Team 9

Advanced Modeling of Toxicological Profiles

2.3. Tox21 Program

Formally created in 2010 [4], the Tox21 program is a result of the collaboration of multiple United States of America agencies:

1. National Institute of Environmental Health Sciences (NIEHS),
2. National Toxicology Program (NTP),
3. the US Environmental Protection Agency's (EPA),
4. the National Center for Computational Toxicology (NCCT),
5. the National Institutes of Health (NIH),
6. National Center for Advancing Translational Sciences (NCATS), and
7. the FDA [5],

with a formed response to ethical concerns to in vivo toxicological tests on animals and their problem of interpretability as a shift in assessment towards in vitro assays allowing for target-specific, mechanism-based, biological observations. [4, 6]

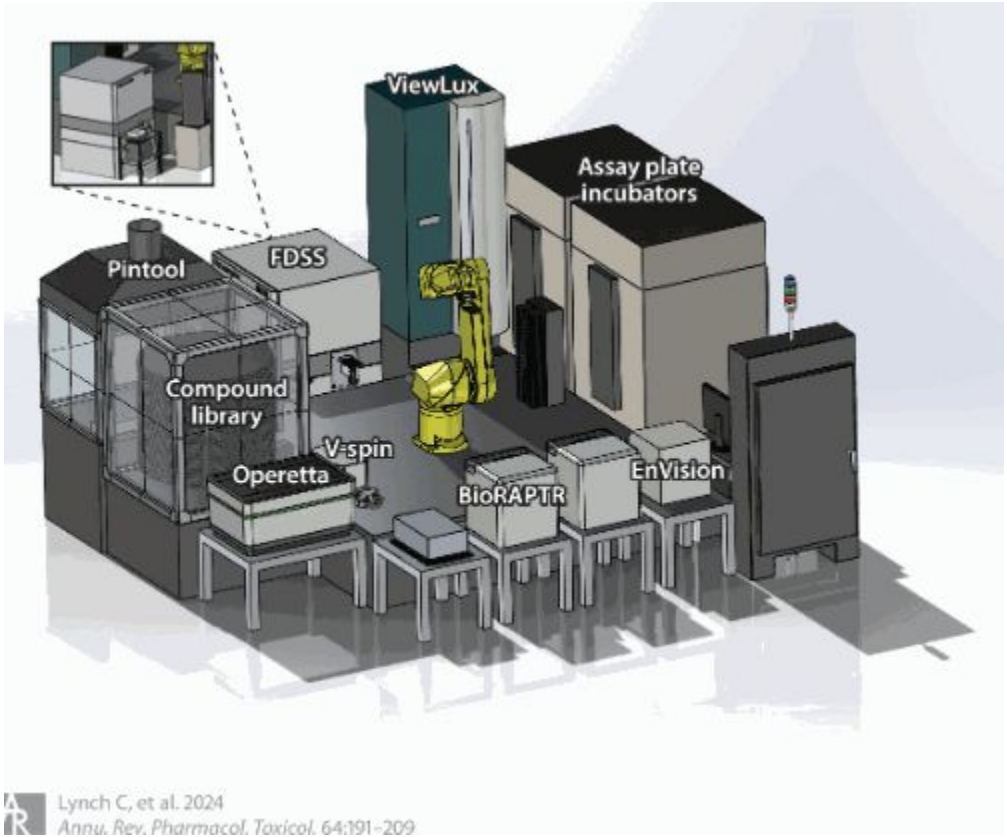
> "*The Tox21 partners agreed to develop a vision and devise an implementation strategy to shift the assessment of chemical hazards away from traditional experimental animal toxicology studies to one based on target-specific, mechanism-based, biological observations largely obtained using in vitro assays.*" [4]

[4] Raymond R. Tice, Christopher P. Austin, Robert J. Kavlock, and John R. Bucher. Improving the human hazard characterization of chemicals: A tox21 update. *Environmental Health Perspectives*, 121(7):756–765, 2013.

