Differentially Expressed Proteins and Genes

Next, we will filter out certain cells and re-split the data. Re-splitting the data can be useful if you want to perform differential analysis and downstream analysis separately for proteins and genes. For your own analyses, re-splitting the data is optional. You could just as well continue with differential analysis with the merged data if you prefer.

Filter Groups

Because we have classified our cells, we can now filter based on those classifications. This can be used to focus on a single cell type for re-clustering and sub-classification or to exclude cells that are not of interest for downstream analysis.

- Click the **Merged counts** data node
- Click **Filtering**
- Click **Filter cells**
- Set to **exclude Cell type is Doublets** using the drop-down menus
- Click **OR**
- Set the second filter to **exclude Cell type is N/A** using the drop-down menus
- Click **Finish** to apply the filter (Figure 1)

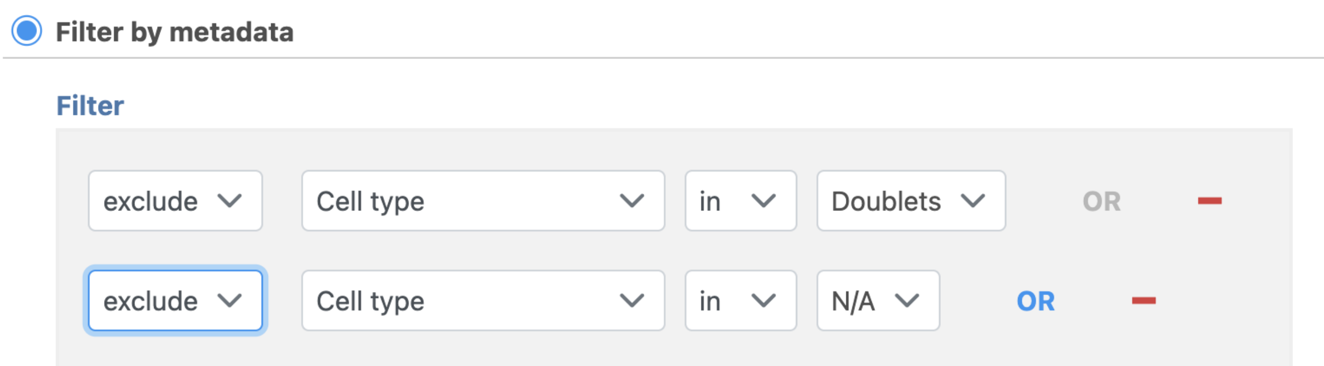


Figure 19. Set up the Filter groups task to exlcude Doublets and cells that are not classified

This produces a *Filtered counts* data node (Figure 2).

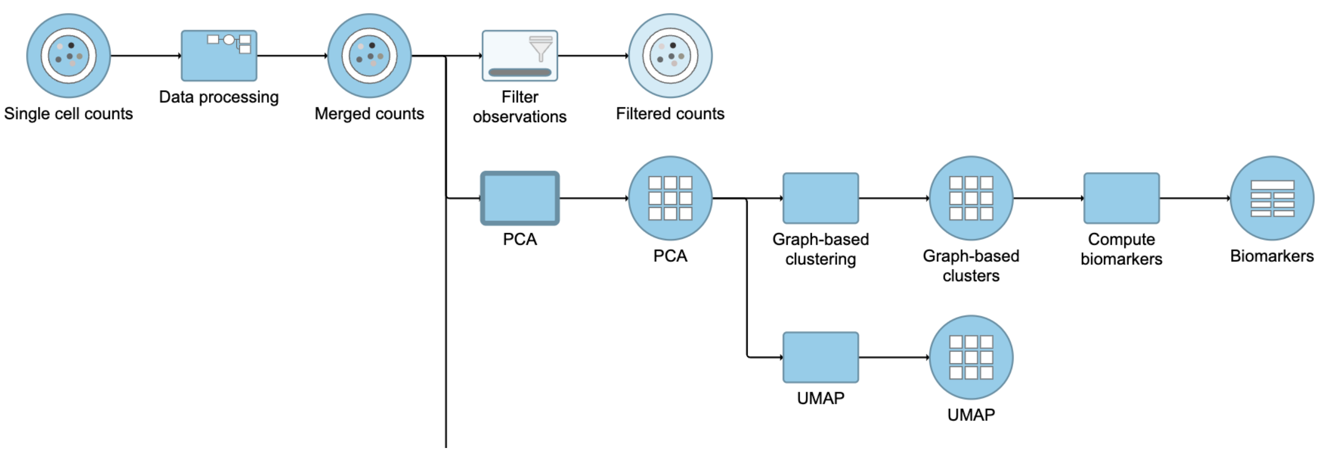


Figure 20. Filter groups output

Re-split the Matrix

- Click the **Filtered counts** data node

- Click **Pre-analysis tools**
- Click **Split by feature type**

This will produce two data nodes, one for each data type (Figure 3). The split data nodes will both retain cell classification information.

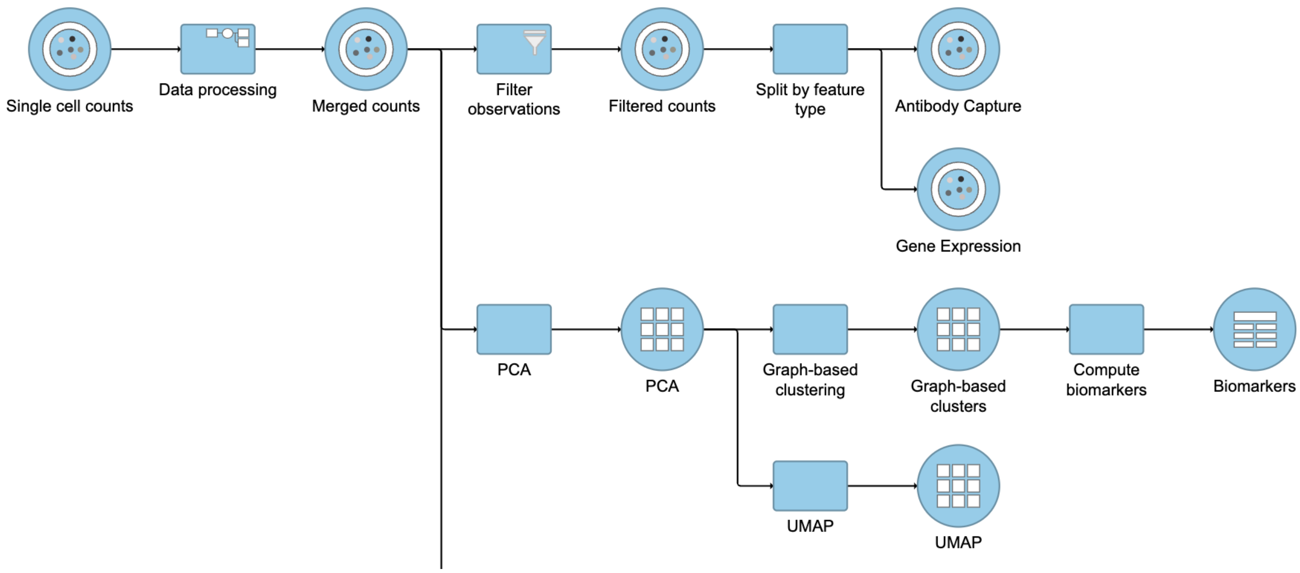


Figure 21. It is possible to re-split the merged matrix once again

Differential Analysis and Visualization - Protein Data

Once we have classified our cells, we can use this information to perform comparisons between cell types or between experimental groups for a cell type. In this project, we only have a single sample, so we will compare cell types.

