# **Differentially Expressed Proteins and Genes**

- Filter Groups
- Re-split the Matrix
- Differential Analysis and Visualization Protein Data
- Differential Analysis, Visualization, and Pathway analysis Gene Expression Data

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)

### Filter by metadata

Filter						
exclude 🗸	Cell type	~	in 🗸	Doublets 🗸	OR	-
exclude 🗸	Cell type	~	in 🗸	N/A 🗸	or –	

#### 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:	Optio	onal co	umns											
Filter	Clear all					Activated B cells vs Mature B cells								
C Antibody Captur	e ID 🖣		View			Antibody Capture ID 1	P-value ≜=	EDB stop up 1	Patio 1	Fold change 1				
P-value	•		view				P-value  F	Low step ab 11	Katio  ↓	Fold change				
□ FDR step up	•	1	.::	ıllı	∷	CD25_TotalSeqB	0	0	1.74	1.74				
Ratio	•													
Fold change	•	2		մե	∷≡	CD45RA_TotalSeqB	0	0	0.24	-4.24				
LSMean	•	3	.::	all.		PD-1_TotalSeqB	0	0	4.76	4.76				
Save filte	er													
Saved filters	☆ ▼	4	.::	ıllı	II	TIGIT_TotalSeqB	5.03E-293	2.14E-292	1.78	1.78				
(No saved filters avai	lable)	5	•••	մե	∷	lgG1_control_TotalSeqB	2.33E-167	7.92E-167	0.92	-1.09				
Generate filtere	d node	6	.••	ıtlı	≣	CD56_TotalSeqB	1.36E-134	3.84E-134	0.91	-1.10				
🕒 Save as manag	ed list	7	.::	մե	II	lgG2a_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



Cell type

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)

					1	Heatmap
Z-s	score					
-1.2	0 1.2					
						D0 00000000000000000000000000000000000
				0000		De Concernation CD16_TotalSeqB
						Control_TotalSeqB
L LL		👁 Heatmap			×	CDER TrialSoop
		Oslar	Cine			CD30_IntalSetb
		Color by Matrix	Size by None		_	CD19_TotalSeqB
]"L			Size by None		•	•••••••••••••••••••••••••••••••••
		Palette	•			
	_00000000000000000000000000000000000000	Missing color				Control_TotalSeqB
						Control_TotalSeqB
		Range Min -1.2				
		Max 🦲 1.2				CD25_IOIalSeqB
		Shape				CD8a_TotalSeqB
		Heatmap 🔿 Rectangle 🔍 Circle				CD4 TotalSeaB
		************************		0000	.000	<b>JO</b> COCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOC
니거그						Be Constant
41						CD127_lotaiSeqB
						Debbase of the second
						TIGIT TotalSegB
		•••••••••••••••••••••••		0000	.000	COCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOC

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).

		Heatmap
Z-score		
-1.2 0 1.2		
		CD15_TotalSeqB
		CD16_TotalSeqB
	e Haatman	Control_TotalSeqB
		CD56_TotalSeqB
	Color Size Color by Matrix Size by None	CD19_TotalSeqB
	Palette	CD45RA_TotalSeqB
	Log scale ()	Constant and the second s
	Range Min -1.2	IgG2b_control_TotalSeq8
	Max 💶 1.2	CD25_TotalSeqB
	Shape	CD8a_TotalSeqB
	Heatmap O Rectangle  © Circle	CD4_TotalSeqB
		CD3_TotalSeqB
	******	CD45RO_TotalSeqB
		CD127_TotalSeqB
		CD14_TotalSeqB
		TIGIT_TotalSeqB
		PD-1_TotalSeqB

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)

#### Gene list

Results:	Optio	onal co	lumns												
Filter	Clear all						Activated B cells vs Mature B cells								
Gene Express	ion ID 🖣	View					Gene Expression ID 1	Gene name 1	P-value ↑ <del></del>	FDR step up 1	Ratio ↑	Fold change ↑			
Gene name	•						· · · · · · · · · · · · · · · · · · ·	1.			1.	0 14			
P-value	•	1	-5-	÷.	ıllı	II	RPL28	RPL28	0	0	0.59	-1.68			
□ FDR step up	•	0	,		.0		55040	55040	0	0	0.01	1.00			
Ratio	•	2	-}-	.*:	ıllı	E	RPS18	RPS18	0	0	0.61	-1.63			
Fold change	•	3	-5-	.::	dh	:=	TNFRSF13B	TNFRSF13B	0	0	34.77	34.77			
LSMean															
Low expresse	ed •	4	-5-	÷.	ıllı	:=	TMSB10	TMSB10	0	0	0.38	-2.62			
Save filter		5	-5-	.:.	ıll.	:=	RPS19	RPS19	0	0	0.56	-1.78			
Saved filters	\$ <b>v</b>														
(No saved filters a	available)	6	-5-	÷.	ıllı	:=	RPS27	RPS27	0	0	0.66	-1.53			
☐ Generate filte	ered node	7	-5-	.÷:	ıll.	∷	ARID3A	ARID3A	0	0	27.56	27.56			
🖺 Save as man	aged list	8	-5-	÷.	մե	∷	IGHA1	IGHA1	0	0	28.51	28.51			
		9	-5-	÷.	մե	∷⊒	RPS12	RPS12	0	0	0.49	-2.06			
		10	-\$-	.÷.	ıllı	∷	IGHD	IGHD	0	0	0.08	-12.02			
		11	-5-	.÷:	ıllı	∷	SAMSN1	SAMSN1	0	0	5.23	5.23			
		12	-5-	.÷.	ıllı	:=	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 k 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 in the volcano plot (Figure 14).