[5] R. Huang. A quantitative high-throughput screening data analysis pipeline for activity profiling. In H. Zhu and M. Xia, editors, *High-Throughput Screening Assays in Toxicology*, volume 1473 of *Methods in Molecular Biology*. Humana Press, 1 edition, 2016.

[6] Caitlin Lynch, Srilatha Sakamuru, Masato Ooka, Ruili Huang, Carleen Klumpp-Thomas, Paul Shinn, David Gerhold, Anna Rossoshek, Sam Michael, Warren Casey, Michael F. Santillo, Suzanne Fitzpatrick, Russell S. Thomas, Anton Simeonov, and Menghang Xia. High-throughput screening to advance in vitro toxicology: Accomplishments, challenges, and future directions. *Annual Review of Pharmacology and Toxicology*, 2024.

Robotic Platform for Tox21 High-Throughput Screens



Lynch C, et al. 2024
Annu. Rev. Pharmacol. Toxicol. 64:191–209

The methodology of HTS for the Tox21 Program has been prepared by the NCATS, which consisted
1. preparation of over 600,000 compound samples in multiple concentrations, within HTS-ready format,
2. Electronic monitoring of each sample related activities, as each mother-daughter plate has been registered using the ActivityBase sample registration system (IDBS, United States), each used laboratory equipment, each robotic process is scheduled and maintenance has been registered to a laboratory information management system (LIMS) to ensure high-quality level.
3. In-lab compound management is done online via fully automated compound management SampleStores (Azenta Life Sciences, United Kingdom),
4. A Bioek (NXP , FXP), with i7 liquid handles (Beckman Coulter, United States) have been used for sample to plate format distribution,
5. each compound has been dissolved within dimethyl sulfoxide (DMSO) and stored in 15 concentrations, where each concentration consists of up to 10-20 mM and stored at room temperature (at this point sample is ready for in vitro studies),
6. To avoid degradation, insolubility, or volatility, a quality and stability control have been performed via gas chromatography-mass spectroscopy, liquid chromatography mass spectroscopy, and nuclear magnetic resonance spectroscopy
Available and possible in vitro studies in NCATS Tox21 robot house are:
1. absorbance, luminescence, and fluorescence using ViewLux and EnVision
2. cell quantification of "cellular features" using Operetta imager,
3. real-time (kinetic measurements) fluorescence measures in adherent or suspended cells using Functional Drug Screening System 7000EX (Hamamatsu, Japan)
To ensure high quality, all compounds were required to satisfy the following criteria: signal to-background ratios (≥ 3), coefficients of variation (≤ 10%), and Z' factors (≥ 0.5), and when all data have been collected activity is calculated in the following manner:

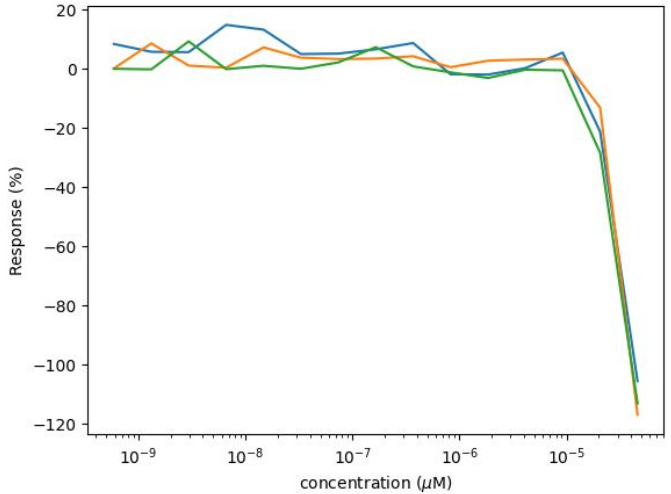
$$\text{DATA}_i = \frac{V_{\text{ligand}} - V_{\text{DMSO}}}{V_{\text{positivie}} - V_{\text{DMSO}}} \cdot 100\%,$$