- Click the **Antibody Capture** data node
- Click **Statistics**
- Click **Differential analysis**
- Click **ANOVA** then click **Next**

The first step is to choose which attributes we want to consider in the statistical test.

- Click **Cell type**
- Click **Add factor**
- Click **Next**

Next, we will set up the comparison we want to make. Here, we will compare the Activated and Mature B cells.

- Drag **Activated B cells** in the top panel
- Drag **Mature B cells** in the bottom panel
- Click **Add comparison**

The comparison should appear in the table as *Activated B cells vs. Mature B cells*.

- Click **Finish** to run the statistical test (Figure 4)

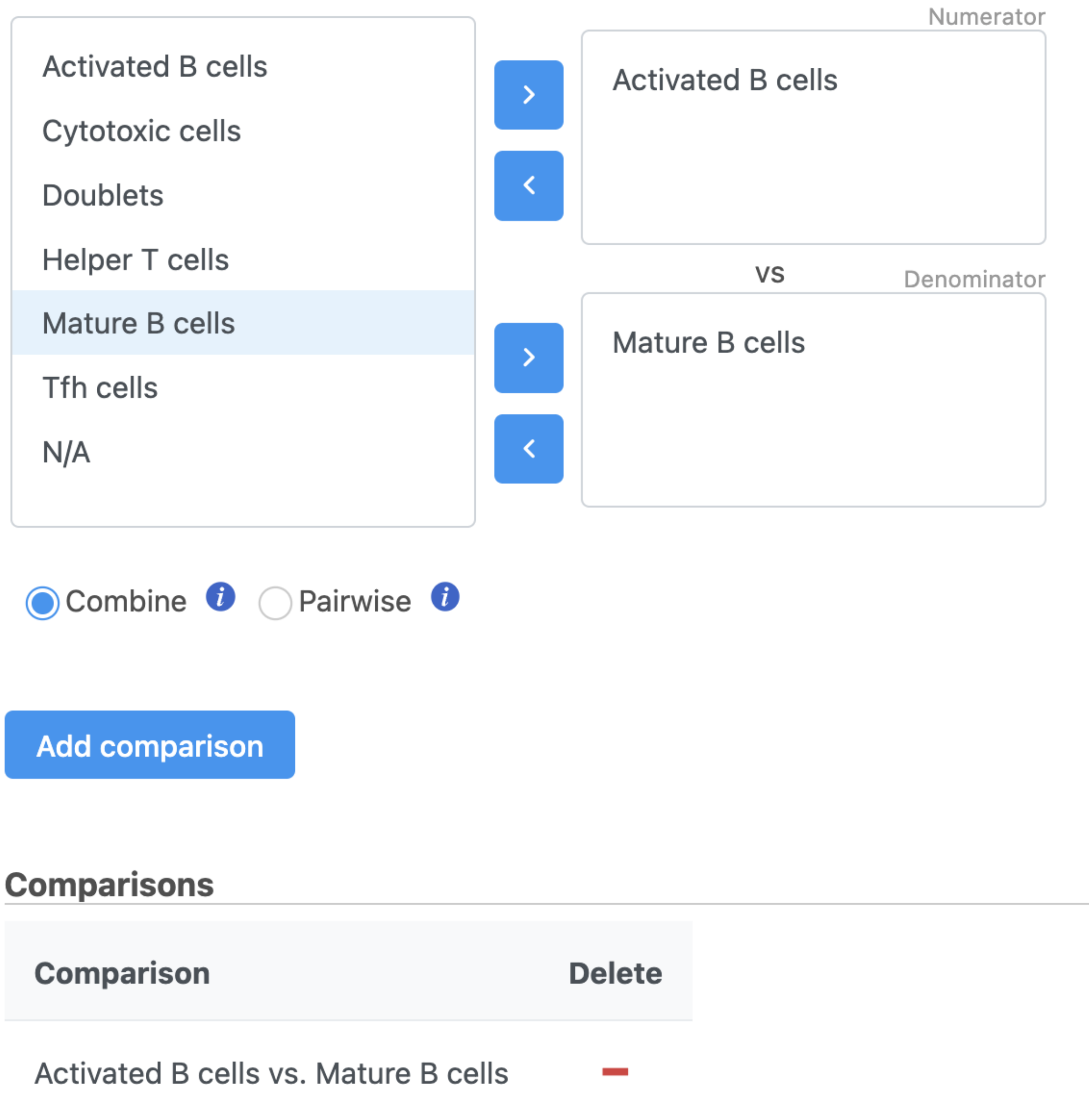
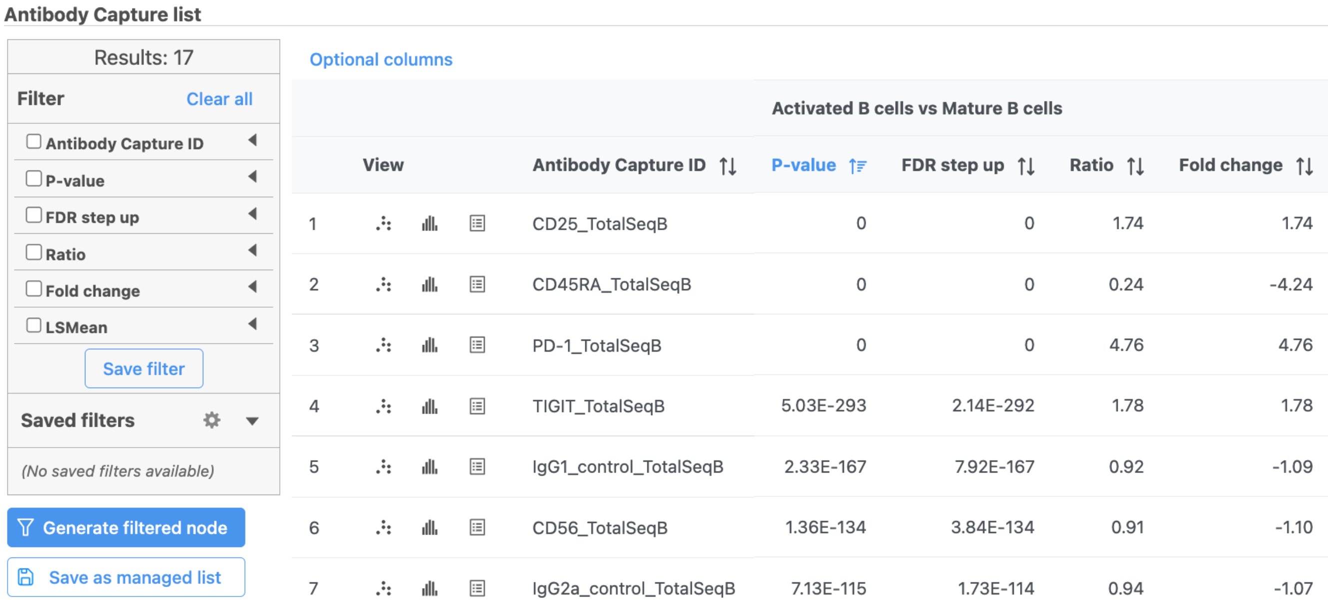


Figure 22. Setting up a comparison for differentially expressed proteins

The *ANOVA* task produces an *ANOVA* data node.

