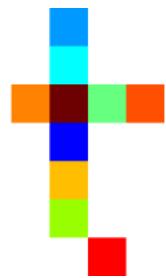


A02 - Automated Annotation



- 1 The automation of cluster naming requires an essential file called the "cluster_annot_ref". The annotation algorithm uses it to decide the cluster names.

Go to the "FCS Annotations" folder and you will find two files like this.

- "cluster_annot_ref_v1.csv"
- "cluster_annot_ref_v2.csv"

Click on the cluster_annot_ref_v1.csv file

It gives pos and neg marker definition for each population.

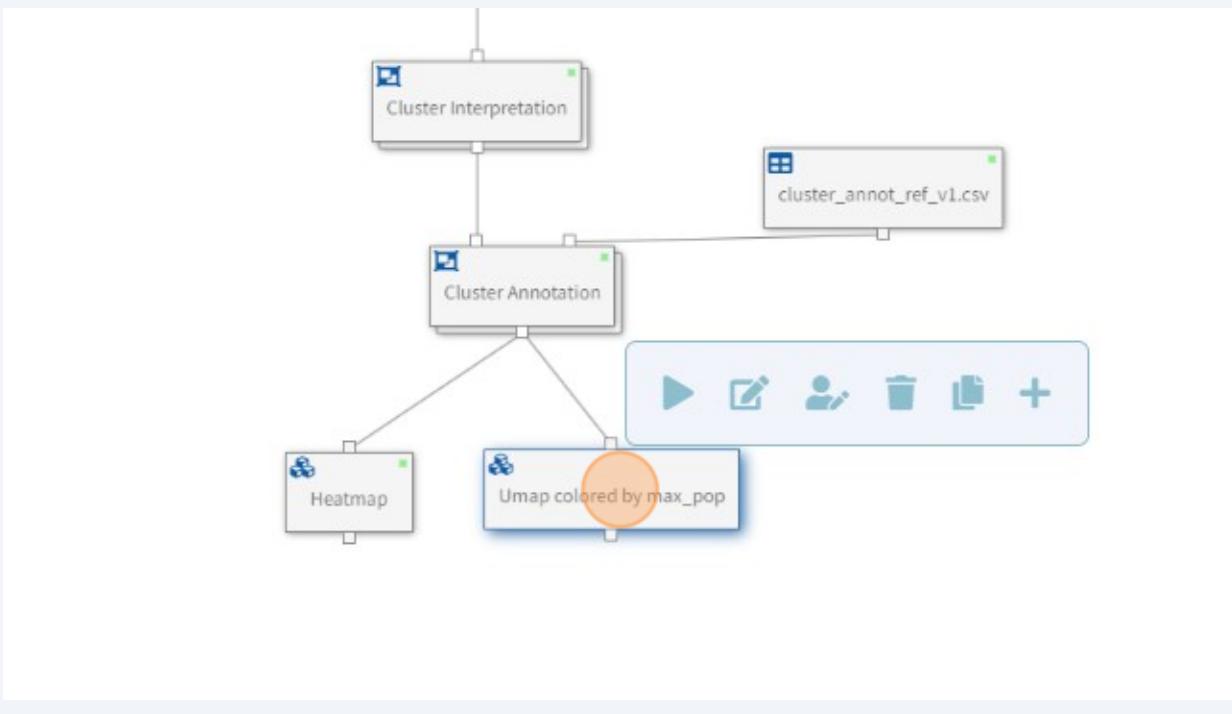
Rows: 3 / 3		
	Previous	Next
population	pos_markers	neg_nr
(character)	(character)	(chara
CD4 T cells	CD4_CD3	
CD8 T cells	CD8_CD3	
Non T cells		CD3

- 2 We will be using the "[Automated Cluster Annotation](#)" workflow.

This workflow has been run with v1 of the reference file.

We will now look at a few plots to check the performance of it.

- 3 Double-click here.



- 4 This opens the projection.

This umap is colored by the population names (population.max_pop) that were automatically assigned to the clusters.

