp2) model:	1
exponential-3 (exp3) model:	1
exponential-4 (exp4) model:	69
exponential-5 (exp5) model:	5

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*	177.75
Neutral control median absolute deviation, by plate: *nmad*	26.687
Coefficient of variation (CV%) in neutral control wells: *(nmad/nmed)*100*	13.46%
POSITIVE CONTROL (well type = "p")	
Positive control well median response value, by plate: *pmed*	NA
Positive control well median absolute deviation, by plate: *pmad*	NA
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($pmad^2$ + $nmad^2$)*	NA
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: 5.

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:** Rotroff DM, Beam AL, Dix DJ, Farmer A, Freeman KM, Houck KA, Judson RS, LeCluyse EL, Martin MT, Reif DM, Ferguson SS. Xenobiotic-metabolizing enzyme and transporter gene expression in primary cultures of human hepatocytes modulated by ToxCast chemicals. J Toxicol Environ Health B Crit Rev. 2010 Feb;13(2-4):329-46. doi: 10.1080/10937404.2010.483949. PMID: 20574906.

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.