

Enhancing a Curated High-throughput Screening Pipeline to Improve Toxicological Interpretations of In Vitro Data

V.D. Hull¹, B.N. Hill¹, A. Unnikrishnan¹, A.L. Karmaus², E.N. Reinke¹, N.C. Kleinstreuer^{3,4}, H.T. Hogberg⁴

¹Inotiv, United States; ²Syngenta Crop Protection, LLC, United States; ³NIH/OD/DPCPSI, United States;

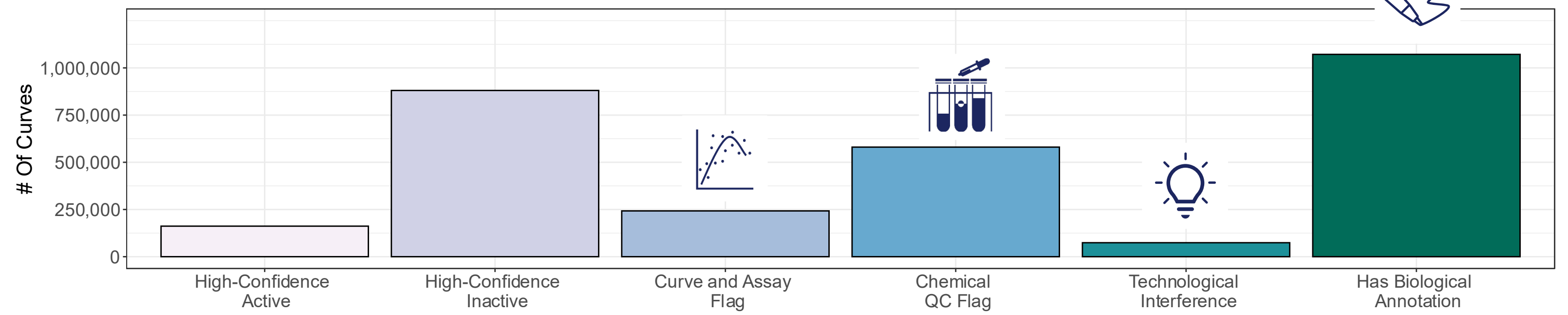
⁴NIH/NIEHS/DTT/NICEATM, United States

Introduction

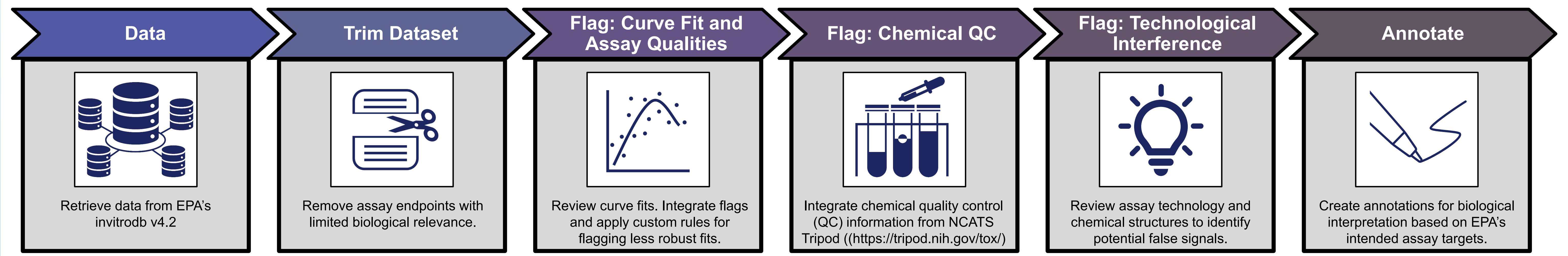
- Testing of thousands of chemicals in the ToxCast and Tox21 high-throughput screening (HTS) programs has generated millions of data points that can be difficult to analyze in detail.
- To increase confidence in HTS bioactivity calls, the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) has reviewed these data to generate a curated HTS (cHTS) data set.
- The curation process includes flagging data points based on chemical quality, curve fit, or potential technological interference.
- Here we describe recent improvements to our HTS curation process that have been made to accommodate major updates in concentration-response modeling in the U.S. Environmental Protection Agency (EPA) ToxCast/Tox21 database (Feshuk et al. 2022).

Summary

- Processing pipeline contains 9,614 chemicals and 1,376 assay endpoints, yielding a total of 1,865,056 concentration-response curves.