- Double-click the **ANOVA** data node to open the task report

The report lists each feature tested, giving p-value, false discovery rate adjusted p-value (FDR step up), and fold change values for each comparison (Figure 5).



Antibody Capture list

Results: 17

Filter — Clear all: Antibody Capture ID, P-value, FDR step up, Ratio, Fold change, LSMean. Save filter

Saved filters: (No saved filters available)

Generate filtered node

Save as managed list

Optional columns

	View	Antibody Capture ID	P-value	FDR step up	Ratio	Fold change
			Activated B cells vs Mature B cells			
1		CD25_TotalSeqB	0	0	1.74	1.74
2		CD45RA_TotalSeqB	0	0	0.24	-4.24
3		PD-1_TotalSeqB	0	0	4.76	4.76
4		TIGIT_TotalSeqB	5.03E-293	2.14E-292	1.78	1.78
5		IgG1_control_TotalSeqB	2.33E-167	7.92E-167	0.92	-1.09
6		CD56_TotalSeqB	1.36E-134	3.84E-134	0.91	-1.10
7		IgG2a_control_TotalSeqB	7.13E-115	1.73E-114	0.94	-1.07

Figure 23. GSA report for protein expression data

In addition to the listed information, we can access dot and violin plots for each gene or protein from this table.

- Click in the *CD45RA_TotalSeqB* row

This opens a dot plot in a new data viewer session, showing CD45A expression for cells in each of the classifications (Figure 6). First, we exclude *Doublets* and *N/A* cells from the plot:

- Open **Select and filter**, select **Criteria**
- Drag "Cell type" from the legend title to the **Add criteria** box
- Uncheck **Doublets** and **N/A**
- Click to include selected points

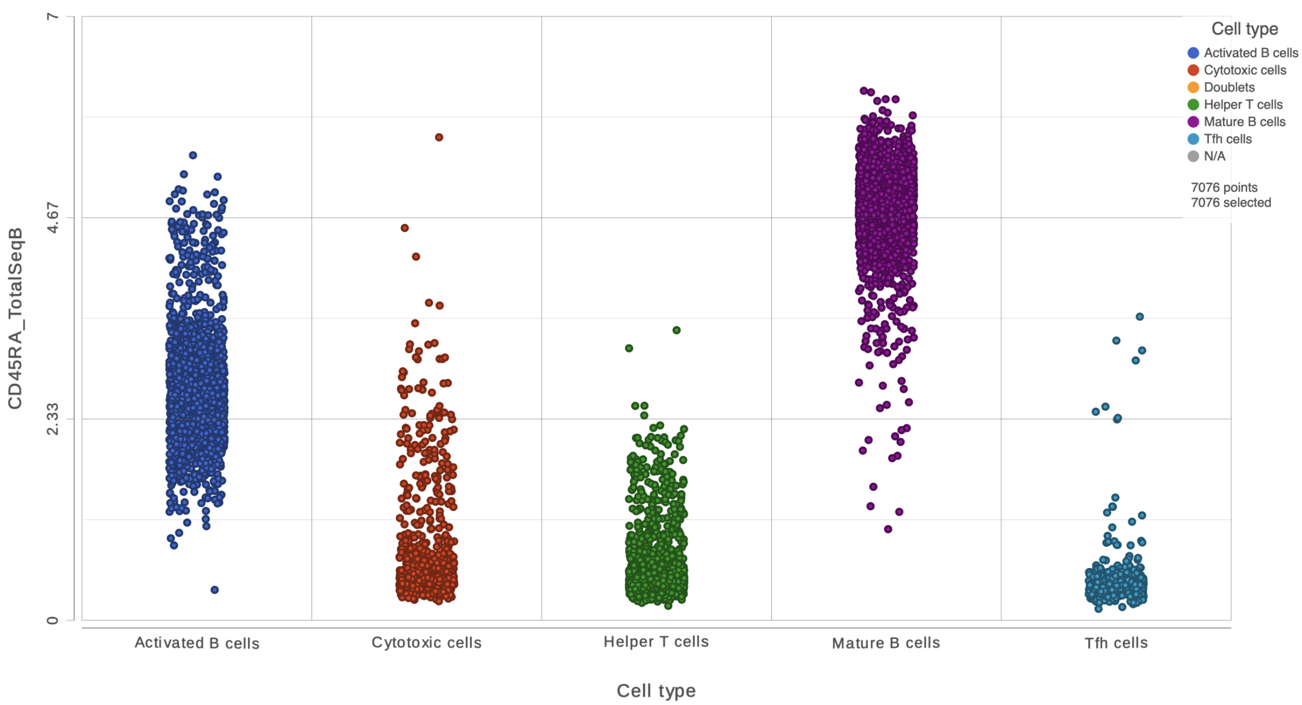


Figure 24. CD45RA dot plot for all cells

We can use the *Configuration* panel on the left to edit this plot.

- Open the **Style** icon
- Switch on **Violins** under *Summary*
- Switch on **Overlay** under *Summary*
- Switch on **Colored** under *Summary*
- Select the *Graph-based clustering* node in the **Color by** section
- **Color by** Graph-based clusters under **Color** and use the slider to decrease the **Opacity**
- Open the **Axes** icon
- Select the *Graph-based clustering* node in the **X axis** section
- Change the *X axis data* to Graph-based clusters
- Use the slider to increase the **Jitter** on the *X axis* (Figure 7)

