EPA United States Environmental Protection Agency

# Chemotype-Enrichment Workflow:

## A univariate analysis workflow for exploring chemical feature enrichments across EPA's ToxCast chemical-assay landscape

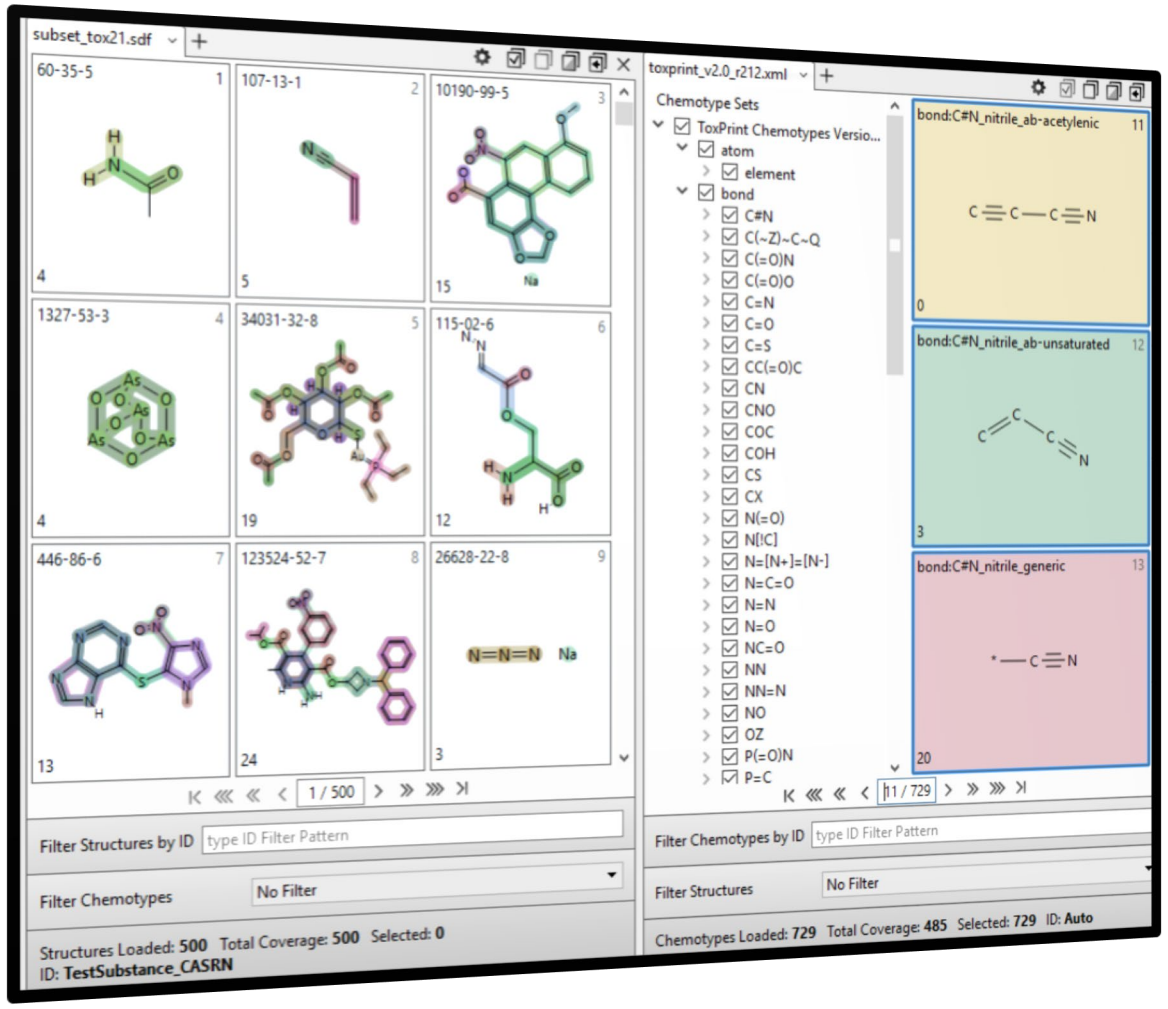
Ryan Lougee
(USEPA ORISE)

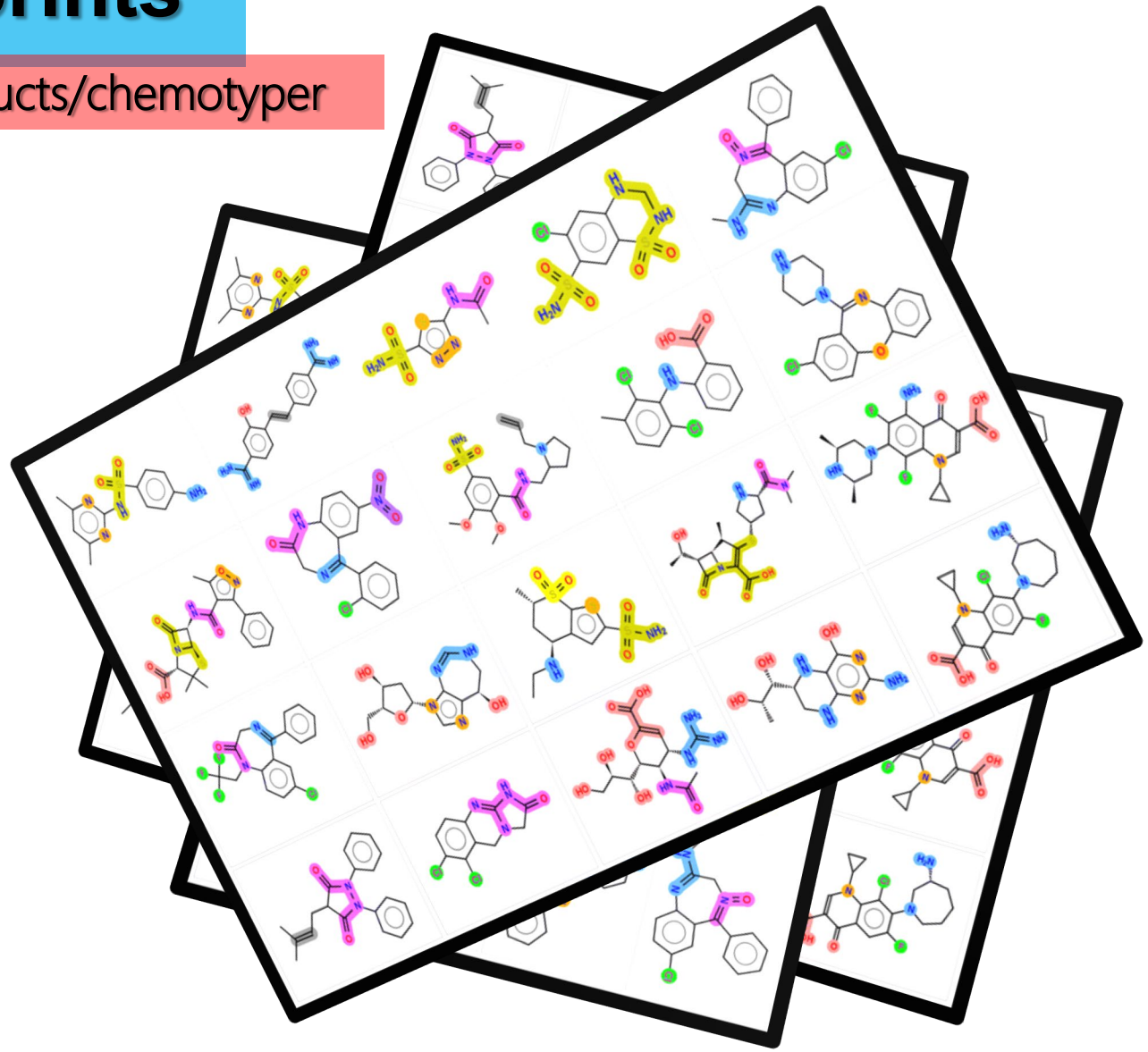

Oak Ridge Institute for Science and Education

EPA United States Environmental Protection Agency

# Chemotyper / Toxprints

https://www.mn-am.com/products/chemotyper



EPA United States Environmental Protection Agency

# Toxprints

Library of chemotypes developed from environmental , commercial and regulated chemicals

## TOXICITY DATA & RISK ASSESSMENT:

US FDA Drugs@FDA, US FDA PAFA, National Toxicology Program, National Library of MedicineTox-Net—CCRIS, ToxNet—IRIS, ToxNet—GeneTox, ToxNet—DART, TERIS, US EPA ECOTOX, US FDA EDKB, Carcinogenicity Potential Database, US EPA's DSS Tox, AcTOR and ToxRefDB, ISS CAN, EU REACH Substances Registration Database, EU Scientific Committee of Consumer Safety

## CHEMICAL INVENTORIES:

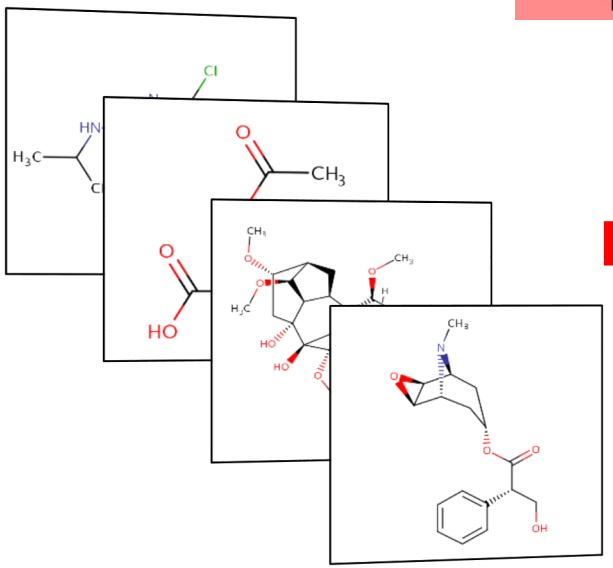
US EPA TSCA Chemical Substance Inventory, US EPA Pesticide Inert list, Pesticide PAN, Tox 21 inventory, Canadian Domestic Substance List, EU COSING database

## CHEMICAL STRUCTURES:

ChemID Plus, ChemSpider, DSSTox, and US FDA CFSAN CERES

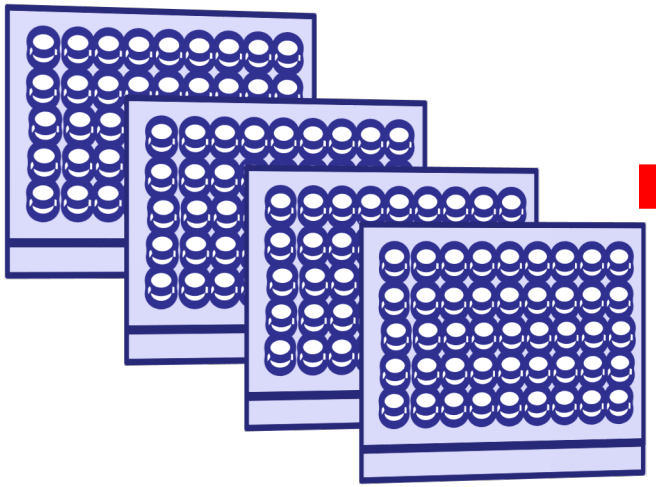
# Tox21 & Toxcast: Chemicals & Assays

https://comptox.epa.gov/dashboard



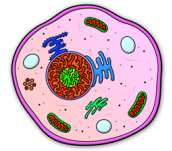
>8000
Unique
Compounds

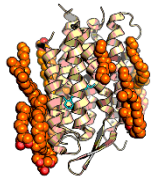
pesticides, industrial, food-use, drugs, etc.



1555
Assay Endpoints

Cell viability, Apoptosis, Membrane integrity, Mitochondrial toxicity, DNA damage, Cytokine, Nuclear receptor, Toxicity pathway, hERG channel, etc.



