Toxygates User Guide

November 2016 update

Johan Nyström-Persson, Level Five Co., Ltd. (johan@level-five.jp)

This is a task-oriented guide to using Toxygates, an interactive toxicogenomics platform based on the Open TG-GATEs dataset. We hope that you will find Toxygates useful.

Contents

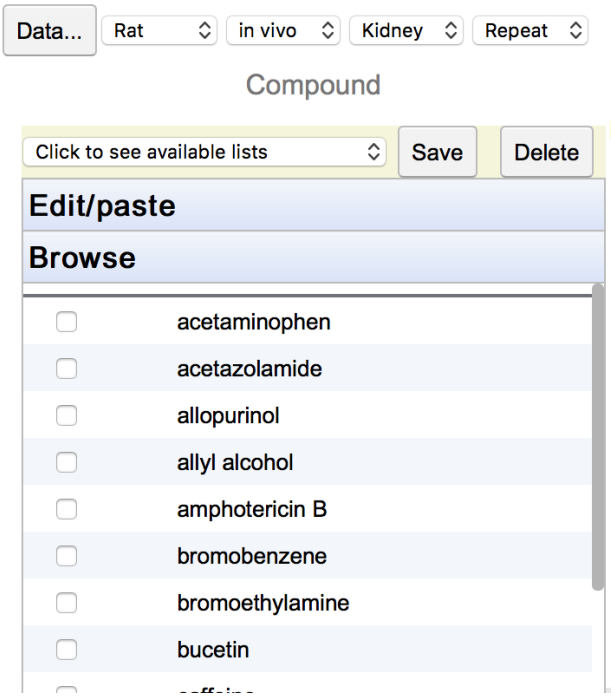
1. Basic concepts

Toxygates is split into several different **screens**, such as *Start*, *Sample groups, View data* and so on. The navigation links to move between screens (if they are available) are located at the top of the page.

To inspect data (such as Open TG-GATEs) in Toxygates, it is necessary to first define ***Sample groups***.

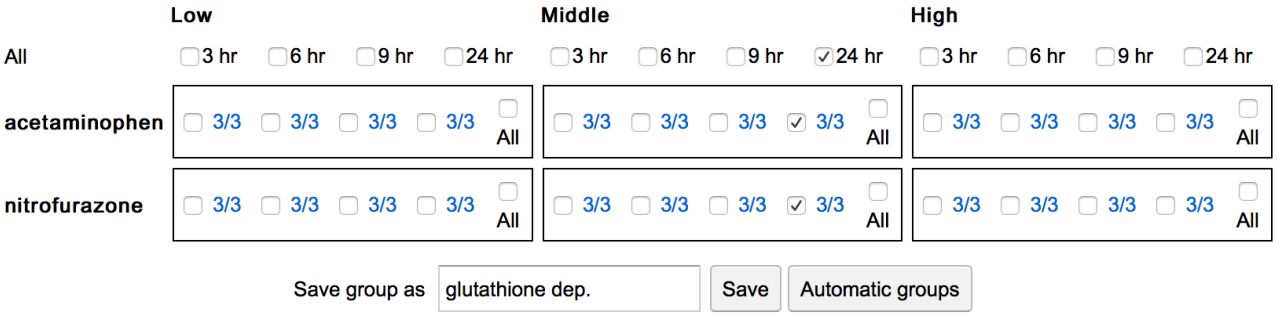
First select the **species, test type, repeat type and tissue** you are interested in. This can be done at the top left corner of the *Sample groups* screen.

Then select the **compounds** you are interested in from the compound list.



When you have selected one or several compounds, you can select **time and dose combinations** on the right hand side. In Open TG-GATEs, usually, each time and dose combination will correspond to 3 treated and 3 control samples (in some cases less).