where V_{ligand} denotes compound well values, $V_{\text{positivie}}$, and V_{DMSO} respectively denote median wells values of positive controls, and DMSO. Furthermore, to ensure reeds independence from the test environment background patterns and subtle abnormalities such as tip effects or blotting from cell dispensing have been removed. [6]

protocol: tox21-err-p1

Assay overview: Estrogen related receptors (ERRs), the first orphan nuclear receptors discovered, play an important role in the control of cellular energy metabolism. ERRs are required for high-energy production in response to the environmental challenges. ERR-alpha was also identified as an adverse marker for breast cancer progression; ERR-alpha-positive tumors have a poor prognosis. To screen and to identify environmental compounds that perturb the ERR signaling pathways could provide valuable information for potential therapeutic or preventive measures in the treatment of metabolic disease. For identifying ERR agonists/antagonist, ERR cell line (provided by National Toxicology Program) has been used to screen Tox21 10K library and to differentiate true ERR antagonists or agonist from cytotoxic substances, the assay is multiplexed with cell viability assay.[5]
Agonist/Antagonist Assay Protocol Summary: ERR cells were dispensed at 2,000 cells/5 ul/well in 1536-well white plates using a Multidrop dispenser. After the assay plates were incubated at a 37 C/5% CO2 incubator for 6 hours, 23 nL of compounds dissolved in DMSO, positive and negative controls or DMSO only was transferred to the assay plate by a pin tool. The plates were incubated at 37 C for 18 hours. 4 ul/well of One-Glo reagent was added into the assay plates using a Flying Reagent Dispenser. After 30-minute incubation at room temperature, the luminescence intensity in the plates was measured using a ViewLux plate reader.[5]
detailed protocol: https://tripod.nih.gov/tox/apps/assays/slp/tox21-err-p1.pdf
Viability Assay Protocol Summary: The ERR cells were dispensed at 2,000 cells/5 ul/well in 1536-well white plates using a Multidrop dispenser. After the assay plates were incubated at a 37 C/5% CO2 incubator for 6 hours, 23 nL of compounds dissolved in DMSO, positive and negative controls or DMSO only was transferred to the assay plate by a pin tool. The plates were incubated at 37 C for 17.5 hours. 1 ul/well of CellTiter-Fluor reagent was added into the assay plates using a Flying Reagent Dispenser. After 30-minute incubation at 37 C/5% CO2, the fluorescence intensity in the plates was measured using a ViewLux plate reader[5]

Assay Target	Cell Lines	Species	Tissue of Origin	Assay Readout	Assay Provider	Toxicity Pathway
ERR regulated genes	ERR Hek293T	Human	Embryonic kidney	Luminescence	NTP	Energy homeostasis

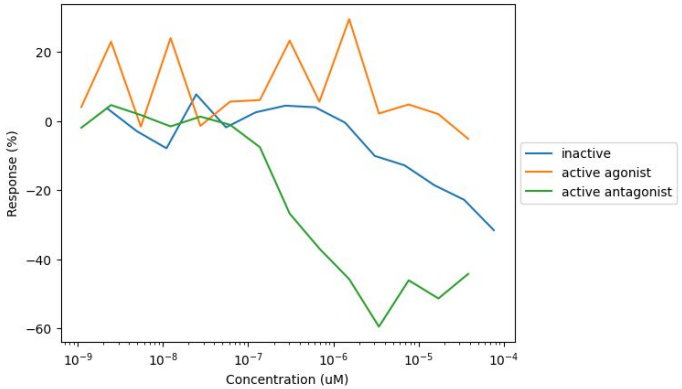


sample excluded: sample with reed error: (SAMPLE_ID: NCGC00255714-01)

[11] Promega Corporation. CellTiter-Glo® luminescent cell viability assay technical bulletin. Technical Report TB288, Promega Corporation, Madison, WI, 2023.