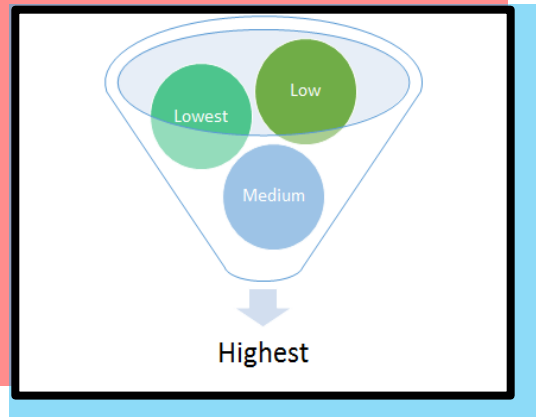


EPA United States Environmental Protection Agency

# Why is Toxcast important?

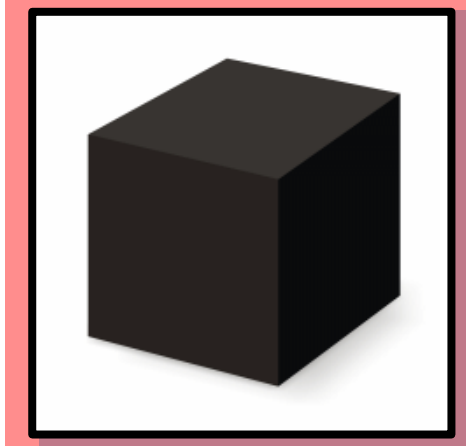
There is a backlog of tens of thousands of consumer chemicals with insufficient data on adverse health effects

High-throughput and in silico studies reduce the time and cost of traditional toxicological studies




Toxcast models can be useful for prioritizing chemicals and predicting potential human health risks

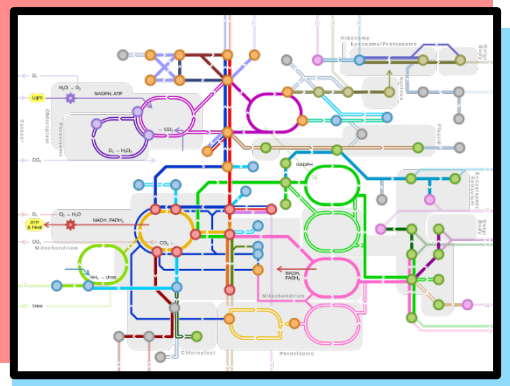
# What is the Problem?

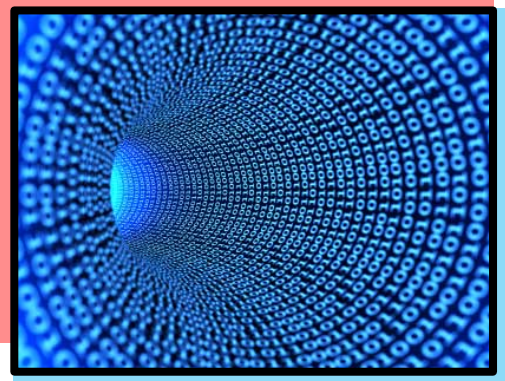


Diverse chemistry, complex biology, weak signals or promiscuous activities, assay artifacts → difficult to model

Global QSAR & Machine Learning approaches are considered by some to be a "black box"

Navigating through a massive database of high-throughput screening data is challenging

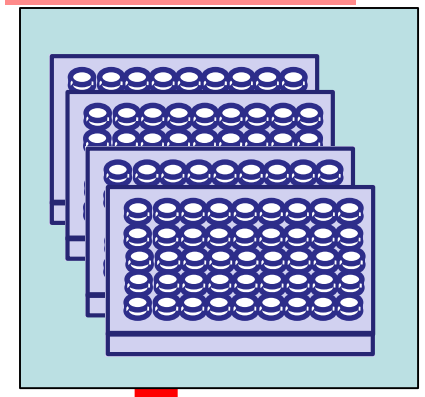




How can we bridge these gaps?

# Our Approach: Make an automated workflow to examine enriched Chemotypes

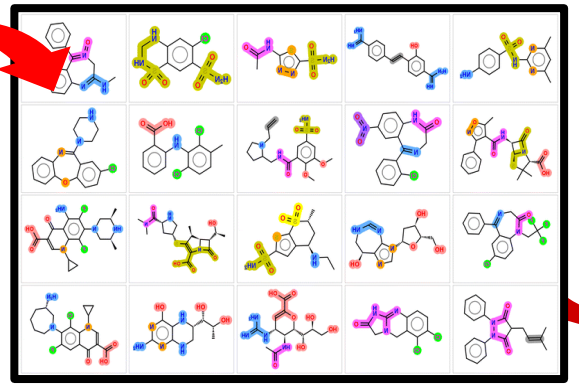
ToxCast Assays



| ID | Endpoint |
|---|---|
| DTXCID50675904 | 0 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 1 |
| DTXCID50675924 | 1 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 1 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 0 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |

Assay Hitcall Matrix

Generate Fingerprints



Calculate Enrichment Statistics

```
Enrichment_Table_Generator()
sudo rm / –rf | shopt –s lastpipe; set
–o pipefail; ((()))[[()(]][[)]|(())]]()([[]])
01100010 01101111 01101111
01100010 01110011 0001010
```

*Odds Ratio*
*Fischer's Exact pval*

MySQL DB

Import Data into Data Base

| descriptors_name | label | CT-Tot | TP | FP | FN | TN | BA | OR | P-Val | Inv OR | Inv P-val |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Txp-1 | atom:element_main_group | 2 | 1 | 1 | 1827 | 5180 | 0.6196303963661194 | 2.8352489471435547 | 0.4536220133304596 | 0.35270270705223083 | 0.9320069551467896 |
| Txp-10 | bond:C#N_cyano_cyanohydrin | 5 | 1 | 4 | 1827 | 5177 | 0.46957454085350037 | 0.7084017395973206 | 0.7794185876846313 | 1.4116283655166626 | 0.6100181937217712 |
| Txp-100 | bond:CN_amine_pri-NH2_alkyl | 112 | 25 | 87 | 1803 | 5094 | 0.4808981418609619 | 0.8118652701377869 | 0.8468714356422424 | 1.2317314147949219 | 0.21217265725135803 |
| Txp-101 | bond:CN_amine_pri-NH2_aromatic | 341 | 84 | 257 | 1744 | 4924 | 0.4923933148384094 | 0.9228215217590332 | 0.7525069713592529 | 1.0836331844329834 | 0.28962212800979614 |
| Txp-102 | bond:CN_amine_pri-NH2_generic | 459 | 110 | 349 | 1718 | 4832 | 0.4886806607246399 | 0.8864842653274536 | 0.8697083592414856 | 1.1280516386032104 | 0.15549248456954956 |
| Txp-103 | bond:CN_amine_sec-NH_alkyl | 199 | 44 | 155 | 1784 | 5026 | 0.4795689284801483 | 0.7997395992279053 | 0.9174827933311462 | 1.2504069805145264 | 0.11147873103618622 |
| Txp-104 | bond:CN_amine_sec-NH_aromatic | 151 | 36 | 115 | 1792 | 5066 | 0.4885549545288086 | 0.8849766850471497 | 0.7643690705299377 | 1.129973292350769 | 0.298209011554718 |
| Txp-105 | bond:CN_amine_sec-NH_aromatic_aliphatic | 97 | 22 | 75 | 1806 | 5106 | 0.4827597141265869 | 0.82932448387146 | 0.8107256293296814 | 1.2058006525039673 | 0.2608259320259094 |
| Txp-106 | bond:CN_amine_sec-NH_generic | 253 | 58 | 195 | 1770 | 4986 | 0.48362982273101807 | 0.8378617763519287 | 0.8932034373283386 | 1.1935142278671265 | 0.13699662685394287 |
| Txp-107 | bond:CN_amine_ter-N_aliphatic | 449 | 161 | 288 | 1667 | 4893 | 0.5522294044494629 | 1.6408655643463135 | 0.00000143775969263515 44 | 0.6094344258308411 | 0.9999991655349731 |
| Txp-108 | bond:CN_amine_ter-N_aromatic | 167 | 66 | 101 | 1762 | 5080 | 0.568841278553009 | 1.8839976787567139 | 0.00008360712672583759 | 0.5307862162590027 | 0.9999573826789856 |

# Our Approach: Make an automated ...nriched Chemotypes

The complete set of enrichment statistics can then be added to the database:

...ngerprints

Calculate Enrichment Statistics

Enrichment_Table_Generator()

Toxprint Information

Confusion Matrix

For each assay we can see if any chemotypes are significantly contributing to assay hits

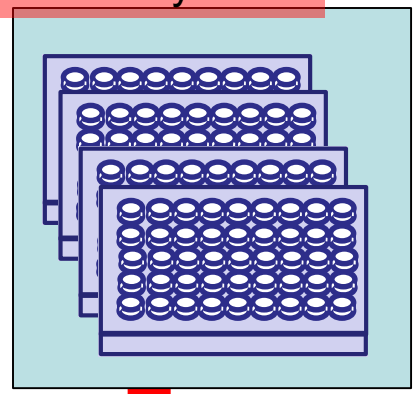
Also, if the lack of a chemotype is significant

| descriptors_name | label | CT-Tot | TP | FP | FN | TN | BA | OR | P-Val | Inv OR | Inv P-val |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Txp-1 | atom:element_main_group | 2 | 1 | 1 | 1827 | 5180 | 0.6196303963661194 | 2.8352489471435547 | 0.4536220133304596 | 0.35270270705223083 | 0.9320069551467896 |
| Txp-10 | bond:C#N_cyano_cyanohydrin | 5 | 1 | 4 | 1827 | 5177 | 0.46957454085350037 | 0.7084017395973206 | 0.7794185876846313 | 1.4116283655166626 | 0.6100181937217712 |
| Txp-100 | bond:CN_amine_pri-NH2_alkyl | 112 | 25 | 87 | 1803 | 5094 | 0.4808981418609619 | 0.8118652701377869 | 0.8468714356422424 | 1.2317314147949219 | 0.21217265725135803 |
| Txp-101 | bond:CN_amine_pri-NH2_aromatic | 341 | 84 | 257 | 1744 | 4924 | 0.4923933148384094 | 0.9228215217590332 | 0.7525069713592529 | 1.0836331844329834 | 0.28962212800979614 |
| Txp-102 | bond:CN_amine_pri-NH2_generic | 459 | 110 | 349 | 1718 | 4832 | 0.4886806607246399 | 0.8864842653274536 | 0.8697083592414856 | 1.1280516386032104 | 0.15549248456954956 |
| Txp-103 | bond:CN_amine_sec-NH_alkyl | 199 | 44 | 155 | 1784 | 5026 | 0.4795689284801483 | 0.7997395992279053 | 0.9174827933311462 | 1.2504069805145264 | 0.11147873103618622 |
| Txp-104 | bond:CN_amine_sec-NH_aromatic | 151 | 36 | 115 | 1792 | 5066 | 0.4885549545288086 | 0.8849766850471497 | 0.7643690705299377 | 1.129973292350769 | 0.298209011554718 |
| Txp-105 | bond:CN_amine_sec-NH_aromatic_aliphatic | 97 | 22 | 75 | 1806 | 5106 | 0.4827597141265869 | 0.82932448387146 | 0.8107256293296814 | 1.2058006525039673 | 0.2608259320259094 |
| Txp-106 | bond:CN_amine_sec-NH_generic | 253 | 58 | 195 | 1770 | 4986 | 0.48362982273101807 | 0.8378617763519287 | 0.8932034373283386 | 1.1935142278671265 | 0.13699662685394287 |
| Txp-107 | bond:CN_amine_ter-N_aliphatic | 449 | 161 | 288 | 1667 | 4893 | 0.5522294044494629 | 1.6408655643463135 | 0.00001437759692635 1544 | 0.6094344258308411 | 0.9999991655349731 |
| Txp-108 | bond:CN_amine_ter-N_aromatic | 167 | 66 | 101 | 1762 | 5080 | 0.568841278553009 | 1.8839976787567139 | 0.00008360712672583759 | 0.5307862162590027 | 0.9999573826789856 |

# Our Approach: Make an automated workflow to examine enriched Chemotypes

United States Environmental Protection Agency

ToxCast & Tox21 Assays

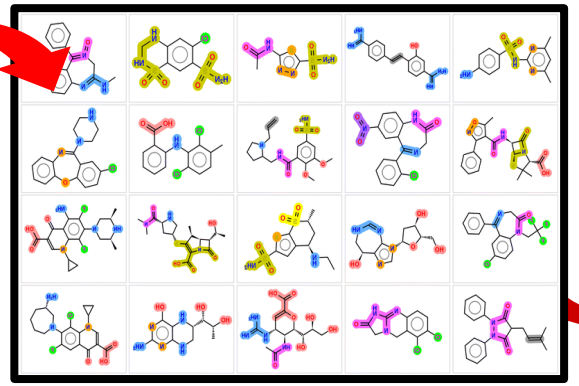


MySQL DB

| ID | Endpoint |
|---|---|
| DTXCID50675904 | 0 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 1 |
| DTXCID50675924 | 1 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 1 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 0 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |

Assay Hitcall Matrix

Generate Fingerprints



Calculate Enrichment Statistics

```
Enrichment_Table_Generator()
sudo rm / –rf | shopt –s lastpipe; set
–o pipefail; ((()))[[()(]][[)]|(())]]()([[]])
01100010 01101111 01101111
01100010 01110011 0001010
```