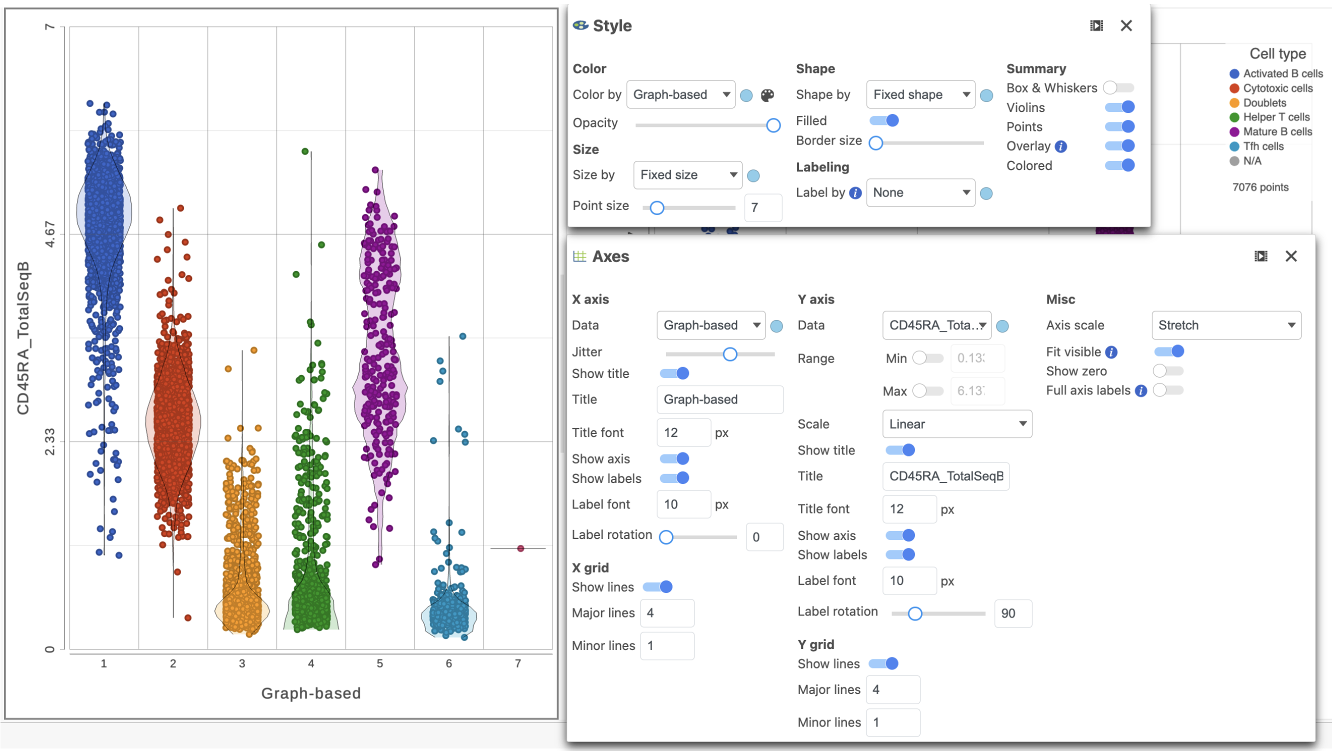


Figure 25. Configure the dot plot using the tools on the left

- Click the **project name** to return to the *Analyses* tab

To visualize all of the proteins at the same time, we can make a hierarchical clustering heat map.

- Click the **ANOVA** data node
- Click **Exploratory analysis** in the toolbox
- Click **Hierarchical clustering/heatmap**
- In the *Cell order* section, choose **Graph-based clusters** from the *Assign order* drop-down list
- Click **Finish** to run with the other default settings
- Double-click the **Hierarchical clustering** task node to open the heatmap

The heatmap can easily be customized using the tools on the left.

- Open the **Axes** icon
- Switch off *Show* **Row labels**
- Increase the **Font** to 16 (Figure 8)

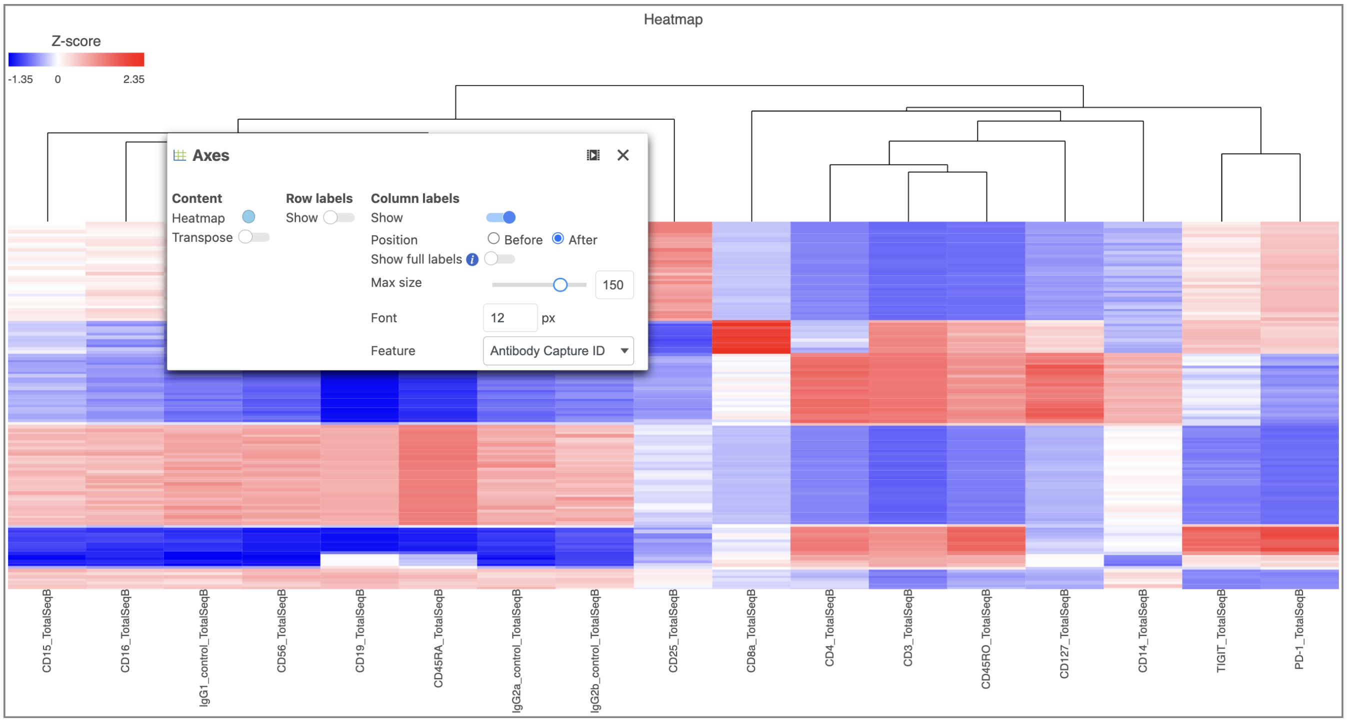


Figure 26. Heatmap showing altered Axes labels

- Activate the **Transpose** switch which will switch the Row and Column labels, so now the Row labels will be shown (Figure 9)

