

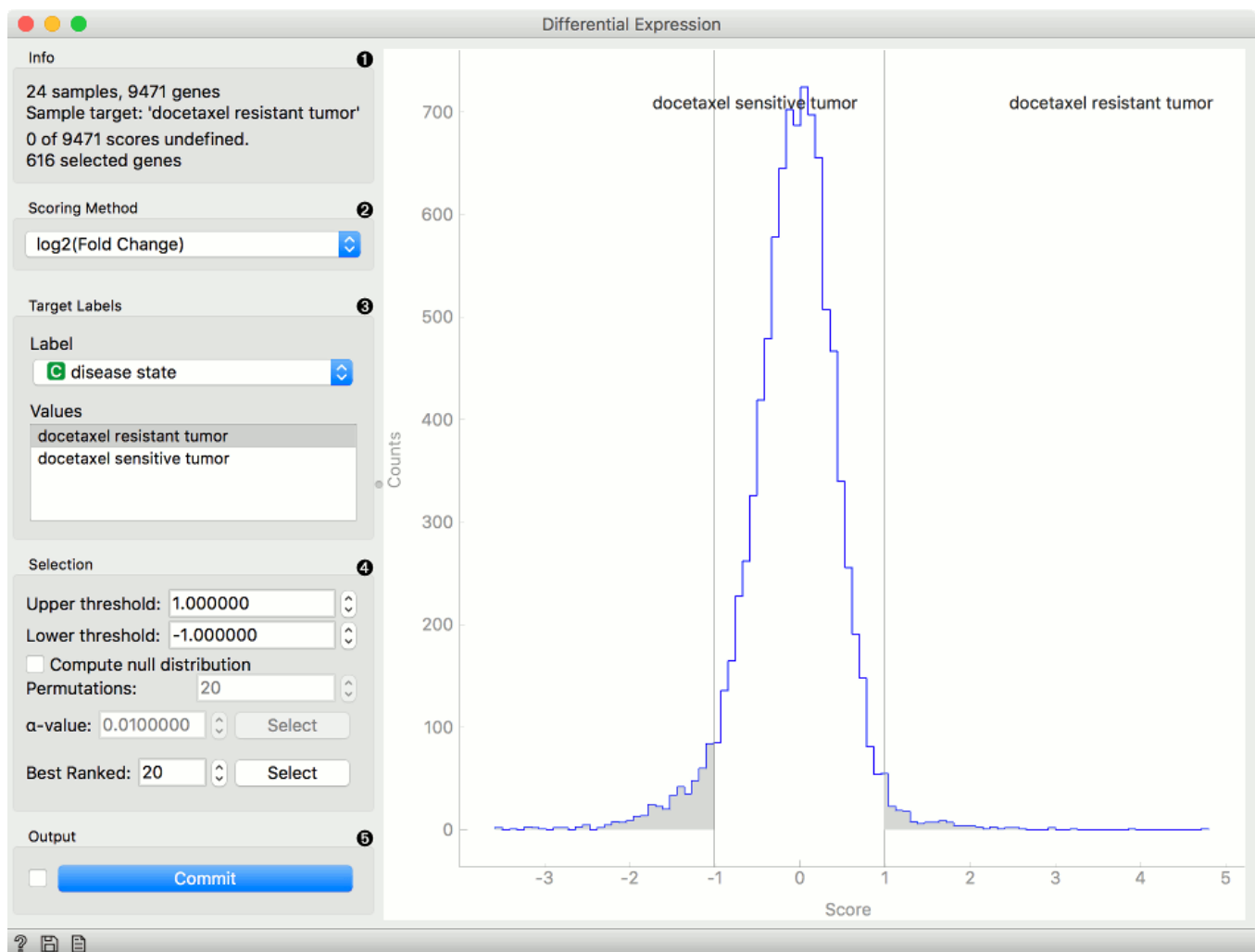
Differential Expression

Plots differential gene expression for selected experiments.

Inputs - Data: Data set.

Outputs - Data Subset: Differentially expressed genes. - Remaining Data Subset: Genes that were not differentially expressed. - Selected Genes: Genes from the select data with scores appended.

This widget plots a [differential gene expression](#) graph for a sample target. It takes gene expression data as an input (from **dictyExpress**, **GEO Data Sets**, etc.) and outputs a selected data subset (normally the most interesting genes).