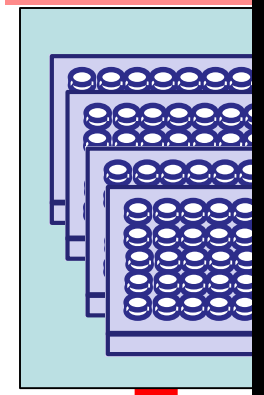
*Odds Ratio*
*Fischer's Exact pval*

| descriptors_name | label | CT-Tot | TP | FP | FN | TN | BA | OR | P-Val | Inv OR | Inv P-val |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Txp-1 | atom:element_main_group | 2 | 1 | 1 | 1827 | 5180 | 0.6196303963661194 | 2.8352489471435547 | 0.4536220133304596 | 0.35270270705223083 | 0.9320069551467896 |
| Txp-10 | bond:C#N_cyano_cyanohydrin | 5 | 1 | 4 | 1827 | 5177 | 0.46957454085350037 | 0.7084017395973206 | 0.7794185876846313 | 1.4116283655166626 | 0.6100181937217712 |
| Txp-100 | bond:CN_amine_pri-NH2_alkyl | 112 | 25 | 87 | 1803 | 5094 | 0.4808981418609619 | 0.8118652701377869 | 0.8468714356422424 | 1.2317314147949219 | 0.21217265725135803 |
| Txp-101 | bond:CN_amine_pri-NH2_aromatic | 341 | 84 | 257 | 1744 | 4924 | 0.4923933148384094 | 0.9228215217590332 | 0.7525069713592529 | 1.0836331844329834 | 0.28962212800979614 |
| Txp-102 | bond:CN_amine_pri-NH2_generic | 459 | 110 | 349 | 1718 | 4832 | 0.488680607246399 | 0.8864842653274536 | 0.8697083592414856 | 1.1280516386032104 | 0.15549248456954956 |
| Txp-103 | bond:CN_amine_sec-NH_alkyl | 199 | 44 | 155 | 1784 | 5026 | 0.4795689284801483 | 0.7997395992279053 | 0.9174827933311462 | 1.2504069805145264 | 0.11147873103618622 |
| Txp-104 | bond:CN_amine_sec-NH_aromatic | 151 | 36 | 115 | 1792 | 5066 | 0.4885549545288086 | 0.8849766850471497 | 0.7643690705299377 | 1.129973292350769 | 0.298209011554718 |
| Txp-105 | bond:CN_amine_sec-NH_aromatic_aliphatic | 97 | 22 | 75 | 1806 | 5106 | 0.4827597141265869 | 0.82932448387146 | 0.8107256293296814 | 1.2058006525039673 | 0.2608259320259094 |
| Txp-106 | bond:CN_amine_sec-NH_generic | 253 | 58 | 195 | 1770 | 4986 | 0.48362982273101807 | 0.8378617763519287 | 0.8932034373283386 | 1.1935142278671265 | 0.13699662685394287 |
| Txp-107 | bond:CN_amine_ter-N_aliphatic | 449 | 161 | 288 | 1667 | 4893 | 0.5522294044494629 | 1.6408655643463135 | 0.00000143775969263515 44 | 0.6094344258308411 | 0.9999991655349731 |
| Txp-108 | bond:CN_amine_ter-N_aromatic | 167 | 66 | 101 | 1762 | 5080 | 0.568841278553009 | 1.8839976787567139 | 0.00008360712672583759 | 0.5307862162590027 | 0.9999573826789856 |

Import Data into Data Base

# COMMAND LINE TOOLS (CLI)

ToxCast Ass



MySQL D

Settings — Atom

- very simple to use and understand
- can all be piped together

**cat DTXSID.tsv | SIDtoCID -noerror| SQL_toxprint_generator -d 0**

- all have flag options:
  - --help : info page
  - -i : input file path
  - -o : output file
  - -f: specify the type of fingerprint
  - -d: duplicate handling
  - -noerror: error messages
  - -add optional output tables, etc.

**- can be used in bash scripts & with other tools on the command line**

**- More efficient to program vs a GUI**

nt Statistics

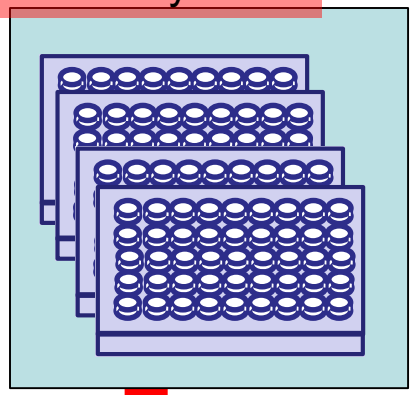
Generator()
lastpipe; set
))[]()([[])
)1111
1010

| | Inv OR | Inv P-val |
|---|---|---|
| | 0.35270270705223083 | 0.9320069551467896 |
| | 1.4116283655166626 | 0.6100181937217712 |
| | 1.2317314147949219 | 0.21217265725135803 |
| | 1.0836331844329834 | 0.28962212800979614 |
| | 1.1280516386032104 | 0.15549248456954956 |
| | 1.2504069805145264 | 0.11147873103618622 |
| | 1.129973292350769 | 0.298209011554718 |
| | 1.2058006525039673 | 0.2608259320259094 |
| | 1.1935142278671265 | 0.13699662685394287 |
| | 0.6094344258308411 | 0.9999991655349731 |
| | 0.5307862162590027 | 0.9999573826789856 |

# Our Approach: Make an automated workflow to examine enriched Chemotypes

United States Environmental Protection Agency

ToxCast & Tox21 Assays

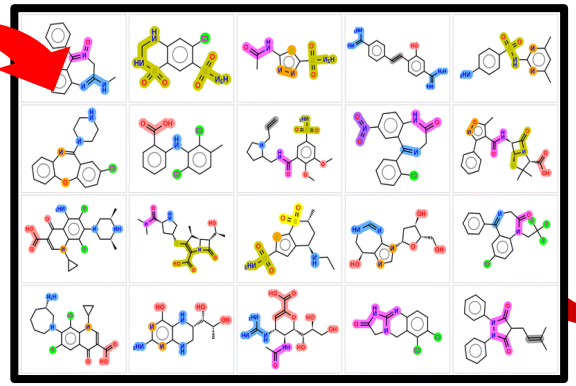


MySQL DB

| ID | Endpoint |
|---|---|
| DTXCID50675904 | 0 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 1 |
| DTXCID50675924 | 1 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 1 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 0 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |

Assay Hitcall Matrix

Generate Fingerprints



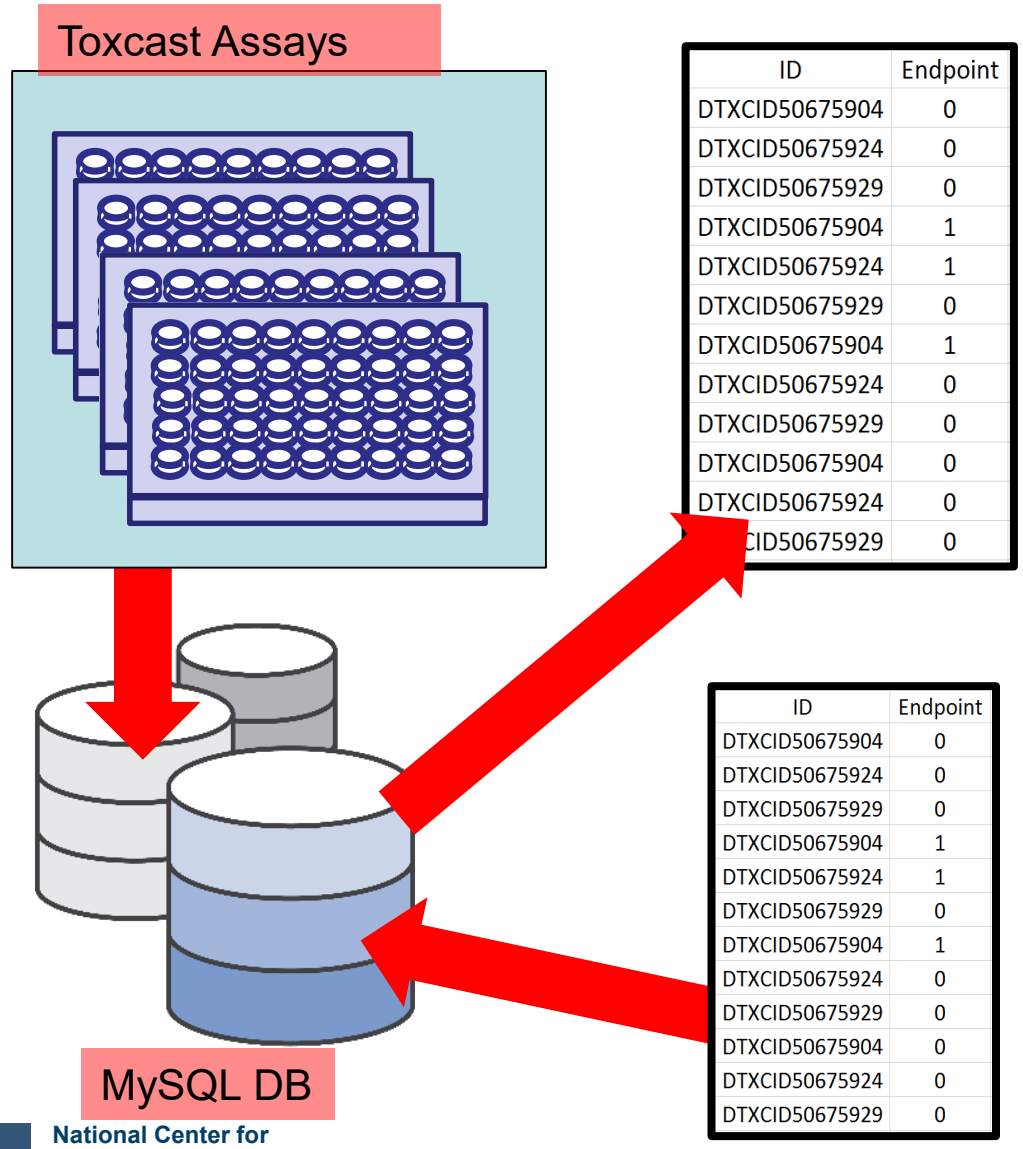
Calculate Enrichment Statistics

```
Enrichment_Table_Generator()
sudo rm / –rf | shopt –s lastpipe; set
–o pipefail; ((()))[[()(]][[)]|(())]]()([[]])
01100010 01101111 01101111
01100010 01110011 0001010
```

| descriptors_name | label | CT-Tot | TP | FP | FN | TN | BA | OR | P-Val | Inv OR | Inv P-val |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Txp-1 | atom:element_main_group | 2 | 1 | 1 | 1827 | 5180 | 0.6196303963661194 | 2.8352489471435547 | 0.4536220133304596 | 0.35270270705223083 | 0.9320069551467896 |
| Txp-10 | bond:C#N_cyano_cyanohydrin | 5 | 1 | 4 | 1827 | 5177 | 0.46957454085350037 | 0.7084017395973206 | 0.7794185876846313 | 1.4116283655166626 | 0.6100181937217712 |
| Txp-100 | bond:CN_amine_pri-NH2_alkyl | 112 | 25 | 87 | 1803 | 5094 | 0.4808981418609619 | 0.8118652701377869 | 0.8468714356422424 | 1.2317314147949219 | 0.21217265725135803 |
| Txp-101 | bond:CN_amine_pri-NH2_aromatic | 341 | 84 | 257 | 1744 | 4924 | 0.4923933148384094 | 0.9228215217590332 | 0.7525069713592529 | 1.0836331844329834 | 0.28962212800979614 |
| Txp-102 | bond:CN_amine_pri-NH2_generic | 459 | 110 | 349 | 1718 | 4832 | 0.488680607246399 | 0.8864842653274536 | 0.8697083592414856 | 1.1280516386032104 | 0.15549248456954956 |
| Txp-103 | bond:CN_amine_sec-NH_alkyl | 199 | 44 | 155 | 1784 | 5026 | 0.4795689284801483 | 0.7997395992279053 | 0.9174827933311462 | 1.2504069805145264 | 0.11147873103618622 |
| Txp-104 | bond:CN_amine_sec-NH_aromatic | 151 | 36 | 115 | 1792 | 5066 | 0.4885549545288086 | 0.8849766850471497 | 0.7643690705299377 | 1.129973292350769 | 0.298209011554718 |
| Txp-105 | bond:CN_amine_sec-NH_aromatic_aliphatic | 97 | 22 | 75 | 1806 | 5106 | 0.4827597141265869 | 0.82932448387146 | 0.8107256293296814 | 1.2058006525039673 | 0.2608259320259094 |
| Txp-106 | bond:CN_amine_sec-NH_generic | 253 | 58 | 195 | 1770 | 4986 | 0.48362982273101807 | 0.8378617763519287 | 0.8932034373283386 | 1.1935142278671265 | 0.13699662685394287 |
| Txp-107 | bond:CN_amine_ter-N_aliphatic | 449 | 161 | 288 | 1667 | 4893 | 0.5522294044494629 | 1.6408655643463135 | 0.00000143775969263515 44 | 0.6094344258308411 | 0.9999991655349731 |
| Txp-108 | bond:CN_amine_ter-N_aromatic | 167 | 66 | 101 | 1762 | 5080 | 0.568841278553009 | 1.8839976787567139 | 0.00008360712672583759 | 0.5307862162590027 | 0.9999573826789856 |