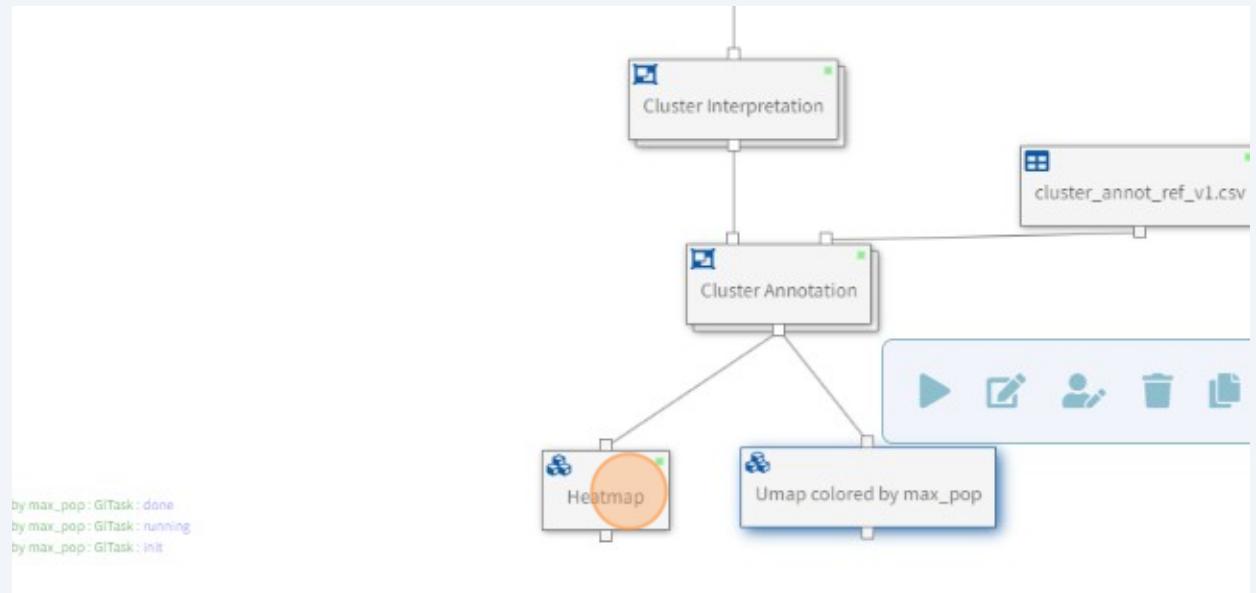




Discuss! How does this compare to the manual annotation you did earlier

5

Double click on the Heatmap

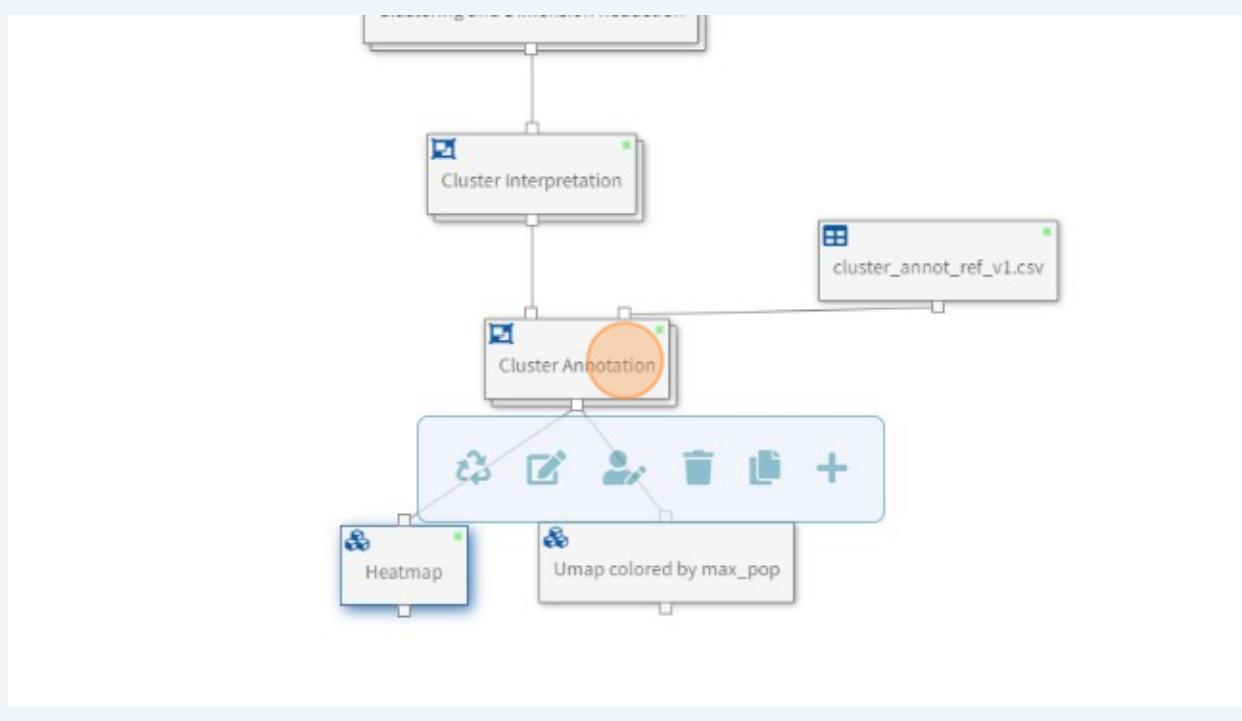


6 The heatmap gives an overview of the annotation.

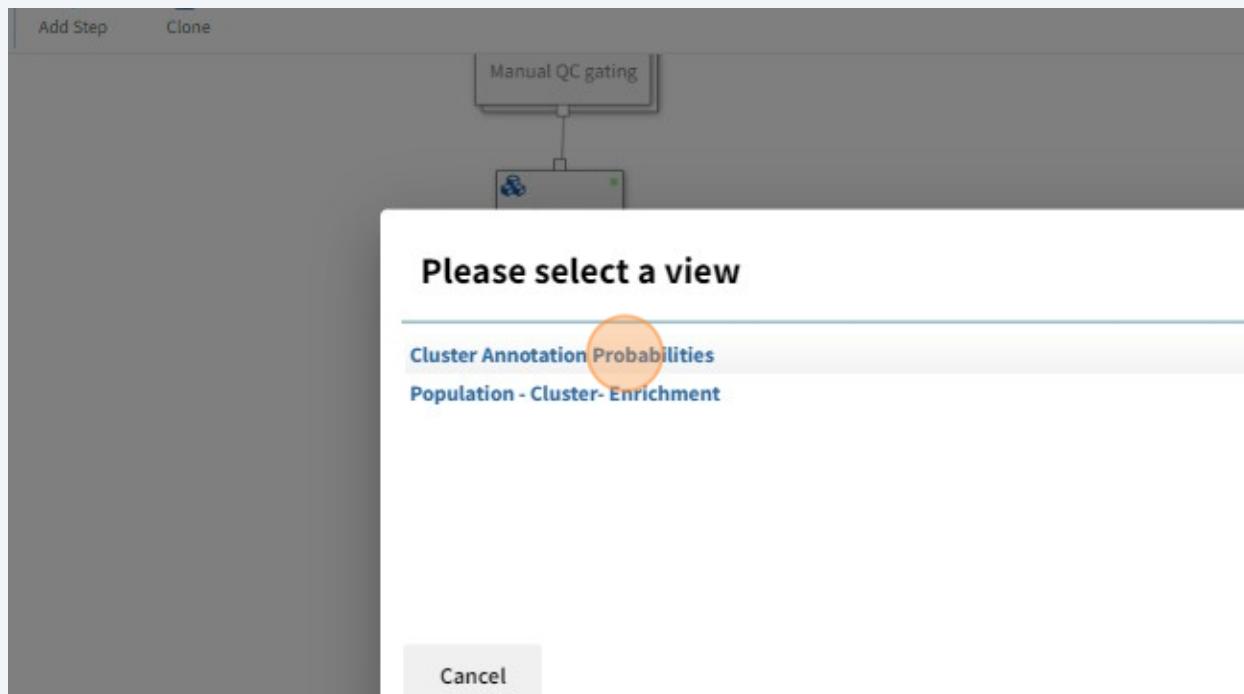


i Discuss! How does it handle the signaling (pErk, IFNg) markers?

