

Increased assay information

The number of publicly available assay endpoints in the upcoming version 4 release has increased to nearly 600 assays, 1300 components, and 1900 endpoints, with updates in assay annotations to enable better data aggregation.

Assay lists for some groups of assay (e.g., ToxCast estrogen and androgen receptor model assay endpoints) and intended target family groups are in the database and will be used by the CompTox Chemicals Dashboard to enable users to compare similar biological information more easily.

Assay endpoints now cover ~1400 unique gene targets in invitrodb version 3.3

Intended target family is one way to understand biological target (incomplete list here):

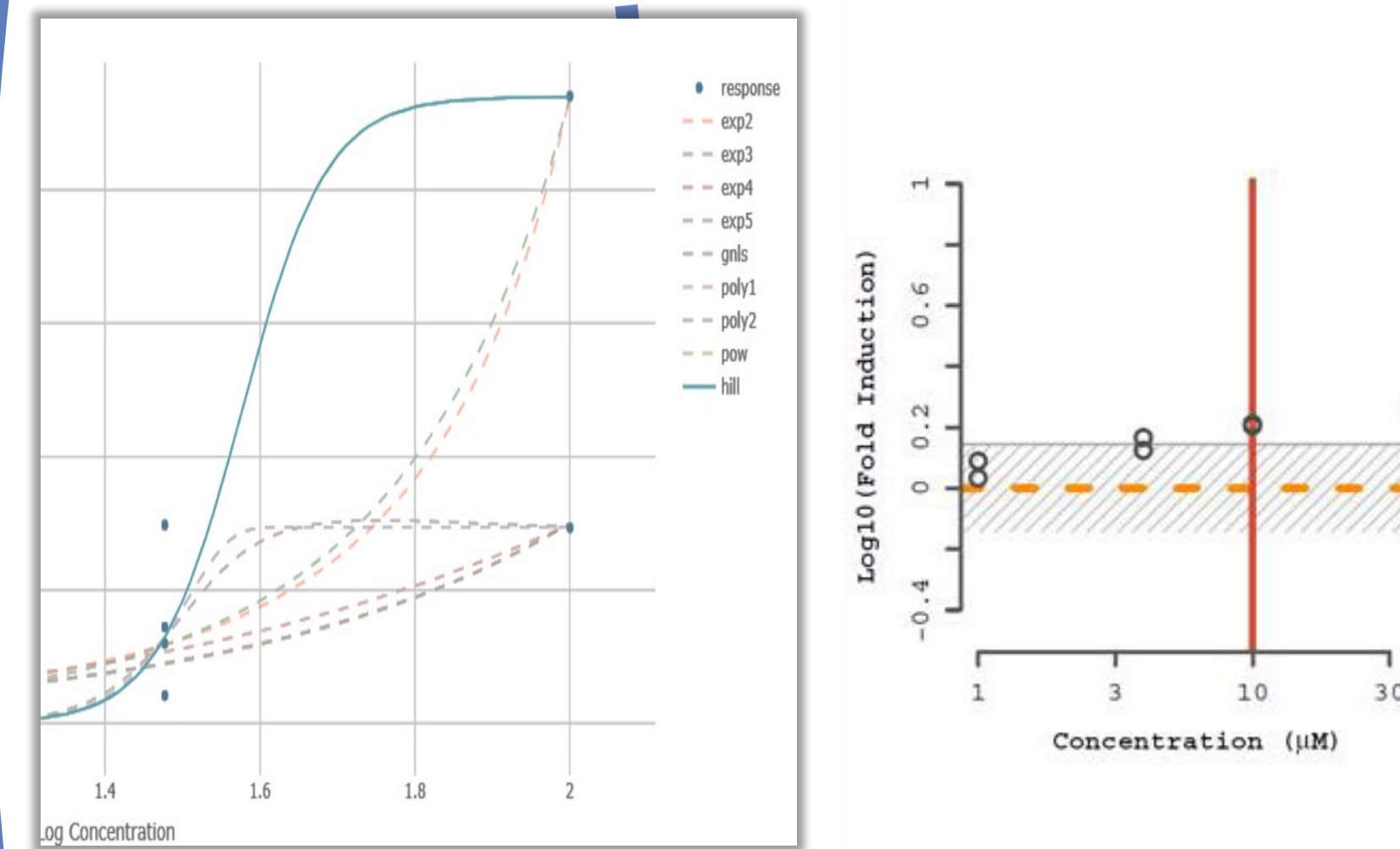
- Apolipoprotein
- Apoptosis
- Background measurement
- Catalase
- Cell adhesion
- Cell cycle
- Cell morphology
- CYP
- Cytokine
- Deiodinase
- DNA binding
- Esterase
- Filaments
- GPCR
- Growth factor
- Histones
- Hydrolase
- Ion channel
- Kinase
- Ligase
- Lyase
- Malformation (zebrafish)
- Membrane protein
- Metabolite (Stemina metabolomics)
- Mitochondria
- Methyltransferase
- microRNA
- Mutagenicity response
- Nuclear receptor
- Oxidoreductase
- Phosphatase
- Protease/inhibitor
- Steroid hormone
- Transferase
- Transporter