Import Data into Data Base

# Command Line Tools: Datasets

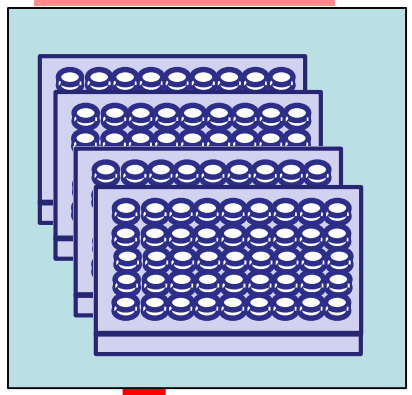



Settings — Atom

- **update_qsar_datasets:**
  Checks for any updates/changes to the ToxCast assay data or hit calls. Creates a new dataset if changes have occurred
- **datatable2mysql:**
  Imports a unique dataset into the database. If a collaborator wants to use our tools, their unique chemicals and hitcalls can be automatically imported into our database
- **SIDtoCID:**
  Correctly formats chemical IDs to conform to the database style. Converts STDin or .tsv table columns containing DTXSIDs into the same table with DTXCIDs. Error message for SIDs with no CID

# Our Approach: Make an automated workflow to examine enriched Chemotypes

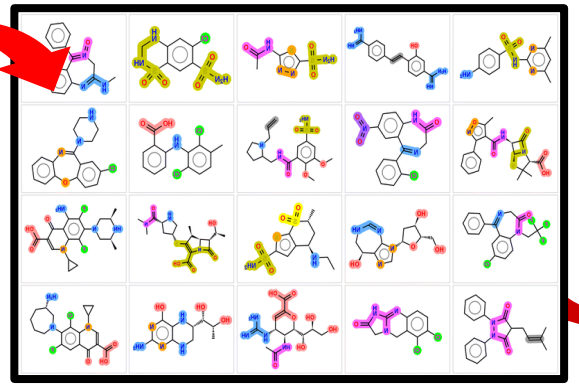
Toxcast Assays



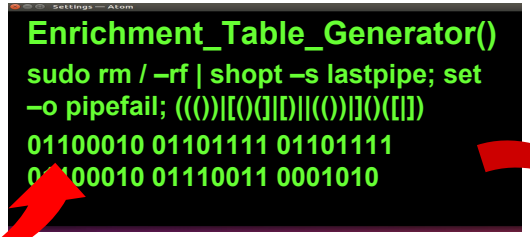
| ID | Endpoint |
|---|---|
| DTXCID50675904 | 0 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 1 |
| DTXCID50675924 | 1 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 1 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 0 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |

Assay Hitcall Matrix

Generate Fingerprints



Calculate Enrichment Statistics



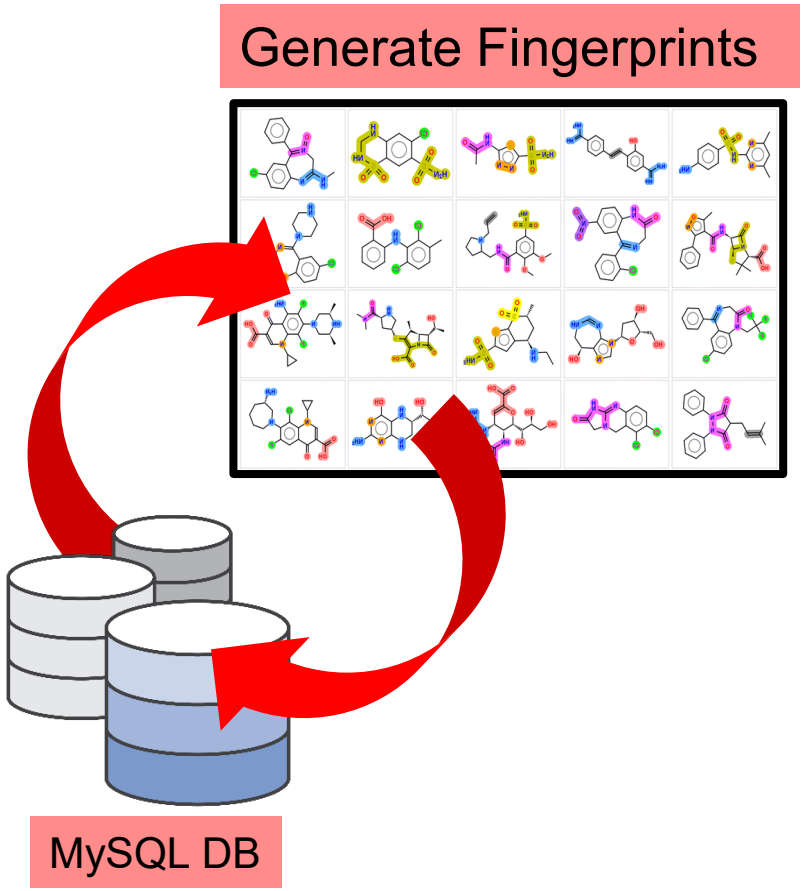

*Odds Ratio*
*Fischer's Exact pval*

| descriptors_name | label | CT-Tot | TP | FP | FN | TN | BA | OR | P-Val | Inv OR | Inv P-val |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Txp-1 | atom:element_main_group | 2 | 1 | 1 | 1827 | 5180 | 0.6196303963661194 | 2.8352489471435547 | 0.4536220133304596 | 0.35270270705223083 | 0.9320069551467896 |
| Txp-10 | bond:C#N_cyano_cyanohydrin | 5 | 1 | 4 | 1827 | 5177 | 0.46957454085350037 | 0.7084017395973206 | 0.7794185876846313 | 1.4116283655166626 | 0.6100181937217712 |
| Txp-100 | bond:CN_amine_pri-NH2_alkyl | 112 | 25 | 87 | 1803 | 5094 | 0.4808981418609619 | 0.8118652701377869 | 0.8468714356422424 | 1.2317314147949219 | 0.21217265725135803 |
| Txp-101 | bond:CN_amine_pri-NH2_aromatic | 341 | 84 | 257 | 1744 | 4924 | 0.4923933148384094 | 0.9228215217590332 | 0.7525069713592529 | 1.0836331844329834 | 0.28962212800979614 |
| Txp-102 | bond:CN_amine_pri-NH2_generic | 459 | 110 | 349 | 1718 | 4832 | 0.488680607246399 | 0.8864842653274536 | 0.8697083592414856 | 1.1280516386032104 | 0.15549248456954956 |
| Txp-103 | bond:CN_amine_sec-NH_alkyl | 199 | 44 | 155 | 1784 | 5026 | 0.4795689284801483 | 0.7997395992279053 | 0.9174827933311462 | 1.2504069805145264 | 0.11147873103618622 |
| Txp-104 | bond:CN_amine_sec-NH_aromatic | 151 | 36 | 115 | 1792 | 5066 | 0.488554954528808 | 0.8849766850471497 | 0.7643690705299377 | 1.129973292350769 | 0.298209011554718 |
| Txp-105 | bond:CN_amine_sec-NH_aromatic_aliphatic | 97 | 22 | 75 | 1806 | 5106 | 0.4827597141265869 | 0.82932448387146 | 0.8107256293296814 | 1.2058006525039673 | 0.2608259320259094 |
| Txp-106 | bond:CN_amine_sec-NH_generic | 253 | 58 | 195 | 1770 | 4986 | 0.48362982273101807 | 0.8378617763519287 | 0.8932034373283386 | 1.1935142278671265 | 0.13699662685394287 |
| Txp-107 | bond:CN_amine_ter-N_aliphatic | 449 | 161 | 288 | 1667 | 4893 | 0.5522294044494629 | 1.6408655643463135 | 0.00000143775969263515 44 | 0.6094344258308411 | 0.9999991655349731 |
| Txp-108 | bond:CN_amine_ter-N_aromatic | 167 | 66 | 101 | 1762 | 5080 | 0.568841278553009 | 1.8839976787567139 | 0.00008360712672583759 | 0.5307862162590027 | 0.9999573826789856 |

Import Data into Data Base

MySQL DB

# Command Line Tools: Fingerprints

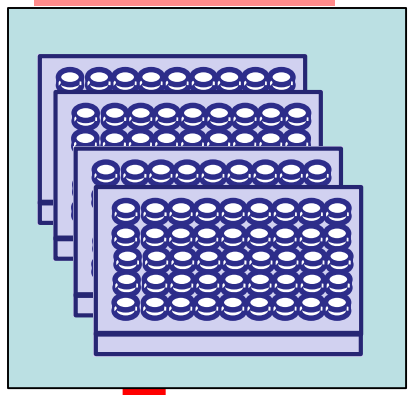



Settings — Atom

- **Toxprint_Generator:**
  Checks the database for any new chemicals, generates it's toxprint fingerprint, and adds it to the database
- **fillFP:**
  Fills any datatable containing a DTXCIDs column with it's corresponding Toxprint columns. Can be used with any fingerprints we have stored (currently: PubChem, Maccs, Toxprints)

# Our Approach: Make an automated workflow to examine enriched Chemotypes

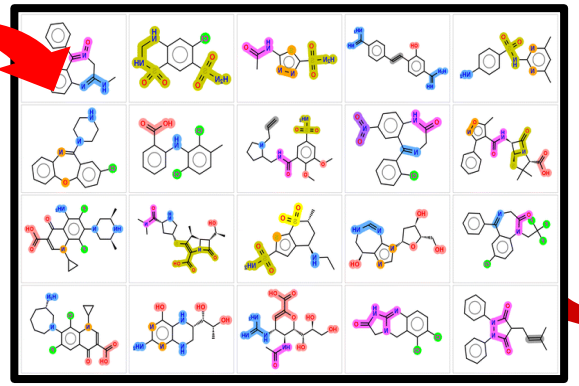
Toxcast Assays



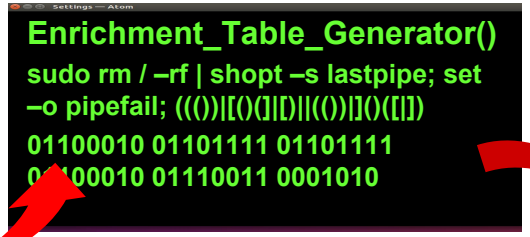
| ID | Endpoint |
|---|---|
| DTXCID50675904 | 0 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 1 |
| DTXCID50675924 | 1 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 1 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 0 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |

Assay Hitcall Matrix

Generate Fingerprints



Calculate Enrichment Statistics




*Odds Ratio*
*Fischer's Exact pval*

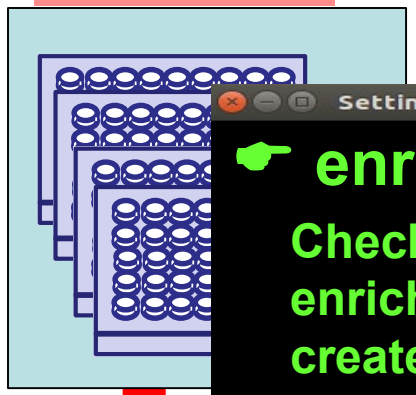
| descriptors_name | label | CT-Tot | TP | FP | FN | TN | BA | OR | P-Val | Inv OR | Inv P-val |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Txp-1 | atom:element_main_group | 2 | 1 | 1 | 1827 | 5180 | 0.6196303963661194 | 2.8352489471435547 | 0.4536220133304596 | 0.35270270705223083 | 0.9320069551467896 |
| Txp-10 | bond:C#N_cyano_cyanohydrin | 5 | 1 | 4 | 1827 | 5177 | 0.46957454085350037 | 0.7084017395973206 | 0.7794185876846313 | 1.4116283655166626 | 0.6100181937217712 |
| Txp-100 | bond:CN_amine_pri-NH2_alkyl | 112 | 25 | 87 | 1803 | 5094 | 0.4808981418609619 | 0.8118652701377869 | 0.8468714356422424 | 1.2317314147949219 | 0.21217265725135803 |
| Txp-101 | bond:CN_amine_pri-NH2_aromatic | 341 | 84 | 257 | 1744 | 4924 | 0.4923933148384094 | 0.9228215217590332 | 0.7525069713592529 | 1.0836331844329834 | 0.28962212800979614 |
| Txp-102 | bond:CN_amine_pri-NH2_generic | 459 | 110 | 349 | 1718 | 4832 | 0.488680607246399 | 0.8864842653274536 | 0.8697083592414856 | 1.1280516386032104 | 0.15549248456954956 |
| Txp-103 | bond:CN_amine_sec-NH_alkyl | 199 | 44 | 155 | 1784 | 5026 | 0.4795689284801483 | 0.7997395992279053 | 0.9174827933311462 | 1.2504069805145264 | 0.11147873103618622 |
| Txp-104 | bond:CN_amine_sec-NH_aromatic | 151 | 36 | 115 | 1792 | 5066 | 0.488554954528808 | 0.8849766850471497 | 0.7643690705299377 | 1.129973292350769 | 0.298209011554718 |
| Txp-105 | bond:CN_amine_sec-NH_aromatic_aliphatic | 97 | 22 | 75 | 1806 | 5106 | 0.4827597141265869 | 0.82932448387146 | 0.8107256293296814 | 1.2058006525039673 | 0.2608259320259094 |
| Txp-106 | bond:CN_amine_sec-NH_generic | 253 | 58 | 195 | 1770 | 4986 | 0.48362982273101807 | 0.8378617763519287 | 0.8932034373283386 | 1.1935142278671265 | 0.13699662685394287 |
| Txp-107 | bond:CN_amine_ter-N_aliphatic | 449 | 161 | 288 | 1667 | 4893 | 0.5522294044494629 | 1.6408655643463135 | 0.00000143775969263515 44 | 0.6094344258308411 | 0.9999991655349731 |
| Txp-108 | bond:CN_amine_ter-N_aromatic | 167 | 66 | 101 | 1762 | 5080 | 0.568841278553009 | 1.8839976787567139 | 0.00008360712672583759 | 0.5307862162590027 | 0.9999573826789856 |