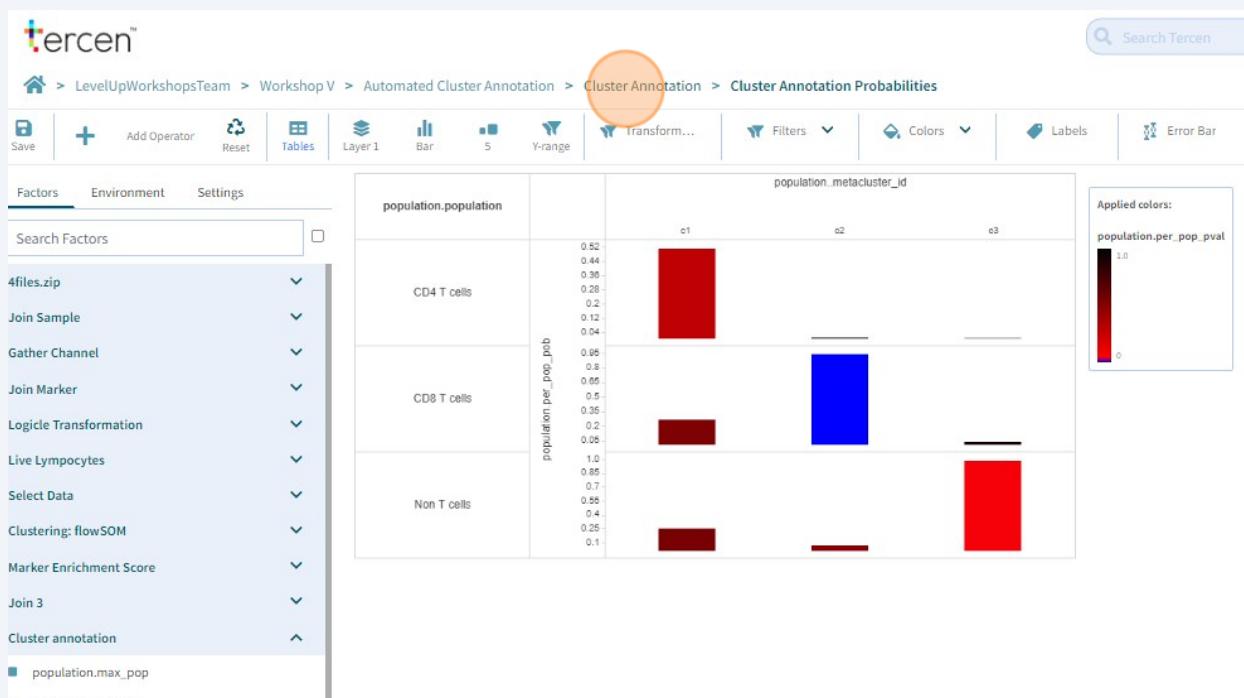
7 Double-click here.



8 Click "Cluster Annotation Probabilities"



9 Click "Cluster Annotation Probabilities"





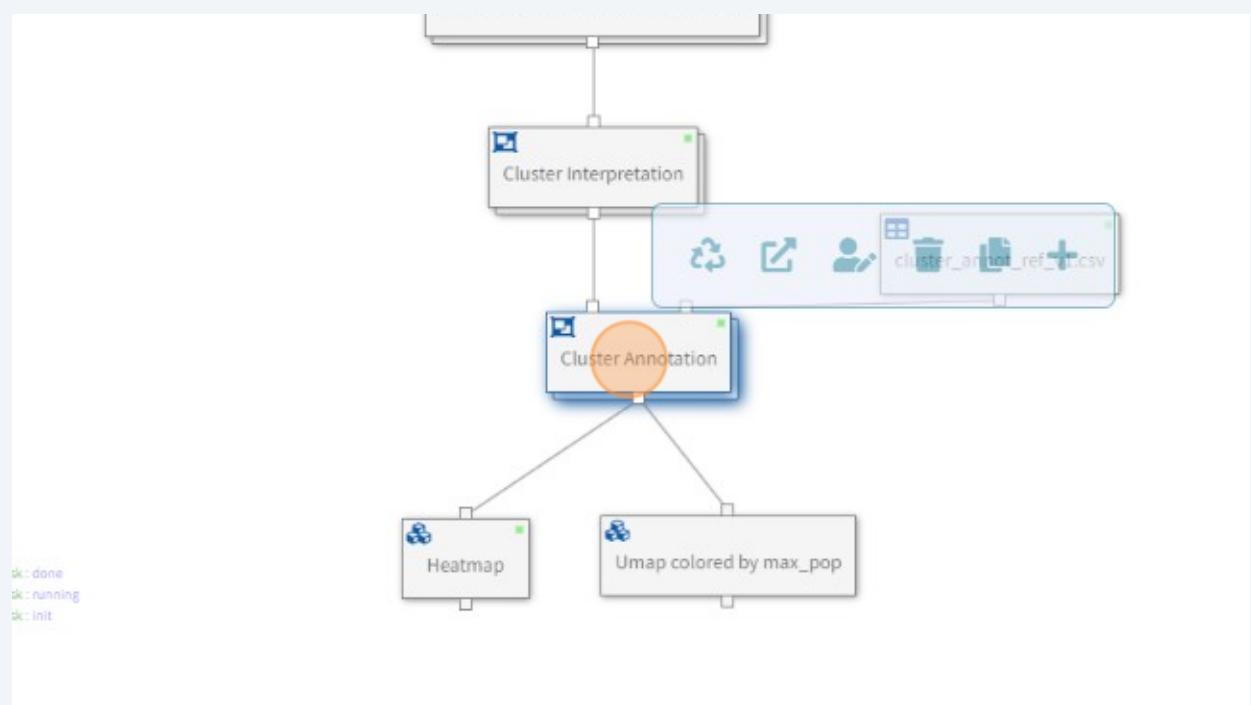
Tip! This diagnostic plot shows how strong the relationship between a cluster and a population was. The bigger the bar, the more probable the cluster belongs to the population. The color is the p-value. The better the value, the more blue it is.

Each population is assigned the cluster with the highest probability.