Curation Pipeline



Pipeline Details

Data Retrieval and Trimming

- Data were retrieved from the EPA invitrodb v4.2 database and analyzed using the ToxCast Analysis Pipeline (tcpl R package v3.2.0; Filer 2024).
- Assay endpoints that were determined to have limited biological relevance were excluded prior to curation. Types of data excluded are summarized below.

Data Excluded	Characterization/Reasons
Background Readout	Denoted as "background measurement" or "background control" in assay metadata.
Channel Readouts for Tox21 Assays	Used to calculate a ratio endpoint; only the ratio is maintained.
Specific Attagene Transcriptional Activation Assay Endpoints	Excluded Attagene transcriptional activation assay endpoints analyzed in the down direction. These assays are only appropriately interpreted as increasing signal (up direction). An exception was made for Attagene cytotoxicity assays, which are measured as cell counts and can be analyzed in the down direction.
Specific Stemina Assays	Excluded Stemina assays that evaluated the test substance at a concentration greater than 300 μ M due to potential cellular stress.
Specific NovaScreen Assays	Excluded NovaScreen assays analyzed in the up direction. These assays are cell-free evaluations of chemical interactions with fluorescent or radiolabeled proteins and can only be interpreted in the down direction.

Flag: Curve Fit and Assay Qualities

- HTS data from EPA's invitrodb include flags for curve fit.
- During curation, NICEATM adds novel flags. These are evaluated along with the EPA flags to refine confidence of fit.

Flag Type	Flag Reasons
Point-of-departure (POD) Extrapolation	Activity concentration at cutoff (ACC) is extrapolated above highest tested concentration. Activity concentration at 50% of maximal activity (AC50) is extrapolated above highest tested concentration. Benchmark dose is extrapolated above highest tested concentration.
Assay Source Specific Flags	NovaScreen cell-free biochemical assay has an active call with less than 50% efficacy. Tox21 assay has a gain-loss model as the winning model and only the lowest concentration tested exceeds the activity cutoff.
Model Specific Flags	Curve with borderline activity has polynomial-quadratic model as winning model. Curve with active call has gain-loss model as winning model and AC50 extrapolated below the tested concentration range.
Directionality	Direction of curve does not match the intended direction of analysis, resulting in a negative hit call. Down-direction (i.e., inhibition, antagonism, loss-of-signal, etc.) call had a gain-loss model as the winning model.
Overall Fit	The tcpl R package assigned four or more flags to the curve fit, even if none of the flags resulted in a Flag-Omit call.

Flag: Chemical QC

- Chemical quality data are pulled from the National Center for Advancing Translational Science (NCATS) Tripod website.
- The chemical quality is assessed when compounds are freshly taken out of the freezer (T0) and after compounds are kept at room temperature for four months (T4).
- QC call is determined as detailed in the table below.

QC Grade*	NICEATM Interpretation
A	
B	
C	Pass
Z	
Ac	
BC	
Cc	Caution for low concentration
D	
Fc	
F	
Fns	Caution for low purity
I	
M	
W	Not determined/inconclusive
U	
ND	

*See <https://tripod.nih.gov/tox/samples> for definitions of each grade.

	QC T4 Interpretation			
	Caution concentration	Caution purity	Not determined/Inconclusive	Pass
Caution concentration	Flag for QC	Flag for QC	Flag for QC	Pass
Caution purity	Flag for QC	Flag for QC	Flag for QC	Pass
Not determined/Inconclusive	Flag for QC	Flag for QC	Pass	Pass
Pass	Flag for QC	Flag for QC	Pass	Pass

Flag: Technological Interference

Determine if chemical was tested in Tox21 assays that included tests for luciferase inhibition and autofluorescence interference for red, blue, and green wavelengths (Borrel et al. 2020a).

If chemicals were not experimentally tested, use the InterPred Tool to predict luciferase and fluorescence interference (Borrel et al. 2020b).

Identify chemicals that were experimentally determined or predicted to be autofluorescent or luciferase inhibiting.

Flag concentration-response data if the chemical may be causing confounding interference with the assay's detection technology.