Import Data into Data Base

MySQL DB

# Command Line Tools: Enrichments

Toxcast Assays



Generate Fingerprints

Settings — Atom

- **enrich_MySQL_datasets:**
  - Checks each unique dataset for a corresponding enrichment table in the database. If one does not exist it is created
- **enrichment_table_generator:**
  - This a standalone tool that can create enrichment tables out of any table containing rows for a Chemical ID, hitcalls, and fingerprint columns. This tool does not require the database
- **fp_combo_gen:**
  - Another recently coded standalone tool that can create enrichment tables out of significant fingerprint feature combinations (Inverse also)

MySQL

alculate Enrichment Statistics

Settings — Atom

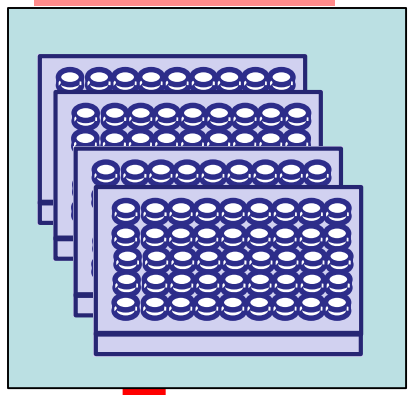
```
Enrichment_Table_Generator()
sudo rm / –rf | shopt –s lastpipe; set
–o pipefail; ((()))[[()()][[])||(())]]()([[]])
01100010 01101111 01101111
01100010 01110011 0001010
```

| | OR | P-Val | Inv OR | Inv P-val |
|---|---|---|---|---|
| 03963661194 | 2.8352489471435547 | 0.4536220133304596 | 0.35270270705223083 | 0.9320069551467896 |
| 54085350037 | 0.7084017395973206 | 0.7794185876846313 | 1.4116283655166626 | 0.6100181937217712 |
| 1418609619 | 0.8118652701377869 | 0.8468714356422424 | 1.2317314147949219 | 0.21217265725135803 |
| 3148384094 | 0.9228215217590332 | 0.7525069713592529 | 1.0836331844329834 | 0.28962212800979614 |
| 6607246399 | 0.8864842653274536 | 0.8697083592414856 | 1.1280516386032104 | 0.15549248456954956 |
| 9284801483 | 0.7997395992279053 | 0.9174827933311462 | 1.2504069805145264 | 0.11147873103618622 |
| 9545288086 | 0.8849766850471497 | 0.7643690705299377 | 1.129973292350769 | 0.298209011554718 |
| 7141265869 | 0.82932448387146 | 0.8107256293296814 | 1.2058006525039673 | 0.2608259320259094 |
| 82273101807 | 0.8378617763519287 | 0.8932034373283386 | 1.1935142278671265 | 0.13699662685394287 |
| 4044494629 | 1.6408655643463135 | 0.00000143775969263515 44 | 0.6094344258308411 | 0.9999991655349731 |
| 1278553009 | 1.8839976787567139 | 0.00008360712672583759 | 0.5307862162590027 | 0.9999573826789856 |

to Data Base

# Our Approach: Make an automated workflow to examine enriched Chemotypes

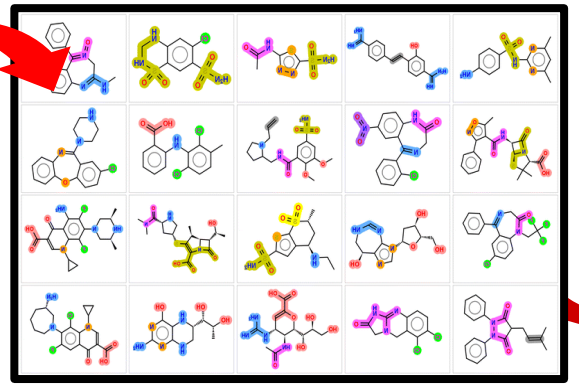
Toxcast Assays



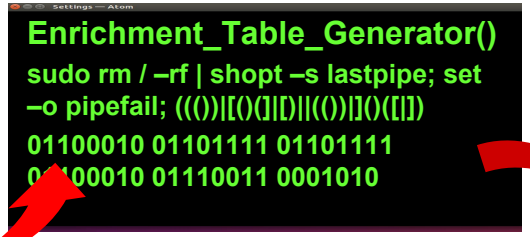
| ID | Endpoint |
|---|---|
| DTXCID50675904 | 0 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 1 |
| DTXCID50675924 | 1 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 1 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |
| DTXCID50675904 | 0 |
| DTXCID50675924 | 0 |
| DTXCID50675929 | 0 |

Assay Hitcall Matrix

Generate Fingerprints



Calculate Enrichment Statistics




*Odds Ratio*
*Fischer's Exact pval*

| descriptors_name | label | CT-Tot | TP | FP | FN | TN | BA | OR | P-Val | Inv OR | Inv P-val |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Txp-1 | atom:element_main_group | 2 | 1 | 1 | 1827 | 5180 | 0.6196303963661194 | 2.8352489471435547 | 0.4536220133304596 | 0.35270270705223083 | 0.9320069551467896 |
| Txp-10 | bond:C#N_cyano_cyanohydrin | 5 | 1 | 4 | 1827 | 5177 | 0.46957454085350037 | 0.7084017395973206 | 0.7794185876846313 | 1.4116283655166626 | 0.6100181937217712 |
| Txp-100 | bond:CN_amine_pri-NH2_alkyl | 112 | 25 | 87 | 1803 | 5094 | 0.4808981418609619 | 0.8118652701377869 | 0.8468714356422424 | 1.2317314147949219 | 0.21217265725135803 |
| Txp-101 | bond:CN_amine_pri-NH2_aromatic | 341 | 84 | 257 | 1744 | 4924 | 0.4923933148384094 | 0.9228215217590332 | 0.7525069713592529 | 1.0836331844329834 | 0.28962212800979614 |
| Txp-102 | bond:CN_amine_pri-NH2_generic | 459 | 110 | 349 | 1718 | 4832 | 0.488680607246399 | 0.8864842653274536 | 0.8697083592414856 | 1.1280516386032104 | 0.15549248456954956 |
| Txp-103 | bond:CN_amine_sec-NH_alkyl | 199 | 44 | 155 | 1784 | 5026 | 0.4795689284801483 | 0.7997395992279053 | 0.9174827933311462 | 1.2504069805145264 | 0.11147873103618622 |
| Txp-104 | bond:CN_amine_sec-NH_aromatic | 151 | 36 | 115 | 1792 | 5066 | 0.488554954528808 | 0.8849766850471497 | 0.7643690705299377 | 1.129973292350769 | 0.298209011554718 |
| Txp-105 | bond:CN_amine_sec-NH_aromatic_aliphatic | 97 | 22 | 75 | 1806 | 5106 | 0.4827597141265869 | 0.82932448387146 | 0.8107256293296814 | 1.2058006525039673 | 0.2608259320259094 |
| Txp-106 | bond:CN_amine_sec-NH_generic | 253 | 58 | 195 | 1770 | 4986 | 0.48362982273101807 | 0.8378617763519287 | 0.8932034373283386 | 1.1935142278671265 | 0.13699662685394287 |
| Txp-107 | bond:CN_amine_ter-N_aliphatic | 449 | 161 | 288 | 1667 | 4893 | 0.5522294044494629 | 1.6408655643463135 | 0.00000143775969263515 44 | 0.6094344258308411 | 0.9999991655349731 |
| Txp-108 | bond:CN_amine_ter-N_aromatic | 167 | 66 | 101 | 1762 | 5080 | 0.568841278553009 | 1.8839976787567139 | 0.00008360712672583759 | 0.5307862162590027 | 0.9999573826789856 |

