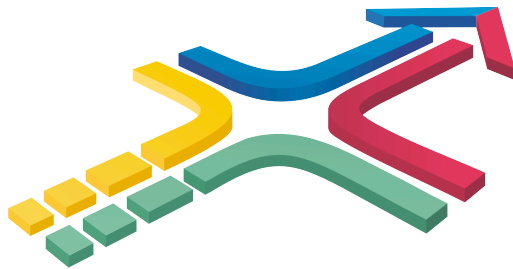


The geneXplain platform



**Workshop SW2: Pathway Analysis in
Transcriptomics, Proteomics and Metabolomics**

Saturday, March 17, 2012

geneXplain GmbH
Am Exer 10b
D-38302 Wolfenbüttel
Germany

E-mail: olga.kel-margoulis@genexplain.com,
alexander.kel@genexplain.com

URL: <http://www.genexplain.com>

1. Launch pre-defined workflow “Find master regulators in networks (GeneWays)”

1.1. Run workflow

This workflow helps searching for master regulators in a network. Master regulators are common nodes upstream of the molecules encoded by a given set of genes. They can be considered as biomarkers or as potential drug targets.

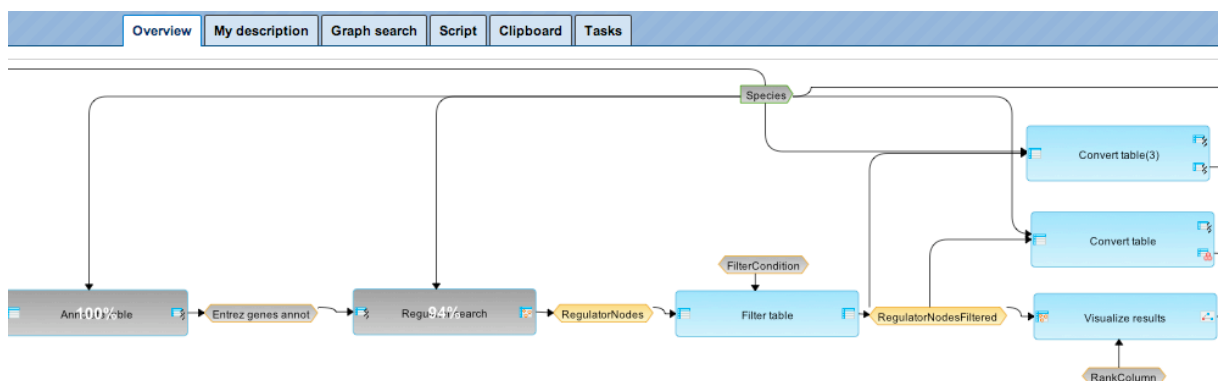
Open the form to start the workflow.

Select the input file as a gene table (📄) or protein table (📄) which may contain any common identifiers (Refseq, UniProt, etc), and select the corresponding species. In the field “ResultFolder” specify the path where the resulting folder will be located.

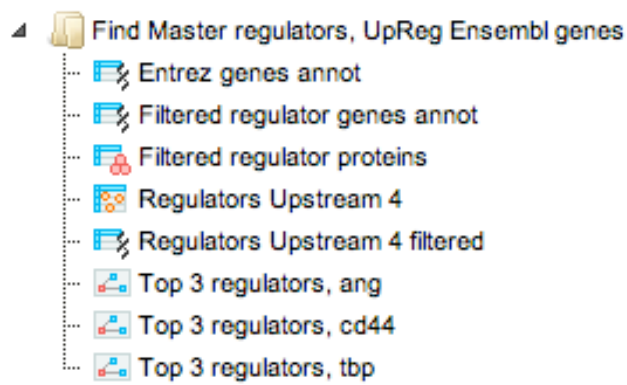
With a click over the default path, a new window will appear; select the location and press “ok”.

Then “Run” workflow.

While the workflow is running, observe the progress in the Operations Field under the tab “Overview”. On the workflow diagram, you can see what steps are already completed, and where the process is now.



When the workflow is completed, the output is stored in a newly generated folder, which contains several files:



1.2. Have a look into the resulting files

The table *Entrez genes annot* (📄) corresponds to your input table converted from the input type of identifiers into the Entrez Gene_IDs identifiers and annotated with gene symbols and gene descriptions.