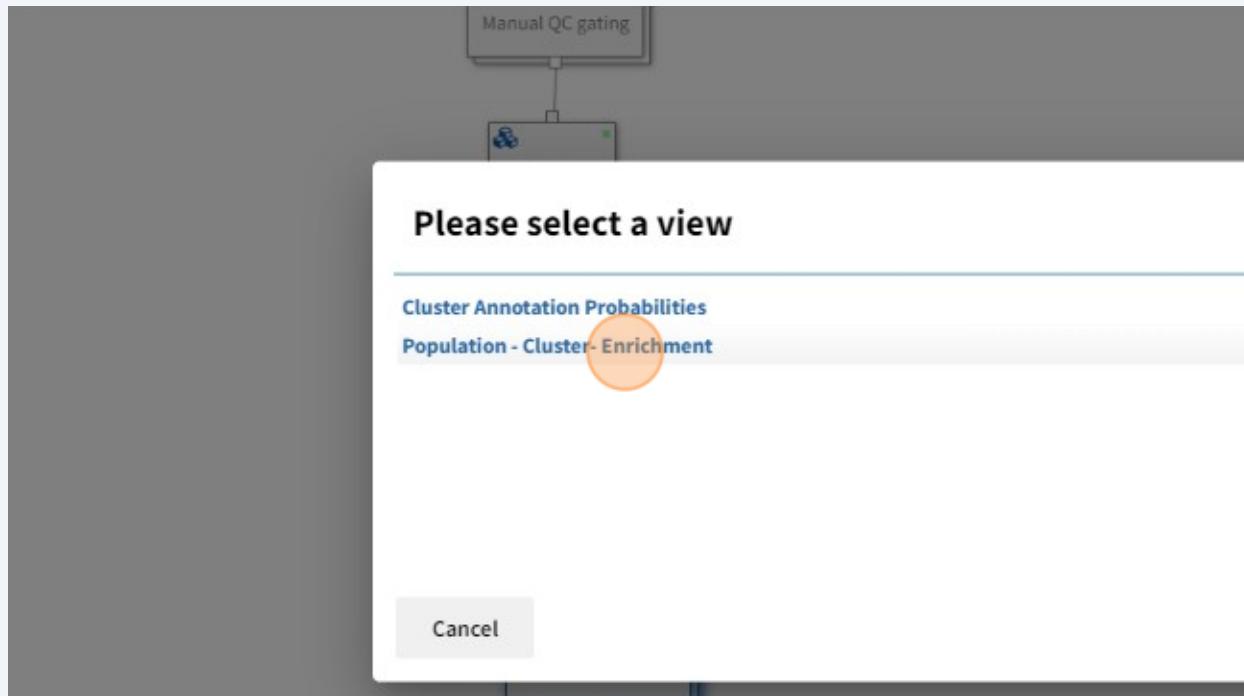
This is why one of the results of auto annotation is a the factor is called "max_pop"

10

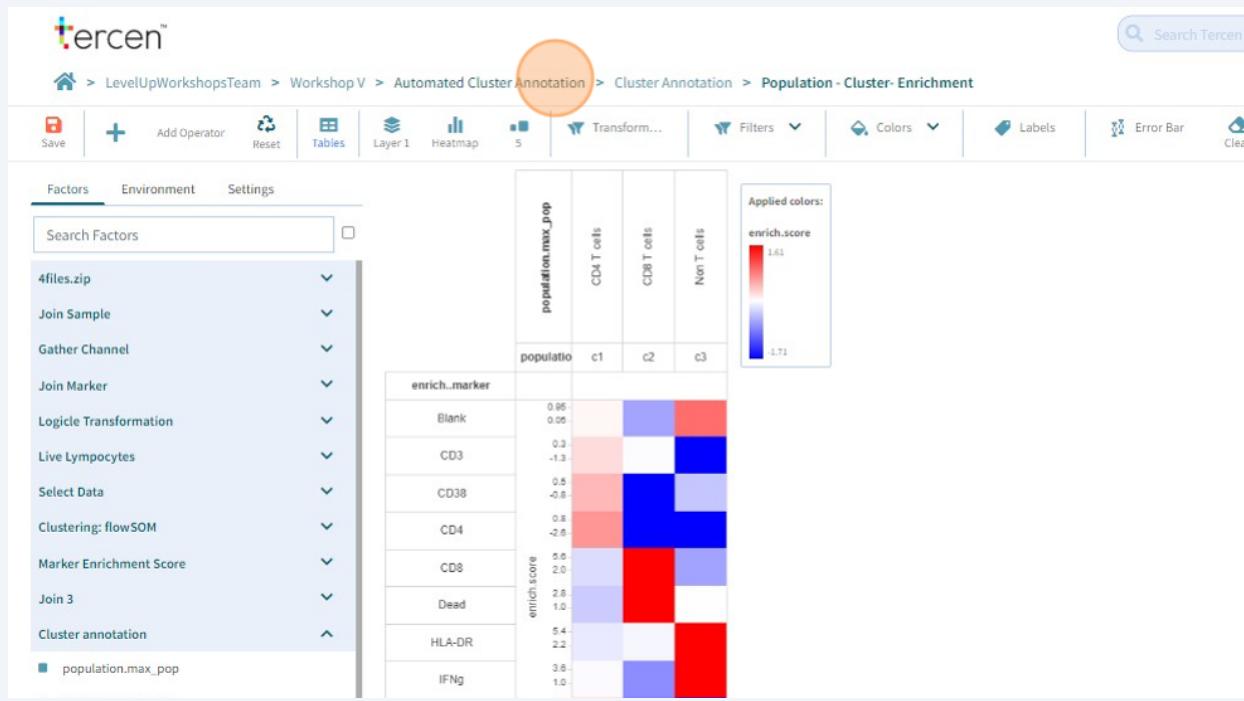
Double-click here. We will look at another plot.



11 Click "Population - Cluster- Enrichment"



12 Click "Automated Cluster Annotation"





Tip! This plot allows you to see the enrichment scores with the population names.

Red means enriched, and Blue means depleted.

Notice the difference between CD4 and CD8 markers across the clusters/population names.

13

We will now replace the cluster_anno_ref file with another.

Go to the "FCS Annotations" folder and look at cluster_annot_rev_v2 file

population	pos_markers
(character)	(character)
p-ERK+ IFNg+ CD4 T cells	p-ERK_IFNg_CD4_CD3
p-ERK- IFNg+ CD4 T cells	IFNg_CD4_CD3
p-ERK+ IFNg- CD4 T cells	p-ERK_CD4_CD3
p-ERK- IFNg- CD4 T cells	CD4_CD3
p-ERK+ IFNg+ CD8 T cells	p-ERK_IFNg_CD8_CD3
p-ERK- IFNg+ CD8 T cells	IFNg_CD8_CD3
p-ERK+ IFNg- CD8 T cells	p-ERK_CD8_CD3
p-ERK- IFNg- CD8 T cells	CD8_CD3
Non T cells	
Non CD4	
Non CD8	



Discuss! The v2 file is very different from v1.

There are many more populations defined.

What is different between the two v1 and v2?