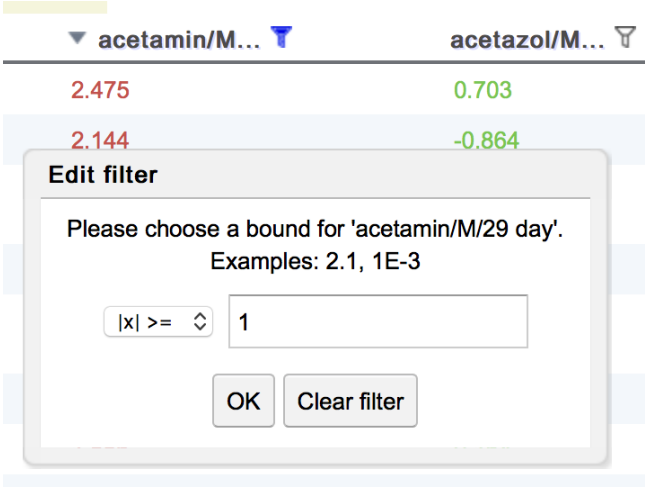
Check the boxes you wish to inspect and save the group. The selected samples will be saved together and later averaged when displayed. If you wish to contrast two different time points or dose levels, etc, you should save them as two groups with different names.

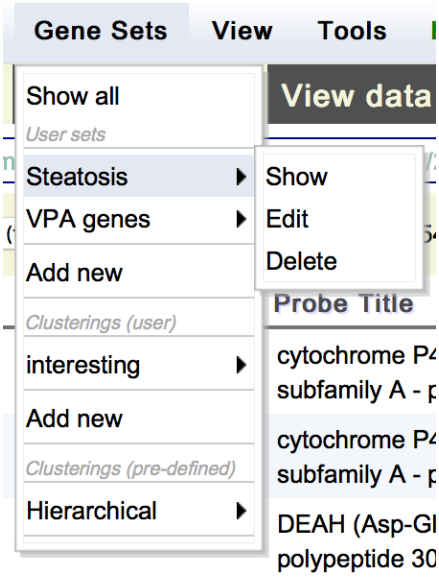


A **group name** will be automatically suggested, but you may enter another name if you wish. Saved groups will be stored in your browser, and will remain the next time you use Toxygates.

Once at least one group has been defined and saved, it is possible to proceed to the ***View data* screen**. Here you can see the data you have selected. Your sample groups will correspond to columns in the table, and genes (probes) will correspond to rows.

Genes may be **filtered** by clicking on the filter icon next to each column header. Here you can enter a numerical threshold for the column. This will exclude all genes that do not pass the threshold. A blue icon indicates that the filter is active.



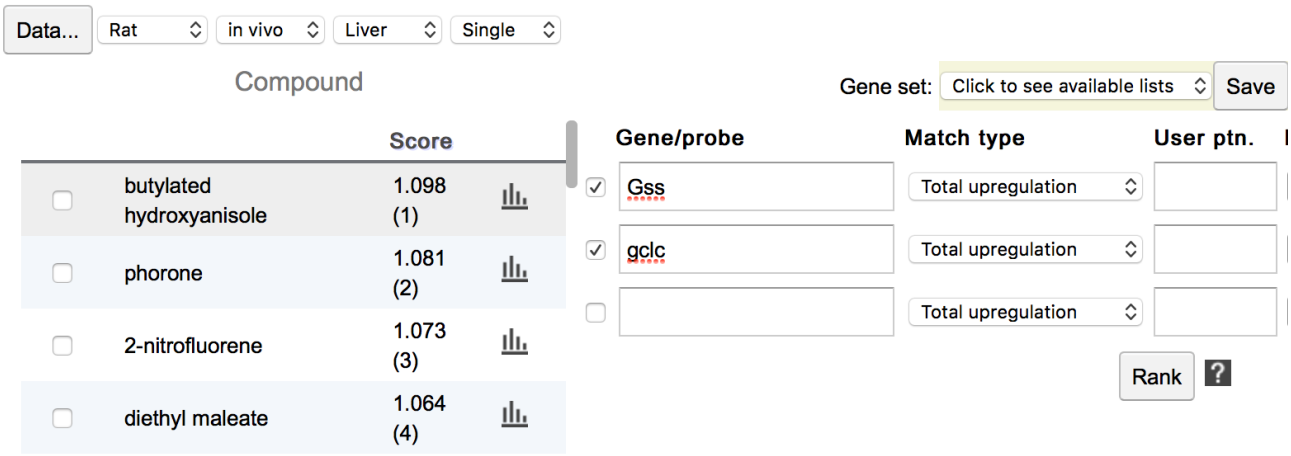


The currently displayed genes may be saved as a *gene set* and given a name. No more than 1000 genes can be saved in this way.