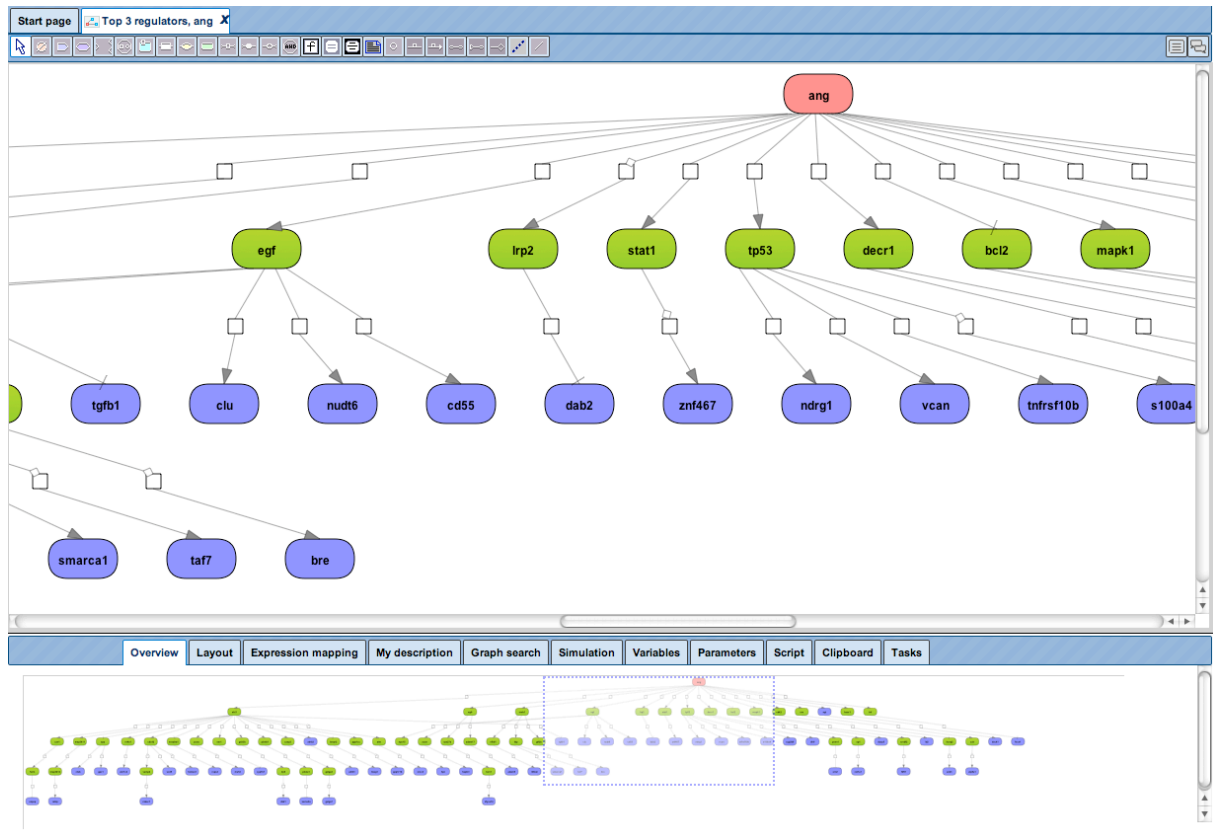
The table *Regulators upstream 4* (📄) is the resulting list of the master regulator molecules at 4 steps upstream of the input molecules. Each master regulator molecule is characterized by Score, Z-score and FDR. The number of the molecules from the input set that can be reached from the master regulator is shown in the column **Reached from set**. Total number of the molecules that can be reached from the master regulator in the network, without regard to the input list, is shown in the column **Reachable total**. Having this table open in the Work space you can find additional options available, specific for this kind of table. You can visualize network of the selected master regulators (🌐), save network as a list of genes in the Tree (📄), or save hits of this network as a list of genes in the Tree (📄).

The tables *Filtered regulator genes annot* (📄) and *Filtered regulator proteins* (📄) result from the table *Regulators upstream 4* (📄) upon filtering it by Z-score >1 and further conversion correspondingly into Ensemble genes or UniProt IDs.