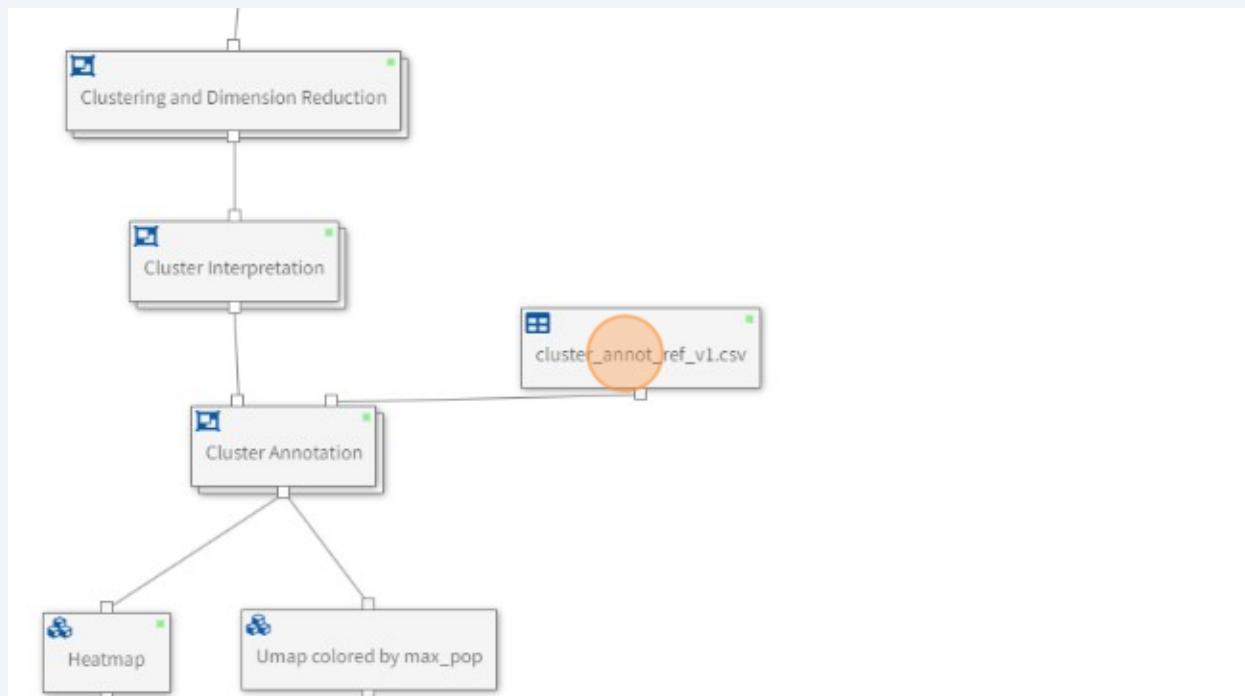
What do you think v2 trying to capture?

To use v2, will we have to also change the number of clusters?

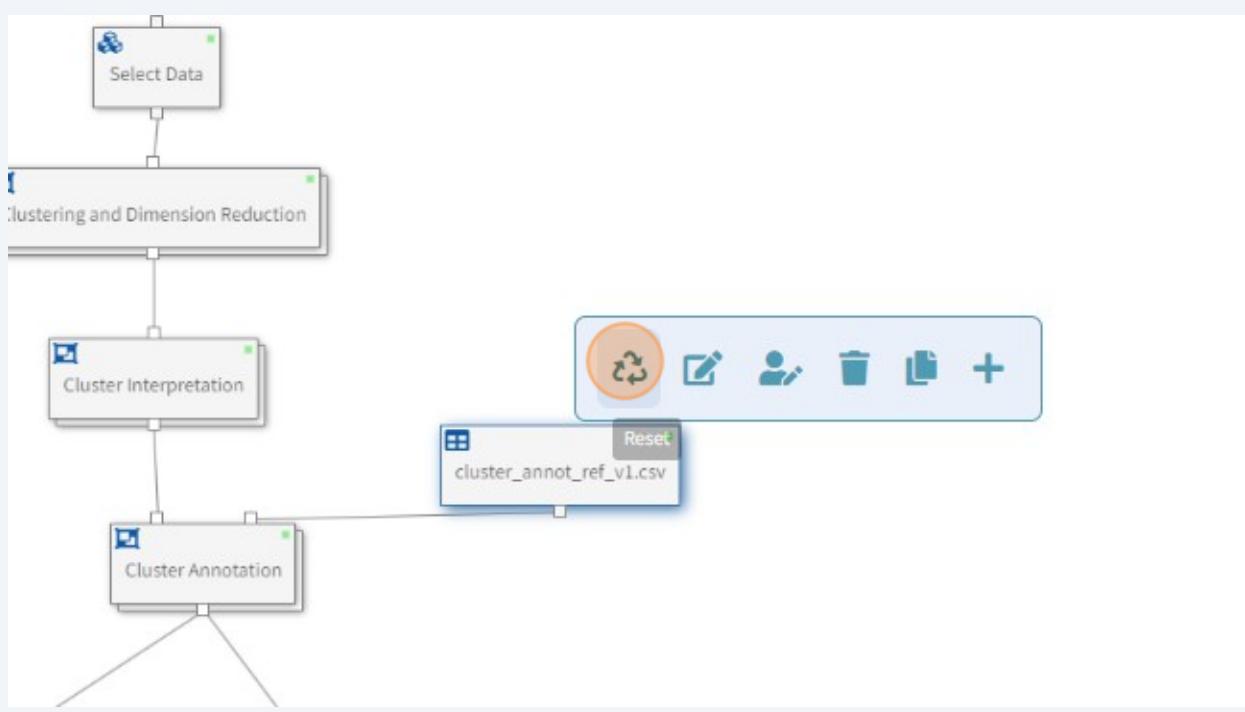
14

Lets begin using v2 of the cluster_anno_ref.