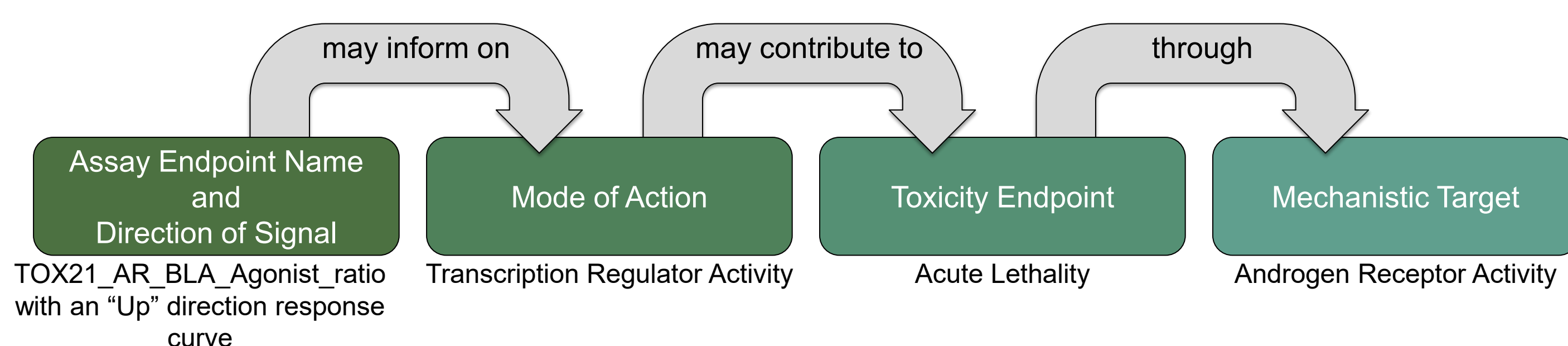
Identify which assays detect output using luciferase or red, blue, or green fluorescence.

Use assay descriptions in invitrodb, protocols on the NCATS Tripod website, and relevant citations to determine the detection technology used by assay.

Biological Annotations

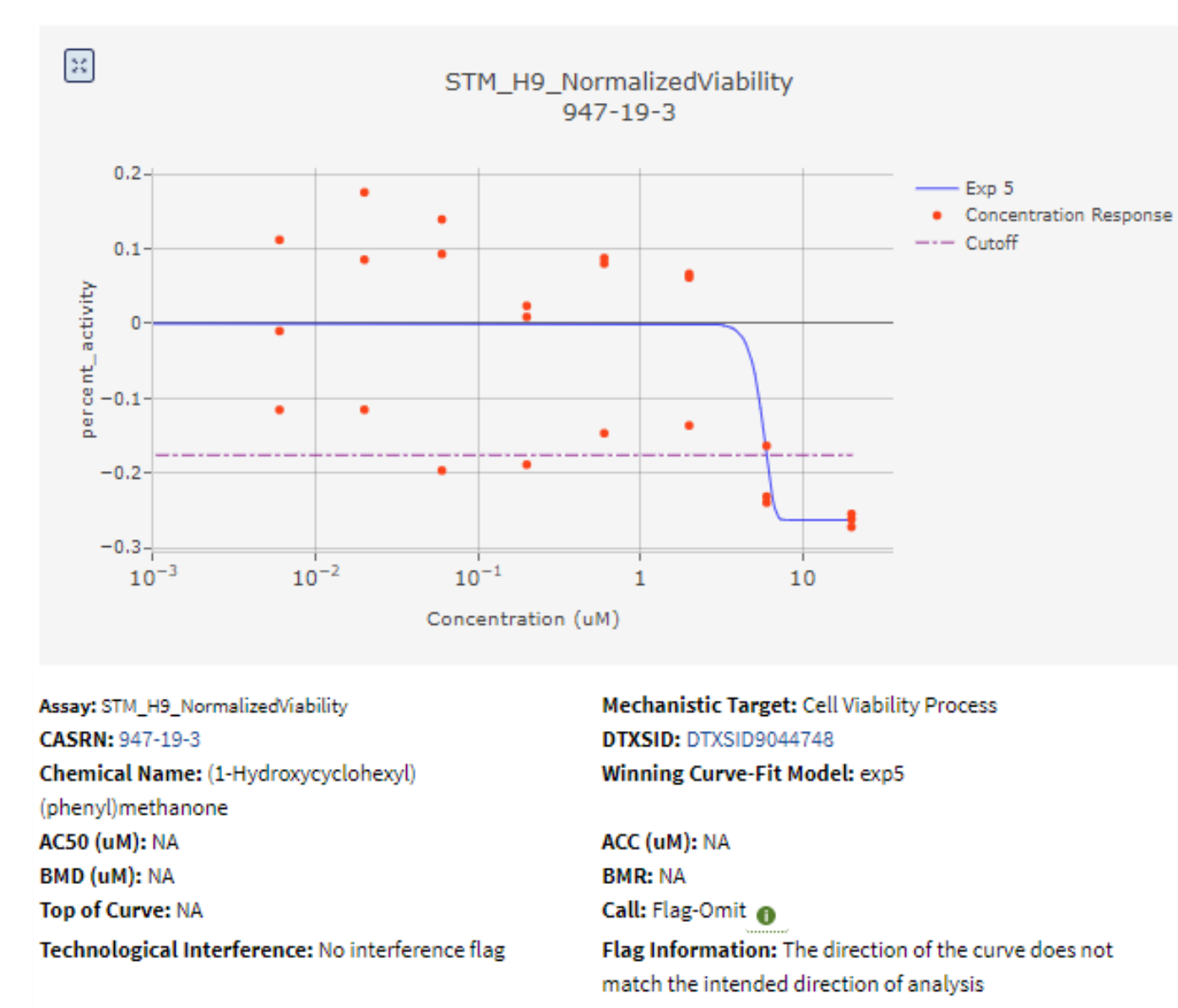
- The resulting cHTS assay inventory is annotated to mechanistic targets that facilitate linkage to modes of action and subsequently to toxicological outcomes of regulatory interest (Hill et al. 2024).
- All mechanistic target and mode-of-action terms were mapped to terms used by the Open Biological and Biomedical Ontology Foundry (Jackson et al. 2021).
- Annotation considers the direction of the curve, since an increased vs. decreased response can change the biological interpretation. For example, assay APR_HepG2_CellLoss_72hr is annotated as "Cell Population Proliferation" when the response is in the up direction and "Cell Viability Process" when the response is in the down direction.