Three diagrams *Top 3 regulators,...* (🌐) present network visualization for three top master regulators, upon sorting the table *Regulators upstream 4* (📄) by the column Score. If you are interested to visualize network for any other master regulator, you can do this manually and save a new diagram into the Tree.

1.3. Layout networks of the master regulators


By double click on the diagram "Top 3 regulators, ang" open this network diagram in the Work Space.

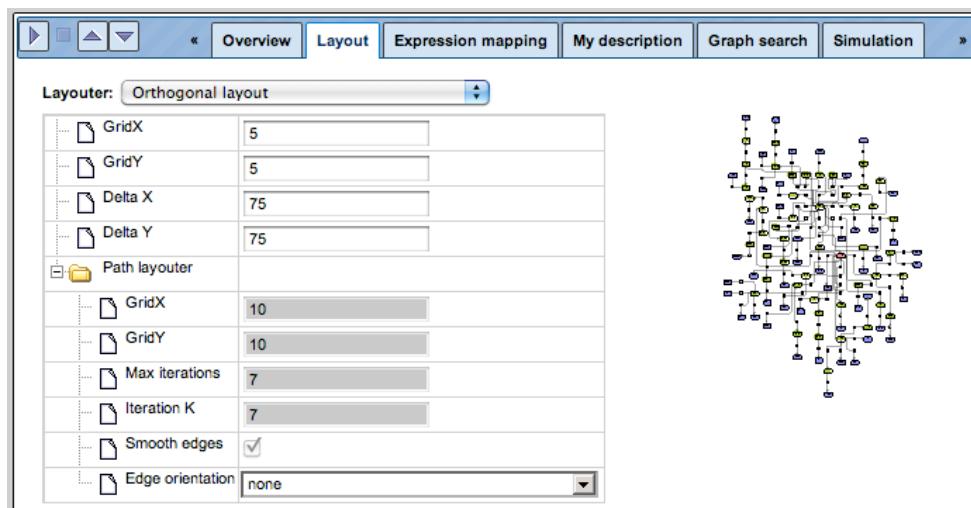



Master regulatory molecule is shown by red color; from genes/molecules we started from – by blue, and the connecting molecules added by the algorithm, are green. Here, all the underlying reactions are based on the GeneWays database.

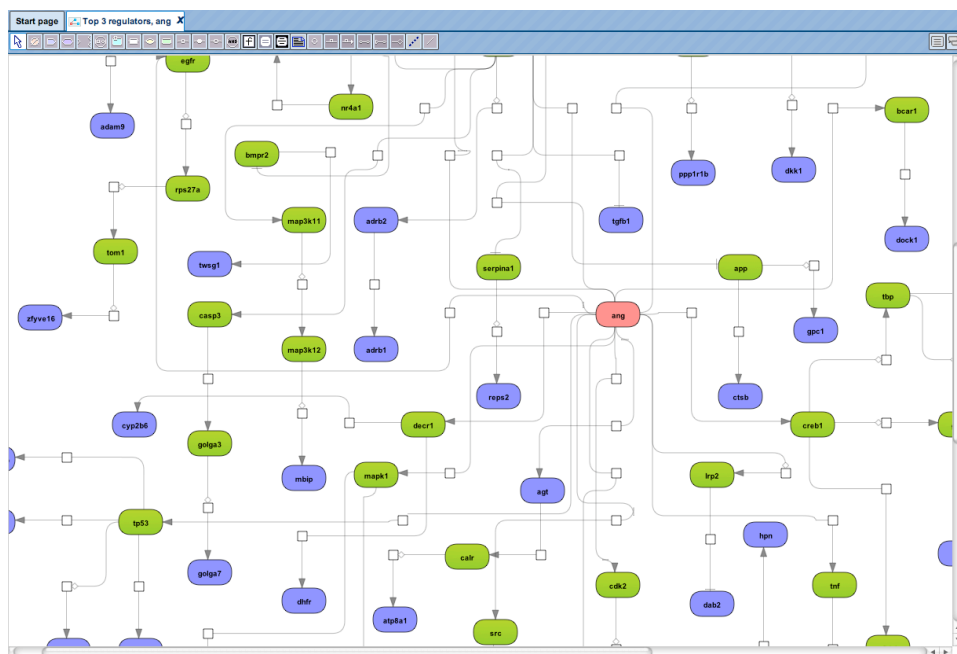
GeneWays is a database about genes and their functional interactions. The underlying data have been retrieved from the original scientific literature by a sophisticated text mining system applied to more than 360,000 full text papers and of more than eight million publication abstracts [Iossifov I., Rodriguez-Esteban R., Mayzus I., Millen K.J., and Rzhetsky A. Looking at cerebellar malformations through text-mined interactomes of mice and humans. *PLoS Comput Biol.* 2009, 5:e1000559. PubMed PMID: 19893633].

