IUF_NPC5_oligodendrocyte_differentiation_120hr

1. General Information

1.1 Assay Title: Leibniz Research Institute for Environmental Medicine (IUF) 120-hour Oligodendrocyte Differentiation Assay (NPC5)

1.2 Assay Summary: IUF_NPC2-5 is a cell-based, multiplexed assay that uses hNPC, a human primary neural progenitor cells, with measurements taken at 72 or 120 hours after chemical dosing in a 96-well plate. Human neural progenitor cells (hNPC) are generated from human fetal brain cortex at gestational week (GW) 16-19. In the neural progenitor cell migration and differentiation assay (NPC2-5), hNPCâ€™s grown as spheres are plated on an extracellular matrix and migrate and differentiate out of the sphere core. The processes radial glia migration, neuronal and oligodendrocyte migration as well as neuronal differentiation, neuronal morphology and oligodendrocyte differentiation are studied using automated phase contrast imaging and automated fluorescence imaging in combination with high content image analysis. In parallel, the cell viability is assessed using an alamar blue viability assay and the cytotoxicity using a lactate dehydrogenase dependent membrane integrity assay. IUF_NPC5_oligodendrocyte_differentiation_120hr is one of 12 assay components measured from the IUF_NPC2-5 assay. Oligodendrocyte differentiation at 120 hr is determined as the number of all O4-positive cells in percent of the number of Hoechst-positive nuclei in the total neurosphere migration area (see IUF_NPC2a_ glia_migration_120hr) after 120 hr of migration/differentiation. Oligodendrocytes are automatically identified using a convolutional neural network (CNN) that was trained using manually annotated images of differentiated oligodendrocytes. Data from the IUF_NPC5_oligodendrocyte_differentiation_120hr component was analyzed at the endpoint IUF_NPC5_oligodendrocyte_differentiation_120hr in the positive analysis fitting direction relative to DMSO as the negative control and baseline of activity. Using a type of differentiation reporter, loss-of-signal activity can be used to understand developmental effects.

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: Leibniz Research Institute for Environmental Medicine (IUF) is a German research institution in Düsseldorf. The IUF research mission is the molecular prevention of environmentally induced health disorders.

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: For the source cells, Lonza holds donor consent and legal authorization that provides permission for all research use.

1.9 Assay Throughput: 96-well plate. The methods described are set up in a 96 well plate format with automated image acquisition and analysis and data evaluation. Pipetting steps such as coating of 96 well plates, compound dilutions, feeding, cell viability and cytotoxicity assay can be automated using a liquid handling system. In the fully automated set up 10 plates with 8 conditions and 5 replicates per condition can be run in one week. This results in the generation of 400 data points for each endpoint within one week (excluding all controls). The throughput is therefore estimated as medium to high.

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: Changes in the oligodendrocyte differentiation are indicative of neurodevelopment.

The human developing central nervous system is more vulnerable to the adverse effects of chemical agents than the adult brain. At present, due to the knowledge gap concerning hazard identification for human neurodevelopmental toxicity (DNT), there is an urgent need for testing and subsequently regulation of chemicals for their potential to interfere with the developing nervous system. Primary human neural progenitor cells (hNPC) cultivated as three-dimensional floating spheres are able to represent several key processes during brain development. In the NPC1 assay, hNPC are exposed to the test compound for 72 hours. Cell proliferation is assessed as an increase in sphere area using automated phase contrast imaging and as incorporation of Bromodeoxyuridine (BrdU) using a luminescence-based cell proliferation ELISA. In parallel, the cell viability is assessed using an alamar blue viability assay. In the neural progenitor cell migration and differentiation assay (NPC2-5), hNPC are plated on an extracellular matrix, and migrate and differentiate out of the sphere core. Thereby the processes radial glia migration, migration of neurons and oligodendrocytes as well as differentiation into neurons and oligodendrocytes can be studied in combination with general cell viability and cytotoxicity. Cell migration and differentiation are critical process during brain development that, if disturbed lead to alterations in brain development and causes cognitive dysfunction. Currently these processes are assessed in the OECD TG426 by neuropathological evaluation of certain brain regions as well as neurobehavioral tests.