protocol: tox21-spec-hek293-p1

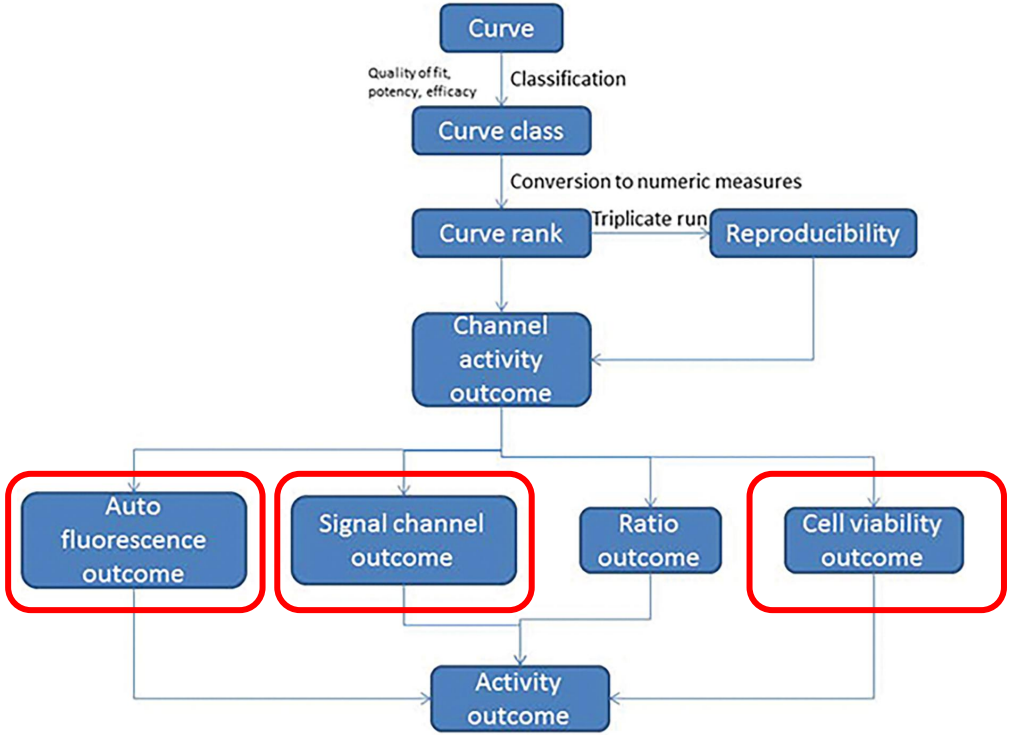
Assay overview: Compound auto fluorescence often interferes with screening of the compound library against various biological targets and signaling pathways. Some of these compounds may have fluorophoric properties. Such properties may interfere with the assay outputs resulting in false positives and/or negatives. To decrease such complications and ensure the compound activities are relevant in terms of biological response, the Tox21 10K compound library was screened for auto fluorescence. From the fluorophores used as reporter signals in our screening assays, the auto fluorescence equivalencies of compounds against three fluorophores have been evaluated: Triamterene, Fluorescein and Rose Bengal sodium for blue, green and red fluorescence, respectively. The fluorescence intensities were measured using Excitation/Emission wavelengths at 540/590nm for red labels, respectively. This assay includes the red fluorescence data.[5]
Viability Assay Protocol Summary: 2,000 HEK293 cells in 5 uL of culture medium containing 10% dialyzed FBS per well were dispensed into black wall/clear bottom 1536-well plates using a Multidrop Combi Dispenser (Thermo Scientific). The assay plates were incubated for 4-5 hr at 37C to allow cell attachment to the well bottom, followed by the transfer of 23 nL compound or DMSO vehicle by a pintool work station (Kalypsys, San Diego, CA). After the assay plates were incubated overnight at 37C and 5% CO2, fluorescence intensities in the assay plates were measured by an Envision (PerkinElmer, Shelton, CT) plate reader using three labels for measuring blue, green and red fluorescence at excitations 590nm respectively.[5]



sample excluded: solutions

[11] Promega Corporation. CellTiter-Glo® luminescent cell viability assay technical bulletin. Technical Report TB288, Promega Corporation, Madison, WI, 2023.