By default, all network diagrams are shown in hierarchical layout, with the master regulatory molecule on top.

To change layout, open tab “Layout” in the Operations Field. When you have selected another layout type, press the “Prepare layout” button (), showing the new layout at the right of the same window in the Operation Field.



Now press "Apply layout" () button to transfer the new layout to the Work Space. Below an example of the same diagram in orthogonal layout is shown.



Next, transfer this diagram into force directed layout, and then return to the hierarchical vertical layout.

1.4. Extension of diagrams via the Graph Search functionality

If you are interested to add more interactions to the molecule of interest onto this diagram, apply Graph Search option.

Select molecule of interest on the diagram opened in the Work Space, in this example TGFB1 is selected, and open the "Graph Search" tab in the Operations Field.

The screenshot shows a network graph with nodes representing genes/proteins and edges representing interactions. The nodes are color-coded: green (src, egfr, lrp2, decr1, stat1, egf, cdk2, acan, nudt6, eps15, hpn), blue (tgfb1, dusp5, dab2, dhfr, cyp2b6, znf467, hpin1, hpn), and purple (ank3, tom1, adam9, adam17). The search configuration panel on the right is titled "Graph search" and includes the following settings:

- Search type: neighbours
- Options:
 - Direction: Down
 - Depth: 1
- Target databases:
 - Biopath
 - EndoNet
 - GO-BKL
 - GeneWays


In the form for Graph Search, specify the search engine as "GeneWays search"

Name	GeneWays
Search engine	GeneWays search


and press  button to add the selected molecule to the elements pane.

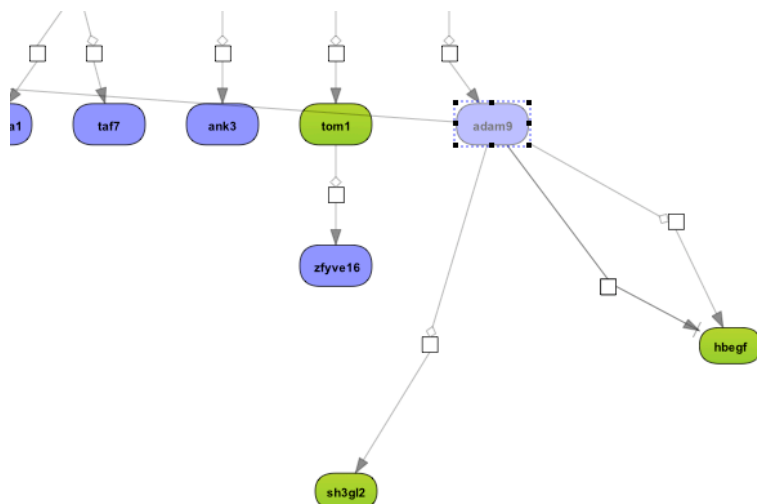
The screenshot shows the software interface with the "Graph search" tab selected. Below the navigation tabs is a table with the following data:

<input checked="" type="checkbox"/> Add	<input checked="" type="checkbox"/> Use	Database	ID	Title	Type	Linked from
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	GeneWays	7040	tgfb1 (h)	Gene	

Apply the  button to start the search, and you will get the list of all interactions where this molecule is involved.


<input checked="" type="checkbox"/> Add	<input checked="" type="checkbox"/> Use	Database	ID	Title	Type	Linked from
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	GeneWays	8754	adam9 (h)	Gene	
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	GeneWays	8754_351_r	adam9 (h) regulates app (h)	Reaction	8754
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	GeneWays	351	app (h)	Gene	8754_351_r
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	GeneWays	8754_1839_a	adam9 (h) activates hbegf (h)	Reaction	8754
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	GeneWays	1839	hbegf (h)	Gene	8754_1839_a
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	GeneWays	8754_1839_i	adam9 (h) inactivates hbegf (h)	Reaction	8754
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	GeneWays	1839	hbegf (h)	Gene	8754_1839_i
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	GeneWays	8754_1839_r	adam9 (h) regulates hbegf (h)	Reaction	8754
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	GeneWays	1839	hbegf (h)	Gene	8754_1839_r
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	GeneWays	8754_6456_r	adam9 (h) regulates sh3gl2 (h)	Reaction	8754
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	GeneWays	6456	sh3gl2 (h)	Gene	8754_6456_r


Now press the  button to add the search items to current diagram.