Such a gene set can also be edited before being saved (for example, by adding GO terms or pathways of interest).

Saved gene sets may be selected from the **Gene sets menu**, which is accessible on the *View data* screen. If you have previously carried out any clusterings and saved them, you will also find those here. The gene sets will be stored in your browser, and remain the next time you return to Toxygates.

If you have selected a gene set from this menu and want to go back to displaying all genes, please select “Show all” at the top of the gene sets menu. To see all genes, you may also need to clear any column filters you have activated.



You can also perform a **Compound ranking** on the screen with the same name. This function allows you to rank compounds according to their behaviour with respect to a set of genes, as seen in expression time series. To use this function, first go to the *compound ranking* screen, then select the organism, tissue and so on that you are interested in (similar to the sample groups screen). Then type in, in the field “Gene/probe”, the gene that you are interested in, for example Gss for glutathione synthetase. You can enter any number of genes (within reasonable limits). Many match types (ranking criteria) are available; the default is “total upregulation”, which ranks the compounds according to how much they upregulate the gene. (For reference, please use the Help menu and select “Help for this screen” while you are performing compound ranking).

When you have specified the rules, click the “Rank” button. When the ranking is complete, you will see the results in the compound list on the left. Each compound will have a number assigned to indicate its relative score. (This number is a relative measure only and has no biological meaning.) By clicking the chart icon next to each compound name, you can see the various time series for the relevant genes for that compound. The dose level that best matches your rules will be highlighted in blue.

This concludes the discussion of basic concepts. Many screens in Toxygates also have a help function, which will give you additional information. You can access it via the menu item “Help for this screen” on the help menu.

