Tox21 Enricher

User's Manual

(Released on 10/29/2017)

http://hurlab.med.und.edu/Tox21Enricher

Junguk Hur, Ph.D.

Department of Biomedical Sciences

University of North Dakota, School of Medicine and Health Sciences

Grand Forks, North Dakota 58202, USA

Email: InformaticsTools@gmail.com

http://hurlab.med.und.edu

Introduction to Tox21 Enricher

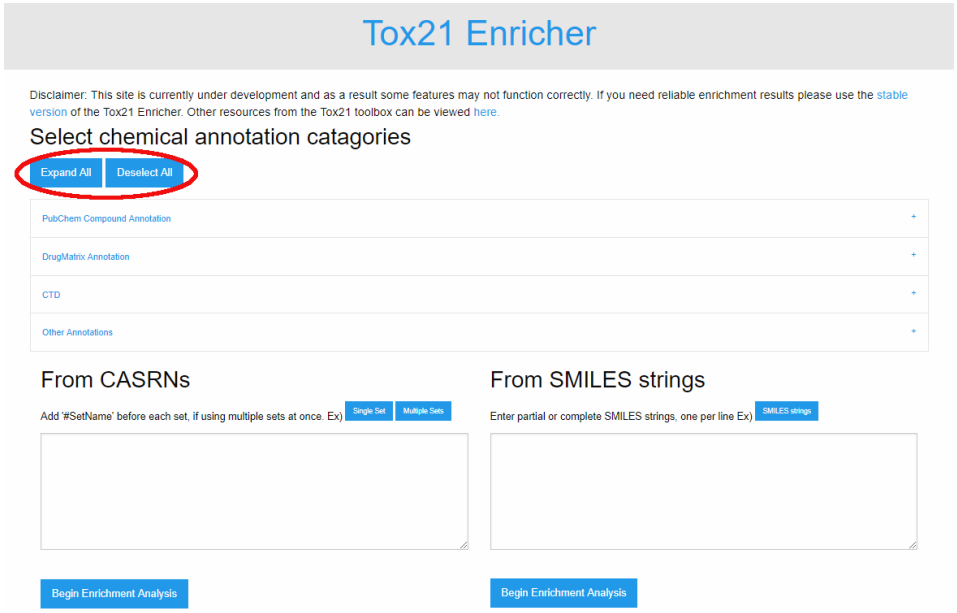
Humans are exposed to tens of thousands of chemicals that are used in our daily life, some at levels that may pose a health risk. For many of these chemicals, there limited toxicological information which makes risk assessment impossible. The United States Toxicology Testing in the 21st Century (Tox21) program was established to develop more efficient and human relevant toxicity assessment methods. The Tox21 program is currently screening over 10,000 chemicals, the Tox21 10K library, using quantitative high-throughput screening (qHTS) of assays that measure effects on toxicity pathways. To date, more than 70 assays have yielded >12 million concentration-response curves by Tox21 researchers. To efficiently apply these data for identifying potential hazardous compounds and for informing further toxicological testing, the United States National Toxicology Program (NTP) has developed several web applications (Tox21 Toolbox: http://ntp.niehs.nih.gov/tbox/), including tools for data visualization (Tox21 Curve Browser) and exploration (Tox21 Activity Profiler).

One critical usage of this dataset is to perform chemical-relational analysis based on the patterns of activity across the Tox21 assays and then to use nearest neighbor based prediction to infer the toxicological properties of untested chemicals via their association with tested chemicals. One approach to inferring the specific properties is to perform chemical annotation enrichment of chemical neighborhoods.