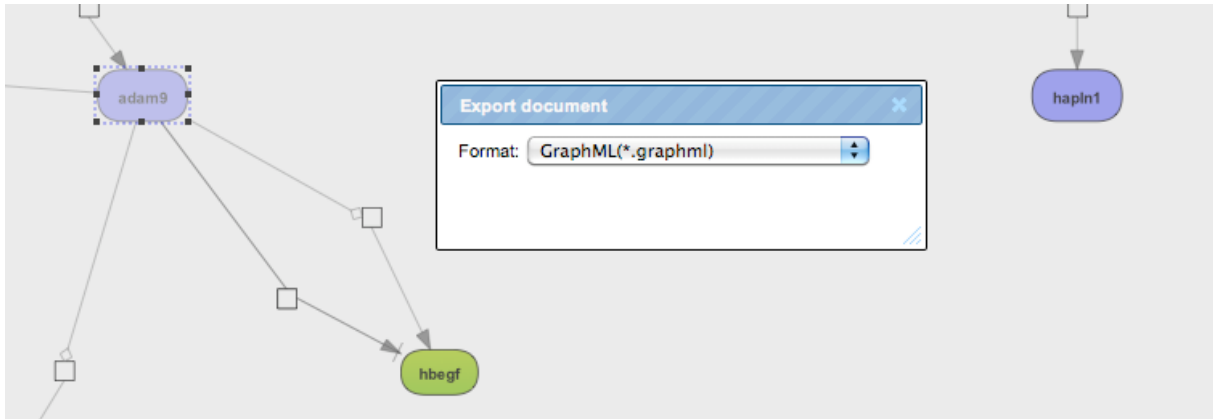


You can make a new layout to get a nice diagram view again.

1.5. Save updated diagram into the tree or export to your local computer

Apply save as  button to save the diagram into your preferable location in the tree.

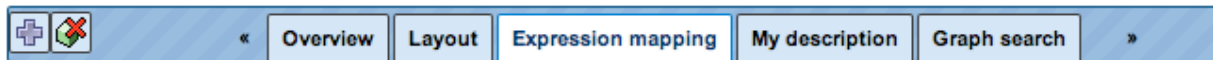
If you would like to export this diagram to your local computer, a number of formats are available. While a diagram is opened in the Work Space, press  (export) button, and a new window with a pull down menu will come up.



Have a look, what formats are available in the menu, and saved a diagram in the format you like.

1.6. Mapping expression data onto the diagram

Open "Expression mapping" tab in the Operations Field.



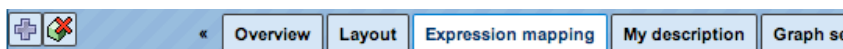
Mapping: (none) ▾

To map expression values on the diagram, drag and drop the corresponding table from the tree area over the diagram. You can then adjust parameters in the form that will be displayed.

data/Projects/ABRF-2012/Data/Colon Cancer/Experiment normalized (MAS5) (Differentially expressed genes Affy)/Probes_Scores and p-value UpDownReg Entrez genes

On the Figure below the file with expression data used is highlighted by blue in the tree area.

Drag and drop this file over the diagram, and the form will automatically appear in the "Expression mapping" tab which was empty before.







