
Tox21 Enricher

User's Manual

(Released on 11/14/2020)

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

*Copyright 2020 University of North Dakota
All rights reserved.*

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>

Table of Contents

I. Introduction to Tox21 Enricher	3
II. Landing Page	4
II.1. Selecting Annotation Categories	5
II.2. Selecting Enrichment Cutoff Value	6
II.3. Selecting Enrichment Analysis Type	7
III.3.a. Enrich From User Provided CASRN List	7
III.3.b. Enrich From Chemicals With Shared Substructures	8
III.3.c. Enrich From Chemicals With Structural Similarity	8
II.4. Performing Enrichment	9
III. Waiting Page	10
III.1. Transaction ID	10
III.2. Queue Position	10
III.3. Status	10
IV. Results Page	11
IV.1. Enrichment Results	11
IV.2. Adjusting Network Generation	14
IV.3. Re-Enriching Selected Chemicals	15
IV.4. Creating the Networks	17
V. Network Page	18

I. Introduction to Tox21 Enricher

Humans are exposed to tens of thousands of chemicals that are used in daily life, some at levels that may pose a health risk. For many of these chemicals, there is 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.

II. Landing Page

Upon visiting the Tox21 Enricher's web page, you will see the landing page.

Tox21 Enricher

Please see [this link](#) for instructions on using this application and the descriptions about the chemical/biological categories. Other resources from the Tox21 toolbox can be viewed [here](#).

Select chemical/biological annotation categories

Expand All

Deselect All

PubChem Compound Annotation	+
DrugMatrix Annotation	+
DrugBank Annotation	+
CTD Annotation	+
Other Annotations	+

Select enrichment cutoff

This will determine the maximum number of results per data set and may affect how many nodes are generated during network generation. (default = 10)

10

Select enrichment type

Note: Please verify you are using the correct chemical identifiers by referencing the [EPA's CompTox Chemicals Dashboard](#).

User Provided CASRN List

Chemicals With Shared Substructures

Chemicals With Structural Similarity

User Provided CASRN List

Add '#SetName' before each set, if using multiple sets at once. Ex)