2.2 Scientific Principles: Primary hNPC are isolated from the fetal brain and represent the process cell migration and differentiation into neurons, oligodendrocytes within this test method. The test method predicts the hazard to induce developmental neurotoxicity in the form of neurophysiological and functional changes in the developing nervous system. The NPC5 oligodendrocyte formation from fetal NPC correspond to oligodendrogenesis during the fetal phase of brain development in vivo.

2.3 Experimental System: suspension hNPC primary cell used. Primary human neural progenitor cells (hNPC) are provided as cryopreserved 3D neurospheres from Lonza, Verviers, Belgium. Material originates from the human brain cortex of different gestational ages (GW16-19). Sex is either specified or determined before the cells are used. Grown in suspension culture and under proliferative conditions (proliferation media and growth factors) these cells represent neural progenitor cell proliferation. 1x106 hNPC are obtained from Lonza (#PT-2599) and expanded. Lonza provides the cells with a viability of at least 20% FACS analysis indicates proliferating neurospheres positive for nestin and Ki67. Following differentiation on a poly D-lysin/laminin matrix in the absence of growth factors, the cells test positive for TUBB3, GFAP, Nestin and O4 (Baumann et al., 2015; Schmuck et al., 2017). Within the first three days after thawing, 100 uL of spheres (at least 20) are plated on poly Dlysin/laminin matrix in an 8-chamber slide with 500 uL N2 Media. On day one after plating greater than 80% of spheres need to be differentiated for the cell to be used in experiments. Differentiation towards the final test system: Cells are frozen in liquid nitrogen and after thawing have to be cultivated in proliferation medium at 37C and 5 % CO2. The medium contains Dulbecco's modified Eagle medium and Hams F12 (3:1) supplemented with 2% B27, 20 ng/mL EGF, 20 ng/mL recombinant human FGF, 1% penicillin and streptomycin. The thawing is performed by repeated addition and removal of proliferation medium until all cells are transferred into suspension in a tissue culture flask. The cells are carefully resuspended and distributed to 10 cm petri dishes filled with fresh, prewarmed proliferation medium. The cells are fed by replacing half the medium with new medium every two to three days. At each feeding day the culture is checked for impurities which are removed into a new petri dish. From this dish mistakenly, sorted spheres can be rescued and placed back in the original culture dish. After 3-4 weeks neurospheres reach the acceptable size of 0.2 – 0.5 mm for passaging by mechanical dissociation. Thereby, neurospheres are cut into small pieces (0.15 - 0.25 mm; depending on the desired sphere size), which round up again to uniform sized neurospheres within 1 day in proliferation medium. By using this method neurospheres are expanded every week. Starting at week 2 poly hema coated dishes are used for the cultivation procedure. Critical consumable: the cultivation medium does not contain serum or serum replacement. The use of epidermal growth factor (EGF) and recombinant human fibroblast growth factor (FGF) is critical for sphere growth. FGF contains 1% bovine serum albumin and is thus prone to batch effects. Critical handling: The thawing media contains DMSO in a concentration that affects cell health which is why thawed cells should quickly be diluted in proliferation media (30 mL of media for one vial of cells). It is recommended to add FGF into medium directly before thawing. At the end of week two of the expansion period (see below), the spheres should be transferred to petri dishes coated with poly hema to prevent cell attachment. attached cells that are not differentiated can be detached using a 1000 uL pipet. To

avoid repeated attachment, all cells should be transferred to a new poly hema coated petri dish. Medium that contains FGF should not be stored longer than 1 week at 4C. During the first two weeks, medium should be removed using a 1000 uL pipet to keep the accidental removal of small spheres to a minimum. The neurospheres should be well distributed in the petri dish to circumvent aggregation, which is especially important after mechanical dissociations.

2.4 Metabolic Competence: Primary hNPC under proliferating and differentiating conditions do not express CYP1A1 and CYP1B1 (Gassmann et al., 2010). Primary hNPC during differentiation, have the capacity to up-regulate glutathione-dependent protective strategies upon reactive oxygen species (ROS) exposure Masjosthusmann et al, 2019). Gene expression levels of genes involved in the antioxidative defense (glutathione peroxidase 1 (GPX1), superoxide dismutase 1 (SOD1), catalase (CAT)) were comparable between the in vitro system and developing human brains in vivo and show similar expression levels (Masjosthusmann et al., 2019). Other metabolic pathways are not characterized.