Mapping: Probes_Scores and p-value UpDownReg Entrez genes ▾

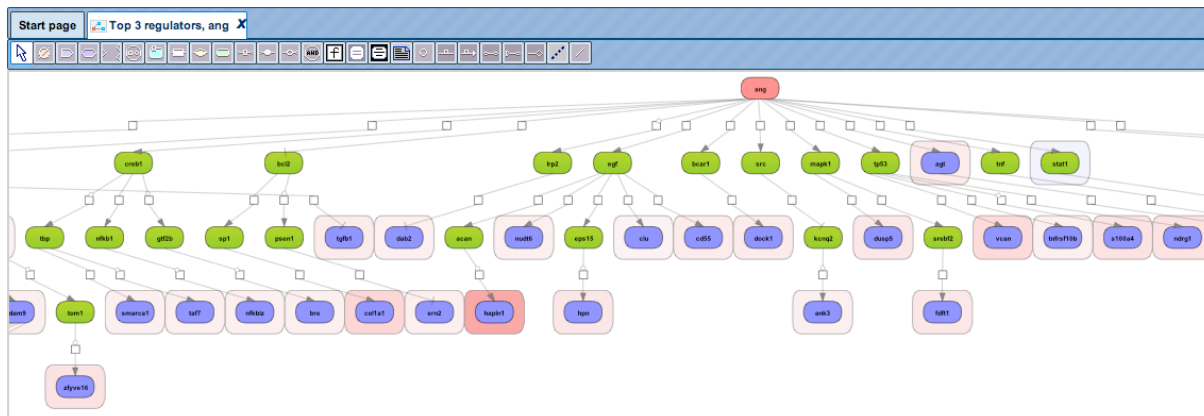
Data source	
Table	Probes_Scores and p-value UpDownReg Entrez genes
Columns	[3] (all columns), -log(P-value), LogFoldChange ▾
Type	Outline fill ▾
Auto min/max	<input checked="" type="checkbox"/>
Minimum value	-8.04489441467489
Start-color	
Maximum value	8.04489441467489
End-color	
Use color for zero	<input checked="" type="checkbox"/>
Zero color	

In this form, in the field "Table" the path to the table is inserted automatically.

In the next field, "Columns", select the column whose expression values you would like to map, LogFoldChange. You can also adjust colors for upregulated and down-regulated genes as it is shown below.

 « Overview Layout Expression mapping My description Graphs	
Mapping: Probes_Scores and p-value UpDownReg Entrez genes	
Data source	
Table	...cores and p-value UpDownReg Entrez genes
Columns	LogFoldChange
Type	Outline fill
Auto min/max	<input checked="" type="checkbox"/>
Minimum value	-6.45346607083302
Start-color	
Maximum value	6.45346607083302
End-color	
Use color for zero	<input checked="" type="checkbox"/>
Zero color	

After pressing  icon you can see the expression mapped. Differentially expressed molecules are highlighted by the colour layer on the diagram, which shown below on the Figure.

