

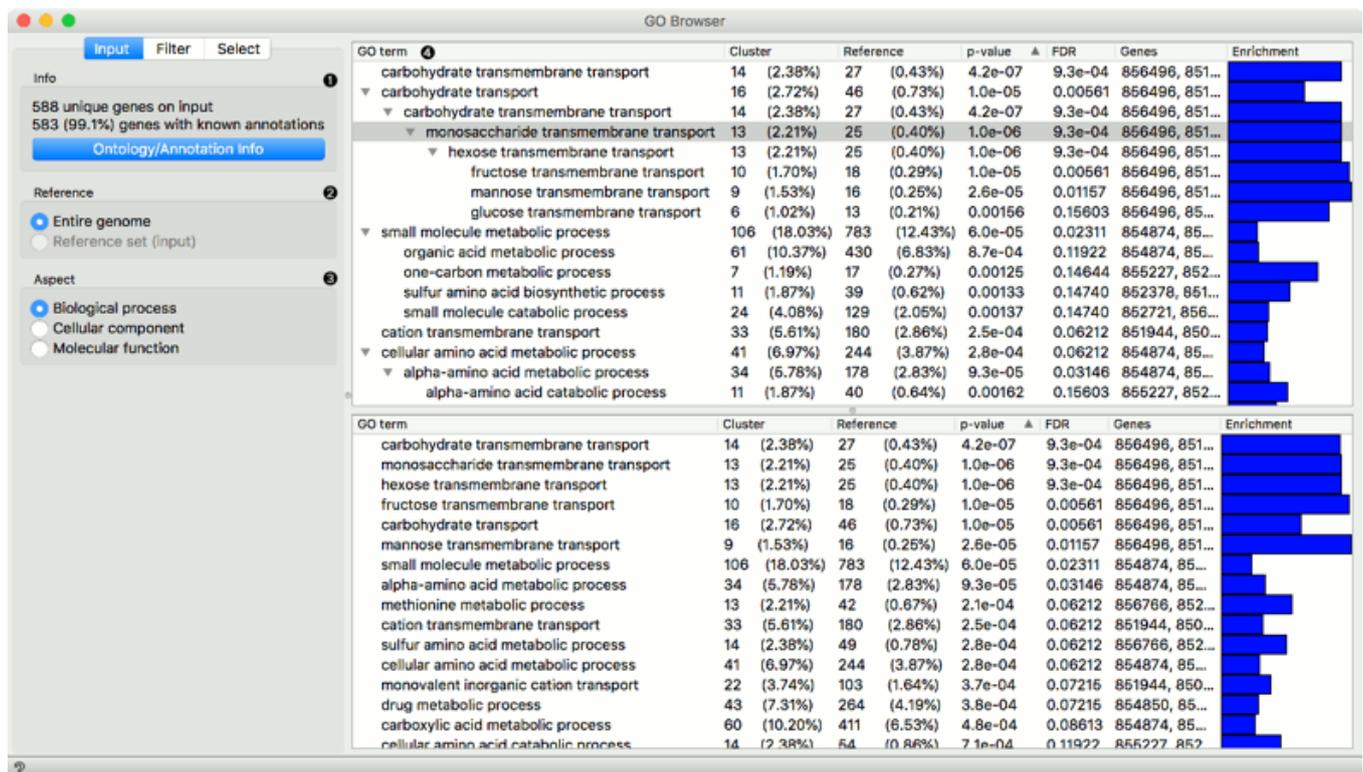
GO Browser

Provides access to Gene Ontology database.

Inputs - Cluster Data: Data on clustered genes. - Reference Data: Data with genes for the reference set (optional).

Outputs - Data on Selected Genes: Data on genes from the selected GO node. - Enrichment Report: Data on GO enrichment analysis.