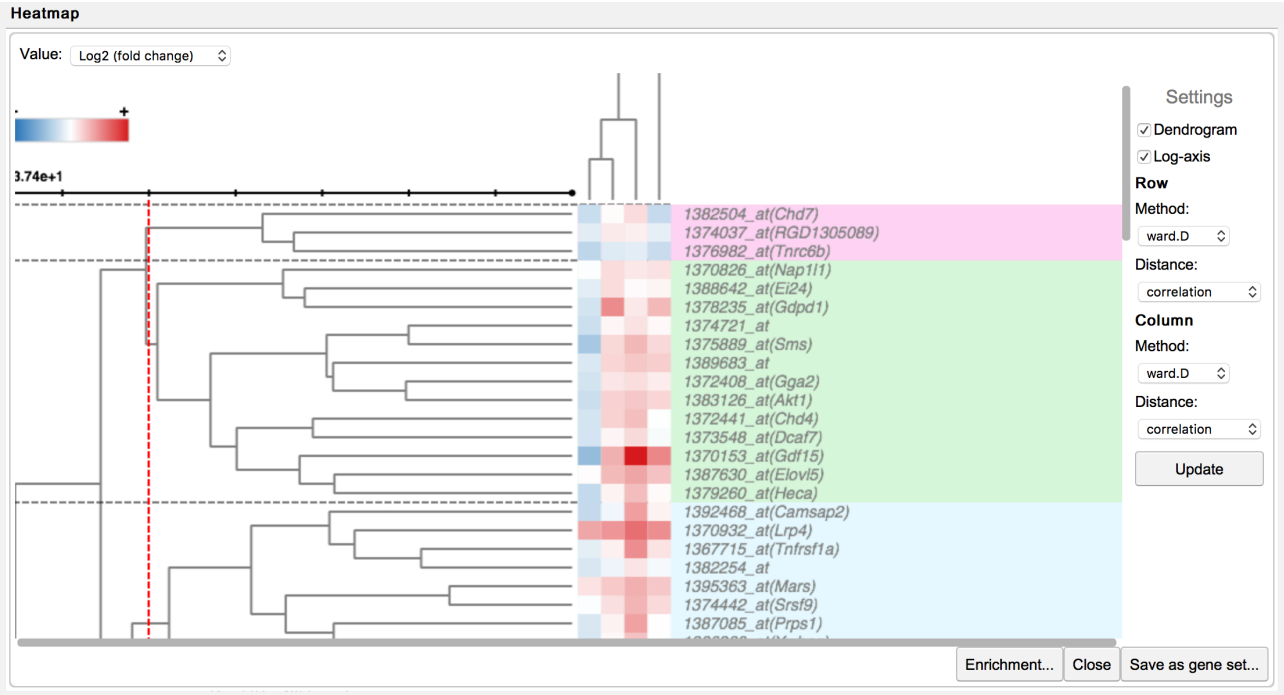
2. Clustering and heat maps

To perform a hierarchical clustering, you must first select the samples (compounds) and genes you are interested in. The Open TG-GATEs dataset is too large to be clustered all at once by Toxygates. To select compounds and samples, follow the steps in section 1 above, save your sample groups and then go to the *View data* screen. At least two groups must be defined.
Most likely, you will at first see too many genes being displayed. For example, the default number is 31,042 genes (probes) for rat samples. There are several ways to select a more specific gene set. Here we will show how to perform *p*-value filtering. This option is only available when your groups contain exactly one time/dose/compound combination, so for this guide, please do not combine multiple doses, times or compounds into one group.

On the data screen, check the “p-value columns” box at the top of the screen. Additional columns will appear. The first column is group 1, the second is *p*-values for group 1, the second is group 2, the third is *p*-values for group 2, and so on.

Click the filter icon to set a filter. Type in a value, such as 0.1. Genes with a larger associated *p*-value will now be excluded. The filter icon will become blue to indicate that the filter is active.

Your total gene count may still be more than 1000. You may filter the second group in the same way as the first group, or you may tighten the filter on the first group, for example to 0.05.



When you are happy with the total number of genes, select “Show heat map…” on the Tools menu.

After a little while, the heat map will appear. You may inspect it, and change parameters such as distance metric and clustering method on the right hand side. If you change the parameters, please click “Update” to see the result.

To define clusters, select a cutoff point by clicking on the dendrogram on the y-axis. This will partition the genes into gene sets corresponding to your selected clusters.

You can test your selected clusters for enrichment of various kinds, for example of pathways and GO terms, by clicking the Enrichment button. This may take a little while to compute. The result

table will show the best matching annotation for each cluster. The enrichment is performed by using TargetMine (see Section 4 below).

Best enrichment results

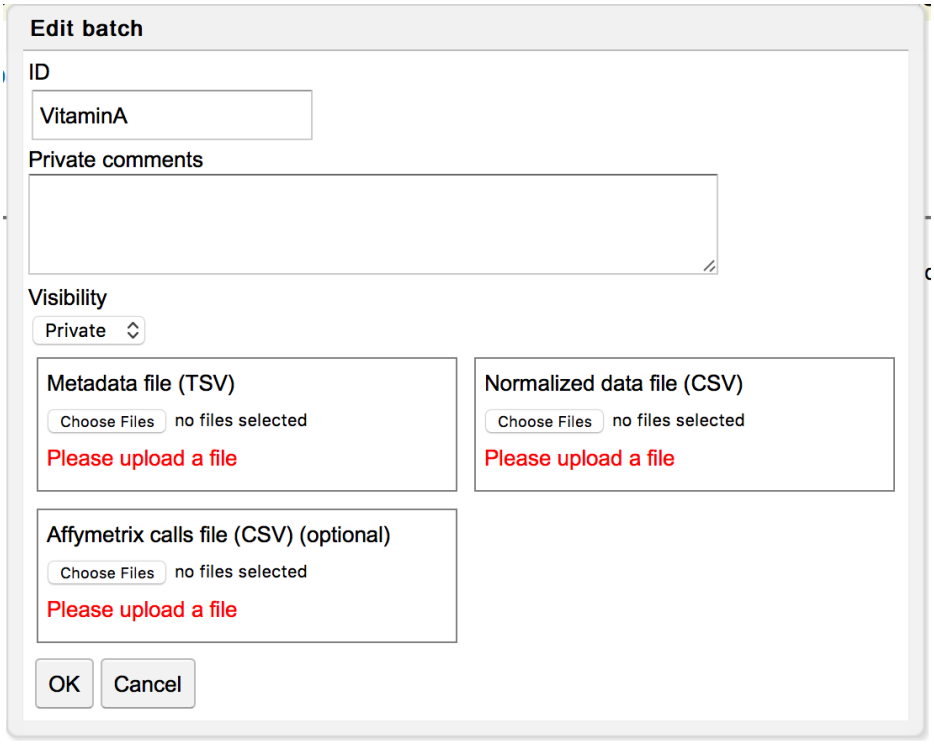
Cluster	Size	ID	Description	p-value	Matches
Cluster 1	181	R-RNO-72689	Formation of a pool of free 40S subunits	0.000183	8
Cluster 2	293	(No result)			
Cluster 3	142	(No result)			