How activity is measured



R. Huang. A quantitative high-throughput screening data analysis pipeline for activity profiling. In H. Zhu and M. Xia, editors, *High-Throughput Screening Assays in Toxicology*, volume 1473 of *Methods in Molecular Biology*. Humana Press, 1 edition, 2016.

Historical approaches to regression problems no Tox21 DataSet, source: https://paperswithcode.com/dataset/tox21-1

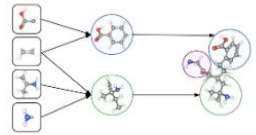
Tox21

Edit

The **Tox21** data set comprises 12,060 training samples and 647 test samples that represent chemical compounds. There are 801 "dense features" that represent chemical descriptors, such as molecular weight, solubility or surface area, and 272,776 "sparse features" that represent chemical substructures (ECFP10, DFS6, DFS8; stored in Matrix Market Format). Machine learning methods can either use sparse or dense data or combine them. For each sample there are 12 binary labels that represent the outcome (active/inactive) of 12 different toxicological experiments. Note that the label matrix contains many missing values (NAs). The original data source and Tox21 challenge site is https://tripod.nih.gov/tox21/challenge/.

Source: Tox21 Machine Learning Data Set

Homepage



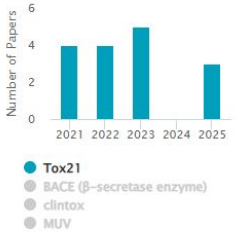
Source: https://www.frontiersin.org/articles/10.3...

Benchmarks

Edit

Trend	Task	Dataset Variant	Best Model	Paper	Code
	Molecular Property Prediction	Tox21	Deep-CBN		
	Drug Discovery	Tox21	elEmBERT-V1		
	Graph Regression	Tox21	CensNet		
	Graph Classification	Tox21	GMT		
	Molecular Property Prediction (1-shot))	Tox21	Meta-MGNN		

Usage



License

Edit

Unknown

Modalities

Edit

Historical approaches to regression problems no Tox21 DataSet

Meta-MGNN [14] utilizes a composition of 3 cross-entropy loss functions calculated with task-aware attention

> "*It calculates the average of all the molecular embedding from the query set of the same task to represent this task.*" [14]

is based upon 12 binary labels (active/inactive).

> "*Toxicity on 12 biological targets, including nuclear receptors and stress response pathways.*" [14]

Convolution with Edge-Node Switching in Graph Neural Networks (CensNet) [13] has been applied as a regression solution to the Tox21 problem, where molecular graphs have been fitted to rounded to active or inactive (binary) biological activity across 15 assays with MSE loss and RMSE performance metric:

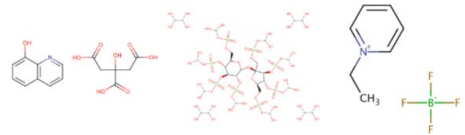
> "*Each compound is associated with 12 binary labels that represent the outcome (active/inactive) of 12 different toxicological experiments.*" [13]
> "*The regression task is similar to graph classification but using different loss function and performance metric.*" [13]

Dataset Overview:

No compounds: 7048 (no substances)
Features: 'SAMPLE_DATA_TYPE', 'CURVE_CLASS2', 'PUBCHEM_SID', 'SAMPLE_ID', 'SAMPLE_NAME', 'TOX21_ID', 'PUBCHEM_CID', 'canonical_smiles', 'isomeric_smiles', 'iupac_name', 'PubChemFingerprint', 'similarity_order',
Labels: 'DATA0', 'DATA1', 'DATA2', 'DATA3', 'DATA4', 'DATA5', 'DATA6', 'DATA7', 'DATA8', 'DATA9', 'DATA10', 'DATA11', 'DATA12', 'DATA13', 'DATA14', 'CONC0', 'CONC1', 'CONC2', 'CONC3', 'CONC4', 'CONC5', 'CONC6', 'CONC7', 'CONC8', 'CONC9', 'CONC10', 'CONC11', 'CONC12', 'CONC13', 'CONC14'