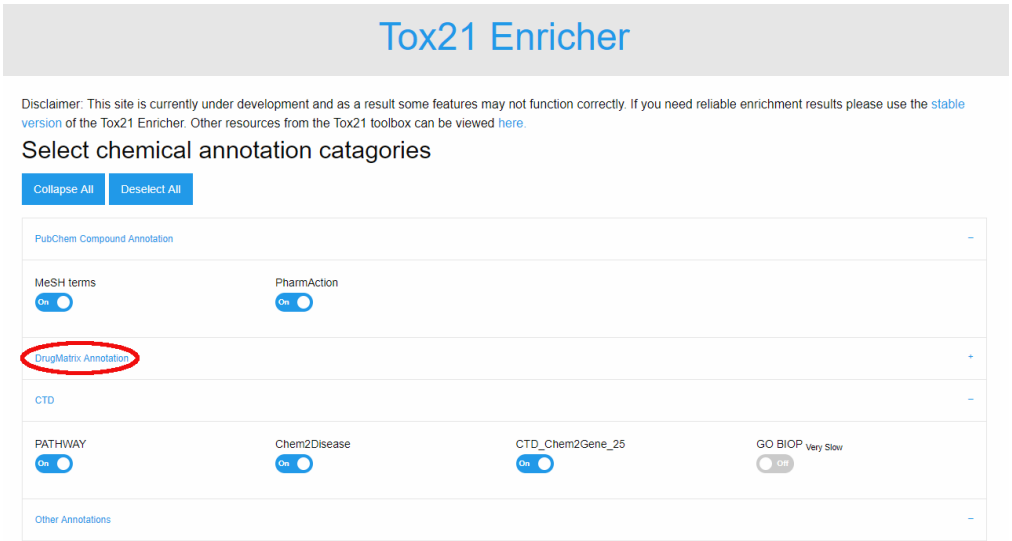
Here, we present Tox21 Enricher, a web-based chemical annotation enrichment tool for Tox21 assay data. Tox21 Enricher identifies significantly over-represented chemical biological annotations among sets of chemicals (neighborhoods), which facilitates the identification of the toxicological properties and mechanisms in the chemical sets.

Landing Page

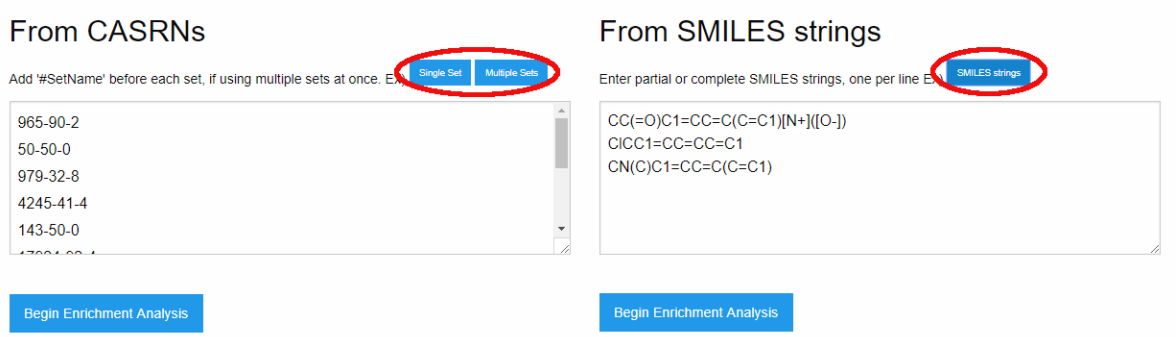
Upon visiting the Tox21 Enricher's web page, you will see the landing page. This page has buttons for expanding the annotation category group accordions and deselecting/selecting all enrichment categories. These buttons are circled in the image below.



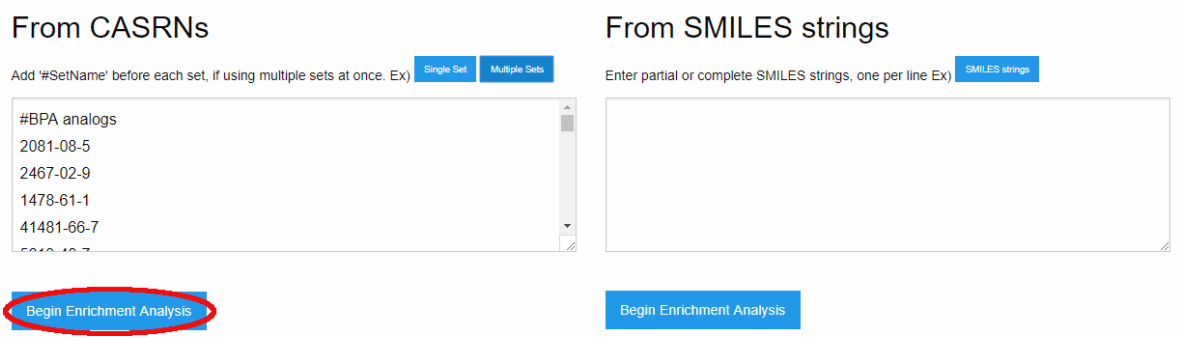
Once the "Expand All" button is clicked, each of the annotation category accordions will expand to show the categories they contain. **Each accordion may also be expanded/collapsed individually as shown below.**