Single Set

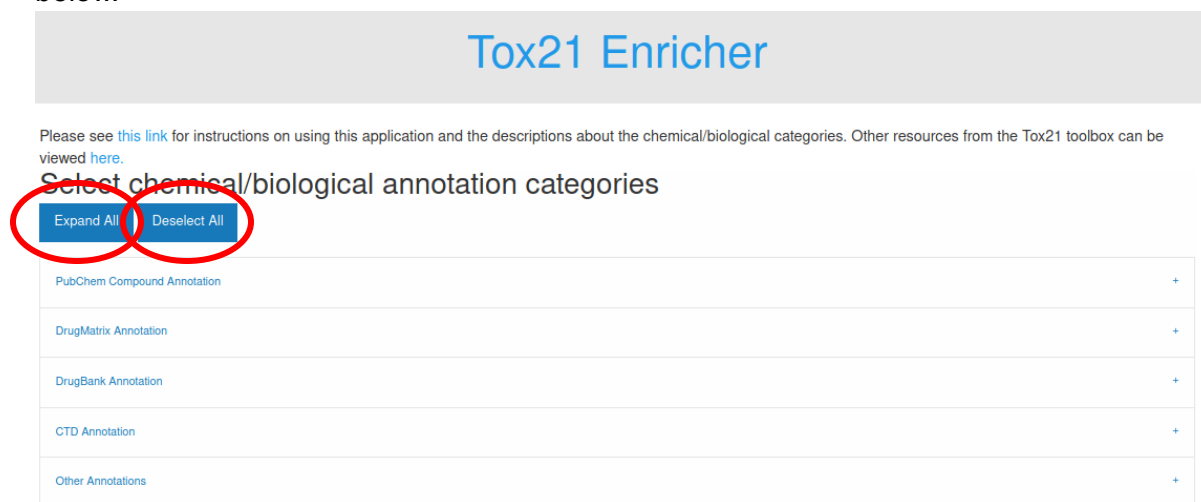
Multiple Sets

Clear Input

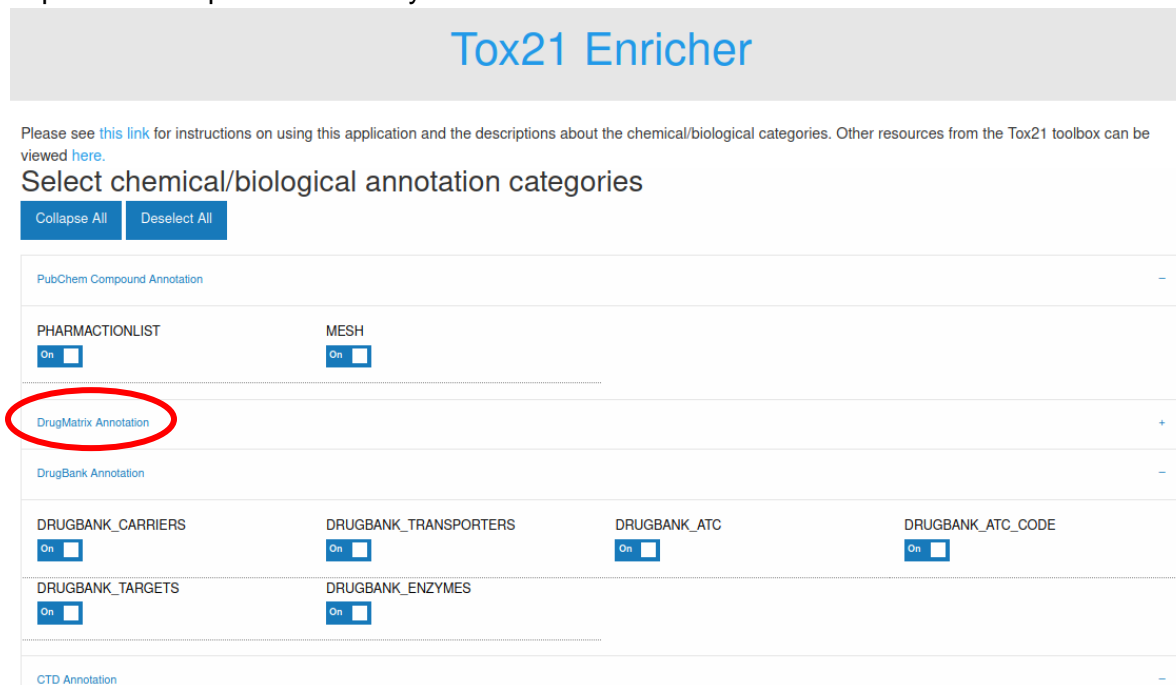
Begin Enrichment Analysis

II.1. Selecting Annotation Categories

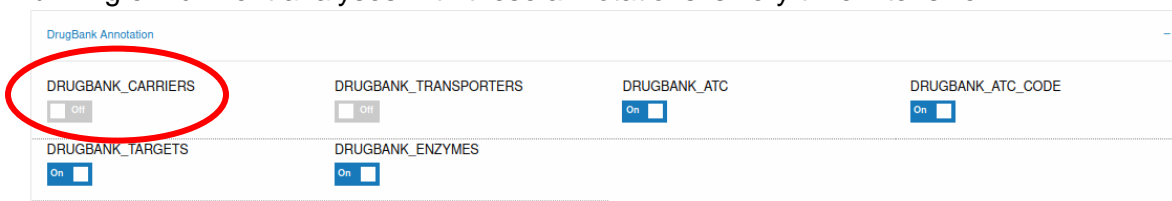
This page has buttons for expanding the annotation category group accordions and deselecting/selecting all enrichment categories. These buttons are circled in the image below.