🗎 Sav	e		& Style		I	a × 🗌		Activated B cells vs Mature B cells	
🖒 Save	as	ni Statistics 🛛 🛛 🗙	Color		Shape		8		Significance Up:regulated (1422)
🔸 Und	0	Analytics	Color by	Significance 🔻 🔵 🦉	Shape by Fixed shape	<b>•</b>	1	•••••••••••••••••••••••••••••••••••••••	<ul> <li>Down-regulated (938)</li> <li>Inconclusive (5573)</li> <li>Not significant (14095)</li> </ul>
A Red	0	Regression line	Up	•	Border size				00070 pointe
🖏 Export i	mage	Significance Show lines	Down	<b>•</b>	Labeling			• •	20072 points
📩 Capture	video	X threshold -2 O 2	Inconclusiv		Label by 👔 Name	•			3 9*
To note	book	Y threshold O 0.05	Insignifican	t <b>🔳 🔻</b>	Label font 10 px		75		·•
Setup			Opacity		$\mathbf{D}$				•
🖸 New p	🖽 Axes		Size			×			
🛢 🛛 Get d	a		Size by	Fixed size 🔻 🔵		_			
Tools	X axis		Y Point size	6			(en		
► Select &	Data	Fold change (	D			· ·	-val		
Additio	Range	Min25.1	Range	Min 🔿 🔲 0	Fit visible 👔 📃		10(P		•
Configure		Max () 67.9;		Max 🔿 🔲 1	Full axis labels 🕧 🔵		log		
⊞ Axe	s Scale	Fold change 🔹	Scale	P-value v					•
ai Statis	Show title	-	Show title	-					
<b>C</b> (1)	Title	log2(Fold change)	Title	-log10(P-value)					•
styl	Title font	12 px	Title font	12 px			55		
# Group	Show axis		Show axis						
A8. Descrip	Show labels	a 📕	Show labels	-				1122 BEL	
📼 Backgr	Label font	10 px	Label font	10 px					
	Label rotati	on O0	Label rotation	-O 90					
	X grid		Y grid	-					5
	Show lines		Show lines				-9	-4 0	5 9
	Major lines	4	Major lines 4						
+ Sheet 1 *	Minor lines	1	Minor lines 1						

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 list															
Results	: 2240	Optio	onal co	lumns	;										
Filter	Clear all							Activated B cells vs Mature B cells							
Gene Expres	sion ID 4		View	,			Gene Expression ID ↑↓	Gene name   ↑↓	P-value ↑₹	FDR step up   ↑↓	Ratio ↑↓	Fold change ↑↓	LSMean(Activated B cells) ↑↓		
Gene name     P-value	•	1	4	:.	alt.		TMSB10	TMSB10	0	0	0.38	-2.62	112186		
FDR step up	•	·	.,	•••	11114		TMSDTO	TWODIO	0	0	0.50	-2.02	1,121.00		
All contrasts	O Per contrast	2	-5-	.÷:	dh		YWHAH	YWHAH	0	0	7.54	7.54	471.31		
Less than or	• 0.05	3	-\$-	.:.	dl.		SAMSN1	SAMSN1	0	0	5.23	5.23	517.43		
Ratio	4	4	-5-	.4	dl.		YBX3	YBX3	0	0	0.04	-23.11	42.13		
From -2 to	2	5	-5-	.÷:	dh	II	ACY3	ACY3	0	0	67.92	67.92	251.03		
Exclude range	•	6	-\$-	.÷.	dl.		LITAF	LITAF	0	0	6.05	6.05	974.57		
	ed 4	7	-5-	.::	dl.		SLAMF7	SLAMF7	0	0	23.13	23.13	211.15		
Save	filter	8	-5-	.4	dl.		IGKC	IGKC	0	0	7.94	7.94	13,123.68		
(No saved filters	available)	9	-\$-	.:.	dl.		IGHD	IGHD	0	0	0.08	-12.02	38.28		
	ered node	10	4-	.÷:	dh	II	IGHA1	IGHA1	0	0	28.51	28.51	574.77		
🔒 Save as mar	naged list	11	4-	.÷:	dl.		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

- ☐ 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

Click

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 1	FDR step up ↑↓	Rich factor ↑↓	Genes in set Î↓	Genes in list ↑↓	Genes not in list ↑↓	Genes in list, not in set ↑↓	Genes not in list, not in set ↑↓	0
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).



Back Transcriptional misregulation in cancer

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.