Enhanced curve-fitting and winning model logic at levels 4 and 5

	Tcpl version 2.0	Tcpl version 3.0 (release 2021)
Curve-fitting models	Hill, gain-loss (a modified Hill), and constant	In addition to Hill, gain-loss, and constant, Polynomial 1 (Linear), Polynomial 2 (Quadratic), Power, Exponential 2, Exponential 3, Exponential 4, and Exponential 5 based on BMDExpress and encoded by R package dependency tcplFit2
Defining positive hitcalls	0 = negative, 1 = positive, -1 = not enough data points to fit	No longer required to be binary; defined as product of three probabilities: median response and top of model exceed the cutoff, and AIC is less than the constant model
Additional potency and uncertainty information	ACC = activity concentration at the cutoff for a positive ACB = activity concentration at a definition of baseline (3* baseline median absolute deviation) AC50 = activity concentration at 50% of maximal response ACXX = activity concentration at XX% of maximal response	New for tcplFit2 models: BMD BMDU BMDL New for tcpl models: LOEC
Winning model and how it is tracked	Winning model has lowest AIC, only winning model information summarized for potency values at mc5	Winning model decided by fitting metric and similar summary statistics are saved with the option for additional parameters
Expanded invitrodb schema to allow for enhanced curve-fitting	Level 4 (mc4) and Level 5 (mc5) tables were "wide" and expected only three models to be fit	Addition of mc4_param and mc5_param tables allow for generic fitting and hitcalling Similar summary level statistics are stored at level 4 and 5

Plotting Utility

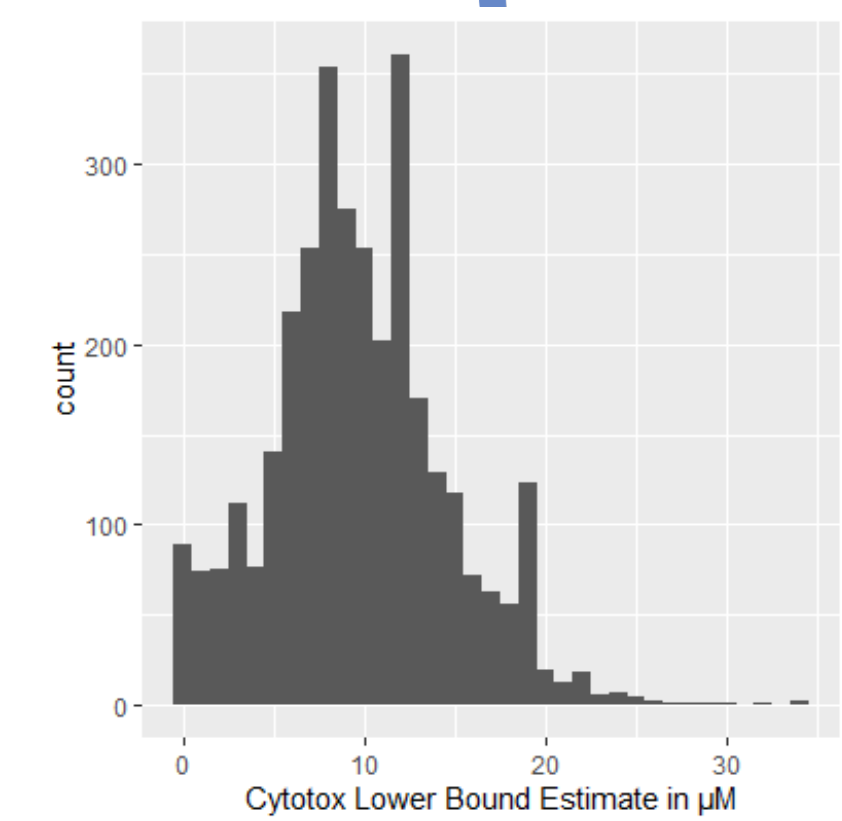
A new standalone plotting utility was built with plotly to display all of the additional curve-fitting models. The utility implements plumber to provide REST api functionality.



This utility also allows for dockerization and hosting so the api can be put into a production environment. Integration of this utility will allow for applications such as the CompTox Chemicals Dashboard to plot up to date bioactivity data

Additional tcpl and database upgrades

The lower bound estimate of the cytotoxicity threshold has been updated, requiring a 5% positive rate in curated cytotoxicity assays (now totaling 88). These data are stored in the cytotox table and can be generated on the fly with function tcplCytoPt()..



Implemented database development tools such as flyway for version tracking and myTap for unit testing before each release

Future direction: Integrate dashboard backend APIs so that no database downloads are required to interact with the richest source of data