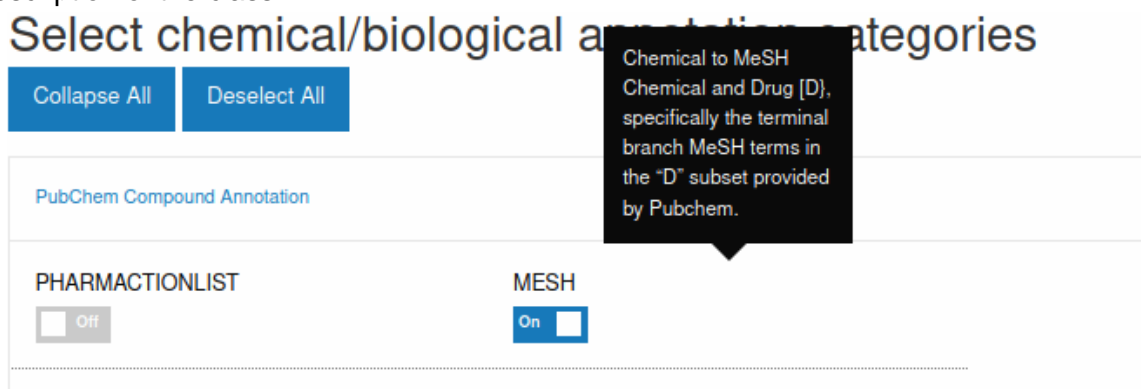
Once the “Expand All” button is clicked, each of the annotation category accordion tabs will expand to show the categories they contain. Each accordion tab may also be expanded/collapsed individually as shown below.



Each annotation class within the accordion tabs can be interacted with. Clicking on the toggle switch for an annotation class will set it to either “On” and “Off.” When a class is set to “On,” Tox21 Enricher will include results from that annotation class, if applicable. When a class is set to “Off,” Tox21 Enricher will ignore results from that class. By default, all annotation classes are set to “On,” *except* for CTD_GOENRICH_BIOPROCESS as running enrichment analyses with these annotations is very time intensive.



Additionally, hovering over each annotation class will display a tooltip with a brief description of the class.



II.2. Selecting Enrichment Cutoff Value

After selecting the appropriate annotation categories, the next step is specifying the enrichment cutoff value. This value determines the maximum number of enrichment results displayed per data set submitted. By extension, this will also determine the maximum number of nodes displayed during network generation. By default, this is set to 10.

Select enrichment cutoff

This will determine the maximum number of results per data set and may affect how many nodes are generated during network generation. (default = 10)



II.3. Selecting Enrichment Analysis Type

After selecting the cutoff value, the final step is selecting the enrichment analysis type and providing input for it. This input can take the form of CASRNs, SMILES, or InChI strings, depending on enrichment type. Each option has a text area for input. Each option also has buttons above their respective text areas that will populate the text areas with example input. The “Clear Input” button will clear the text in the box.

Select enrichment type

Note: Please verify you are using the correct chemical identifiers by referencing the [EPA's CompTox Chemicals Dashboard](#).

The screenshot shows the 'User Provided CASRN List' tab selected. Below the tab name is a text area for input. Above the text area are three buttons: 'Single Set', 'Multiple Sets', and 'Clear Input'. The 'Single Set' button is circled in red. Below the text area is a 'Begin Enrichment Analysis' button.

II.3.a. Enrich From User Provided CASRN List

This enrichment analysis type accepts only CASRN strings as input (one CASRN per line). Multiple sets may be independently enriched at the same time. To enrich multiple sets, separate lines of CASRN strings with set names of the form “#SetName.”

User Provided CASRN List

The screenshot shows the 'User Provided CASRN List' input field. Above the field are three buttons: 'Single Set', 'Multiple Sets', and 'Clear Input'. The 'Single Set' button is circled in red. The input field contains the following text:

```
#Set1
985-90-2
50-50-0
979-32-8
4245-41-4
#Set2
143-50-0
17924-92-4
297-76-7
152-43-2
313-06-4
1056 37 0
```

The first line, '#Set1', and the line '4245-41-4' are circled in red. Below the input field is a 'Begin Enrichment Analysis' button.

Before performing enrichment from a user provided CASRN list, please cross-reference your CASRN inputs with those in the EPA's CompTox Chemicals Dashboard (<https://comptox.epa.gov/dashboard>) to ensure accuracy during the enrichment analysis.

II.3.b. Enrich From Chemicals With Shared Substructures

This enrichment analysis type accepts either SMILES or InChI strings (one per line). Each string is treated as its own data set and will be enriched independently of any other sets provided. By default, this enrichment is in SMILES mode. Clicking the "Switch to InChI input" button will switch to InChI mode, which allows the enrichment to accept InChI strings.

The screenshot shows a web interface titled "Chemicals With Shared Substructures". Below the title, there is a blue button labeled "Switch to InChI input" which is circled in red. Below this button, there is a text input field with the placeholder text "Enter partial or complete SMILES strings, one per line. Ex)". To the right of the input field are two buttons: "SMILES strings" and "Clear Input". Below the input field is a large blue button labeled "Begin Enrichment Analysis".