2.5 Exposure Regime: Culture protocol: After the cell expansion period the cells are cultured for up to four weeks in which they are passaged every week as described in 3.6. Between one to three days after passaging spheres at a size of 0.3 mm are picked for the test method. For the assessment of neural progenitor cell migration and differentiation, the spheres are plated on poly-D-lysine/laminin coated 96-well flat bottom plates in differentiation medium (N2) to initiate migration and differentiation. Thereby one 0.3 mm big sphere is plated in the middle of a well. The differentiation medium consists of DMEM and Ham's F12 at a ratio of 3 to 1 supplemented with 1% N2 and 1% penicillin and streptomycin. Within 5 days NPCs radially migrate out of the sphere core and differentiate into radial glia cells (nestin positive), neurons (TUBB3 positive), oligodendrocytes (O4 positive) and astrocytes (GFAP positive). Cultivation during the test method is performed at 37C and 5% CO2 at a ph of 7.2-7.6. Exposure: 0.3 mm big hNPC are plated in the already prepared test conditions. Exposure starts at day 0 of differentiation and is continued over five days of differentiation until the experiment is terminated. Cells are fed with fresh medium at day 3 of differentiation. Therefore, half of the test condition solution (e.g. solvent control or compound dilution) is replaced by freshly prepared test condition solution. Test Method: Migration distance and cytotoxicity is determined after 72 hours based on brightfield images of each well. The assay is terminated by the assessment of cell viability and cytotoxicity as well as cell fixation after 120 h. Immunocytochemistry is performed for Hoechst positive nuclei, TUBB3 positive neurons and O4 positive oligodendrocytes. On the ICC images, migration after 120 hours, neuronal and oligodendrocyte number, neuronal morphology and neuron/oligodendrocyte specific migration is assessed. immunocytochemical staining (ICC) generates an image for each sphere. The plates are scanned using an automated high content imaging device and all nuclei and their positions are determined automatically based on their intensity and size. Images are imported to the Omnisphero software to run the image analysis that measures the following endpoints. The method is set up for 8 test conditions including 7 compound concentrations and one SC. The test conditions are prepared in a serial dilution from the stock solution. Stock solutions are prepared by diluting the compound in the solvent (e.g. DMSO) in a concentration that allows the preparation of the highest test concentration without exceeding the highest acceptable solvent concentration. For DMSO the highest acceptable solvent concentration is 0.1% which means that the stock concentration needs to be at least 1000x higher than the highest test concentration. Stock solutions in non-sterile solvents (e.g. water or PBS) have to be sterile filtrated using a sterile syringe filter (0.2 um). Stock solutions are aliquoted and stored at -20C. A stock solution is not thawed more than three times. For the preparation of the test condition the stock solution is diluted to the highest test concentration (default 1:1000) in differentiation media. All following dilutions are prepared by serial dilution of the highest concentration in differentiation media with solvent (in the concentration of the highest test concentration). The default serial dilution is 1:3 which covers a concentration range from e.g. 20 uM to 27 nM (729 fold). Depending on the desired concentration range the dilution can be changed to 1:2, 1:5, 1:10 or other. The SC is prepared by adding the solvent to differentiation media in the same concentration as the highest test concentration. 100 uL of the compound dilutions and the SC are added to a 96 well plate. The serial dilution can also be prepared directly in the 96 well plates. hNPC are added to each well after a 15 to 30 min equilibration period at 37C and 5 % CO2.

ASSAY DESIGN SUMMARY

Nominal number of tested concentrations: 7	Target (nominal) number of replicates: 15
Standard minimum concentration tested: 0.143263374485596 µM	Standard maximum concentration tested: 10000 µM
Key positive control: BMP7	Neutral vehicle control: DMSO
Baseline median absolute deviation for the assay (bmad): 28.259	
Response cutoff threshold used to determine hit calls: 56.518	
Detection technology used: High content image analysis (Fluorescence microscopy)	