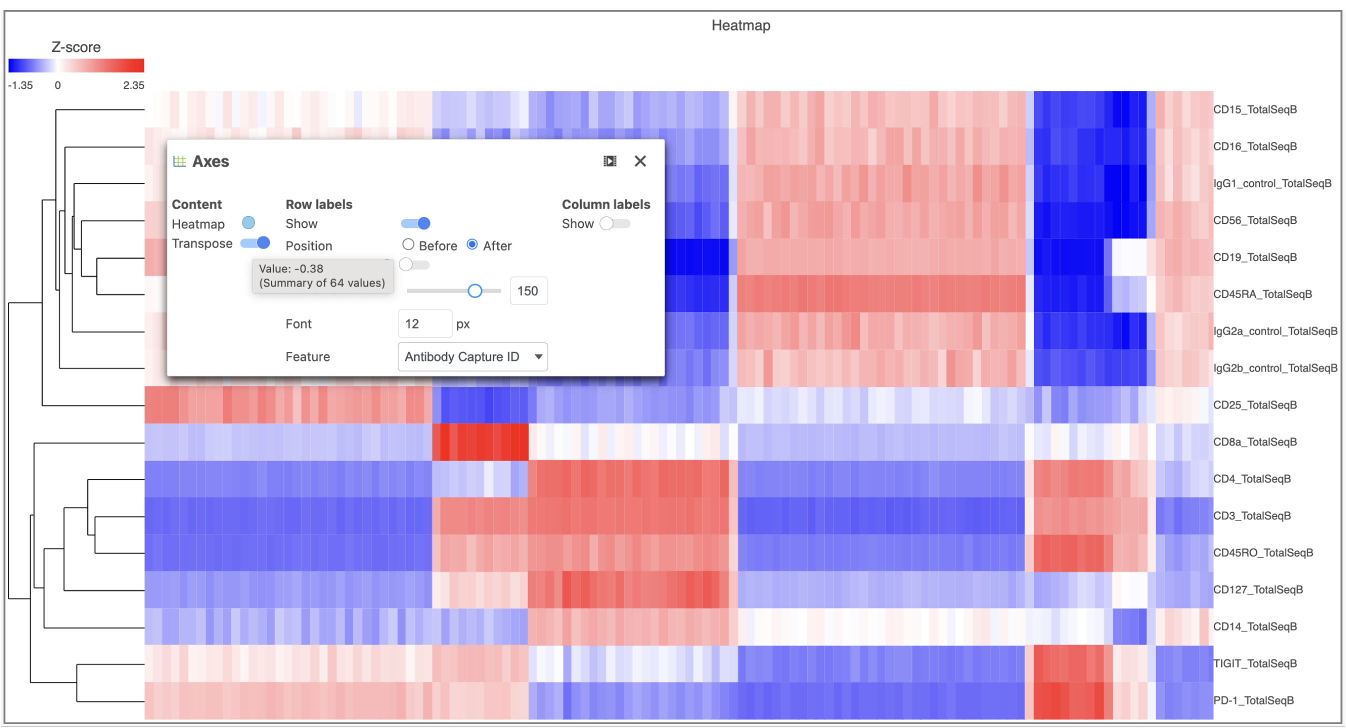


Figure 27. Transpose the Heatmap to switch the columns and rows

- Open the **Dendrograms** icon
- Choose *Row color* **By cluster** and change *Row clusters* to **4**
- Change *Row dendrogram size* to **80** (Figure 10)

Figure 28. Configure the Dendrograms settings

- In the **Heatmap** icon
- Navigate to *Range* under *Color*
- Set the Min and Max to **-1.2** and **1.2**, respectively
- Change the *Shape* to **Circle** (Figure 11)

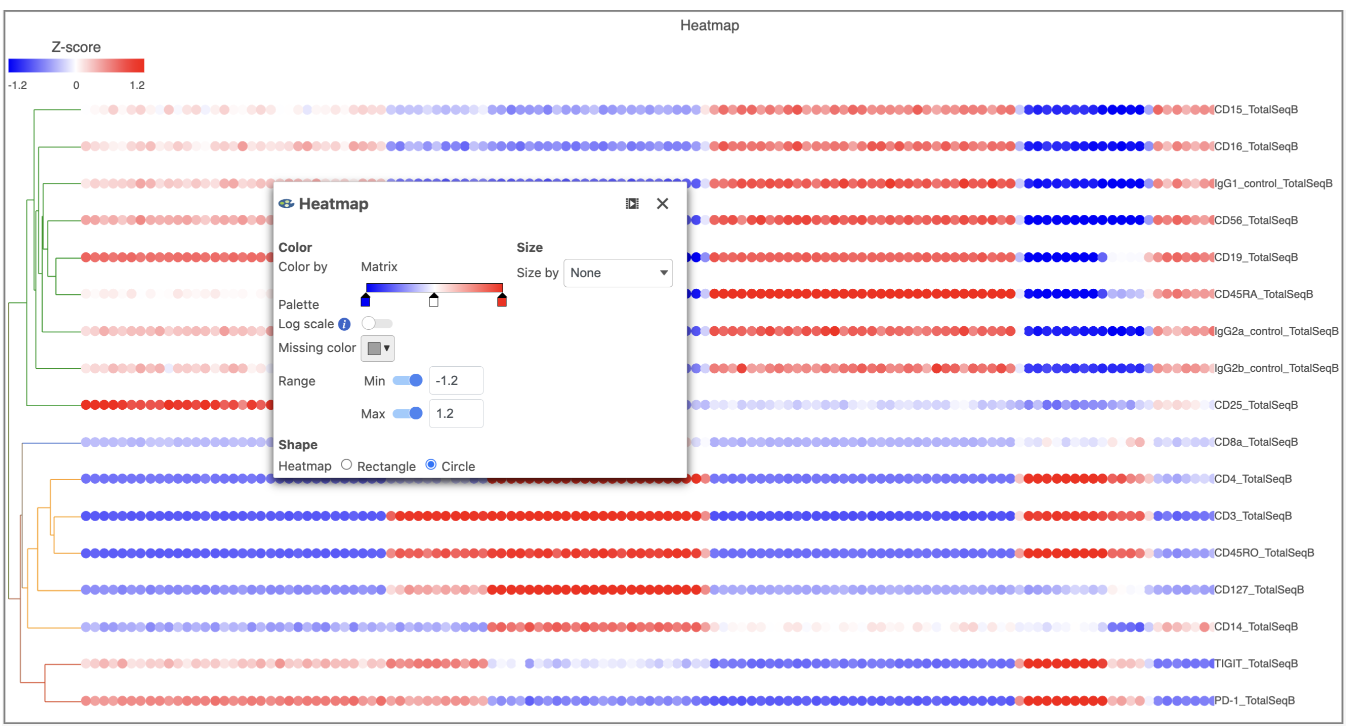


Figure 29. Configure the Heatmap icon

- Switch the *Shape* back to **Rectangle**

- Change the *Color Palette* by clicking on the color squares and selecting colors from the rainbow. Click outside of the selection box to exit this selection. The color options can be dragged alone the Palette to highlight value differences (Figure 12).