II.3.c. Enrich From Chemicals With Structural Similarity

This enrichment analysis type accepts either SMILES or InChI strings (one per line). Each string is treated as its own data set and will be enriched independently of any other sets provided. By default, this enrichment is in SMILES mode. Clicking the "Switch to InChI input" button will switch to InChI mode, which allows the enrichment to accept InChI strings. The Tanimoto similarity threshold may be specified using the slider controls to set the strictness of the search. By default, this is set to 50%.

Chemicals With Structural Similarity

Select Tanimoto similarity threshold (%)



Switch to InChI input

Enter partial or complete SMILES strings, one per line Ex)

SMILES strings

Clear Input

```
COCCOC(=O)CC#N
C1=CC=C(C=C1)CSC#N
C1OC1C1=CC=CC=C1
CC1(C)CC(CC(C)(CN=C=O)C1)N=C=O
OC(=O)C(\Cl)=C(\Cl)C=O
CN(C)C1=CC=C(C=C1)
```

Begin Enrichment Analysis

II.4. Performing Enrichment

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.

User Provided CASRN List

Chemicals With Shared Substructures

Chemicals With Structural Similarity

User Provided CASRN List

Add '#SetName' before each set, if using multiple sets at once. Ex)

Single Set

Multiple Sets

#BPA analogs
2081-08-5
2467-02-9
1478-61-1
41481-66-7
5613-46-7
57-63-6
620-92-8
77-40-7
79-94-7

Begin Enrichment Analysis

III. Waiting Page

After submitting data for enrichment analysis, the waiting page will be loaded. Do not press the “Back” or “Reload” browser buttons at this point, as this will prevent Tox21 Enricher from providing the enrichment results correctly. This page displays a transaction ID assigned to our submitted enrichment, its position in the application’s queue, its running status, and its enrichment analysis type. Clicking the “Show Submitted Items” button will display a list of the chemicals we submitted for this enrichment analysis.

Tox21 Enricher

Enrichment in Progress...

You will be directed to the results page shortly. Please do not use your browser's back button.

Transaction ID	Queue Position	Status	Enrichment Analysis Type
53ebda97-5cd2-469c-a2f1-d41fade00c3b	0	Running	SMILES

[Show Submitted Items](#)

III.1. Transaction ID

The transaction ID is a UUID assigned to each enrichment analysis process. This ID is for Tox21 Enricher to keep track of the enrichment requests it receives. This is not the same as the ID for the result set of the process: this is a separate ID.

III.2. Queue Position

Tox21 Enricher can only simultaneously process a maximum of 5 enrichment analysis processes. Therefore, only transactions in queue positions 0-4 may be running at any given time. Enrichment analysis processes in positions 5 and greater must wait for a currently running process to terminate.

III.3. Status

Enrichment processes may have one of four different statuses at any given time:

- **Initializing** – The enrichment analysis request is being processed.
 - **Running** – The enrichment analysis is currently running.
 - **Waiting** – The enrichment analysis is in a queue position in 5 or greater and is waiting for a running process to terminate.
 - **Complete** – The enrichment analysis is complete and will shortly redirect the user to the results page.
-

IV. Results Page

IV.1. Enrichment Results

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

Tox21 Enricher

Enrichment Results

[Expand All](#)

Set: BPAanalog	+
Set: Flameretardants	+
Set: PAH	+

[Chart Full Heat Map](#) [Cluster Heat Map](#)

[Begin Chart Full Network Creation](#) [Begin Cluster Network Creation](#)

[Download Full Result Set \(.zip\)](#)

Adjust Network Generation

This will determine the maximum number of results per data set and may affect how many nodes are generated during network generation. (default = 10)

[Regenerate Network](#)

The results shown above are that of our multi-set input that we submitted in our landing page example. Hovering over any of the items in the result sets will display a tooltip describing what the item is.

Enrichment Results

[Expand All](#)

Set: BPAanalog

A list of the top 10 most significant annotations for each annotation class (.xls format).

