




# Tox21 Data Challenge 2014

Predictions on the final evaluation set submitted by all challenge participants are available here **tox21-challenge.zip** (final-results/tox21-challenge.zip)

Results for the final evaluation dataset are available here 

## Resources



- Tox21 at NCATS (<http://www.ncats.nih.gov/research/reengineering/tox21/tox21.html>)
- Tox21 at the Environmental Protection Agency (<http://epa.gov/ncct/Tox21/>)
- Tox21 at the National Toxicology Program/National Institute of Environmental Health Sciences (<http://ntp.niehs.nih.gov/?objectid=05F80E15-F1F6-975E-77DDEDBDF3B941CD>)

- Tox21 Robot at NCATS (<http://www.ncats.nih.gov/research/reengineering/tox21/projects/compound-toxicity.html>)

## Key Dates

**August 18, 2014**

NCATS begins accepting submissions

**November 14, 2014 (11:59 p.m. ET)**

Registration and submission deadline

**January 12, 2015**















Winners announced

[Register Now » \(registration.jsp\)](#)

## Training Datasets

The complete training dataset is available here ([download?id=tox21\\_10k\\_data\\_all sdf&sec=](#)). For individual datasets, please use the links below. In the datasets, "1" means active, "0" means inactive.

### Nuclear Receptor Signaling Panel

Assay	SDF	SMILES
AR ( <a href="https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743040">https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743040</a> )		
AhR ( <a href="https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743122">https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743122</a> )		
AR-LBD ( <a href="https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743053">https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743053</a> )		
ER ( <a href="https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743079">https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743079</a> )		
ER-LBD ( <a href="https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743077">https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743077</a> )		
aromatase ( <a href="https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743139">https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743139</a> )		
PPAR-gamma ( <a href="https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743140">https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743140</a> )		

### Stress Response Panel

Assay	SDF	SMILES
ARE ( <a href="https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743219">https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743219</a> )		

**Assay****SDF****SMILES**ATAD5 (<https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=720516>)HSE (<https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=743228>)MMP (<https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=720637>)p53 (<https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=720552>)**Final Evaluation**

The final evaluation dataset is now available for download as either SMILES (download?id=tox21\_10k\_challenge\_scoresmls&sec=) or SDF (download?id=tox21\_10k\_challenge\_scoresdf&sec=). Results submitted for this dataset will be used to determine the final ranking of the competition. Note that you can submit multiple times but only the latest submission will be used for scoring. You can continue submitting to the leaderboard to test your model until October 13<sup>th</sup>, 2014, after which time the leaderboard will be closed and submissions thereafter will be toward the final evaluation set.

**Testing Dataset**

The testing dataset is available for download here (download?id=tox21\_10k\_challenge\_testsmiles&sec=). Please note this dataset is only used to evaluate performance for the leaderboard; a separate dataset will be used to determine the winner.

Results for the testing dataset are now available for download (download?id=tox21\_10k\_challenge\_testsdf&sec=).

**Public Domain Code**

The following are links to code developed by our group that might be useful for the challenge. Please feel free to contact us (mailto:NCATS9800tox21challenge@mail.nih.gov) for any questions about the code.

- LyChI (<https://github.com/ncats/lychi>) is a structure standardizer that can be used to remove salts and solvents.
- PCFP (<https://bitbucket.org/caodac/pcfp>) is our implementation of the PubChem fingerprint ([ftp://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem\\_fingerprints.txt](ftp://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem_fingerprints.txt)).
- A number of molecular descriptors (e.g., topological indices, group contribution, solvent accessible surface area, etc.) are available as part of the library synthesizer (<https://bitbucket.org/caodac/reagents>) code base.
- Tox21Baseline (<https://spotlite.nih.gov/opensource/tox21baseline>) is a simple implementation of the naive Bayes ([http://en.wikipedia.org/wiki/Naive\\_Bayes\\_classifier](http://en.wikipedia.org/wiki/Naive_Bayes_classifier)) classifier—utilizing LyChI for

structure standardization and PCFP for descriptor/feature extraction—to provide baseline models for the 12 datasets in the challenge.

NCATS Home (<http://www.ncats.nih.gov>) | Privacy Notice (<http://www.nih.gov/about/privacy.htm>) | Disclaimer (<http://www.nih.gov/about/disclaim.htm>)

National Center for Advancing Translational Sciences (NCATS), 6701 Democracy Boulevard, Bethesda MD 20892-4874 ☎  
301-435-0888