After selecting the appropriate annotation categories, the next step is providing input for the enrichment analysis. This input can take the form of either CASRNs or SMILES strings. Each option has a text area for input. Each option also has buttons above their respective text areas that will populate example input.

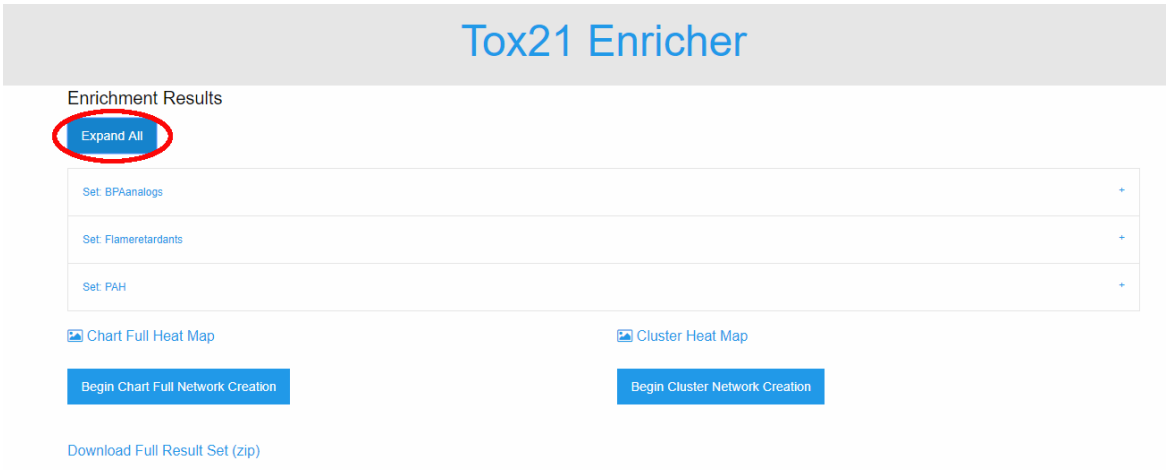


After input has been entered, it is time to begin enrichment. This can be done by clicking on the appropriate “Begin Enrichment Analysis” button for whichever input you are using. The button that is clicked will use the data in its respective text area, so you must ensure you are using the right button. This can be seen using the multi-set CASRN example input below.