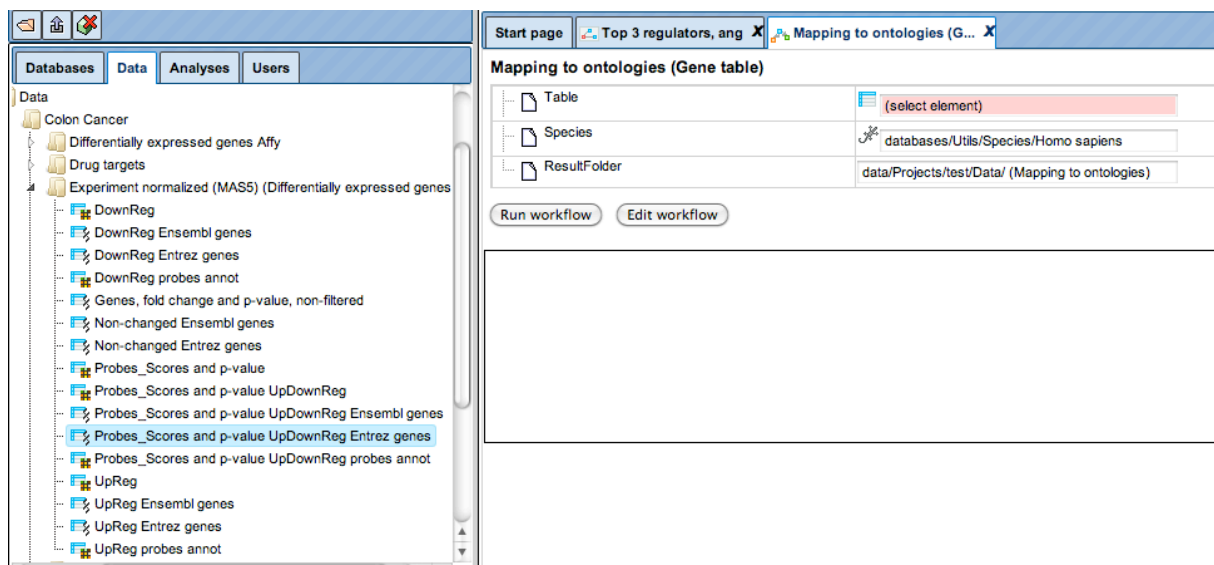


The expression values can be inferred by the gradation in colour. More intensive color corresponds to a bigger fold change value whereas the lighter shade corresponds to a smaller fold change value. In this example all the upregulated genes are shown with a colour gradient from white to red whereas down-regulated genes are shown in with a colour gradient from white to blue.

2. Launch pre-defined workflow “Mapping to ontologies”

2.1. Run workflow

All genes or proteins in the selected table can be mapped to GO categories, Reactome pathways, HumanCyc pathways, and TF classification. For example, you can use tables with upregulated or downregulated genes, or any table with genes or proteins as input into the workflow „Mapping to ontologies“.



Select the input file is a gene table (📄) or protein table (📄) which may contain any common identifiers, and select the corresponding species. In the field “ResultFolder” specify the path where the resulting folder will be located.

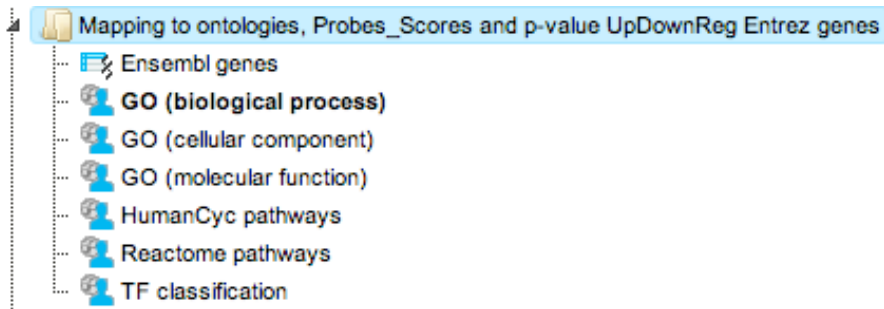
With a click over the default path, a new window will appear; select the location and press “ok”.

Then “Run” workflow.

While the workflow is running, observe the progress in the Operations Field under the tab “Overview”. On the workflow diagram, you can see what steps are already completed, and where the process is now.

2.2. Have a look into the resulting files

When the workflow is completed, a new folder is formed and several files are deposited by the program:




The results tables are automatically open in the Work Space and look like (e.g., *GO (biological process)*) may look like this:

ID	Level	Title	Number of hits	Group size	Expected hits	P-value	Hit names
GO:0002474	5	antigen processing and presentation of peptide antigen via MHC class I	85	181	16.00867	1.17E-41	B2M, BCAP31, BLMH, CTSS, ERAP1, ...
GO:0002478	5	antigen processing and presentation of exogenous peptide antigen	82	174	15.38955	2.2843E-40	B2M, CD74, CTSS, HLA-A, HLA-A (ENSG00000206503), ...
GO:0019884	4	antigen processing and presentation of exogenous antigen	82	176	15.56644	6.7579E-40	B2M, CD74, CTSS, HLA-A, HLA-A (ENSG00000206503), ...
GO:0048002	4	antigen processing and presentation of peptide antigen	86	212	18.75049	4.5646E-36	B2M, BCAP31, BLMH, CD74, CTSS, ...
GO:0019060	4	intracellular transport of viral proteins in host cell	18	19	1.68047	1.7542E-18	DERL1, IFIT1, TAP1, TAP1 (ENSG00000168394), TAP1 (ENSG00000206297), ...

The output tabulates how many and which genes from your list ("hits") fall into which category, how many known genes are in this category, how many hits would have been expected by chance, and what is the P-value for the found number of hits being obtained by chance.

2.3. Further possible actions on the results

Any table can be filtered by any field. Filtering conditions can be set up under the tab "Filters" in the Operations Field. Try to apply a filter.

Filtered table can be saved in the tree as a new table, with the icon  in the Operations Field. Save filtered table on your computer.

Genes/proteins falling into one category, can be saved as a new table in the tree. Choose a category of interest and save genes as a new table in the tree.