

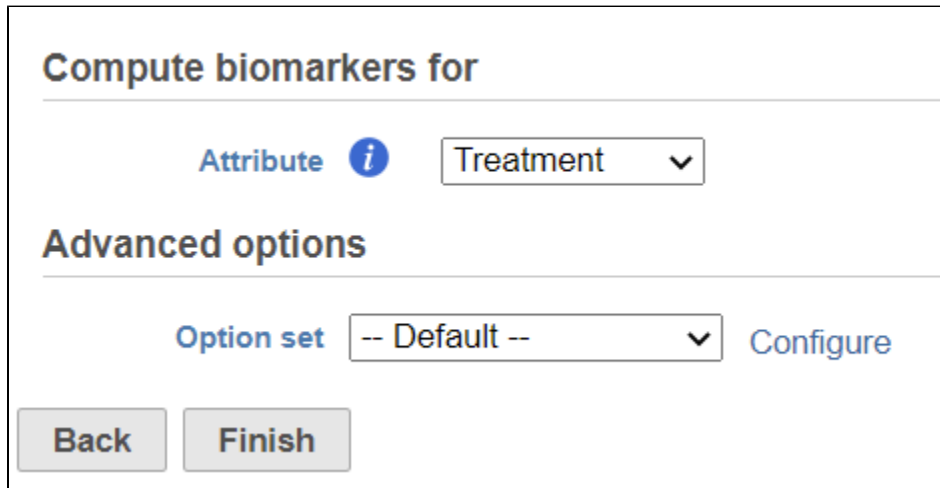
Compute biomarkers

This task can be invoked from count matrix data node or clustering task report (*Statistics > Compute biomarkers*). It performs Student's t-tests on the selected attribute, comparing one subgroup at a time vs all the others combined. By default, the up-regulated genes are reported as biomarkers.

- [Compute biomarker dialog](#)

Compute biomarker dialog

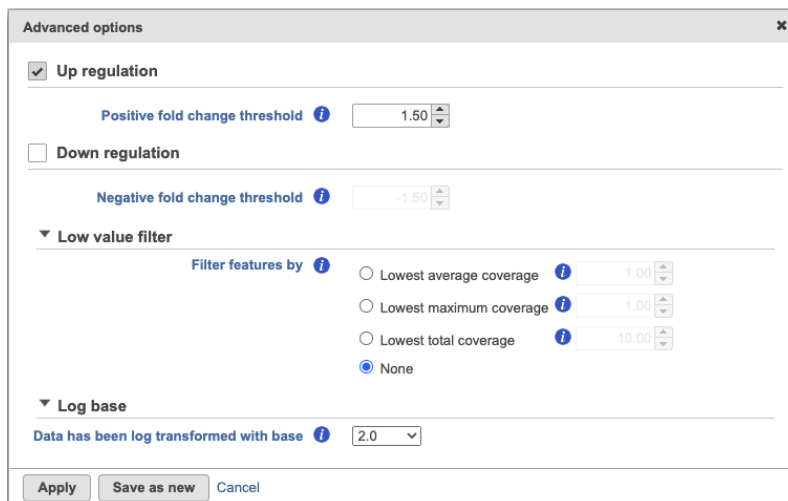
In the set-up dialog, select the attribute from the drop down list. The available attributes are categorical attributes which can be seen on the *Data* tab (i.e. project-level attributes) as well as and data node-specific annotation, e.g. graph-based clustering result (Figure 1). If the task is run on graph-based clustering output data node, the calculation is using upstream data node which contains feature counts – typically the input data node of PCA.



The dialog box is titled "Compute biomarkers for". It has a section for "Attribute" with a dropdown menu currently set to "Treatment". Below this is a section titled "Advanced options". Inside "Advanced options", there is an "Option set" dropdown menu set to "-- Default --" and a "Configure" button. At the bottom of the dialog are two buttons: "Back" and "Finish".

Figure 4. Compute biomarker dialog: selecting attribute

Click on the **Configure** under *Advanced options* to change the criteria for the output features (Figure 2).



The "Advanced options" dialog box contains several sections. The "Up regulation" section is checked, with a "Positive fold change threshold" of 1.50. The "Down regulation" section is unchecked. The "Low value filter" section is expanded, showing "Filter features by" with three radio button options: "Lowest average coverage" (1.00), "Lowest maximum coverage" (1.00), and "Lowest total coverage" (10.00). The "None" option is selected. The "Log base" section is expanded, showing "Data has been log transformed with base" set to 2.0. At the bottom are "Apply", "Save as new", and "Cancel" buttons.