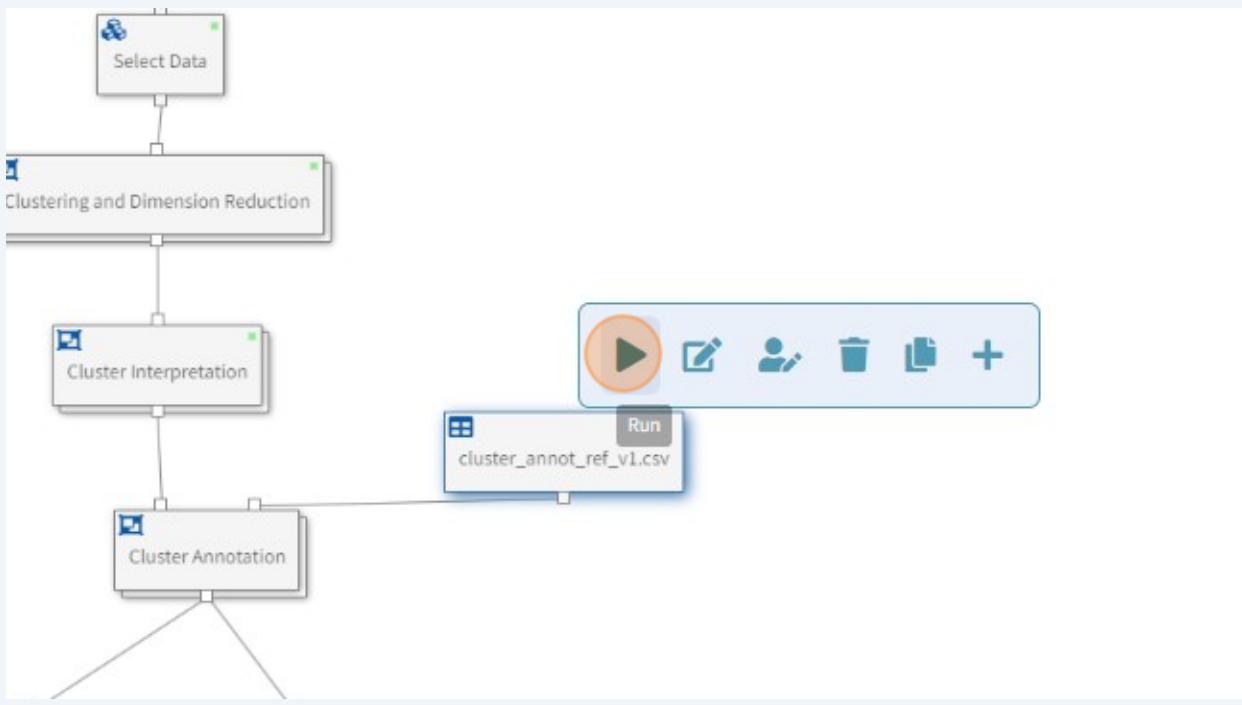
15 Click here.



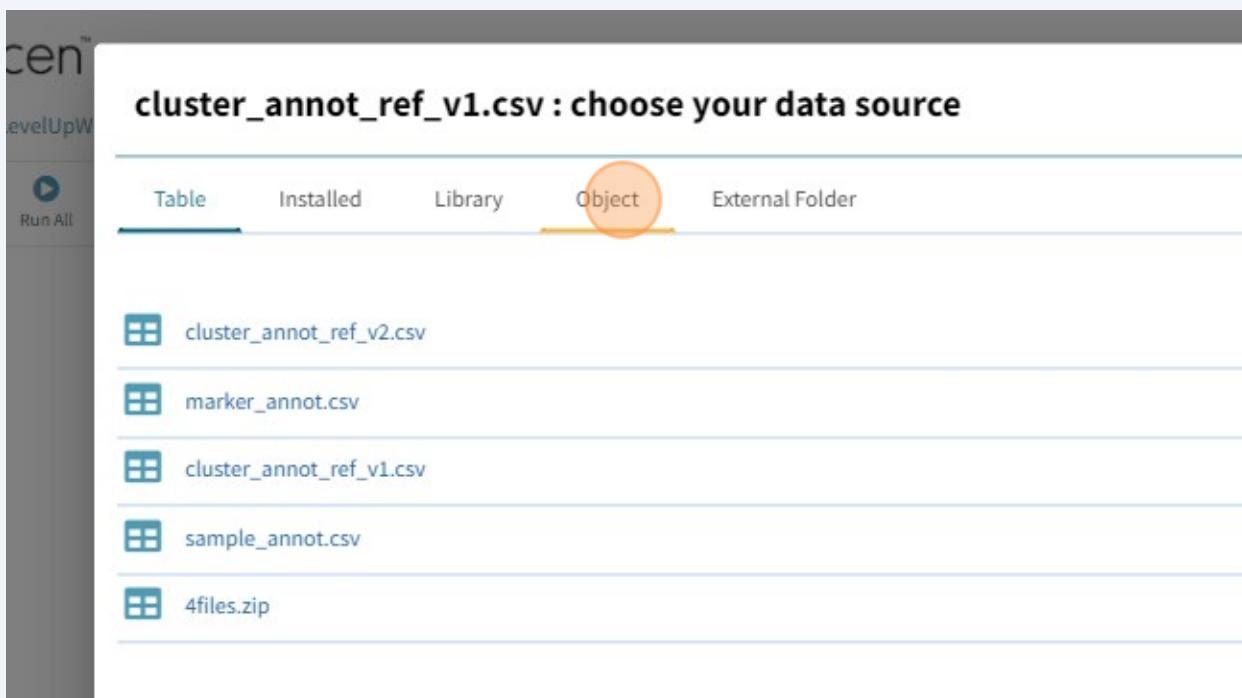
16 Click here.



17 Click here.

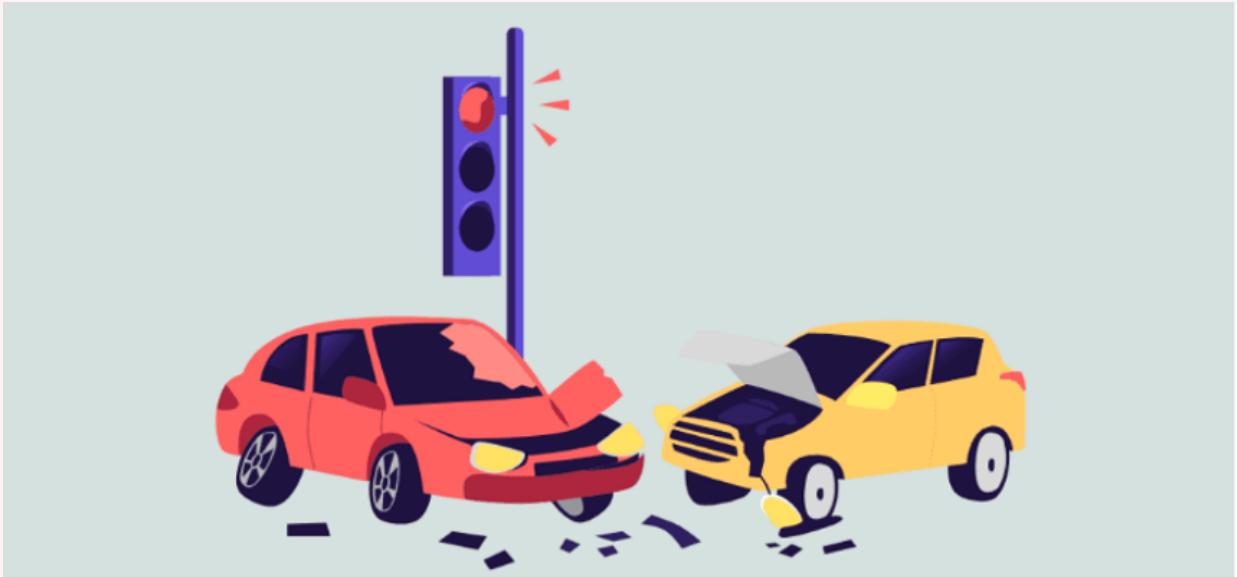


18 Click "Object"





Alert! Alert! Notice how we had to go the Object tab.
This is because the algorithm is expecting an object file NOT a table!