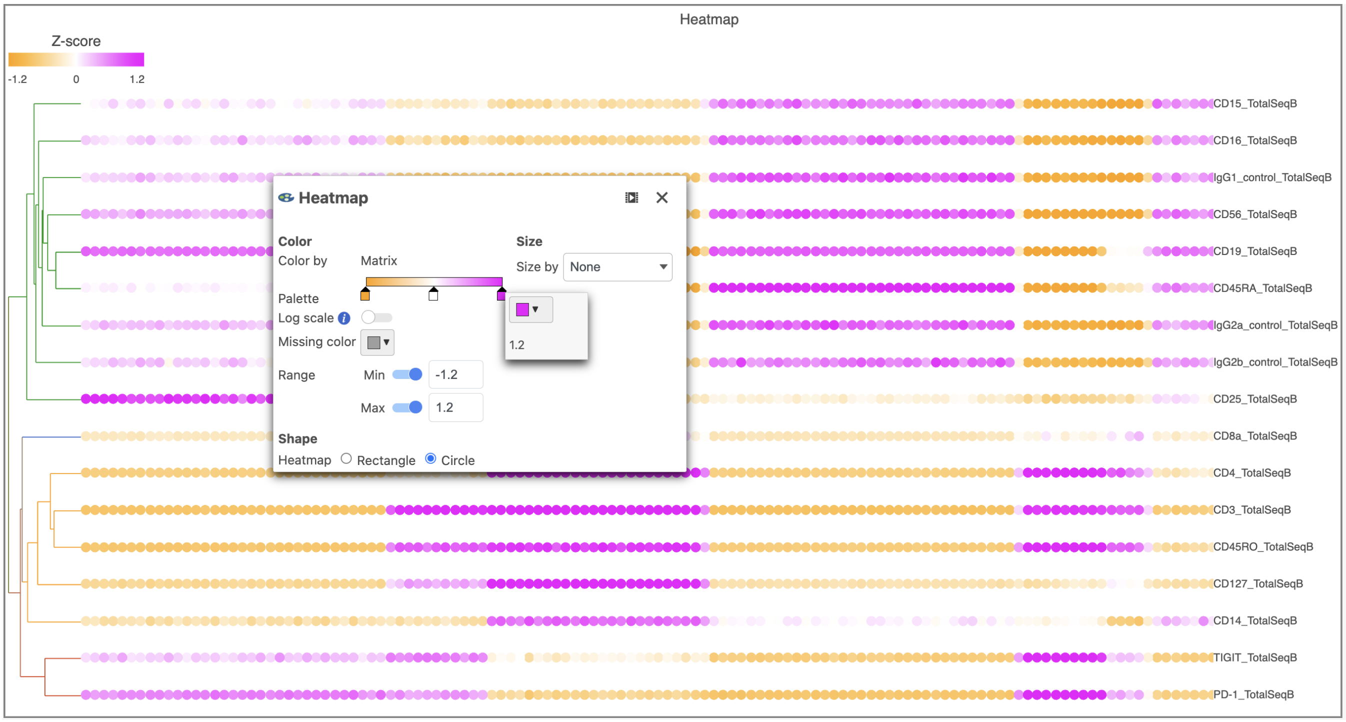


Figure 30. Heatmap showing expression of protein markers after changing the Heatmap settings further

Feel free to explore the other tool options on the left to customize the plot further.

Differential Analysis, Visualization, and Pathway analysis - Gene Expression Data

We can use a similar approach to analyze the gene expression data.

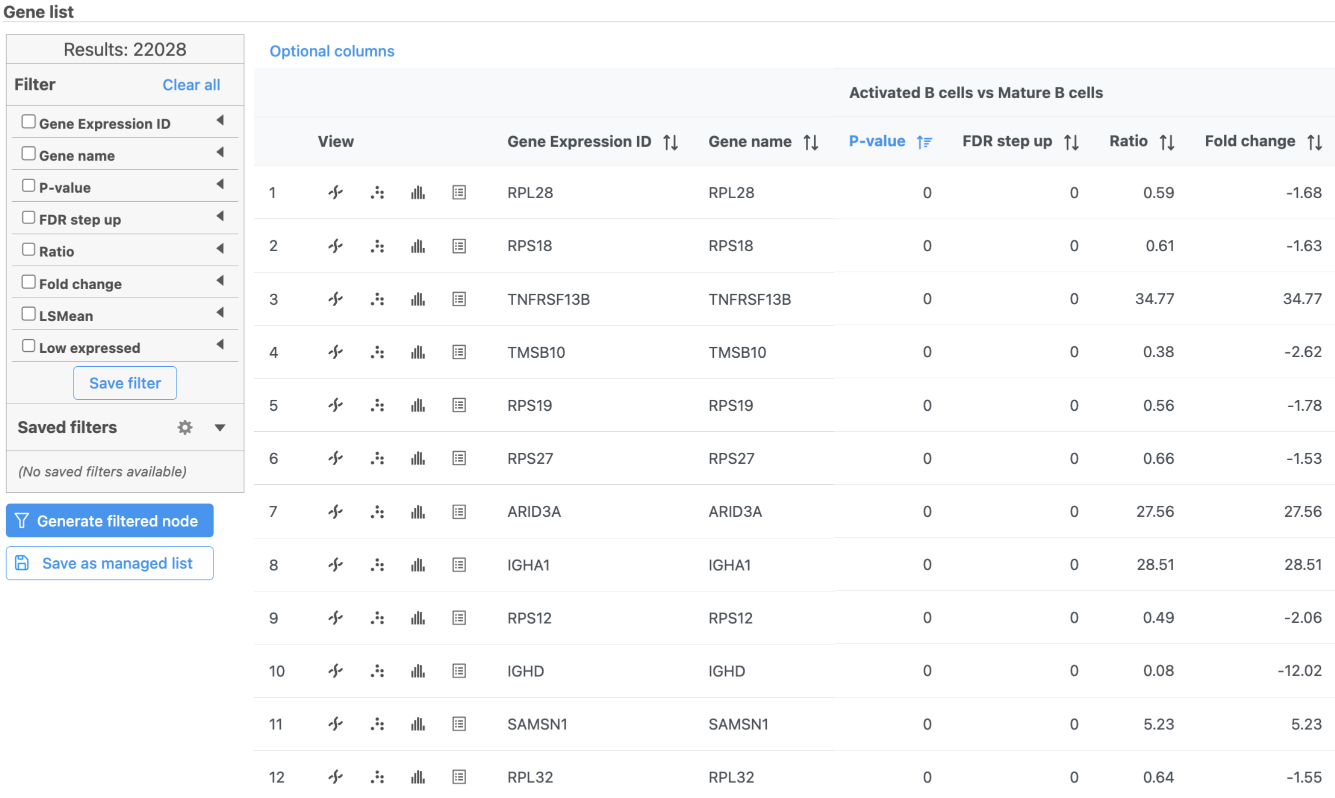
- Click the **project name** to return to the *Analyses* tab
- Click the **Gene Expression** data node
- Click the **Antibody Capture** data node
- Click **Statistics**
- Click **Differential analysis**
- Click **ANOVA** then click **Next**
- Click **Cell type**
- Click **Add factor**
- Click **Next**
- Drag **Activated B cells** in the top panel
- Drag **Mature B cells** in the bottom panel
- Click **Add comparison**

The comparison should appear in the table as *Activated B cells vs. Mature B cells*.

- Click **Finish** to run the statistical test

As before, this will generate an *ANOVA* task node and n *ANOVA* data node.

- Double-click the **ANOVA** task node to open the task report (Figure 13)



	View	Gene Expression ID	Gene name	P-value	FDR step up	Ratio	Fold change
				Activated B cells vs Mature B cells			
1		RPL28	RPL28	0	0	0.59	-1.68
2		RPS18	RPS18	0	0	0.61	-1.63
3		TNFRSF13B	TNFRSF13B	0	0	34.77	34.77
4		TMSB10	TMSB10	0	0	0.38	-2.62
5		RPS19	RPS19	0	0	0.56	-1.78
6		RPS27	RPS27	0	0	0.66	-1.53
7		ARID3A	ARID3A	0	0	27.56	27.56
8		IGHA1	IGHA1	0	0	28.51	28.51
9		RPS12	RPS12	0	0	0.49	-2.06
10		IGHD	IGHD	0	0	0.08	-12.02
11		SAMSN1	SAMSN1	0	0	5.23	5.23
12		RPL32	RPL32	0	0	0.64	-1.55

Figure 31. GSA report for the gene expression data

Because more than 20,000 genes have been analyzed, it is useful to use a volcano plot to get an idea about the overall changes.

- Click in the top right corner of the table to open a volcano plot

The Volcano plot opens in a new data viewer session, in a new tab in the web browser. It shows each gene as a point with cutoff lines set for P-value (y-axis) and fold-change (x-axis). By default, the P-value cutoff is set to 0.05 and the fold-change cutoff is set at |2| (Figure 14).

The plot can be configured using various tools on the left. For example, the **Style** icon can be used to change the appearance of the points. The X and Y-axes can be changed in the **Axes** icon. The **Statistics** icon can be used to set different Fold-change and P-value thresholds for coloring up/down-regulated genes. The in plot controls can be used to transpose the volcano plot (Figure 14).