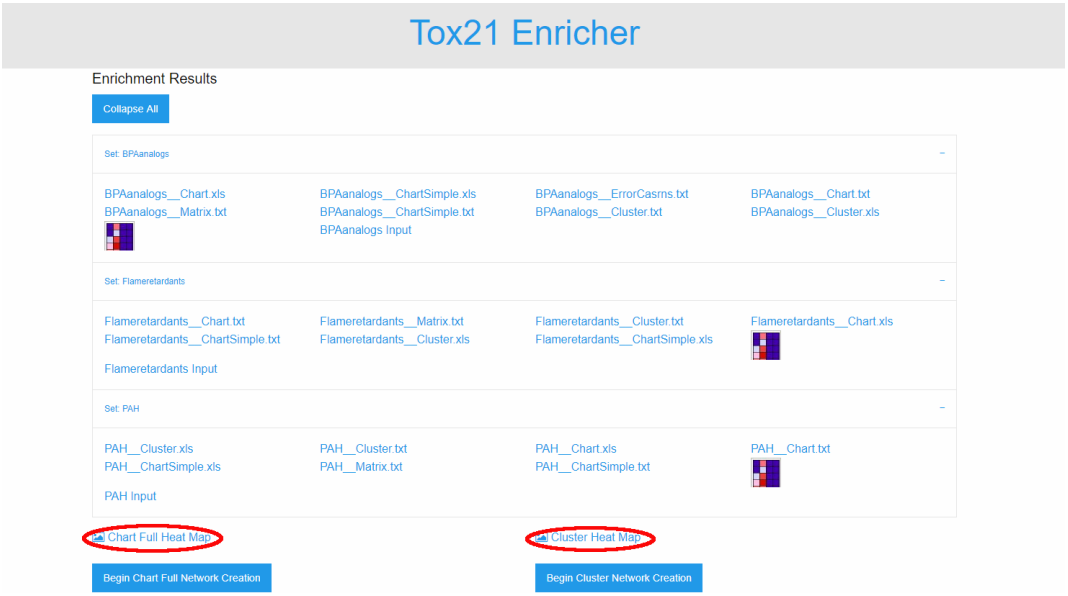


Results Page

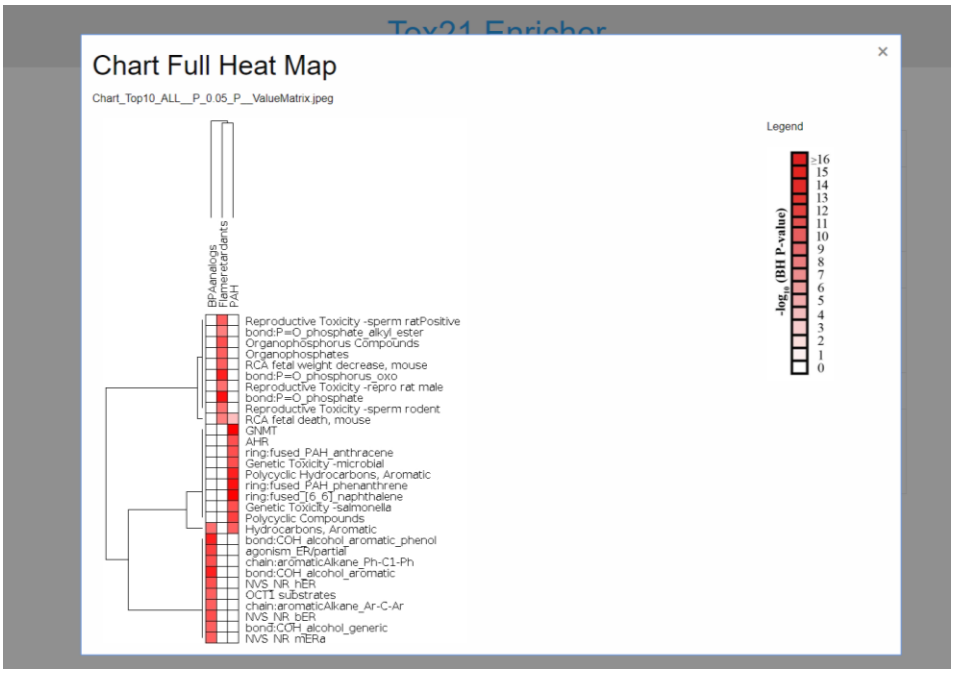
Once enrichment is complete, a results page will be loaded. Like the landing page, results on this page will be placed in collapsed accordions. This accordion can be clicked on to expand it, or the “Expand” button may be used, as on the landing page. This button is highlighted in the screenshot below.



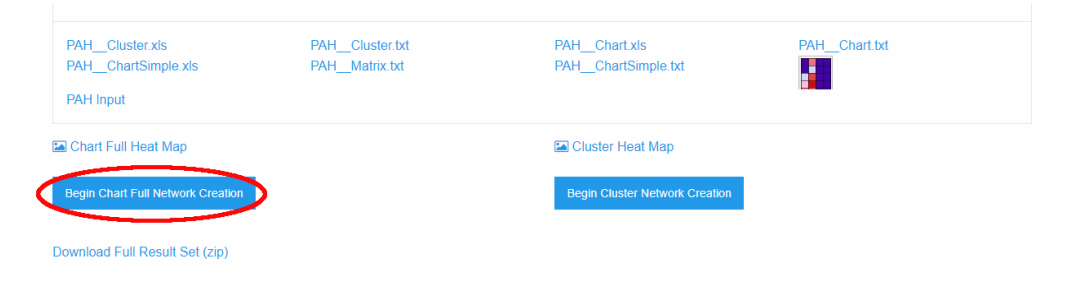
The results shown above are of our multi-set input that we submitted in our last landing page example. In the screenshot below, we can see the results for each set expanded. The cluster and chart full heat map image links can also be seen highlighted.