19 Select the cluster_annot_ref_v2.csv

cluster_annot_ref_v1.csv : choose your data source

Table Installed Library **Object** External Folder

Automated Cluster Annotation	A template to perform clustering (Phenograph) and dimension reduction (tSNE) on flow cytometry data.	3 minutes ago
Basic Flow Template_2	A template to perform clustering (Phenograph) and dimension reduction (tSNE) on flow cytometry data.	38 minutes ago
Basic Flow Template	A template to perform clustering (Phenograph) and dimension reduction (tSNE) on flow cytometry data.	1 hours ago
Manual Cluster Annotation Workflow	A template to perform clustering (Phenograph) and dimension reduction (tSNE) on flow cytometry data.	1 hours ago
cluster_annot_manual.csv		2 hours ago
cluster_annot_ref_v2.csv		1 days ago
marker_annot.csv		1 days ago
cluster_annot_ref_v1.csv		1 days ago
sample_annot.csv		1 days ago
4files.zip		4 days ago
a7e9f2a3-a31f-41ad-b200-c1d11d5c3312		4 days ago
47e3e288-5789-4d4e-8f0f-837fc11a2bb9		4 days ago

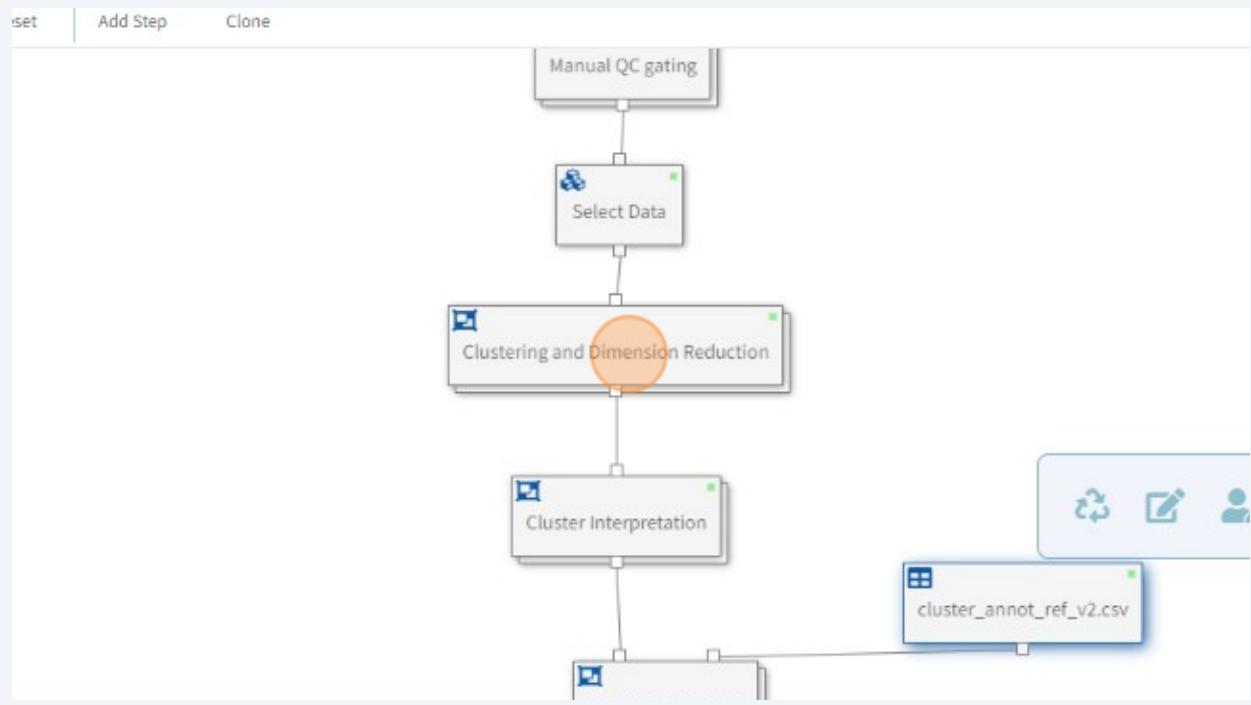
Cancel OK

20

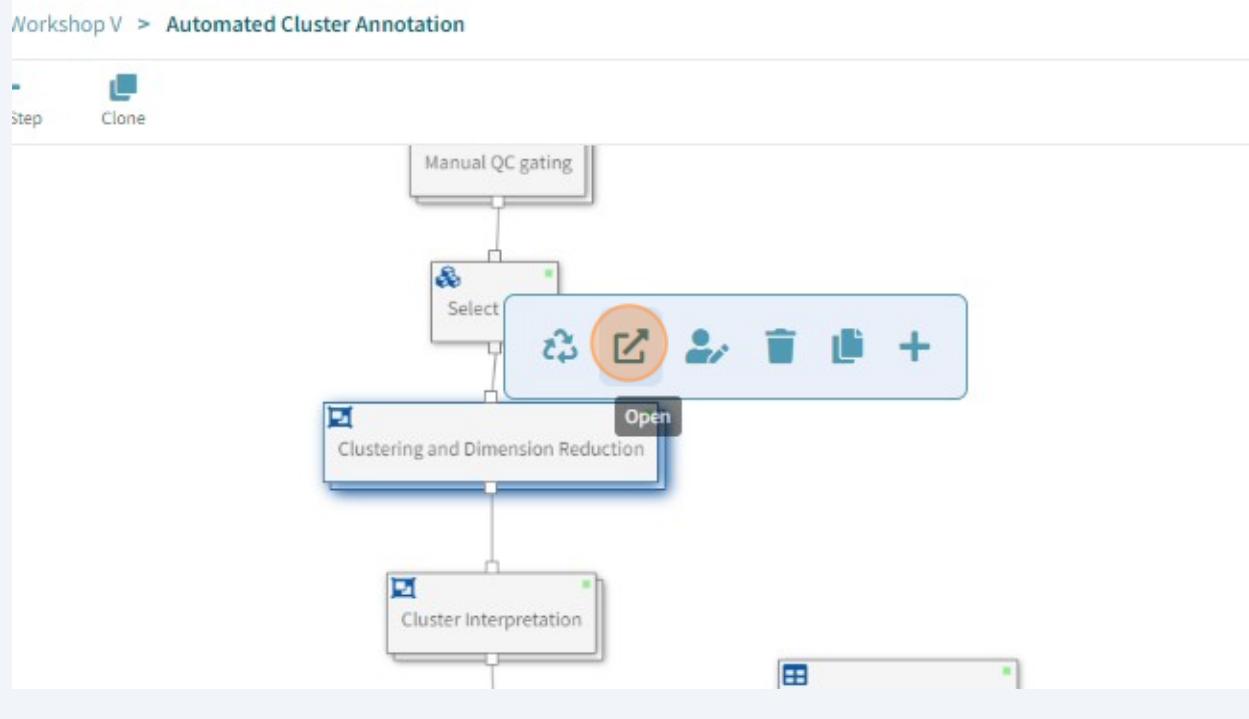
We require to also change the clustering approach by changing the number of clusters and markers. Lets do this now.