1. Information of the data input and output. The first line shows the number of samples and genes in the data set. The second line displays the selected sample target (read around which the graph is plotted). The third line shows the number of undefined genes (missing data) and the fourth the number of genes in the output.
2. Select the plotting method in *Scoring method*:
 - [Fold change](#): final to initial value ratio
 - **log2 (fold change)**: binary logarithmic transformation of fold change values
 - [T-test](#): parametric test of null hypothesis
 - **T-test (P-value)**: parametric test of null hypothesis with [P-value](#) as criterium
 - [ANOVA](#): variance distribution
 - **ANOVA (P-value)**: variance distribution with P-value as criterium
 - [Signal to NoiseRatio](#): biological signal to noise ratio
 - [Mann-Whitney](#): non-parametric test of null hypothesis with P-value as criterium
 - [Hypergeometric test](#): for binary expression data.
3. Select *Target Labels*. Labels depend on the attributes in the input. In *Values* you can change the sample target (default value is the first value on the list, alphabetically or numerically).
4. *Selection* box controls the output data.
 - By setting the *Lower threshold* and *Upper threshold* values you are outputting the data outside this interval (the most interesting expression levels). You can also manually place the threshold lines by dragging left or right in the plot.
 - If you click *Compute null distribution* box, the widget will calculate null distribution and display it in the plot. *Permutations*

field allows you to set the precision of null distribution (the more permutations the more precise the distribution), while [alpha-value](#) will be the allowed probability of false positives. Press *Select* to output this data.

- The final option is to set the number of best ranked genes and output them with *Select*.

5. When *Auto commit is on* is ticked, the widget will automatically apply the changes. Alternatively press *Commit*. If the *Add gene scores to output* is ticked, the widget will append an additional column with gene scores to the data.

Example

From the GEO Data Sets widget, we selected *Breast cancer and docetaxel treatment* (GDS360) with 14 treatment resistant and 10 treatment sensitive tumors. Then we used the **Differential Expression** widget to select the most interesting genes. We left the upper and lower threshold at default (1 and -1) and output the data. Then we observed the selected data subset in a **Data Table**. The table shows selected genes with an additional gene score label.

The screenshot shows the Orange Data Mining interface. A workflow is visible at the top: GEO Data Sets → Expression Data → Data → Differential Expression → Data subset → Data → Data Table. The Differential Expression widget is the central focus, displaying a histogram of log2(Fold Change) scores. The histogram has a peak at 0, with shaded regions for scores less than -1 and greater than 1. The y-axis is labeled 'Counts' and ranges from 0 to 700. The x-axis is labeled 'Score' and ranges from -3 to 5. The widget's settings are as follows:

- Info: 24 samples, 9471 genes; Sample target: 'docetaxel resistant tumor'; 0 of 9471 scores undefined; 616 selected genes.
- Scoring Method: log2(Fold Change)
- Target Labels: Label: disease state; Values: docetaxel resistant tumor, docetaxel sensitive tumor.
- Selection: Upper threshold: 1.000000; Lower threshold: -1.000000; Compute null distribution: ; Permutations: 20; α -value: 0.0100000; Best Ranked: 20.
- Output: Commit

The Data Table widget shows the following data:

Entrez ID	disease state	31420_at	32021_at	32815_at
1	docetaxel re...	24.600	218.470	1456.180
2	docetaxel re...	15.783	64.094	485.903
3	docetaxel re...	20.674	96.231	212.544
4	docetaxel re...	33.528	203.803	851.863
5	docetaxel re...	19.285	18.150	88.688
6	docetaxel re...	23.315	30.530	455.855
7	docetaxel re...	32.803	18.340	245.478
8	docetaxel re...	21.213	18.296	37.780
9	docetaxel re...	13.874	26.760	35.388
10	docetaxel re...	19.860	41.213	216.432
11	docetaxel re...	16.489	17.549	49.898
12	docetaxel re...	39.145	8.952	101.579
13	docetaxel re...	38.401	5.898	97.033
14	docetaxel re...	20.296	10.399	121.908
15	docetaxel se...	5.898	34.057	88.894
16	docetaxel se...	14.613	27.602	50.225
17	docetaxel se...	12.013	35.015	138.814
18	docetaxel se...	14.685	5.898	86.063
19	docetaxel se...	9.627	31.793	66.641
20	docetaxel se...	24.067	14.232	46.463