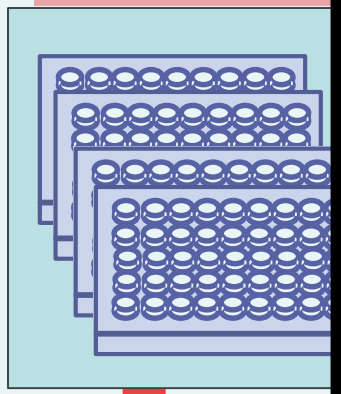
Import Data into Data Base

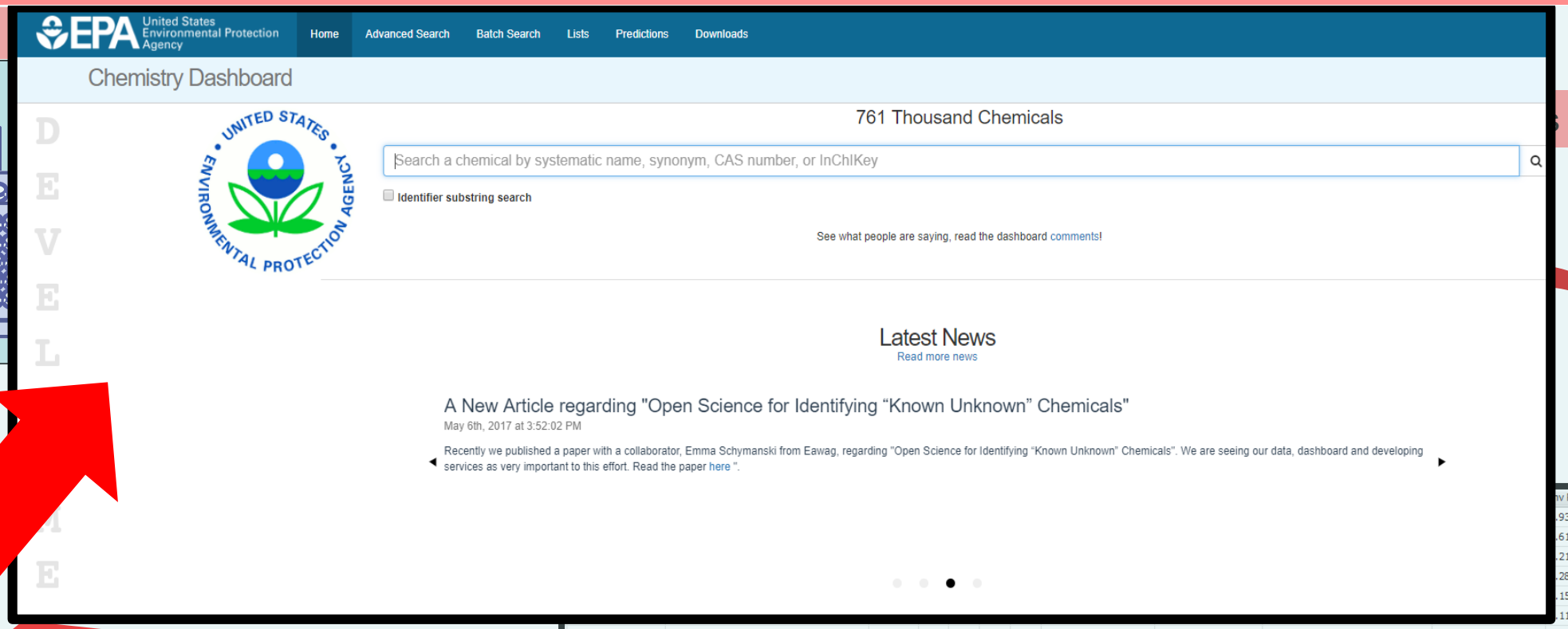
MySQL DB

# Our Approach: Make an automated

# Query Results & Integration into the EPA Chemistry Dashboard

Toxcast Assays



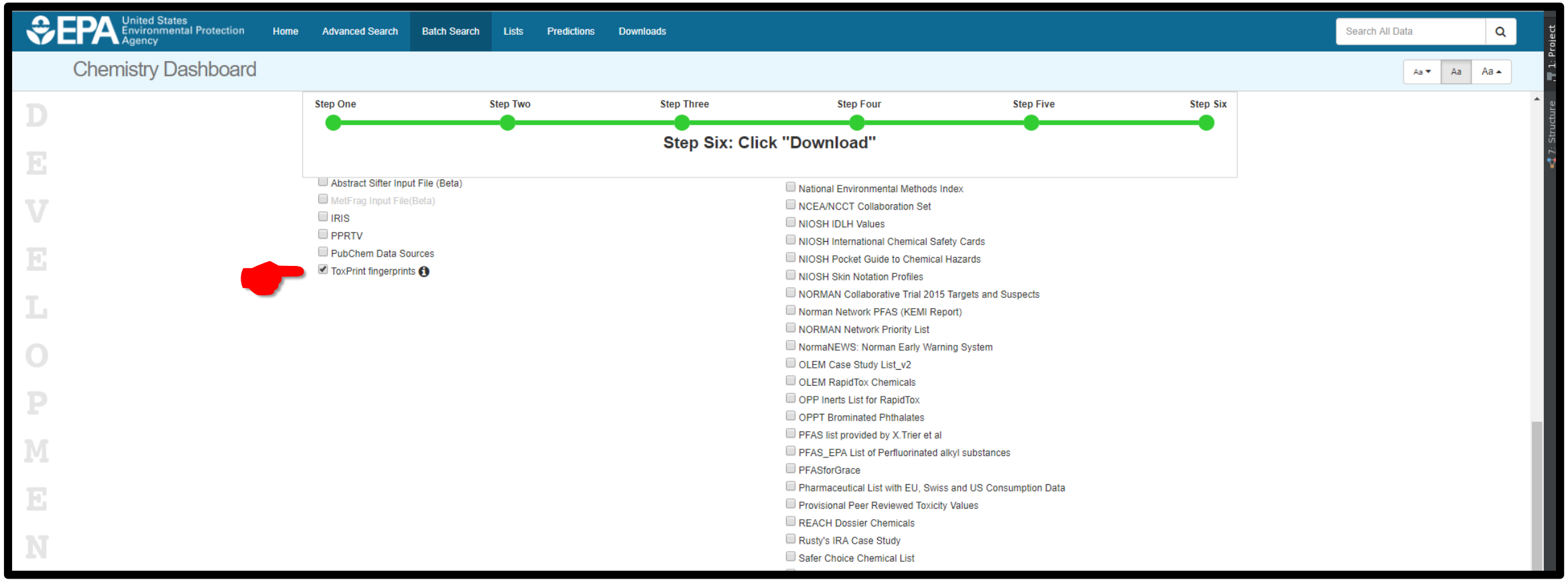



MySQL DB

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Txp-104 | bond:CN_amine_sec-NH_aromatic | 151 | 36 | 115 | 1792 | 5066 | 0.4885549545288086 | 0.8849766850471497 | 0.7643690705299377 | 1.129973292350769 | 0.298209011554718 |
| Txp-105 | bond:CN_amine_sec-NH_aromatic_aliphatic | 97 | 22 | 75 | 1806 | 5106 | 0.4827597141265869 | 0.82932448387146 | 0.810725629329681 4 | 1.2058006525039673 | 0.2608259320259094 |
| Txp-106 | bond:CN_amine_sec-NH_generic | 253 | 58 | 195 | 1770 | 4986 | 0.48362982273101807 | 0.8378617763519287 | 0.8932034373283386 | 1.1935142278671265 | 0.136996626853942 87 |
| Txp-107 | bond:CN_amine_ter-N_aliphatic | 449 | 161 | 288 | 1667 | 4893 | 0.5522294044494629 | 1.6408655643463135 | 0.0000014377596926351544 | 0.6094344258308411 | 0.9999991655349731 |
| Txp-108 | bond:CN_amine_ter-N_aromatic | 167 | 66 | 101 | 1762 | 5080 | 0.568841278553009 | 1.8839976787567139 | 0.00008360712672583759 | 0.5307862162590027 | 0.9999573826789856 |

Import Data into Data Base

# Query Results & Integration into the EPA Comptox Chemistry Dashboard




ToxPrint fingerprints have been computed for all 760k compounds in DSSTox and are available for download in the EPA Comptox Chemistry Dashboard

MOCK UP

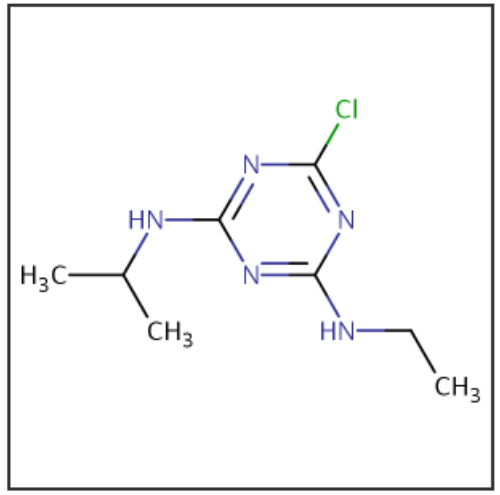




# Atrazine

1912-24-9 | DTXSID9020112

Searched by Approved Name: Found 1 result for 'Atrazine'.




## Wikipedia

Atrazine is an herbicide of the triazine class. Atrazine is used to prevent pre- and postemergence broadleaf weeds in crops such as maize (corn) and sugarcane ar such as golf courses and residential lawns. It is one of the most widely used herbicides in US and Australian agriculture. It was banned in the European Union in O 2003, when the EU found groundwater levels exceeding the limits set by regulators, and Syngenta could neither show that this could be prevented nor that...Read

## Intrinsic Properties

## Structural Identifiers

## Linked Substances

## Presence in Lists

## Record Information

## Quality Control Notes

Executive Summary (Beta) | Chemical Properties | Env. Fate/Transport | Hazard | **Chemotype** | Exposure | Bioassays | Similar Compounds | Related Substances | Synonyms | Literature | Links | Comments

## Enrichments

Download as: TSV | Excel

| descriptors_name | descriptors_label |
|---|---|
| Pubchem-81 | >= 1 Sn |
| Pubchem-81 | >= 1 Sn |
| Txp-338 | bond:metal_group_III_other_Sn_generic |
| Txp-340 | bond:metal_group_III_other_Sn_organo |

## Txp-338

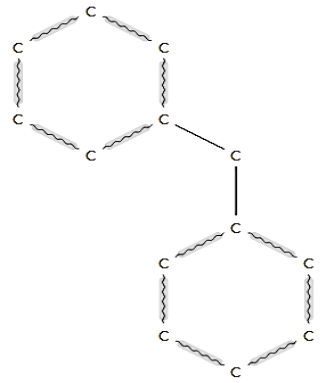
| Assay Name | Balanced Accuracy |
|---|---|
| ACEA_AR_antagonist_80hr | |
| ACEA_AR_antagonist_AUC_viability | |
| ACEA_ER_AUC_viability | |
| ACEA_ER_AUC_viability | |

## chain:oxy-alkaneLinear_ethyleneOxide_EO6

Descriptor Name: Txp-500

Searched by Approved Name: Found 1 result for 'Atrazine'.

Chain:aromaticAlkane_Ph-C1-Ph



***Significance Threshold:***

*Odds Ratio ≥ 3*

*Fischer's Exact pval ≤ 0.05*

Information

First, get the installation plan for the printer signed off by the IT department, they need to confirm that the hardware and specifications are compatible with the rest of the system.

Create a plan for installing the printer and identify the potential risks and instruments used and have the plan signed off by the person in charge of health and safety for that work area. Pay particular attention to the weight, bulk and center of gravity of the printer for the install, routing of the cables so they are not a trip hazard, and run in accordance with category 5 ethernet cabling standards and mains electricity safety standards. Also plan for, and have the electrical safety team sign to confirm, the earth-leakage tests on the equipment and the suitability of the cable and mains supply for the power requirements and expected duty-cycle of the printer. Note that certain printers include heaters and the vents must be accounted for.

Your plan must also include details of the people involved in the installation and the skills required of them, as the activities undertaken by employees are monitored by the Investors In People certification department in conjunction with the campaign for appropriate job titles.

A document must be created to assess the expected ramifications of the installation and how the use of the printer will fit into existing office procedures, with citings from relevant procedure documentation and updates to said documentation all in accordance with ISO 9000 business quality assurance.

A business continuity plan should be drawn up detailing the expected availability of support, warranty, drivers, replacement parts, supplies and compatible paper, where each of these items can be sourced from and what the average delay of obtaining said items is so that in the event of an emergency information will be available so that business may continue as smoothly as possible.

Union approval must be sought where the the procedure(s) affected by the printer involve employees of the company.

Submit forms when completed.

Read the networking documentation describing the assignment of internal IP addresses on the LAN, and submit a request for an available IP address along with updated documentation. Check with the DHCP lease assignments that such a request isn't already in use, and reserve a space in the DHCP scope. Update the DHCP Server action log to reflect the changes.

Check the connectivity of the RJ45 ethernet socket in the intended destination, and acquire a category-5 patch lead in accordance with the colour scheme of the patch panel. Patch the socket and update the patch panel documentation.

With the electron flow in the supply circuits suppressed by the application of an air-gap, connect the 8 wire twisted pair endpoint to the command decoding circuitry of the networked device. Cease the application of the air-gap and allow the devices' internal testing to run completely. Navigate through the presented operations and ensure the static internet protocol location identifier is not the selected option.

Dispose of the packaging in an environmentally sound fashion, and file the packaged software and instruction booklets

## Enrichments

Compounds | Assays | Info

| Assay Name * | descriptors_name | descriptors_label | CT-Tot | TP | FP | FN | TN | BA | OR | P-Val | Inv OR | Inv P-val |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ACEA_AR_agonist_AUC_viability | Txp-479 | chain:aromaticAlkane_Ph-C1-Ph | 32 | 20 | 12 | 198 | 559 | 0.6817206144332886 | 4.705387115478516 | 0.00002670356661838O | 0.21252235770225525 | 0.9999947547912598 |
| ACEA_AR_antagonist_80hr | Txp-479 | chain:aromaticAlkane_Ph-C1-Ph | 32 | 23 | 9 | 259 | 498 | 0.6883049607276917 | 4.91377067565918 | 0.0000238816137425601 48 | 0.20350968837738037 | 0.9999955296516418 |
| ACEA_AR_antagonist_AUC_viability | Txp-479 | chain:aromaticAlkane_Ph-C1-Ph | 32 | 23 | 9 | 265 | 492 | 0.6843419671058655 | 4.744654178619385 | 0.00003576712333597243 | 0.21076351404190063 | 0.9999930262565613 |
| ACEA_ER_80hr | Txp-479 | chain:aromaticAlkane_Ph-C1-Ph | 123 | 55 | 68 | 471 | 2516 | 0.64473557472229 | 4.320594310760498 | 0.000000000000152175784451889 78 | 0.23144963383674622 | 1 |
| ACEA_ER_AUC_viability | Txp-479 | chain:aromaticAlkane_Ph-C1-Ph | 123 | 77 | 46 | 1007 | 1980 | 0.6444443464279175 | 3.291308641433716 | 0.000000000146070197604153 4 | 0.30383050441741943 | 1 |
| APR_HepG2_CellCycleArrest_72h_dn | Txp-479 | chain:aromaticAlkane_Ph-C1-Ph | 48 | 23 | 25 | 232 | 792 | 0.6263020634651184 | 3.1406896114349365 | 0.00016134207544382662 | 0.31840139627456665 | 0.9999524354934692 |
| APR_HepG2_CellLoss_24h_dn | Txp-479 | chain:aromaticAlkane_Ph-C1-Ph | 48 | 32 | 16 | 289 | 735 | 0.6922200322151184 | 5.08650541305542 | 0.0000000815861440628396 | 0.19659863412380219 | 1 |
| APR_HepG2_CellLoss_72h_dn | Txp-479 | chain:aromaticAlkane_Ph-C1-Ph | 48 | 42 | 6 | 430 | 594 | 0.7275390625 | 9.669767379760742 | 0.000000000197551475178414 65 | 0.1034151017665863 | 1 |
| APR_HepG2_MicrotubuleCSK_24h_dn | Txp-479 | chain:aromaticAlkane_Ph-C1-Ph | 48 | 10 | 38 | 58 | 966 | 0.5758463740348816 | 4.382940292358398 | 0.0005115841631777585 | 0.22815735638141632 | 0.9998970627784729 |
| APR_HepG2_MitoMass_24h_up | Txp-479 | chain:aromaticAlkane_Ph-C1-Ph | 48 | 5 | 43 | 28 | 996 | 0.5384114384651184 | 4.1362128257751465 | 0.013410350307822227 | 0.2417670637369156 | 0.9974146485328674 |
| APR_HepG2_MitoMembPot_24h_dn | Txp-479 | chain:aromaticAlkane_Ph-C1-Ph | 48 | 20 | 28 | 182 | 842 | 0.6194661259651184 | 3.3045525550842285 | 0.0001516203192295506 6 | 0.302612841129303 | 0.9999576807022095 |