[BPAanalog__Cluster.xls](#) [BPAanalog__Cluster.txt](#) [BPAanalog__ChartSimple.txt](#) [BPAanalog__Cluster.xls](#)

[BPAanalog__ChartSimple.xls](#) [BPAanalog__Chart.txt](#) [BPAanalog__Chart.xls](#) [BPAanalog__ErrorCasrns.txt](#)






[BPAanalog Input](#)

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.

Tox21 Enricher

Enrichment Results

[Collapse All](#)

Set: BPAanalogs			
BPAanalogs_Matrix.txt BPAanalogs_ChartSimple.xls 	BPAanalogs_Cluster.txt BPAanalogs_Chart.txt BPAanalogs Input	BPAanalogs_ChartSimple.txt BPAanalogs_Chart.xls	BPAanalogs_Cluster.xls BPAanalogs_ErrorCasrns.txt
Set: Flameretardants			
Flameretardants_Cluster.xls Flameretardants_Matrix.txt	Flameretardants_Cluster.txt Flameretardants_Chart.txt	Flameretardants_ChartSimple.xls Flameretardants_ChartSimple.txt	Flameretardants_Chart.xls 
Set: PAH			
PAH_Cluster.xls PAH_ChartSimple.txt PAH Input	PAH_Chart.xls PAH_Matrix.txt	PAH_ChartSimple.xls PAH_Cluster.txt	PAH_Chart.txt 
 Chart Full Heat Map		 Cluster Heat Map	
Begin Chart Full Network Creation		Begin Cluster Network Creation	

[Download Full Result Set \(.zip\)](#)

Adjust Network Generation

This will determine the maximum number of results per data set and may affect how many nodes are generated during network generation. (default = 10)

[Regenerate Network](#)

Clicking on one of the heatmap image links will display the corresponding heatmap as seen in the screenshot below.



All the enrichment result files can be downloaded by clicking the “Download Full Result Set (.zip)” link, which will prompt us to download the file **tox21enricher.zip**.

[Chart Full Heat Map](#)

[Cluster Heat Map](#)

[Begin Chart Full Network Creation](#)

[Begin Cluster Network Creation](#)

[Download Full Result Set \(.zip\)](#)

IV.2. Adjusting Network Generation

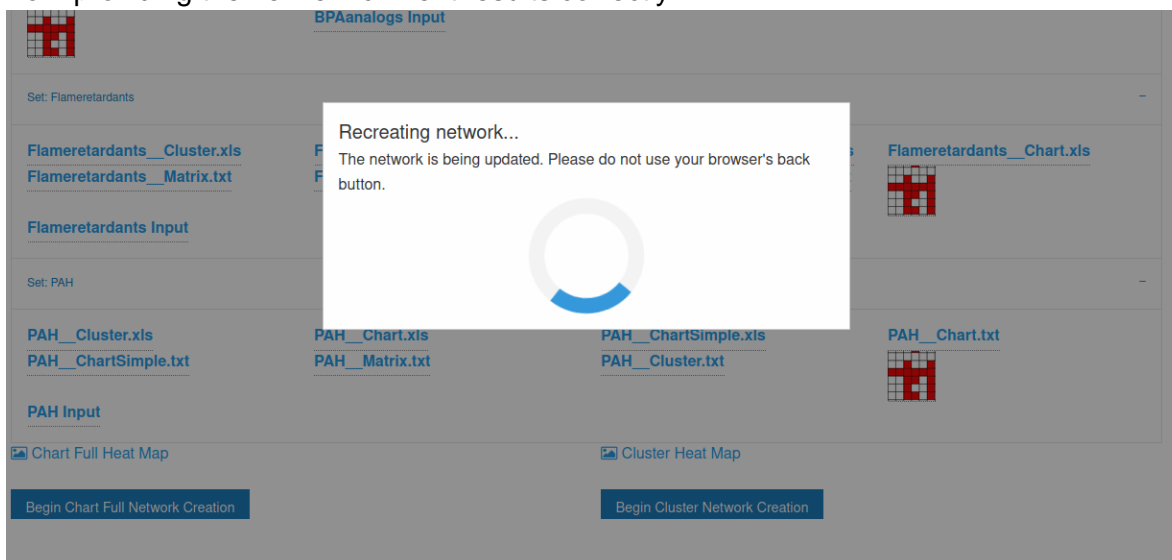
The slider controls can be used to re-perform enrichment with a different cutoff value. Changing this value may change the number of results displayed in the heatmaps and number of nodes in the network.