GO Browser widget provides access to [Gene Ontology database](#). Gene Ontology (GO) classifies genes and gene products to terms organized in a graph structure called an ontology. The widget takes any data on genes as an input (it is best to input statistically significant genes, for example from the output of the **Differential Expression** widget) and shows a ranked list of GO terms with p-values. This is a great tool for finding biological processes that are over- or under-represented in a particular gene set. The user can filter input data by selecting terms in a list.



INPUT tab

1. Information on the input data set. *Ontology/Annotation Info* reports the current status of the GO database.
2. Select the reference. You can either have the *entire genome* as reference or a *reference set* from the input.
3. Select the ontology where you want to calculate the enrichment. There are three *Aspect* options:
 - [Biological process](#)
 - [Cellular component](#)
 - [Molecular function](#)
4. A ranked tree (upper pane) and list (lower pane) of GO terms for the selected aspect:
 - **GO term**
 - **Cluster:** number of genes from the input that are also annotated to a particular GO term (and its proportion in all the genes from that term).
 - **Reference:** number of genes that are annotated to a particular GO term (and its proportion in the entire genome).
 - **P-value:** probability of seeing as many or more genes at random. The closer the p-value is to zero, the more significant a particular GO term is. Value is written in [e notation](#)).
 - **FDR:** [false discovery rate](#) - a multiple testing correction that means a proportion of false discoveries among all discoveries up to that FDR value.
 - **Genes:** genes in a biological process.
 - [Enrichment](#) level

The screenshot shows the 'Filter' tab of the 'GO Browser' widget. The interface is split into two main panels.

Left Panel:

- Filter GO Term Nodes (1):**
 - Genes: # 1
 - p-value: 0.01000000
 - FDR: 0.20000000
 - Significance test (2):**
 - Binomial
 - Hypergeometric
- Evidence codes in annotation (3):**
 - EXP: 0 annots(0 genes)
 - IDA: 13907 annots(4459 genes)
 - IPI: 2127 annots(1388 genes)
 - IMP: 11240 annots(3902 genes)
 - IGI: 3575 annots(1956 genes)
 - IEP: 114 annots(94 genes)
 - ISS: 1657 annots(980 genes)
 - ISA: 168 annots(116 genes)
 - ISO: 8 annots(6 genes)
 - ISM: 1122 annots(1112 genes)
 - IGC: 0 annots(0 genes)
 - RCA: 0 annots(0 genes)
 - TAS: 430 annots(229 genes)
 - NAS: 94 annots(48 genes)
 - IC: 1209 annots(764 genes)
 - ND: 3592 annots(2095 genes)
 - IEA: 34803 annots(5490 genes)
 - NR: 0 annots(0 genes)

Right Panel:

- Annotated genes (1):**
 - Directly or Indirectly
 - Directly
- Output (2):**
 - All selected genes
 - Term-specific genes
 - Common term genes

FILTER tab

1. Filter GO Term Nodes by:

- **Genes** is a minimal number of genes mapped to a term
- **P-value** is a max term p-value
- **FDR:** is a max term [false discovery rate](#)

2. Significance test specifies distribution to use for null hypothesis:

- [Binomial](#): use a binomial distribution
 - [Hypergeometric](#): use a hypergeometric distribution
3. [Evidence codes in annotation](#) show how the annotation to a particular term is supported.

SELECT tab

1. *Annotated genes* outputs genes that are:
 - **Directly or Indirectly** annotated (direct and inherited annotations)
 - **Directly** annotated (inherited annotations won't be in the output)
2. *Output*:
 - **All selected genes**: outputs genes annotated to all selected GO terms
 - **Term-specific genes**: outputs genes that appear in only one of selected GO terms
 - **Common term genes**: outputs genes common to all selected GO terms
 - **Add GO Term as class**: adds GO terms as class attribute

Example

In the example below we have used **GEO Data Sets** widget, in which we have selected *Caffeine effects: time course and dose response* data set, and connected it to a **Differential Expression**. Differential analysis allows us to select genes with the highest statistical relevance (we used ANOVA scoring and agent label) and feed them to **GO Browser**. This widget lists four biological processes for our selected genes. Say we are interested in finding out more about *monosaccharide transmembrane transport* as this term has a high enrichment rate. To learn more about which genes are

annotated to this GO term, select it in the view and observe the results in a **Data Table**, where we see all the genes participating in this process listed. The other output of **GO Browser** widget is enrichment report, which we observe in the second **Data Table**.