Such enrichment testing may, in some situations, indicate whether your clustering is meaningful.

If you wish, you may repeat the steps above. Once you are happy with your clusters, you may click the Save button to save the clusters as gene sets. They will then become accessible on the “Gene sets” menu on the data screen.

3. Uploading your own data

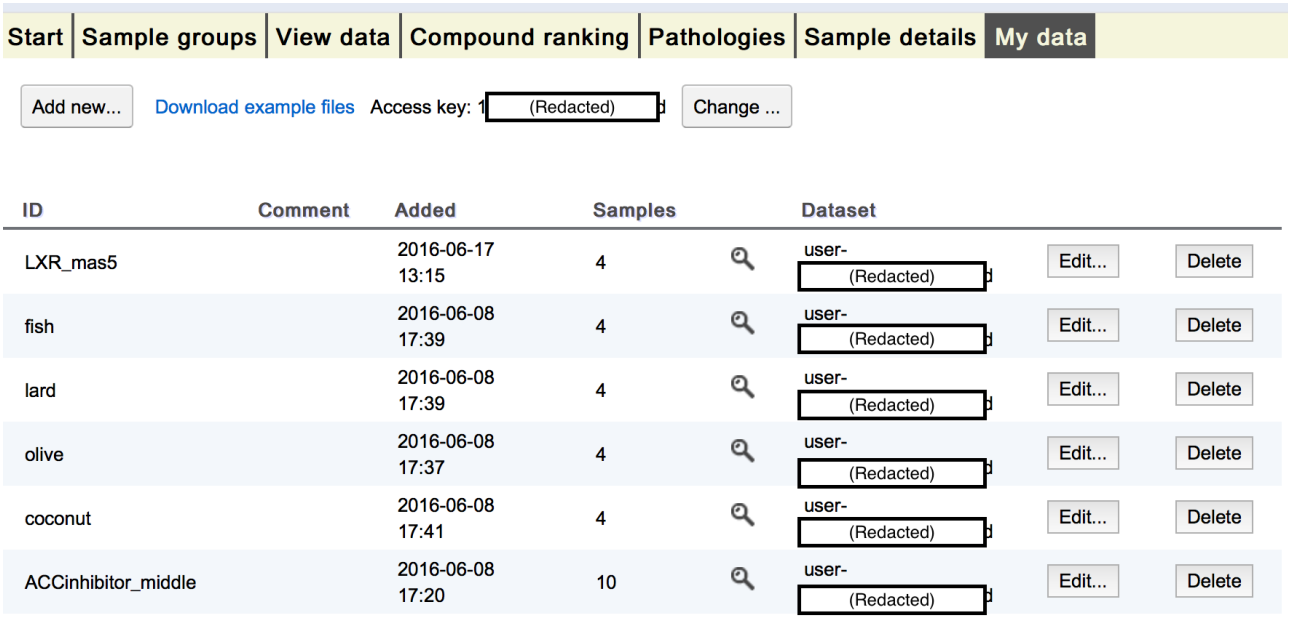
Toxygates supports data upload of pre-normalised microarray data from rat, human and mouse platforms. Normalisation can be carried out using the *affy* package in R (bioconductor).
Uploaded data is managed from the *My data* screen. We recommend that you first download the example files (linked from that screen) to understand the necessary data format.