Task: Create a model that based on provided Canonical Smiles and concentration would predict response (DATA) for 'SAMPLE_DATA_TYPE' (signal pathways (agonist and antagonist), viability, and autofluorescence)
Data availability: https://drive.google.com/drive/u/0/folders/195KAyBS80Qdu5-uTHUWGVScDd4S7jBmM

Examples of compounds present in tox21



(a) 8-Hydroxyquinoline citrate (b) Sucrose octasulfate-aluminum complex (c) 1-Ethylpyridinium tetrafluoroborate

Drawbacks of Tox21

Results based on Tox21 might vary from human reaction to toxins, as Tox21 is a highly sophisticated and well-prepared in vitro model, and thus following might differ from in cito reaction due to sample preparation standardized methodology (needed for standard tests):

· "improper serial dilution of the compound,"

· some compounds had not been solvable in high concentrations,

· as DMSO has been chosen to be the standard solvent, some compounds might highly vary in solubility, or at some concentrations they might not be soluble.

· Some samples are supposed to vary so standardization might affect negatively the standard among samples binding affinity, [9]

"“Perfect” assays do not exist.

· Coverage of all chemicals of interest is incomplete (i.e., volatiles).

· A high throughput system for measuring the free concentration of a compound in vitro is not yet available.

· Xenobiotic metabolism is lacking in virtually all in vitro assays.

· Interactions between cells are poorly captured.

· Distinguishing between statistical and biological significance is difficult.

· Extrapolating from in vitro concentration to in vivo dose or blood levels is not straightforward.

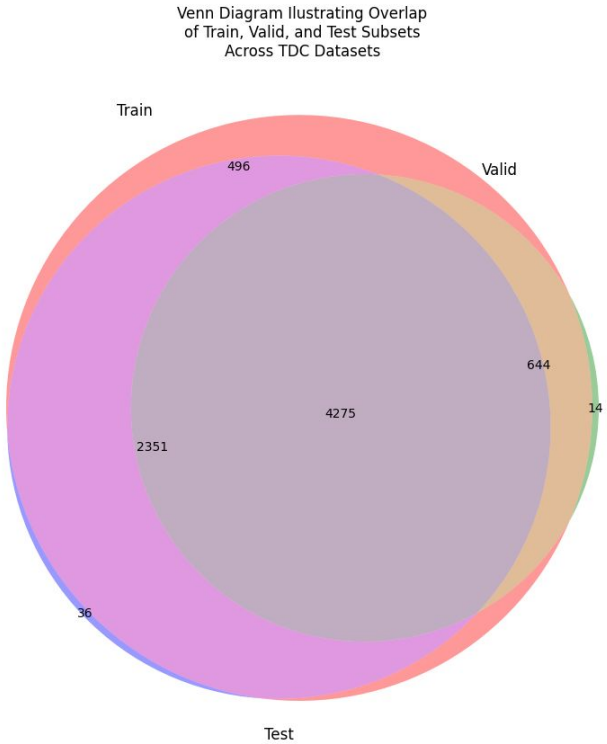
· Assessing the effects of chronic exposure conditions in vitro is not possible.

· Identifying when a perturbation to a gene or pathway would lead to an adverse effect in animals or humans remains a challenge.

· Achieving routine regulatory acceptance of the developed prediction models is years away."[4]

[9] David S. Auld, Mark W. Farmen, Stephen D. Kahl, et al. *Receptor Binding Assays for HTS and Drug Discovery.* Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda (MD), updated 2018 jul 1 edition, 2012. Available from: https://www.ncbi.nlm.nih.gov/books/NBK91992/.

Tox21 alias datasets analysis Therapeutics Data Commons & Tox24



- Problem of data leaking in TDC across assays
- Top models in TDC and Tox24 are:
  - XGBoost or LightGBM,
  - with features as Mordred, Pubchem, MACCS fingerprints
  - with cross-validation based on model random greed search