Adjust Network Generation

This will determine the maximum number of results per data set and may affect how many nodes are generated during network generation. (default = 10)

Regenerate Network

Clicking the “Regenerate Network” button will start the regeneration process. Do not press the “Back” or “Reload” browser buttons at this point, as this will prevent Tox21 Enricher from providing the new enrichment results correctly.



IV.3. Re-Enriching Selected Chemicals

After performing either the **Chemicals With Shared Substructures** or **Chemicals With Structural Similarity** enrichment analysis types, the “Re-enrich Selected Chemicals” table will appear on the results page. Here, we can see a list of all the CASRNs associated with the SMILES or InChI strings we submitted. By default, all chemicals in the table are selected. By clicking the “Expand All” button, we can view all the tabs like on the landing page. By clicking the “Deselect All” button, all the chemicals in the table will be deselected. The “Deselect All With Warnings” button only appears if any of the result chemicals produces a “reactive substructures” warning (discussed later). This button will only select/deselect the chemicals flagged with this warning.

Re-enrich Selected Chemicals





Select CASRNs for re-enrichment: Expand All Deselect All Deselect All With Warnings

Set: Set1	+
Set: Set2	+
Set: Set3	+



Re-Enrich Selected Chemicals

Re-enrich Selected Chemicals

Select CASRNs for re-enrichment: Collapse All Deselect All Deselect All With Warnings


Set: Set1	-				
Select	Chemical Structure	Name	Structural Similarity (Tanimoto)	SMILES	CASRN
<input checked="" type="checkbox"/>		2-Methoxyethyl cyanoacetate	1.00	<chem>COCCOC(=O)CC#N</chem>	10258-54-5
<input checked="" type="checkbox"/>		Butyl cyanoacetate	0.64	<chem>CCCCOC(=O)CC#N</chem>	5459-58-5
<input checked="" type="checkbox"/>		Ethyl cyanoacetate	0.62	<chem>CCOC(=O)CC#N</chem>	105-56-6
<input checked="" type="checkbox"/>		Methyl cyanoacetate	0.60	<chem>COC(=O)CC#N</chem>	105-34-0

By clicking the “Re-Enrich Selected Chemicals” button, a new enrichment analysis is performed using all the selected chemicals. This enrichment analysis is functionally identical to the **User Provided CASRN List** enrichment type and clicking the button will show the waiting page until the enrichment process completes.

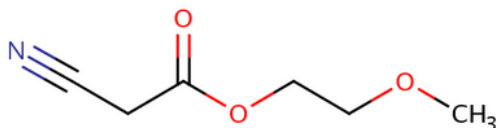
<input checked="" type="checkbox"/>		N,N,3-Trimethylaniline	0.50	<chem>Cc1cccc(N(C)C)c1</chem>	121-72-2
<input checked="" type="checkbox"/>		N,N-dimethylpyridin-4-amine	0.50	<chem>CN(C)c1ccncc1</chem>	1122-58-

Re-Enrich Selected Chemicals

Occasionally, a result chemical may have a “reactive structures” warning. A chemical is flagged with this if it either contains a known hyperreactive structural group, but the user-submitted chemical does not, or if the user-submitted chemical contains a known hyperreactive structural group while the result chemical does not. It is recommended that you deselect chemicals with this warning and perform re-enrichment on your submitted data.

<input type="checkbox"/>		Benzyl sulfide	0.52	<chem>c1ccc(CSCc2ccccc2)cc1</chem>	538-74-9	[Cyanide]
--------------------------	---	----------------	------	------------------------------------	----------	-----------

The chemical structure image for a chemical can be clicked to show additional information for the chemical. The “View at PubChem” link will take you to the chemical’s entry at PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) in a new tab.



2-Methoxyethyl cyanoacetate

IUPAC Name	2-methoxyethyl 2-cyanoacetate
CASRN	10258-54-5
SMILES	<chem>COCCOC(=O)CC#N</chem>
InChI	InChI=1S/C6H9NO3 /c1-9-4-5-10-6(8)2-3-7 /h2,4-5H2,1H3
InChI Key	SGLKIEOMYXTGBH- UHFFFAOYSA-N
Molecular Formula	C6H9NO3
Molecular Weight	143.14056

[View at PubChem](#)

IV.4. Creating the Networks

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.