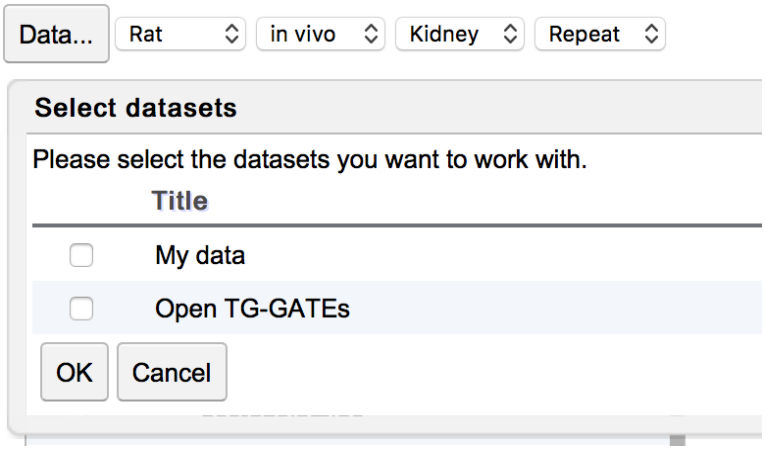
Once you have prepared the necessary files, you can upload them. Data is uploaded as *batches* (sets of samples). To begin an upload, click the “Add new…” button.
Next, you will need to fill out some details, such as a unique string to identify the batch, and proceed to upload the necessary files. If there is any error in the files, you should get a warning

message indicating what is wrong. Once you click “OK”, the upload process starts. It may take several minutes to finish.

After uploading, you can manage your batches from the My data screen. You can inspect the contents of a batch by clicking the magnifying glass. Editing and deleting a batch may be done by using the respective buttons.



When data has been uploaded, you can examine it as follows. Go to the sample group definitions screen (as in section 1) and select the necessary dataset, by using the “Data…” button. Your uploaded samples will be in a dataset called My data, which you should enable.



You can optionally also enable other datasets, such as Open TG-GATEs, if you wish to perform a comparison with this data. Your compound names will be prefixed with the word [user]. For example, if you uploaded Acetaminophen data, it will appear as [user] Acetaminophen to distinguish it from the pre-existing data in Toxygates.
You can now proceed as before to view your data (see section 1) or cluster it (see section 2) by including it in sample groups.

Your data will be associated with an *access key,* which you can see at the top of the *My data* screen. The access key is a password that protects your data from other users. The key is automatically saved in your browser, but we recommend that you write it down, to avoid losing

access to your data. If you know your key, you can access your data on a different computer or in a different web browser by clicking the “Change…” button and entering the key.

4. Other tasks

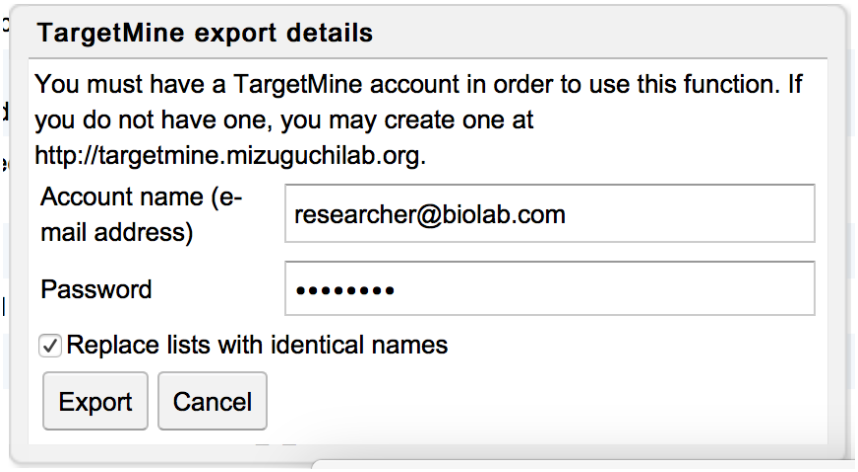
Orthologous data inspection
If you define sample groups (see Section 1 above) from multiple species (for example, one group with human samples and one with rat samples), then the Affymetrix probes (and corresponding genes) will be combined in the data table, based on orthologous amino acid sequences in the relevant proteins. This allows for easier cross-species data analysis. In the example above, two probes have been combined in the first row (1 human, and 1 rat), and 8 in the second row (7 human, 1 rat). Rat probes are prefixed with “Rat” and human probes with “HG”. For each row, the data values displayed will be the median value of all the underlying probes’ expression values.