MOCK UP



## ASSAY: LTEA_HepaRG_FAS_up

Inducible Reporter | mRNA induction

Enrichments | Compounds | Assays | Info

| Assay Name * | descriptors_name | descriptors_label | CT-Tot | TP | FP | FN | TN | BA | OR | P-Val | Inv OR | Inv P-val |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LTEA_HepaRG_FAS_up | Txp-15 | bond:C#N_nitrile | 27 | 7 | 20 | 95 | 930 | 0.5832881927490234 | 3.4263157844543457 | 0.01139814406633377 | 0.2918586730957031 | 0.9972311854362488 |
| LTEA_HepaRG_FAS_up | Txp-183 | bond:CX_halide_aromatic-X_ether_aromatic_(Ph-O-Ph) | 32 | 8 | 24 | 94 | 926 | 0.5789215564727783 | 3.2836878299713135 | 0.008679874241352081 | 0.3045356273651123 | 0.9977613687515259 |
| LTEA_HepaRG_FAS_up | Txp-228 | bond:OZ_oxide_hyroxy | 20 | 5 | 15 | 97 | 935 | 0.5780038833618164 | 3.2130584716796875 | 0.03696121275424957 | 0.311229944229126 | 0.9909928441047668 |
| LTEA_HepaRG_FAS_up | Txp-264 | bond:quatN_alkyl_acyclic | 3 | 2 | 1 | 100 | 949 | 0.7856689095497131 | 18.979999542236328 | 0.026177959516644478 | 0.052687037736177... | 0.999112606048584 |
| LTEA_HepaRG_FAS_up | Txp-276 | bond:S(=O)O_sulfonicAcid_acyclic_(chain) | 14 | 4 | 10 | 98 | 940 | 0.5956509709358215 | 3.8367347717285156 | 0.03891432285308838 | 0.2606382966041565 | 0.9923533201217651 |
| LTEA_HepaRG_FAS_up | Txp-294 | bond:S=O_sulfonyl_S_(connect_Z=2) | 9 | 4 | 5 | 98 | 945 | 0.6752423644065857 | 7.714285850524902 | 0.007168598007410765 | 0.12962962687015533 | 0.9992839097976685 |
| LTEA_HepaRG_FAS_up | Txp-48 | bond:C(=O)O_carboxylicEster_alkenyl | 14 | 4 | 10 | 98 | 940 | 0.5956509709358215 | 3.8367347717285156 | 0.03891432285308838 | 0.2606382966041565 | 0.9923533201217651 |
| LTEA_HepaRG_FAS_up | Txp-488 | chain:aromaticAlkene_Ph-C2_acyclic_generic | 25 | 7 | 18 | 95 | 932 | 0.5937488079071045 | 3.815204620361328 | 0.007264365907758474 | 0.2621091306209564 | 0.9984081387519836 |
| LTEA_HepaRG_FAS_up | Txp-490 | chain:aromaticAlkene_Ph-C2 | 20 | 6 | 14 | 96 | 936 | 0.6034883856773376 | 4.1785712242126465 | 0.009007133543491364 | 0.23931623995304108 | 0.9982352256774902 |
| LTEA_HepaRG_FAS_up | Txp-610 | ring:hetero_[5]_N_imidazole | 25 | 8 | 17 | 94 | 933 | 0.6142356395721436 | 4.670838356018066 | 0.015918648568913334 | 0.21409432590007782 | 0.9997076988220215 |
| LTEA_HepaRG_FAS_up | Txp-611 | ring:hetero_[5]_N_pyrazole | 15 | 5 | 10 | 97 | 940 | 0.6198971271514893 | 4.84536075592041 | 0.010598755441606045 | 0.2063829749822616 | 0.9982573986053467 |
| LTEA_HepaRG_FAS_up | Txp-612 | ring:hetero_[5]_N_pyrrole | 14 | 5 | 9 | 97 | 941 | 0.6318469643592834 | 5.389461517333984 | 0.007646706886589527 | 0.18554729223251343 | 0.9988660216331482 |
| LTEA_HepaRG_FAS_up | Txp-617 | ring:hetero_[5]_N_triazole_(1_2_4-) | 26 | 7 | 19 | 95 | 931 | 0.5883190631866455 | 3.6105263233184814 | 0.009160916320979595 | 0.27696794271469116 | 0.9978834390640259 |
| LTEA_HepaRG_FAS_up | Txp-618 | ring:hetero_[5]_N_triazole_(1_3_4-) | 29 | 7 | 22 | 95 | 928 | 0.5742576122283936 | 3.1081340312957764 | 0.017021438106894493 | 0.3217364549636841 | 0.9954628944396973 |
| LTEA_HepaRG_FAS_up | Txp-624 | ring:hetero_[5]_O_dioxolane_(1_3-) | 13 | 4 | 9 | 98 | 941 | 0.6066854000091553 | 4.267573833465576 | 0.029980719089508057 | 0.23432518541812897 | 0.9946792721748352 |
| LTEA_HepaRG_FAS_up | Txp-640 | ring:hetero_[5_6]_N_indole | 11 | 5 | 6 | 97 | 944 | 0.680682897567749 | 8.109965324401855 | 0.0022387809585779905 | 0.12330508232116699 | 0.9997772574424744 |
| LTEA_HepaRG_FAS_up | Txp-643 | ring:hetero_[5_6]_N_purine | 3 | 2 | 1 | 100 | 949 | 0.7856689095497131 | 18.979999542236328 | 0.026177959516644478 | 0.052687037736177... | 0.999112606048584 |
| LTEA_HepaRG_FAS_up | Txp-657 | ring:hetero_[6]_N_pyrimidine | 35 | 9 | 26 | 93 | 924 | 0.5828487277030945 | 3.4392058849334717 | 0.004390374291688204 | 0.29076477885246277 | 0.9988998770713806 |
| LTEA_HepaRG_FAS_up | Txp-670 | ring:hetero_[6]_Z_1_2- | 8 | 3 | 5 | 99 | 945 | 0.6400862336158752 | 5.7272725105285645 | 0.03454276919364929 | 0.1746031790971756 | 0.9956991672515869 |

**MOCK UP**

# Generate Enrichment Table

Step One — Step Two — Step Three — Step Four — Step Five — Step Six

**Step One: Select Input**

**Select Input Format**

Upload as…

**Select Input Type(s)**

- ☐ Chemical Name
- ☐ CASRN
- ☐ InChIKey ☐ Skeleton
- ☐ DSSTox Substance ID
- ☐ MS-Ready Formula(e)
- ☐ Exact Formula(e)
- ☐ Monoisotopic Mass