Enrichment Results

[Expand All](#)

Set: Set1	+
Set: Set2	+
Set: Set3	+

[Chart Full Heat Map](#)[Cluster Heat Map](#)

[Begin Chart Full Network Creation](#)[Begin Cluster Network Creation](#)

[Download Full Result Set \(.zip\)](#)

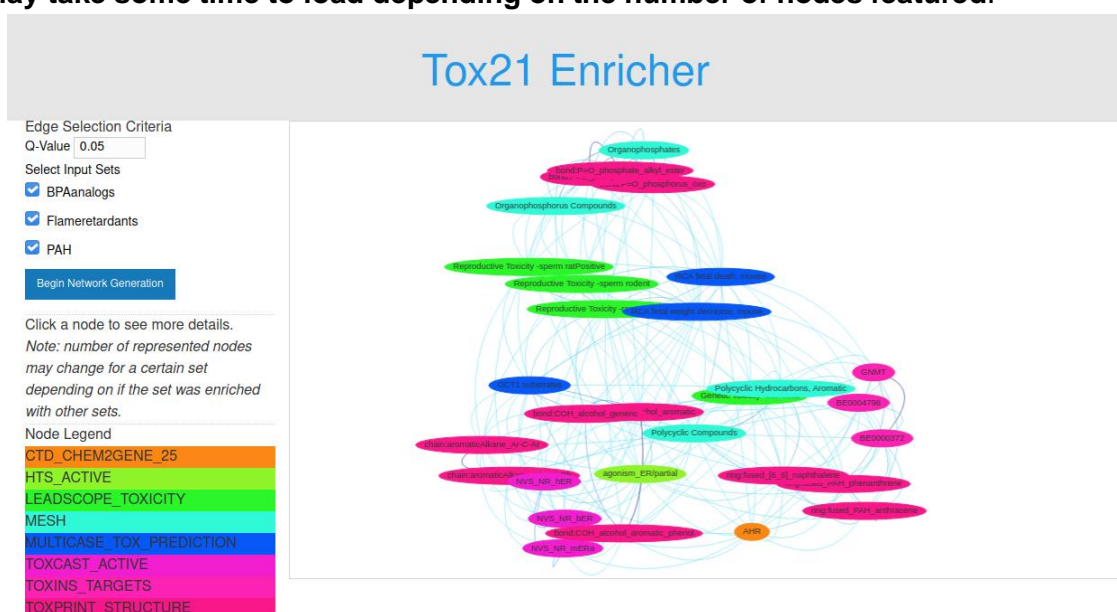
Adjust Network Generation

This will determine the maximum number of results per data set and may affect how many nodes are generated during network generation. (default = 10)

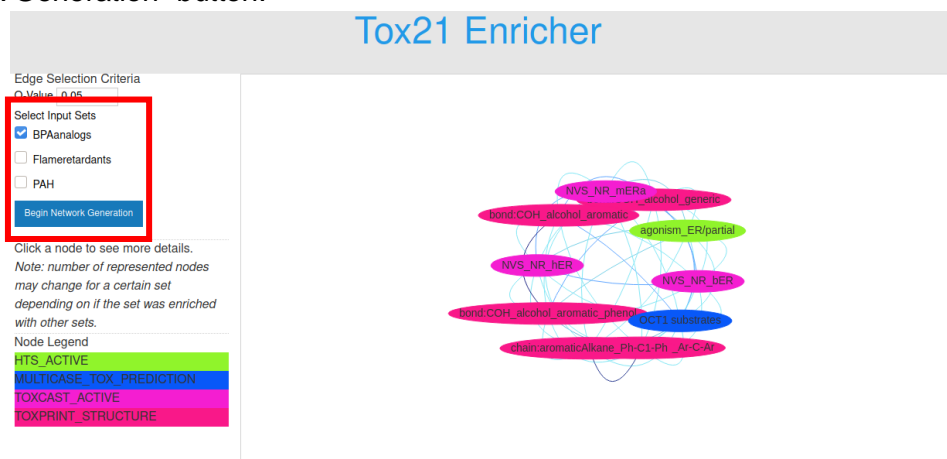
[Regenerate Network](#)

