

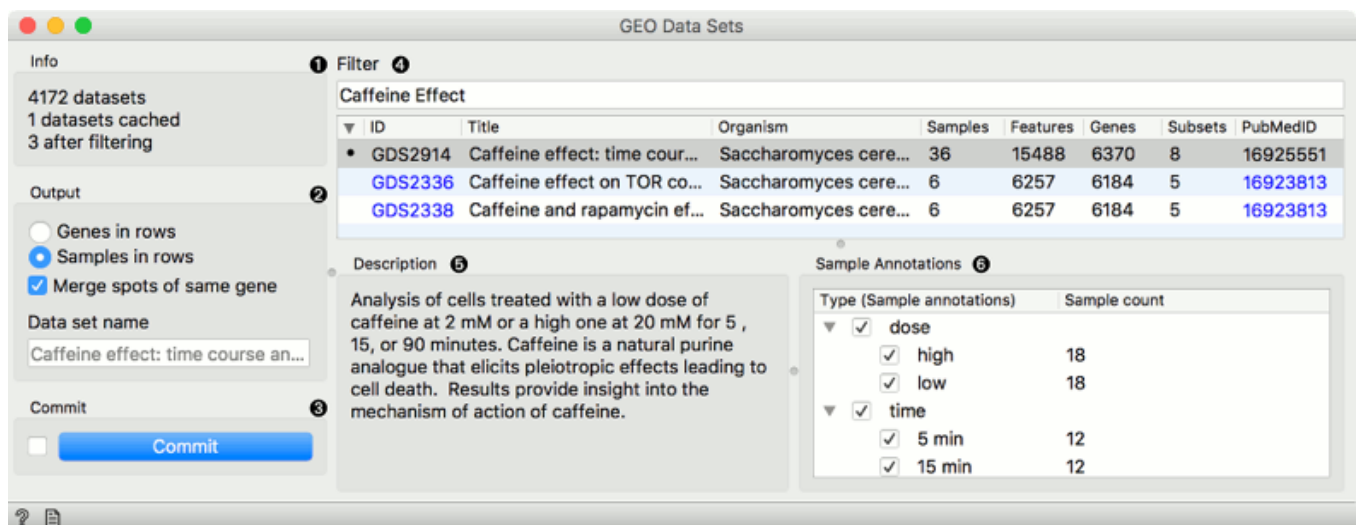
GEO Data Sets

Provides access to data sets from gene expression omnibus [GEO DataSets](#).

Inputs - None

Outputs - Expression data: Data set selected in the widget with genes or samples in rows.

[GEO DataSets](#) is a database of gene expression curated profiles maintained by [NCBI](#) and included in the [Gene Expression Omnibus](#). This Orange widget provides access to all its data sets and outputs a data set selected for further processing. For convenience, each downloaded data set is stored locally.



1. Information on the GEO data set collection. Cached data sets are the ones currently stored on the computer.
2. Output features. If *Samples in rows* is selected, genes (or spots) will be used as attributes. Alternatively samples will be used as attributes. *Merge spots of same gene* averages measures of the same gene. Finally, in the *Data set name* you can rename the output data. GEO title will be used as a default name.
3. If *Auto commit is on*, then the selected data set will be automatically

communicated to other widgets. Alternatively, click *Commit*.

4. *Filter* allows you to search for the data set. Below you see a list of GEO data sets with an ID number (link to the NCBI Data Set Browser), title, organism used in the experiment, number of samples, features, genes, subsets and a reference number for the PubMed journal (link to the article abstract).
5. Short description of the experiment from which the data set is sourced.
6. Select which *Sample Annotations* will be used in the output.

Example

GEO Data Sets is similar to the **File** widget, since it is used to load the data. In the example below we selected *Caffeine effect: time course and dose response* dataset from the GEO data base. Do not forget to press *Commit* to output the data. We can inspect the data in *Data Table*.

The screenshot displays the Orange Data Mining interface. The **GEO Data Sets** widget is shown with a filter dialog open. The filter dialog shows a list of datasets under the heading "Caffeine Effect". The selected dataset is "GDS2338 Caffeine and rapamycin ef...". The description of the dataset is: "Analysis of cells treated with a low dose of caffeine at 2 mM or a high one at 20 mM for 5, 15, or 90 minutes. Caffeine is a natural purine analogue that elicits pleiotropic effects leading to cell death. Results provide insight into the mechanism of action of caffeine." The sample annotations are listed as follows:

Type (Sample annotations)	Sample count
dose	
high	18
low	18
time	
5 min	12
15 min	12
90 min	12
agent	
caffeine	
calcofluor w...	
congo red	

The **Data Table** widget shows the resulting data with columns: Entrez ID, class, dose, time, agent, -- EMPTY, and AAC1 855078. The data is as follows:

Entrez ID	class	dose	time	agent	-- EMPTY	AAC1 855078
1	caffeine[low]	low	5 min	caffeine	0.208	-0.31
2	caffeine[low]	low	5 min	caffeine	0.243	0.03
3	calcofluor w...	low	5 min	calcofluor w...	-0.117	0.13
4	calcofluor w...	low	5 min	calcofluor w...	-0.054	0.20
5	congo red[lo...	low	5 min	congo red	-0.218	0.1
6	congo red[lo...	low	5 min	congo red	0.148	0.02
7	caffeine[low]	low	15 min	caffeine	-0.580	0.20
8	caffeine[low]	low	15 min	caffeine	0.301	0.06
9	calcofluor w...	low	15 min	calcofluor w...	-0.410	0.01
10	calcofluor w...	low	15 min	calcofluor w...	0.251	0.00
11	congo red[lo...	low	15 min	congo red	-0.177	-0.10
12	congo red[lo...	low	15 min	congo red	0.247	0.07
13	caffeine[low]	low	90 min	caffeine	-0.232	-0.37
...