Clicking on the one of the heat map image links will display as seen in the screenshot below.

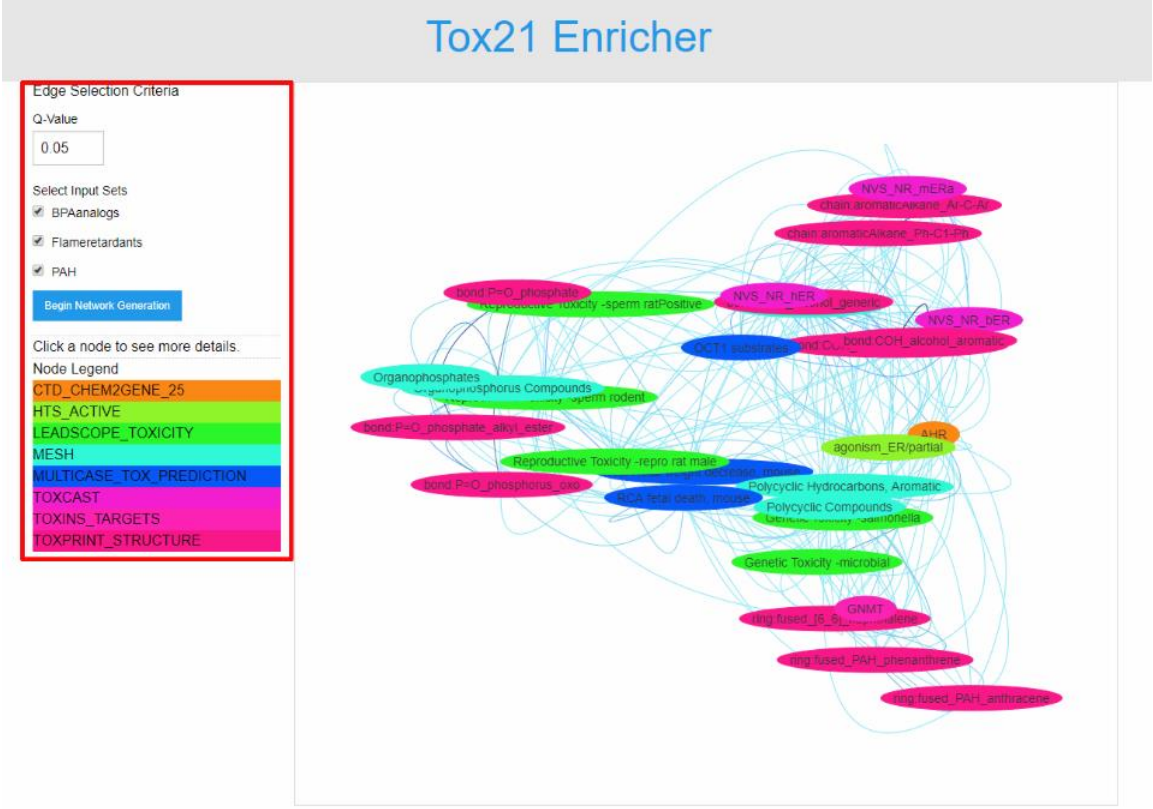


Next, we can click either the “Begin Chart Full Network Creation” button or the “Begin Cluster Network Creation” button. These buttons can be seen highlighted below.



Network Page

After clicking either of the network creation buttons, you will see the network page as shown below. The left side of this page is where we have edge selection criteria, input set select for node data to visualize, and the color key for the current network. In the screenshot below, we are using the Chart Full data for our network. The nodes correspond to significantly enriched annotations and the edges indicates that there are significant overlap between the two annotations in terms of chemical contents. The edge color gradient indicates the degree of overlap based on Jaccard index.