V. Network Page

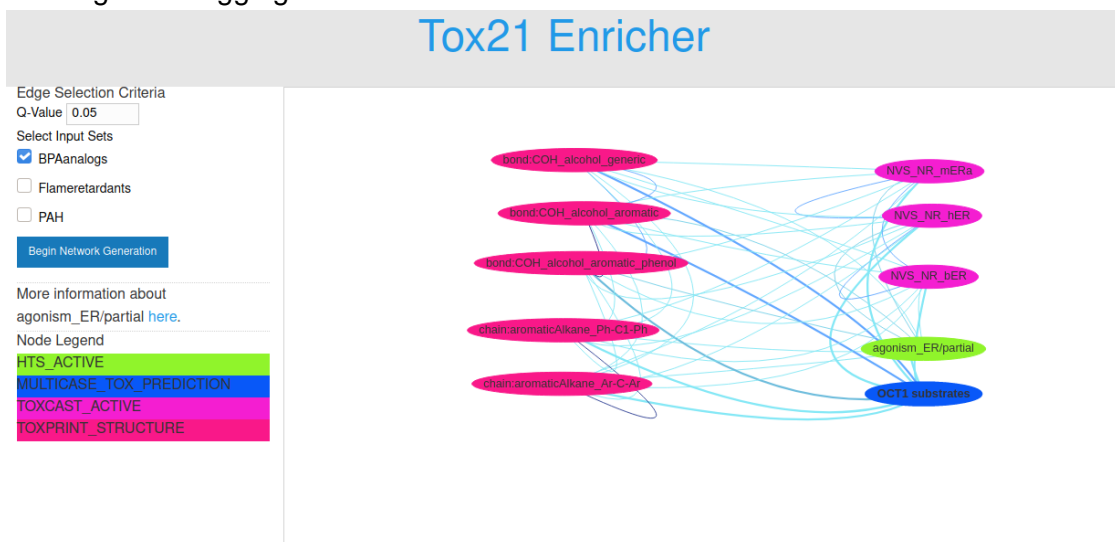
After clicking either of the network creation buttons, you will see the network page as shown below. The left side of this page is where we have edge selection criteria, input set select for node data to visualize, and the color key for the current network. In the screenshot below, we are using the Chart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



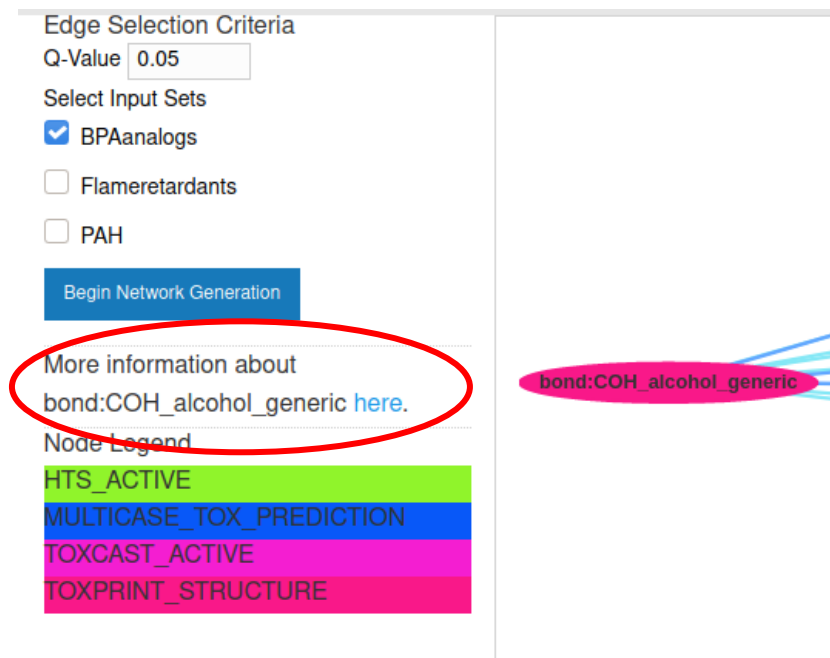
To view a subset of the network, select/deselect sets on the left and click the “Begin Network Generation” button.



Clicking and dragging a node allows the user to move it around in the network box.



Clicking on a node will also present a prompt with a link to get more details about the annotation's class.



END OF THE USER'S MANUAL