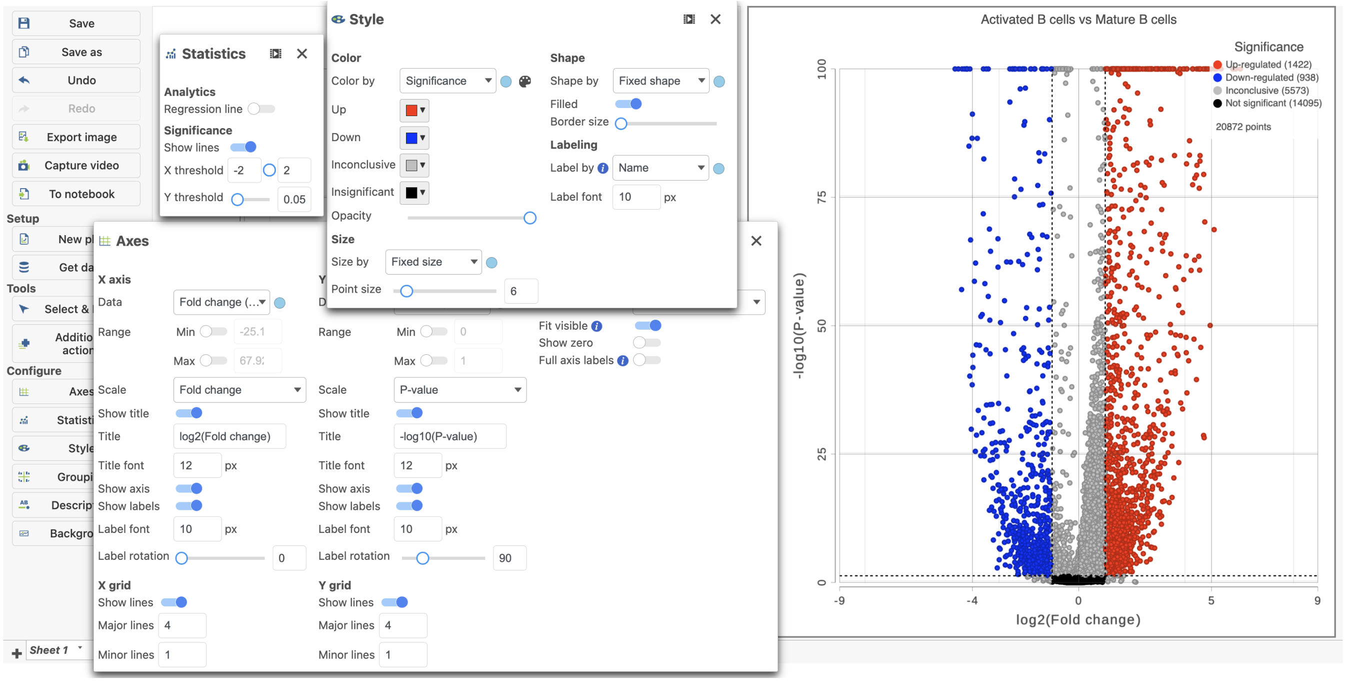


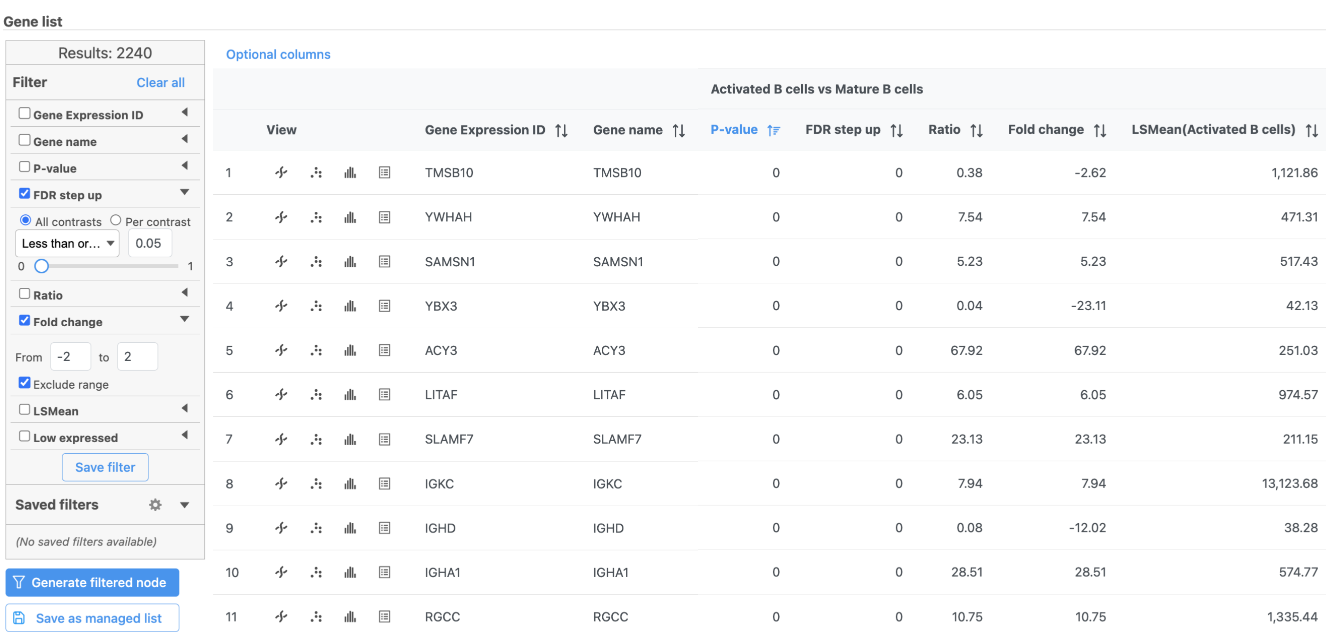
Figure 32. The volcano plot can be Configured using the icons on the left and in plot controls

- Click the **ANOVA report** tab in your web browser to return to the full report

We can filter the full set of genes to include only the significantly different genes using the filter panel on the left.

- Click **FDR step up**
- Type **0.05** for the cutoff and press **Enter** on your keyboard
- Click **Fold change**
- Set to From **-2** to **2** and press **Enter** on your keyboard

The number at the top of the filter will update to show the number of included genes (Figure 15).



	Gene Expression ID	Gene name	P-value	FDR step up	Ratio	Fold change	LSMean(Activated B cells)
1	TMSB10	TMSB10	0	0	0.38	-2.62	1,121.86
2	YWHAH	YWHAH	0	0	7.54	7.54	471.31
3	SAMSN1	SAMSN1	0	0	5.23	5.23	517.43
4	YBX3	YBX3	0	0	0.04	-23.11	42.13
5	ACY3	ACY3	0	0	67.92	67.92	251.03
6	LITAF	LITAF	0	0	6.05	6.05	974.57
7	SLAMF7	SLAMF7	0	0	23.13	23.13	211.15
8	IGKC	IGKC	0	0	7.94	7.94	13,123.68
9	IGHD	IGHD	0	0	0.08	-12.02	38.28
10	IGHA1	IGHA1	0	0	28.51	28.51	574.77
11	RGCC	RGCC	0	0	10.75	10.75	1,335.44

Figure 33. Use the panel on the left to filter the list for significant genes

- Click **Generate filtered node** to create a new data node including only these significantly different genes

A task, *Differential analysis filter*, will run and generate a new *Filtered Feature list* data node. We can get a better idea about the biology underlying these gene expression changes using gene set or pathway enrichment. Note, you need to have the Pathway toolkit enabled to perform the next steps.