Description of Annotations

Pubchem Compound Annotations

- MeSH terms
  Chemical to MeSH Chemical and Drug [D}, specifically the terminal branch MeSH terms in the “D” subset. These terms can be viewed here:
  https://meshb.nlm.nih.gov/treeView

- PharmAction
  Chemical to Pharmacological Action terms derived from the USNLM. Details of these annotations can be found here:
  https://meshb.nlm.nih.gov/record/ui?ui=D020228

DrugMatrix Annotations

Chemical annotations derived from the NTP DrugMatrix Database that reflect a variety of pharmacological and toxicological associations with chemicals. These are hand curated annotations that are based on literature review.

- ACTIVITY_CLASS
  Each compound studied in DrugMatrix is grouped into an activity class, which represents a more generic compound annotation than structure activity class. Compounds are grouped together based on having structure activity class annotations that are related by a common therapeutic activity (i.e. anti-inflammatory) or toxicological activity (DNA damager). Structures are not considered when grouping compounds, such that unrelated structures that act through a common molecular target are grouped together. Likewise, compounds with distinct, but pharmacologically related targets, are also grouped together under a single activity class term. Compounds with a structure activity annotation unrelated by therapeutic or toxicological activity are simply annotated with their structure activity until such time that related compounds are added to the database.

- ADVERSE_EFFECT
  A comprehensive vocabulary list of 1,600 adverse effects was created for the purpose of curating adverse drug effects. Each of the adverse effects was associated with a tissue/organ ID and given an “Adverse Effect Group Name”. Each of the adverse effects is also assigned a severity index based on the seriousness of the adverse effect or clinical event. Genotoxicity, mutagenicity, carcinogenicity, and life threatening or potential life-threatening events are given a score of “SSS”, serious but manageable events are given a score of “SS”, and measurable but not serious events are give a score of “S”. During the curation of adverse drug effects, terms from the list of 1,600 matching those in the PDR, drug package insert, or primary literature for a particular drug are selected and given a frequency score (1- most frequent, 2-less common, 3-rare) according to the information in the documents. The “Adverse Effect” associated with each drug in DrugMatrix™ is thus the curated adverse effect terms for each drug prefixed with

the associated tissue/organ ID and the frequency score (e.g., LIV_1_Increased Liver Enzymes; CVS_3_Bradycardia). If an “SSS” severity label is associated with an adverse effect, then the curated adverse effect appears in “Tissue Toxicity” of the compound, while the “Adverse Effect Group Name” appears in “Known Toxicity” of the compound.

- INDICATION
  Each compound studied in DrugMatrix™ is associated with its approved clinical indications. A comprehensive vocabulary list of 1822 indications was created for this purpose. While curating clinical indications for a given drug, terms from this list of 1822 matching those found in the Physician’s Desk Reference (PDR), drug package insert, subscribed online databases, or primary literature for the drug are selected and curated.

- KNOWN_TOXICTY
  Chemical to high severity Adverse Effect associations. These associations have been curated as described under ADVERSE_EFFECT

- MECHANISM
  Mechanism displays one or more descriptive phrases established in DrugMatrix™ to classify a compound based on its phenotypic or physiological effect such as “Inhibit platelet aggregation”, “Increase lipid catabolism” or “Block neural transmission.” A list of 75 Mechanism terms was established for compound curation. These descriptions are typically less specific than MECH_LEVEL_1

- MECH_LEVEL_1
  The highest-level mechanism description for a chemical that use such terms as “Hormone Modulation” or “Hormone Replacement”

- MECH_LEVEL_2
  A mechanistic description of a chemical that is more specific than MECH_LEVEL_1. For example, Androgen Receptor Antagonist” or “Androgen Receptor Agonist”.