2.6 Response: All endpoints are generated from the same experimental run and from each well/sphere in the 96 well plate. Primary DNT specific endpoints of the test method are: migration distance radial glia at 72 hours (NPC2a), migration distance radial glia at 120 hours (NPC2a), migration distance neurons at 120 hours (NPC2b), migration distance oligodendrocytes at 120 hours (NPC2c), neuronal differentiation at 120 hours (NPC3), neurite length at 120 hours (NPC4), neurite area at 120 hours (NPC4), and oligodendrocyte differentiation 120 hours (NPC5). Secondary endpoints are: cell number at 120 hours (which is used for normalization of neuronal and oligodendrocyte differentiation), cytotoxicity at 72 hours, cytotoxicity at 120 hours, and viability at 120 hours. Oligodendrocyte differentiation at 120 hours is determined as number of all O4 positive cells in percent of the amount of all Hoechst positive nuclei in the migration area after 120 hours of differentiation. The identification of oligodendrocytes is done automatically using a convolutional neural network (CNN). Training of the CNN was done based on manually annotated experiments. The endpoint specific control for oligodendrocyte differentiation is BMP7. BMP7 promotes the BMP signaling cascade which upregulates astroglia differentiation and maturation and inhibits oligodendrocyte formation (Baumann et al., 2015; Gross et al., 1996; Mabie et al., 1997). 100 ng/mL BMP7 inhibit oligodendrocyte differentiation to between 0 and 60 % of SC.

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:

DNT-IVB: Assays included in the developmental neurotoxicity in vitro battery, see https://www.regulations.gov/document/EPA-HQ-OPP-2020-0263-0006

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

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: The raw data format is different depending on the endpoints. For all endpoints assessed in a multiplate reader (viability and cytotoxicity) the raw data format are excel files containing values (one for each endpoint, timepoint and well) measured as relative fluorescence units. These values are transferred from the original excel file into the AXES sheet. The original excel output files is saved for traceability of the data. The migration distance of radial glia after 72 h which is measured manual in ImageJ is directly copied into the AXES sheet as value in µm. Original brightfield images are archived for 10 years. All other raw data is computed from the ICC images in the Omnisphero software and is exported and saved as one csv file. From there the values are again transferred to the AXES sheets. The following data is exported from Omnisphero: number of all cells in migration area, number of neurons all neurons in migration area, number of all oligodendrocytes in migration area, migration distance radial glia (µm), mean migration distance all neurons (µm), mean migration distance all oligodendrocytes (µm), neurite length (µm), and neurite area (pixel). All original ICC images are archived for 10 years. If not otherwise stated, all data processing steps are performed in a R based evaluation tool that was designed for data processing, curve fitting and point of departure evaluation of in vitro concentration response toxicity data. Data processing describes all processing steps of raw data that are necessary to obtain the final response values including the normalization, curve fitting and benchmark concentration calculation. Raw data processing to summary data for the oligodendrocyte differentiation endpoint takes the number of all oligodendrocyte is divided by the number of all cells (in the migration area). oligodendrocyte differentiation % = number of oligodendrocytes / all cells x 100. For the normalization to the solvent control, each replicate data point is normalized to the median of the solvent control in the respective experiment.

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.)

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.), 6: resp.multneg1 (Multiply the normalized response value (resp) by -1; -1*resp.)

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:

21: bmad2 (Add a cutoff value of 2 multiplied by the baseline median absolute deviation (bmad). By default, bmad is calculated using test compound wells (wllt = t) for the endpoint.), 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.), 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: 238		Number of chemicals tested: 223
ACTIVITY HIT CALLS		
Active hit count: hitc≥0.9 52	Inactive hit count: 0≤hitc<0.9 120	NA hit count: *hitc<0* *66*
WINING MODEL SELECTION		
Number of sample-assay endpoints with winning *hill* model:	16	
gain-loss (gnls) model:	25	
power(pow) model:	28	
linear-polynomial (poly1) model:	45	
quadratic-polynomial(poly2) model:	24	
exponential-2 (exp2) model:	3	
exponential-3 (exp3) model:	3	
exponential-4 (exp4) model:	58	
exponential-5 (exp5) model:	36	

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*	0
Neutral control median absolute deviation, by plate: *nmad*	0
Coefficient of variation (CV%) in neutral control wells: *(nmad/nmed)*100*	NA%
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: 36.

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:** Baumann J, Gassmann K, Masjosthusmann S, DeBoer D, Bendt F, Giersiefer S, Fritsche E. Comparative human and rat neurospheres reveal species differences in chemical effects on neurodevelopmental key events. Arch Toxicol. 2016 Jun;90(6):1415-27. doi: 10.1007/s00204-015-1568-8. Epub 2015 Jul 28. PMID: 26216354.

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.