Figure 5. Configure the biomarker filter criteria based on fold change

By default, the result outputs the top 10 features that are up-regulated by at least 1.5 fold change (in linear scale) for each subgroup comparing to the others. The result is displayed in a table with each column is a subgroup name, each row is a feature. Features are ranked by the ascending p-values within each sub-category. An example is shown in Figure 3. If a subgroup has fewer biomarkers than the others, the "extra" fields for that subgroup will be left blank.

	Cluster 8	Cluster 9	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14	Cluster 15	Cluster 16	Cluster 17
1)	SMTN (ENSG000000183963)	SFRP4 (ENSG000000106483)	KLK3 (ENSG000000142515)	SOD2 (ENSG000000112096)	IGLC1 (ENSG000000021675)	SELE (ENSG000000007908)	TF (ENSG000000091513)	PLAZG4D (ENSG000000159337)	MAP3K7CL (ENSG000000156265)	NGFR (ENSG000000064300)
2)	TAGLN (ENSG000000149591)	SNRNP70 (ENSG000000104852)	REXO2 (ENSG000000076043)	MMP7 (ENSG000000137673)	CRISPLD2 (ENSG000000103196)	ACKR1 (ENSG0000000213088)	KLIF5 (ENSG000000102554)	PSCA (ENSG000000167653)	IL32 (ENSG000000008517)	PLP1 (ENSG000000123560)
3)	ACTB (ENSG000000075624)	CLDN3 (ENSG000000165215)	PMEPA1 (ENSG000000124225)	CFTFR (ENSG000000001626)	IER3 (ENSG000000137331)	CRIP1 (ENSG0000000213145)	SLC12A2 (ENSG000000064651)	TGM4 (ENSG000000163810)	MYL12A (ENSG000000101608)	MPZ (ENSG000000158887)
4)	CST3 (ENSG000000101439)	SULF1 (ENSG000000137573)	GALNT7 (ENSG000000109586)	LTF (ENSG000000012223)	HHIP (ENSG000000164161)	C11orf96 (ENSG000000187479)	ELF3 (ENSG000000163435)	SERPINF11 (ENSG0000000206072)	SRPX (ENSG000000101955)	NRXN1 (ENSG000000179915)
5)	PDLIM7 (ENSG000000198923)	PIM3 (ENSG000000198355)	KCNN2 (ENSG000000080709)	UBD (ENSG0000000213886)	KLIF10 (ENSG000000155090)	EMP1 (ENSG000000134531)	DEFB1 (ENSG000000164825)	TRPM8 (ENSG000000144481)	PKA (ENSG000000171033)	S100B (ENSG000000160307)
6)	VCCL (ENSG000000035403)	IL1B (ENSG000000125538)	BCAM (ENSG000000187244)	TGM2 (ENSG000000198959)	BHLHE40 (ENSG000000134107)	HLA-E (ENSG0000000204592)	OLF4M (ENSG000000102837)	FAM3B (ENSG000000183844)	PLCD3 (ENSG000000161714)	SEMA3B (ENSG000000012171)
7)	FLNC (ENSG000000128591)	GOLPH3 (ENSG000000113384)	MAL2 (ENSG000000147676)	GABRP (ENSG0000000094755)	IER5 (ENSG000000162783)	SOC3 (ENSG000000184557)	SOX9 (ENSG000000125398)	DMXL1 (ENSG000000172869)	SMAD5 (ENSG000000113658)	ANGPTL7 (ENSG000000171819)
8)	CKB (ENSG000000166165)	F2R (ENSG000000181104)	ARHGDIIB (ENSG000000111348)	HLA-DRA (ENSG0000000204287)	RASL12 (ENSG000000103710)	LRRC32 (ENSG000000137507)	PITX1 (ENSG000000069011)	SEC14L2 (ENSG000000100003)	ACSS3 (ENSG000000111058)	APOD (ENSG000000189058)
9)	HSPB8 (ENSG000000152137)	WNK4 (ENSG000000126562)	GRIN3A (ENSG000000198785)	TMSB4X (ENSG0000000205542)	ADRA1A (ENSG000000120907)	VCAM1 (ENSG000000162692)	KRT23 (ENSG000000108244)	CRISPLD2 (ENSG000000103196)	ATP2A2 (ENSG000000174437)	CDH19 (ENSG0000000071991)
10)	COL6A2 (ENSG000000142173)	ITGA11 (ENSG000000137809)	FGFR1 (ENSG000000127418)	B2M (ENSG000000166710)	EPPK1 (ENSG0000000261150)	THBD (ENSG000000178726)	KRT13 (ENSG000000171401)	DBI (ENSG000000155368)	PLCB1 (ENSG000000182621)	ITGA6 (ENSG0000000091409)

Download

Figure 6. Biomarkers table (example). Top 10 biomarkers for each cluster are shown. Download link provides the full results table

Additional Assistance



१.

Your Rating: Results: 13 rates