- MECH_LEVEL_3
  A mechanistic description of a chemical that is more specific than MECH_LEVEL_2, for example, “Androgen Receptor”

- MESH_LEVEL_1
  MeSH terms from the highest level in the Chemical and Drugs [D] annotation set. See: https://meshb.nlm.nih.gov/treeView

- MESH_LEVEL_2
  MeSH terms from the second highest level in the Chemical and Drugs [D] annotation set. See: https://meshb.nlm.nih.gov/treeView

- MESH_LEVEL_3
  MeSH terms from the third highest level in the Chemical and Drugs [D] annotation set. See: https://meshb.nlm.nih.gov/treeView

- MODE_CLASS
  One or more descriptive phrases were established in DrugMatrix™ to describe a compound based on how it affects its molecular target. A list of 75 Modes have been created for the assignment of the Mode Class, such as “Enzyme Inhibitor”, “Receptor Agonist, Selective” or “Channel Blocker.”

- PRODUCT_CLASS
  Product class, such as “Hormones, Endocrine and Metabolic”, “Central Nervous System (CNS)”, or “Anti-infectives” is a general industry classification of the drug. A vocabulary list of 24 terms was established for product class. Each therapeutic class is pre-associated with a product class in the curation database so that when a therapeutic class is selected, the product class is determined automatically.

- STRUCTURE_ACTIVITY
  Each compound studied in DrugMatrix is assigned a structure activity class based on the molecular target associated with its approved clinical use, as reported in the literature. When the molecular target is not clearly defined, the compound is classified based on the accepted or most relevant mechanism of action and/or clinical indications. When appropriate or necessary, a compound is classified with both its mechanism of action and molecular target in order to bin compounds into a more general group. When diverse chemical structure types are active against the same protein target, the compounds are sub-grouped into pharmacophore types, and a structure activity class is assigned to each subgroup based on the molecular target and chemical structure type.

- THERAPEUTIC_CLASS
  The purpose of the “therapeutic class” category is to classify a compound based on its therapeutic uses with respect to its indications. A comprehensive list of 120 therapeutic classes, such as “Antidiabetic Agents”, “General Anesthetics, Intravenous” or “Antibacterials, Systemic”, was established for this purpose. During the process of curating drug indications, each indication is associated with an appropriate therapeutic class term from this list of 120. The same therapeutic class may be associated with several indications of a given compound. For example, the therapeutic class “Antibacterials, Systemic” may be associated with the indications “Mycobacterium tuberculosis”, “Erythema Nodosum Leprosum (ENL)”, “Leprosy”, and “Atypical Mycobacterial Diseases” of the same compound.

- TISSUE_TOXICTY
  This category displays one or more common terms used to highlight serious adverse effects associated with a compound, such as “Carcinogenicity” or “Electrolyte Disturbance.” All compounds associated with an adverse drug reaction classified with a severity score of “SSS”, as well as compounds associated with

any cardiovascular toxicity, hepatotoxicity, or nephrotoxicity with a frequency score of 1, are highlighted by the respective “Adverse Effect Group Name”.

- TA_LEVEL_1
  Chemical to drug/therapeutic use category (e.g., Infectious Disease). TA_LEVEL_1 is the highest (ie most abstract) drug use category

- TA_LEVEL_2
  Chemical to drug use/therapeutic category that is more specific than TA_LEVEL_1 (e.g., solid tumors).

- TA_LEVEL_3
  Chemical to drug use/therapeutic category that is more specific than TA_LEVEL_2 (e.g., prostate cancer).

CTD

Chemical annotations in this set were derived from the Comparative Toxciogenomics Database (CTD; http://ctdbase.org), a hand curated database that is hosted by NC State and maintained by the Dr. Carolyn Mattingly’s group.

- Pathway
  Chemical to molecular pathway (e.g., P53 pathway) associations that have been curated through review of the literature. Some associations in this set have been inferred based on bridging linkages through genes.

- Chem2Disease
  Chemical to disease (e.g., prostate cancer) associations that have been curated through review of the literature. Some associations in this set have been inferred based on bridging linkages through genes.

- CTD_Chem2Gene_25
  Chemical to gene (e.g., Aryl Hydrocarbon Receptor (AHR)) associations that have been curated through review of the literature. Associations were limited to the top 25 most strongly associated genes for each chemical.

- GO BIOP
  Chemical to Gene Ontology Biological Processes (e.g., Xenobiotic Metabolism) associations that have been curated through review of the literature. Some associations in this set have been inferred based on bridging linkages through genes.