	Gene Symbol	Probe Title	Probe	acetamid/M…	▲ valproic/H…
	HG-..:ACTN3 (1 probe) Rat..:Actn3 (1 probe)	actinin - alpha 3 (1 probe) actinin alpha 3 (1 probe)	1367962_at 206891_at	(absent)	0.111
	HG-..:LOC1006527 (1 probe) HG-..:LOC1006533 (1 probe) HG-..:SCD (5 probes) Rat..:Scd1 (1 probe)	stearoyl-CoA desaturase (delta-9-desaturase) (4 probes) stearoyl-Coenzyme A desaturase 1 (1 probe) uncharacterized LOC100653313 (1 probe)	1370355_at 200831_s_at 200832_s_at 211162_x_at 211708_s_at 223839_s_at	0.929	0.165
	HG-..:IRS1 (3 probes) Rat..:Irs1 (1 probe)	insulin receptor substrate 1 (4 probes)	1369771_at 204686_at 238933_at 242979_at	1.118	0.197

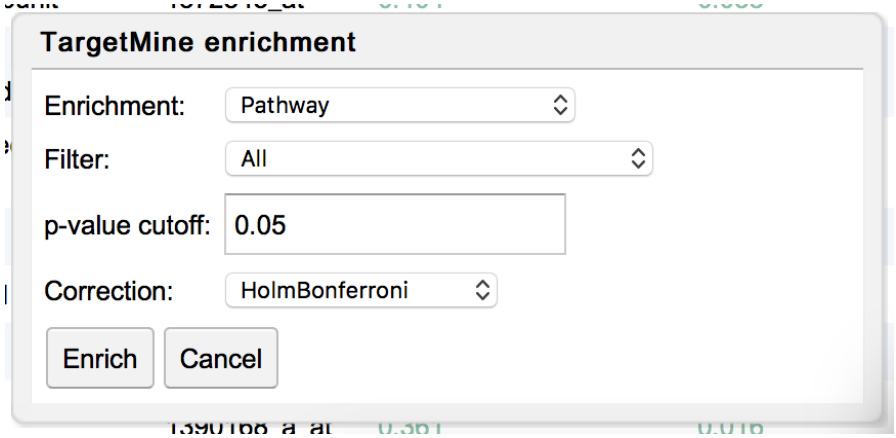
Import/export to TargetMine

TargetMine (**http://targetmine.mizuguchilab.org**) is a data warehouse for drug discovery. If you have a user account, you can synchronise your gene lists in Toxygates with those in TargetMine, through either importing or exporting them. This is done on the Tools menu on the *View data* screen. TargetMine provides many analysis functions not available in Toxygates and may allow you to obtain additional insight about your gene sets.