- Click the **Filtered feature list** data node
- Click **Biological interpretation** in the toolbox
- Click **Pathway enrichment**
- Make sure that **Homo sapiens** is selected in the *Species* drop-down menu
- Click **Finish** to run
- Double-click the **Pathway enrichment** task node to open the task report

The pathway enrichment results list KEGG pathways, giving an enrichment score and p-value for each (Figure 16).

Gene set	Description	Enrichment score	P-value	FDR step up	Rich factor	Genes in set	Genes in list	Genes not in list	Genes in list, not in set	Genes not in list, not in set
path:hsa05202	Transcriptional misregulation in cancer	11.96	6.37E-6	1.93E-3	0.26	167	44	123	870	5,708
path:hsa05200	Pathways in cancer	11.33	1.2E-5	1.93E-3	0.20	482	98	384	816	5,447
path:hsa04068	FoxO signaling pathway	11.00	1.67E-5	1.93E-3	0.28	121	34	87	880	5,744
path:hsa00100	Steroid biosynthesis	8.38	2.28E-4	0.02	0.50	18	9	9	905	5,822
path:hsa04630	JAK-STAT signaling pathway	7.93	3.58E-4	0.02	0.25	134	33	101	881	5,730
path:hsa04210	Apoptosis	7.87	3.81E-4	0.02	0.25	129	32	97	882	5,734
path:hsa05225	Hepatocellular carcinoma	7.80	4.11E-4	0.02	0.24	157	37	120	877	5,711
path:hsa05222	Small cell lung cancer	7.68	4.64E-4	0.02	0.27	88	24	64	890	5,767

Figure 34. Results of pathway enrichment test

To get a better idea about the changes in each enriched pathway, we can view an interactive KEGG pathway map.

- Click **path:hsa05202** in the Transcriptional misregulation in cancer row

The KEGG pathway map shows up-regulated genes from the input list in red and down-regulated genes from the input list in green (Figure 17).

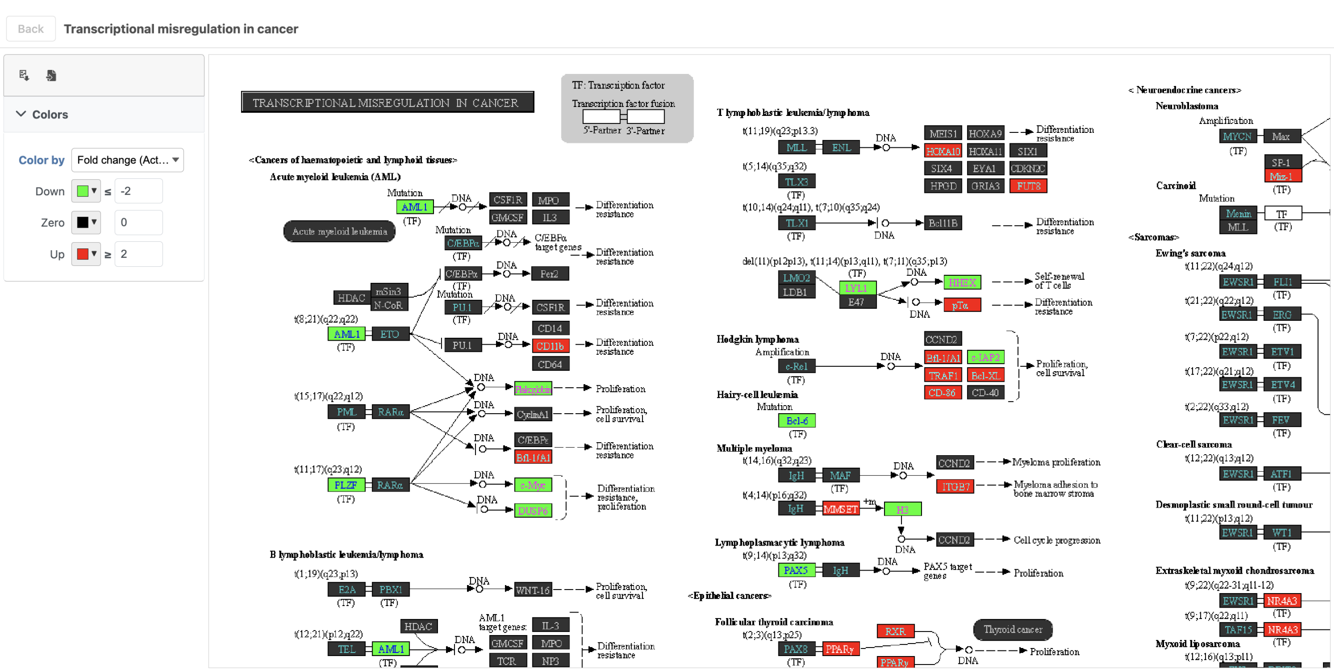


Figure 35. Transcriptional misregulation in cancer pathway with significant genes highlighted in green and red

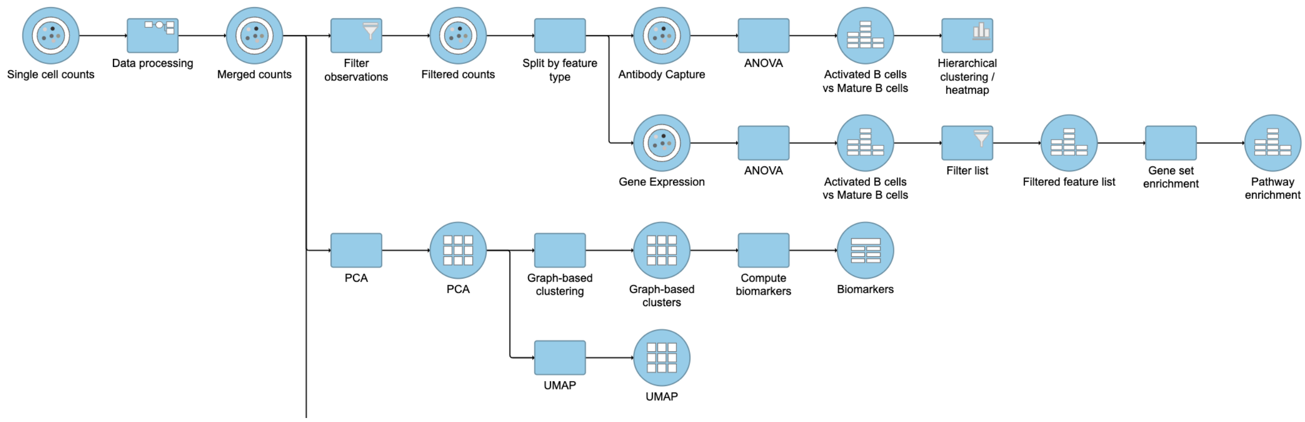


Figure 36. Final CITE-Seq pipeline

Additional Assistance

If you need additional assistance, please visit our support page to submit a help ticket or find phone numbers for regional support.