DrugBank Targets

Chemical to molecular target annotation. The annotations in this set associate chemicals with their known molecular targets/ molecular initiators (e.g., receptors) that have been curated in DrugBank (https://www.drugbank.ca/), a resource maintained by the Dr. David Wishart's Group at University of Alberta. The annotation terms in this set are gene names that correspond to proteins that are known to associate with specific chemicals or macromolecular structures (e.g., DNA) that chemicals and known to bind to.

DrugBank ATC Code

Chemical to ATC codes. The annotations in this set associate chemicals with an effect ontology, known as the Anatomical Therapeutic Chemical Classification System that is maintained by the WHO (https://www.whocc.no/atc_ddd_index/). DrugBank (https://www.drugbank.ca/), a resource maintained by the Dr. David Wishart's Group at University of Alberta, has curated ATC codes to chemicals/drugs in the DrugBank database which were extracted and used as annotations in the Tox21 Enricher application. The annotation terms in this set are ATC names that correspond to therapeutic classifications for drugs.

Toxin Targets

Chemical to molecular target annotation. The annotations in this set associate chemicals with their known molecular targets/ molecular initiators (e.g., receptors) that have been curated in Toxin and Toxin Target Database (http://www.t3db.ca/), a resource maintained by the Dr. David Wishart's Group at University of Alberta. The annotation terms in this set are gene names that correspond to proteins that are known to associate with specific chemicals or macromolecular structures (e.g., DNA) that chemicals and known to bind to.

Leadscope Toxicity

Chemical to SAR-based predicted toxicological effects. The annotations in this set associate chemicals with their predicted toxicology effects are predicted by a suite of Leadscope SAR models. Importantly these are PREDICTIONS, not observed effects. The SAR predictions were generated using the Leadscope (Columbus, OH; https://www.leadscope.com/)Insilico First application (https://www.leadscope.com/isfcui/app#; Leadscope Inc. For the purposes of the Tox21 Enricher any chemical scoring >0.5 of being active was considered active in an SAR model. The annotation terms in this set are SAR model names which are associated with specific toxicological effects.

Multicase Toxicity Prediction

Chemical to SAR-based predicted toxicological effects. The annotations in this set associate chemicals with their predicted toxicology effects are predicted by a suite of Multicase SAR models. Importantly these are PREDICTIONS, not observed effects. The SAR model predictions were kindly provided by the Multicase Inc (Beachwood, OH; http://www.multicase.com/), specifically Dr. Aleck Sedykh. For the purposes of the Tox21

Enricher any chemical scoring >50% probability of being active was considered active in an SAR model. The annotation terms in this set are SAR model names which are associated with specific toxicological effects.

ToxRefDB

Chemical to in vivo toxicological effect annotations. The annotations in the set associate chemicals with their toxicological effects in guideline toxicity studies. These associations were had curated by staff at the US EPA under the direction of Drs. Matt Martin and Richard Judson. ToxRefDB annotations can be accessed here (https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data). The annotation terms in this set are toxicological effect ontology names.

HTS Actives

Chemical to in vitro biological activity annotations. This annotation set associates chemicals with assays from Tox21 (https://ntp.niehs.nih.gov/results/tox21/index.html) where they were observed to be active. The list chemicals in each Tox21 assay were provided by Dr. Jui-Hua Hsieh. All Tox21 assay results used in the Tox21 Enricher can be viewed in the Tox21 Activity Profiler (https://sandbox.ntp.niehs.nih.gov/tox21-activity-browser/). The annotation terms in this set are assay names from Tox21.

ToxCast

Chemical to in vitro biological activity annotations. This annotation associates chemicals with assays from ToxCast where they were observed to be active. A complete listing of all results from ToxCast can be found here: https://actor.epa.gov/dashboard/. The annotation terms in this set are assay names from ToxCast.

ToxPrint Structures

Chemical to substructure/chemotype annotations. A detailed description of the ToxPrint chemotypes can be found here (https://toxprint.org/). In short these are collection of chemical substructures that have been mechanistically linked to toxicological effects such as genotoxicity.

END OF THE USER'S MANUAL