Knowledge Organization Structure

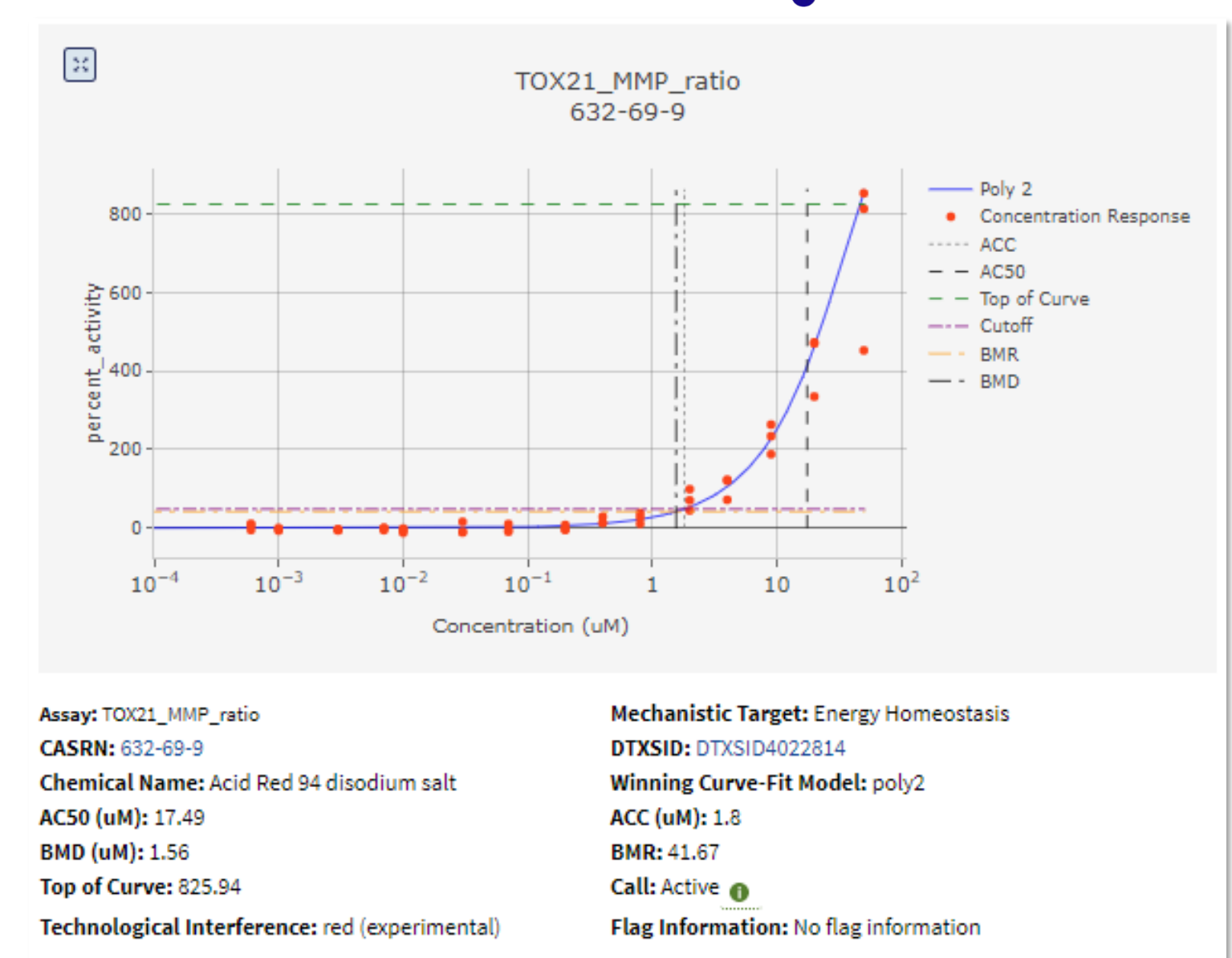


Access, Explore, and Visualize cHTS Data

The cHTS data described in this poster are available in NICEATM's Integrated Chemical Environment (ICE, <https://ice.ntp.niehs.nih.gov/>), an open-access resource for toxicologically relevant data and computational tools.



Assay: STIM_H9_NormalizedViability
CASRN: 947-19-3
Chemical Name: (1-Hydroxycyclohexyl)(phenyl)methanone
AC50 (uM): NA
BMD (uM): NA
Top of Curve: NA
Technological Interference: No interference flag
Mechanistic Target: Cell Viability Process
DTXSID: DTXSID9044748
Winning Curve-Fit Model: exp5
ACC (uM): NA
BMR: NA
Call: Flag-Omit
Flag Information: The direction of the curve does not match the intended direction of analysis



Assay: TOX21_MMP_ratio
CASRN: 632-69-9
Chemical Name: Acid Red 94 disodium salt
AC50 (uM): 17.49
BMD (uM): 1.56
Top of Curve: 825.94
Technological Interference: red (experimental)
Mechanistic Target: Energy Homeostasis
DTXSID: DTXSID4022814
Winning Curve-Fit Model: poly2
ACC (uM): 1.8
BMR: 41.67
Call: Active
Flag Information: No flag information

Conclusion

- HTS methods are often used to characterize a chemical's biological activity.
- To ensure that bioactivity calls are robust, interpretable, and of high confidence, NICEATM adds curve-fit flags, chemical quality flags, and technological interference flags to the cHTS data in ICE.