The screenshot illustrates the workflow in Orange Data Mining. It starts with 'GEO Data Sets' leading to 'Expression Data', which is processed by 'Differential Expression' to produce 'Cluster Data'. This data is then used by 'GO Browser' to generate an 'Enrichment Report' and 'Data on Selected Genes'.

Data on Selected Genes

ANOVA	class	dose	time	agent	GAL2	HXK1	HXT11
Entrez ID					1589571356728	5587854037303	3054747706421
					850770	850614	854009
1	caffeine low ...	low	5 min	caffeine	0.174	-0.064	0.337
2	caffeine low ...	low	5 min	caffeine	0.281	0.590	0.134
3	calcofluor w...	low	5 min	calcofluor w...	0.141	-0.186	0.105
4	calcofluor w...	low	5 min	calcofluor w...	0.049	0.212	-0.115
5	congo red lo...	low	5 min	congo red	0.104	-0.607	0.233
6	congo red lo...	low	5 min	congo red	-0.033	-0.051	-0.027
7	caffeine low ...	low	15 min	caffeine	0.756	0.700	0.671
8	caffeine low ...	low	15 min	caffeine	0.576	0.854	0.472
9	calcofluor w...	low	15 min	calcofluor w...	0.113	-0.137	0.270
10	calcofluor w...	low	15 min	calcofluor w...	-0.262	-0.037	-0.213
11	congo red lo...	low	15 min	congo red	0.035	-0.259	0.110

GO Browser

GO term	Cluster	Reference	p-value	FDR	Genes	Enrichment
carbohydrate transmembrane transport	14 (2.38%)	27 (0.43%)	4.2e-07	9.3e-04	856496, 85...	
carbohydrate transport	16 (2.72%)	46 (0.73%)	1.0e-05	0.00561	856496, 85...	
carbohydrate transmembrane transport	14 (2.38%)	27 (0.43%)	4.2e-07	9.3e-04	856496, 85...	
monosaccharide transmembrane transport	13 (2.17%)	25 (0.40%)	1.0e-06	9.3e-04	856496, 85...	
hexose transmembrane transport	13 (2.21%)	25 (0.40%)	1.0e-06	9.3e-04	856496, 85...	
fructose transmembrane transport	10 (1.70%)	18 (0.29%)	1.0e-05	0.00561	856496, 85...	
mannose transmembrane transport	9 (1.53%)	16 (0.25%)	2.6e-05	0.01157	856496, 85...	
glucose transmembrane transport	6 (1.02%)	13 (0.21%)	0.00156	0.15603	856496, 851...	
small molecule metabolic process	106 (18.03%)	783 (12.43%)	6.0e-05	0.02311	850885, 851...	

Enrichment Report

GO Term Id	GO Term Name	Cluster Frequency	Genes in Cluster	Reference Frequency
O:0034219	carbohydrate transmembrane t...	0.024	14	0.004
O:0008645	hexose transmembrane transp...	0.022	13	0.004
O:0015749	monosaccharide transmembra...	0.022	13	0.004
O:0015755	fructose transmembrane trans...	0.017	10	0.003
O:0008643	carbohydrate transport	0.027	16	0.007
O:0015761	mannose transmembrane trans...	0.015	9	0.003
O:0044281	small molecule metabolic proc...	0.180	106	0.0124
O:1901605	alpha-amino acid metabolic pr...	0.058	34	0.028
O:0006555	methionine metabolic process	0.022	13	0.007
O:0098655	cation transmembrane transport	0.056	33	0.029
O:0000096	sulfur amino acid metabolic pro...	0.024	14	0.008
O:0006520	cellular amino acid metabolic p...	0.070	41	0.039