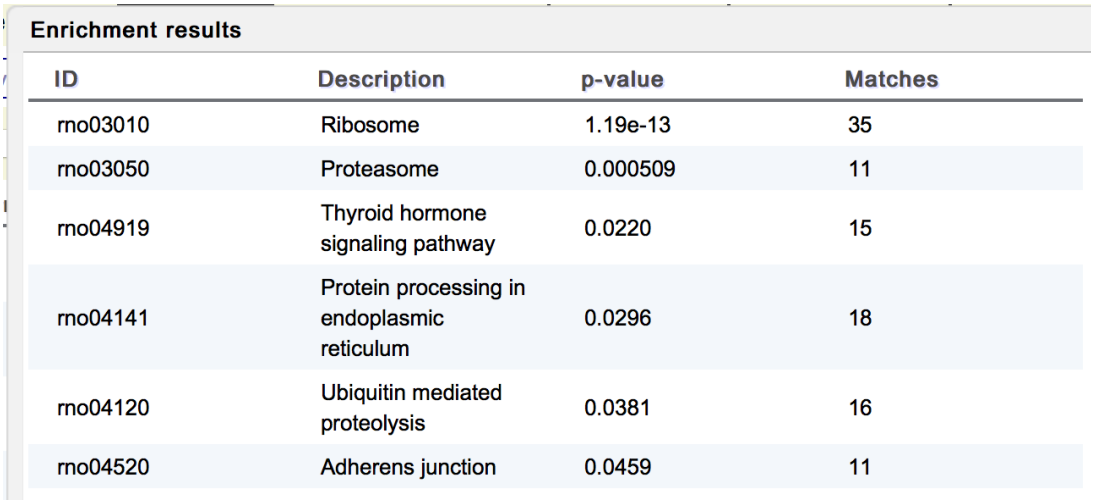
Enrichment testing
By using this function, also available on the Tools/TargetMine data menu on the *View data* screen, you can test your currently displayed gene set for enrichment. Various enrichment types are



available, such as pathways, GO terms, GOSlim terms and integrated pathway clusters (IPCs). For each type, sub-categories and other parameters are also available. Please note that no more than 1000 genes can be tested simultaneously.

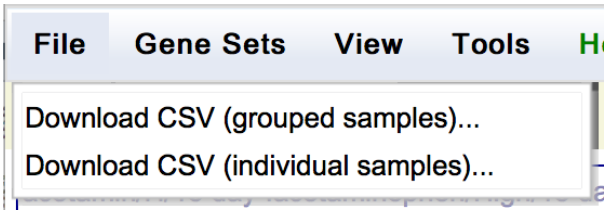


The result of enrichment testing is a list of enriched objects, the number of matching genes, and the corresponding *p*-value for each entry.



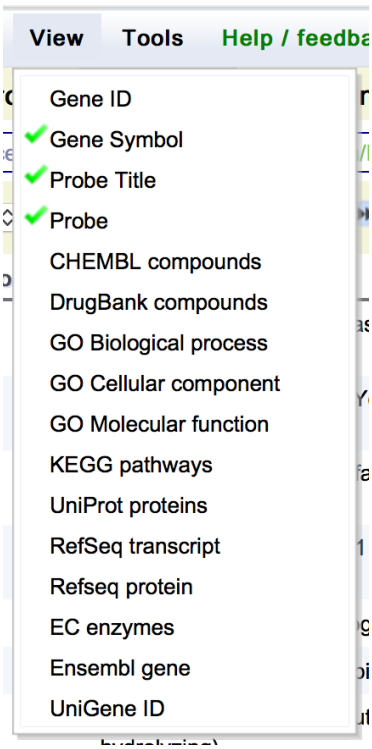
ID	Description	p-value	Matches
rno03010	Ribosome	1.19e-13	35
rno03050	Proteasome	0.000509	11
rno04919	Thyroid hormone signaling pathway	0.0220	15
rno04141	Protein processing in endoplasmic reticulum	0.0296	18
rno04120	Ubiquitin mediated proteolysis	0.0381	16
rno04520	Adherens junction	0.0459	11

Downloading data



From the File menu on the *View data* screen, it is possible to download the data displayed in the data table as a CSV file. Downloading the data as grouped samples will show it as averaged values, as it is displayed in Toxygates. Downloading the data as individual samples, on the other hand, will break out each value of each sample separately, which may be useful to carry out further statistical testing (for example, to compute your own *p*-values).

Viewing pathways, GO terms and other annotations



By using the View menu on the *View data* screen, it is possible to enable or disable additional columns in the data table. This will show you additional information about your probes and genes. In many cases, these additional items will also have hyperlinks, taking you to external web pages with further information about them.

Inspecting sample attributes and biochemical data

On the *Sample details* screen you can see all the attributes for all the samples in the groups you have defined (see Section 1 above). This includes blood biochemical data as well as various measurements.

Viewing pathologies

Pathologies may be displayed on the *Pathologies* screen. Only pathologies for your currently defined sample groups will be visible (see section 1 above).