- Data are annotated using ontologies to facilitate interpretation of cHTS assays allowing for more seamless integration into chemical safety assessments.
- Curation will continue to ensure that the highest quality data are available to stakeholders.

References and Acknowledgments

Borrel et al. 2020a. Sci Rep. 10(1):3986. <https://doi.org/10.1038/s41598-020-60747-3>.
Borrel et al. 2020b. Nucleic Acids Res. 48(W1):W586-W590. <https://doi.org/10.1093/nar/gkaa378>.
Feshuk et al. 2022. Invitrodb version 4.2 release. <https://doi.org/10.23645/epacmptox.6062623.v8>.
Filer 2024. tcpl: ToxCast Data Analysis Pipeline. <https://github.com/usepa/comptox-toxcast-tcpl>.
Hill et al. 2024. Incorporating Ontologies into High Throughput Screening Assay Annotations to Increase Data Use and Interpretation. Poster presentation at SOT 2024.
Jackson et al. (2021). Open Biological and Biomedical Ontology (OBO) Foundry Database. <https://doi.org/10.1093/database/baab069>.

This project was supported by the Intramural Research Program (ES# 103386) at the National Institute of Environmental Health Sciences, National Institutes of Health under contract HHSN273201500010C. Since the poster was written as part of the official duties of the authors, it can be freely copied. We thank Catherine Sprinkle, Inotiv, for editorial input.