**Bin Endpoint**

- ☐ No
- ☐ Yes

Set Threshold

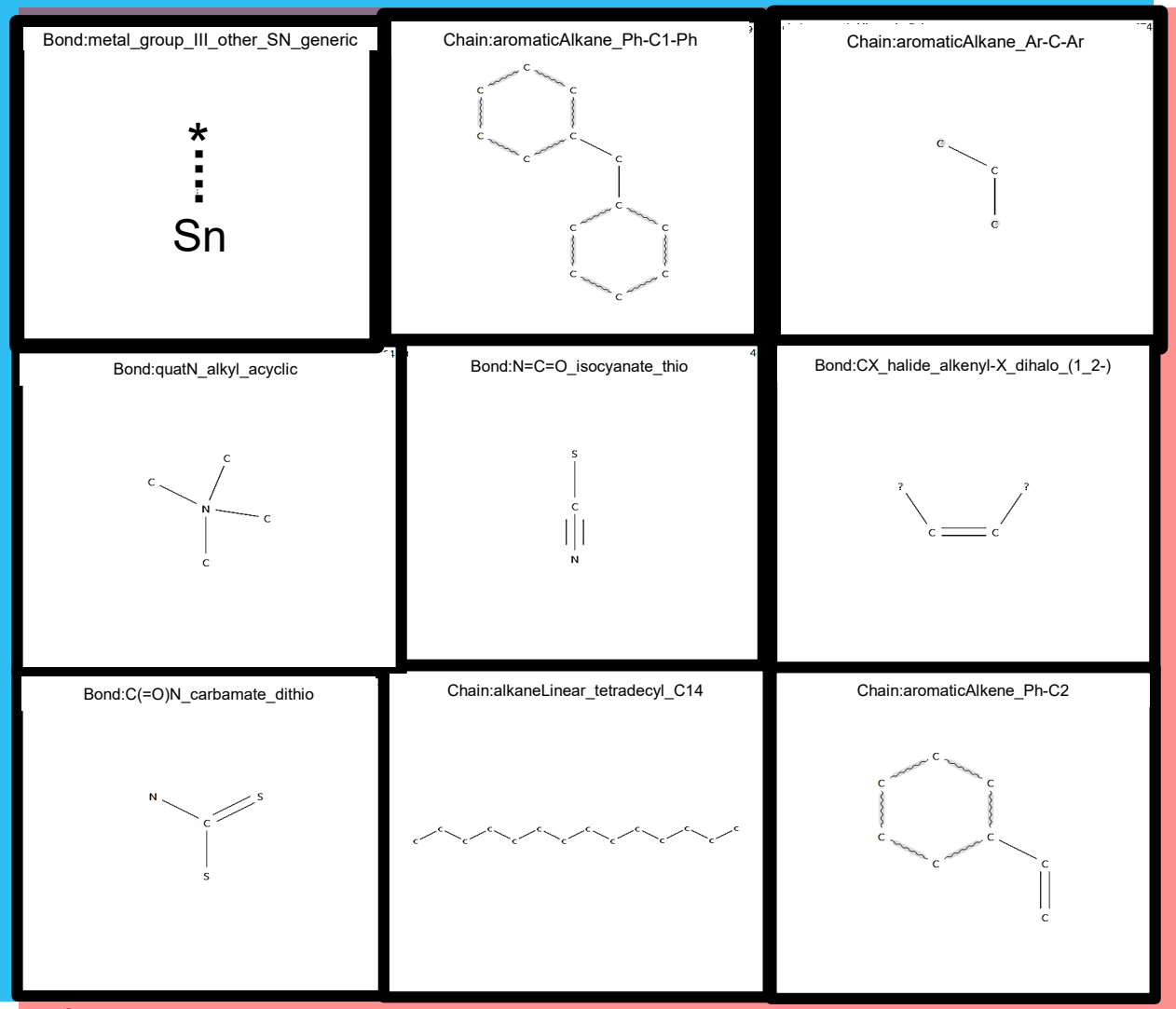
**Example Input**

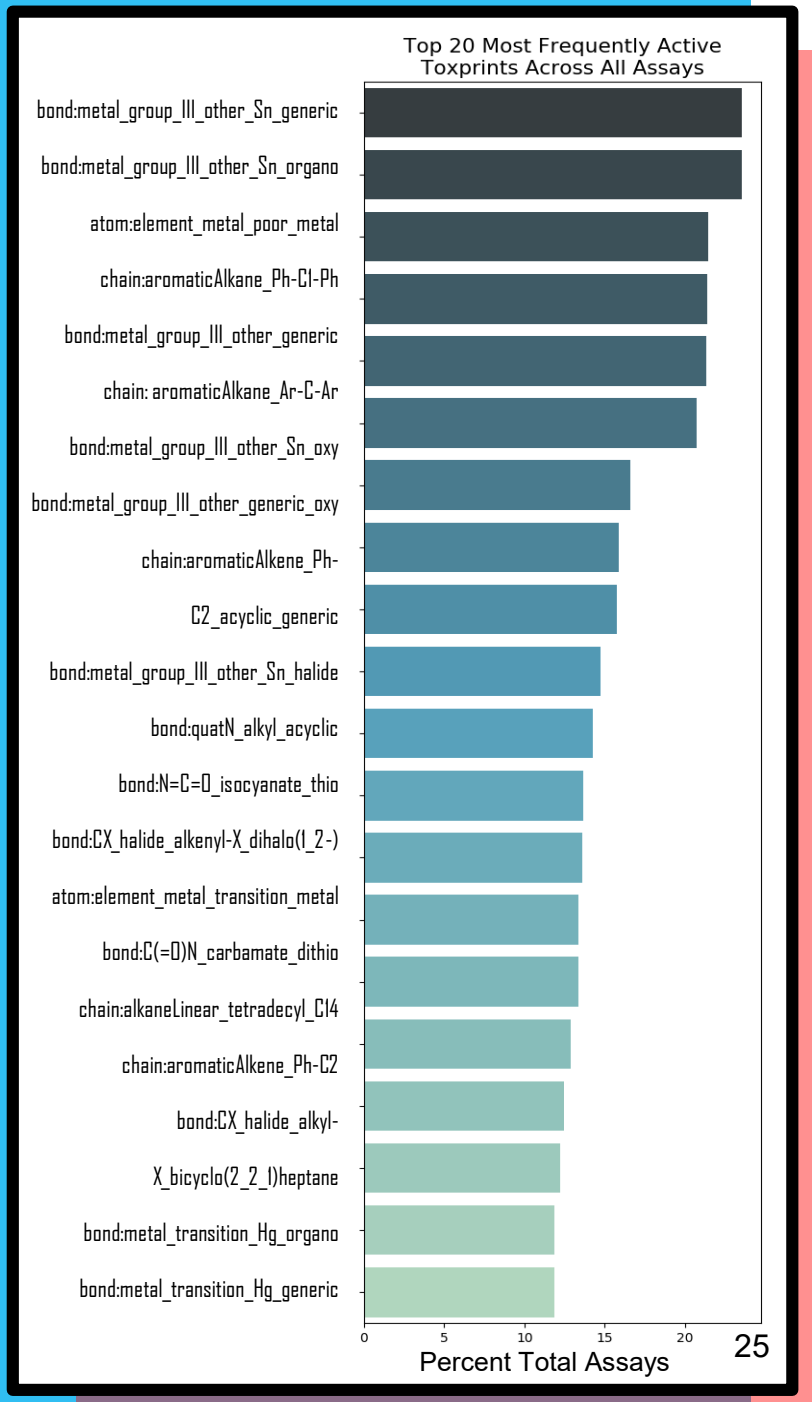
| ID | ENDPOINT |
|---|---|
| DTXSID12344 | 0 |
| DTXSID122546 | 1 |
| DTXSID23462 | 1 |
| DTXSID009546 | 1 |
| DTXSID43668 | 0 |
| DTXSID653 | 0 |
| DTXSID252565 | 1 |
| DTXSID2456678 | 0 |
| DTXSID778335 | 0 |

Display All Chemicals | Download Chemical Data

# Promiscuous Chemotypes

Top 9 Chemotypes Most Frequently seen in assay hits

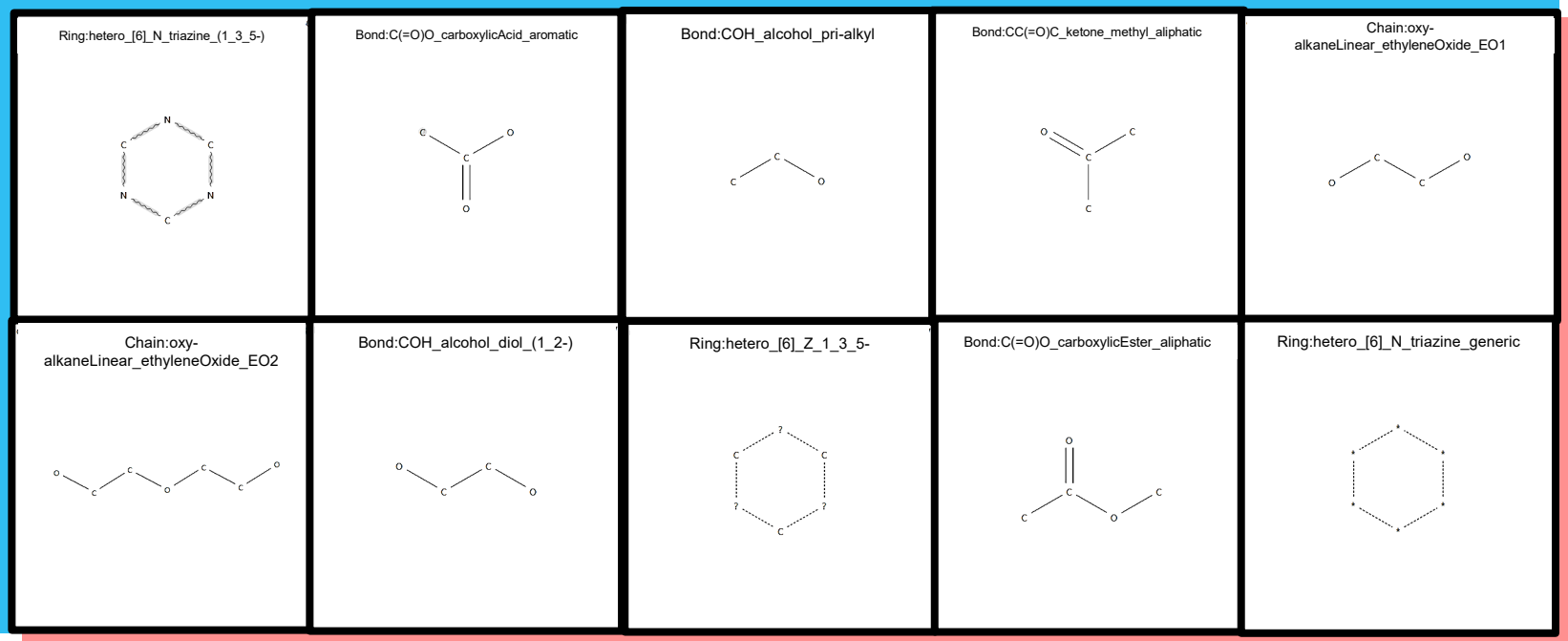


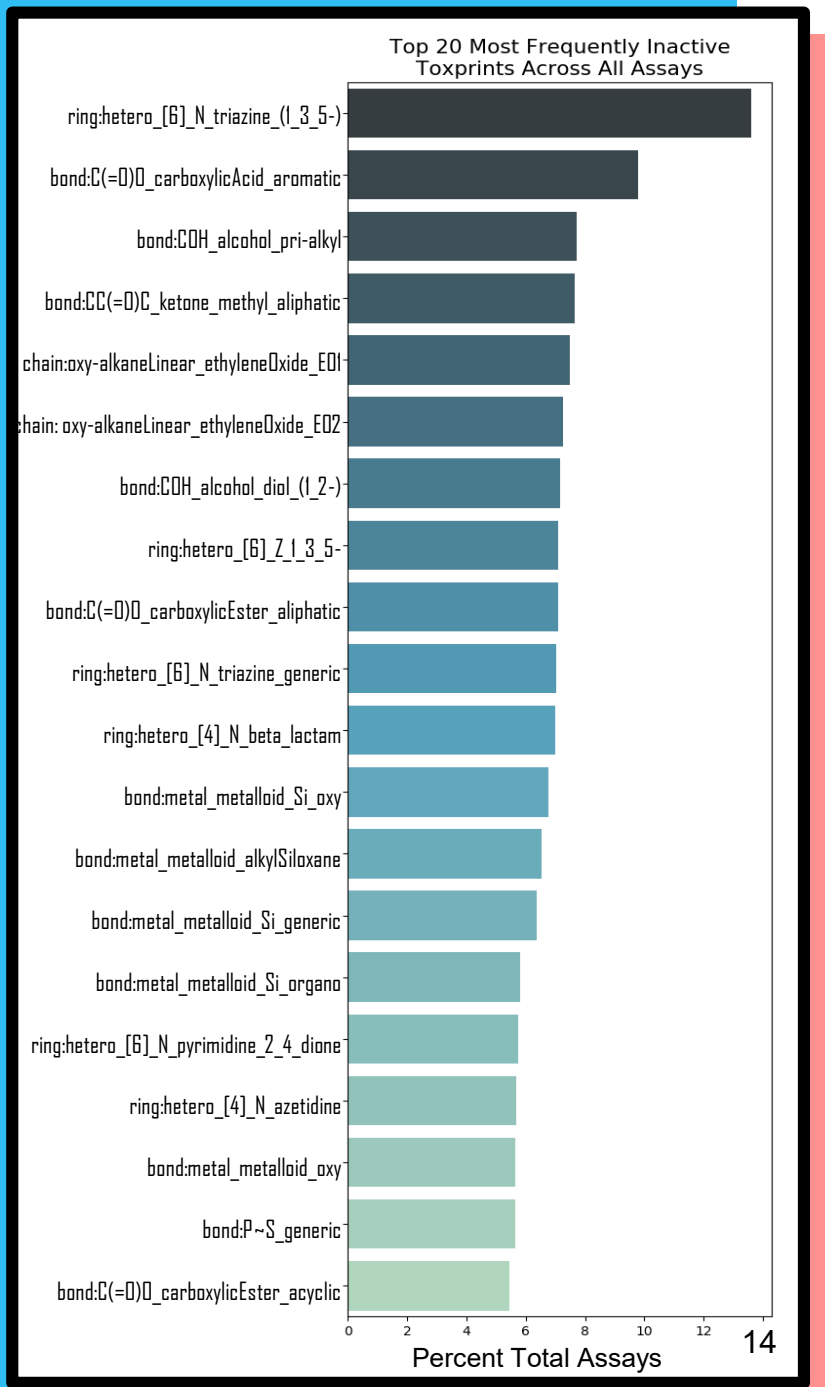



# Inverse Promiscuity

Top 10 Chemotypes Most Infrequently seen in assay hits

- If we take the inverse hitcall, and run the analysis we can see where the lack of a chemotype may be significant
- Triazine rings (~3/10)
- Ethane-1,2-diol like (~4/10)



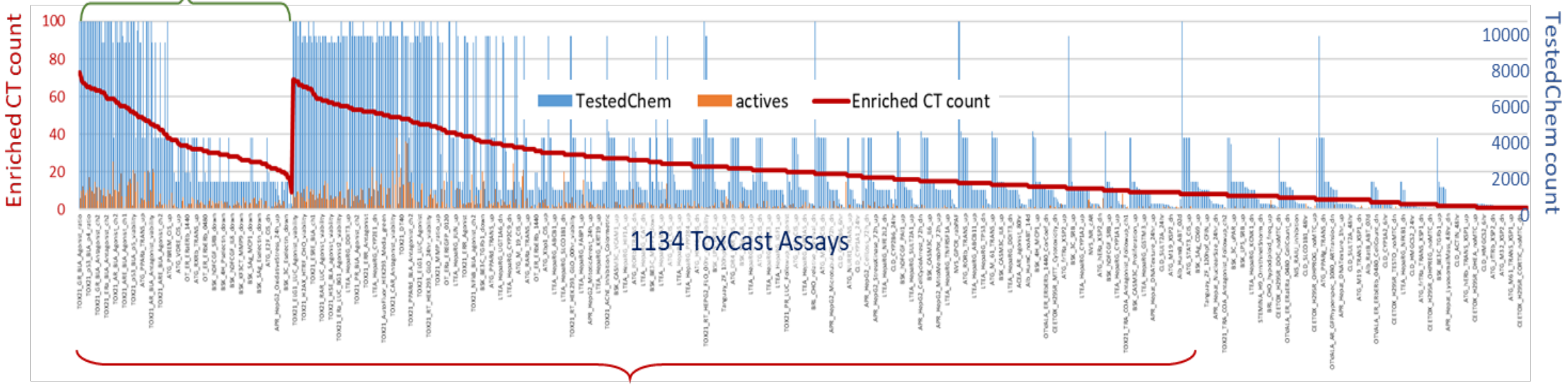



# Assay Space Overview

"Global" QSAR models* for 167 ToxCast Assays
*Avg & Median BA = 0.67*

* Random Forest models based on ToxPrint CT descriptors, validated using independent Test Set & Y randomization, with Training (100 A, 100 I) & Test (25 A, 25 I) Set minimums (J. Fitzpatrick, unpublished)

*Global QSAR models could not be derived for 972/1134 (86%) of ToxCast Assays*




*≥ 10 enriched CTs for 821 (72%) of ToxCast Assays*

➔ *Significant chemical-bioactivity "signal" within local chemistry domains defined by CTs*

EPA United States Environmental Protection Agency

# Concluding Remarks

- A general workflow for Chemotype-Enrichments has been successfully programmed
- The approach is completely general and can be applied to any binary "activity" dataset (e.g., in vivo or in vitro bioassays, functional use categories, etc)
- Elements of the workflow are being integrated into the publicly available USEPA Comptox Chemistry Dashboard
- A suite of command-line tools were developed that could be publicly shared
- The Chemotype-Enrichment workflow shows great promise for elucidating chemical signals across the assay space and supporting Comptox research

EPA United States Environmental Protection Agency

# Acknowledgments


- Ann Richard
- Chris Grulke
- Antony Williams
- Xia Meng Howey
- Nathan Braswell
- Jon Gardener
- NCCT Staff
- NCCT dev team

EPA
United States
Environmental Protection
Agency

# QUESTIONS?