21

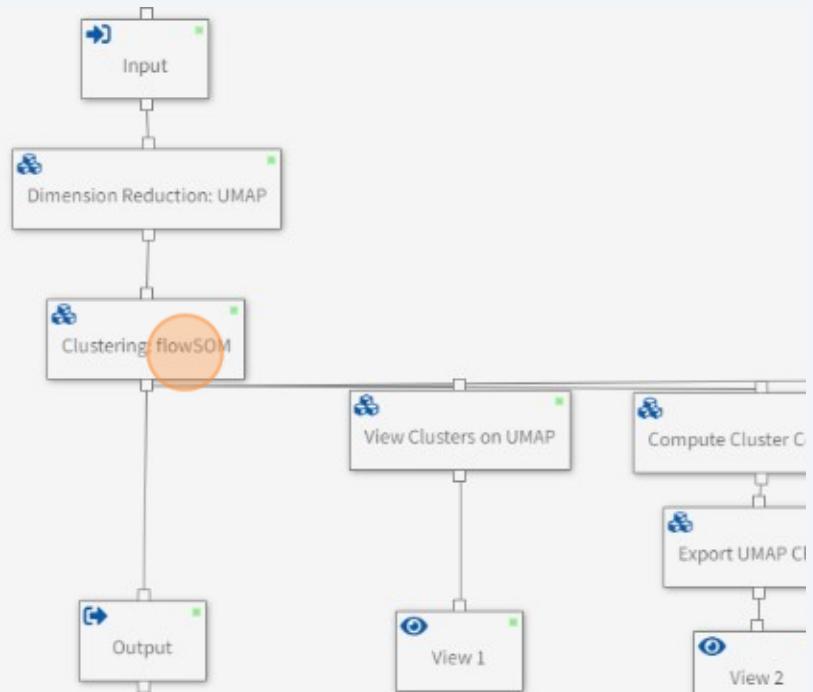
Click here.



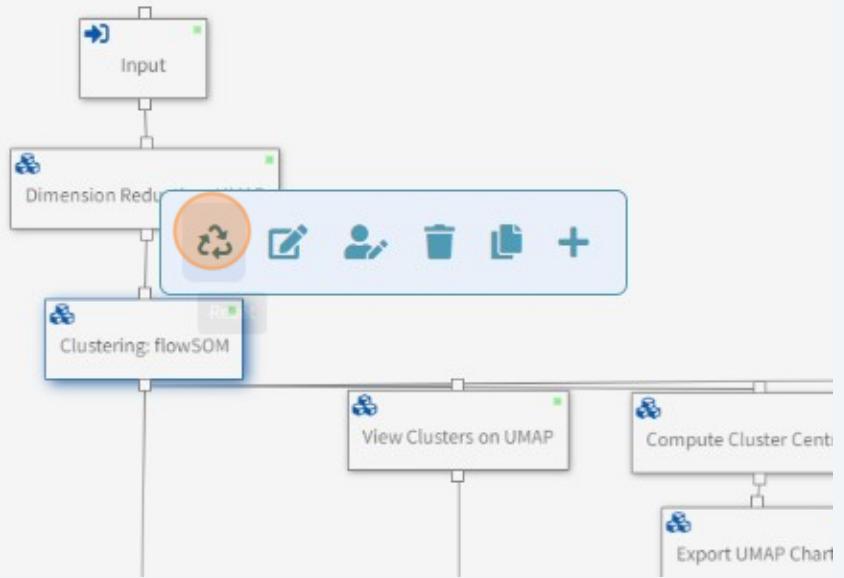
22 Click here.



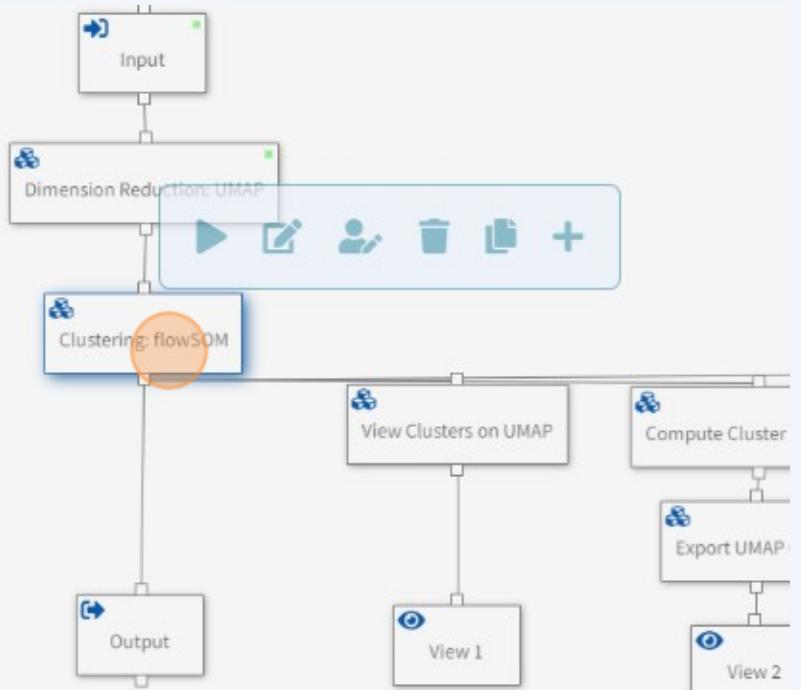
23 Click here.



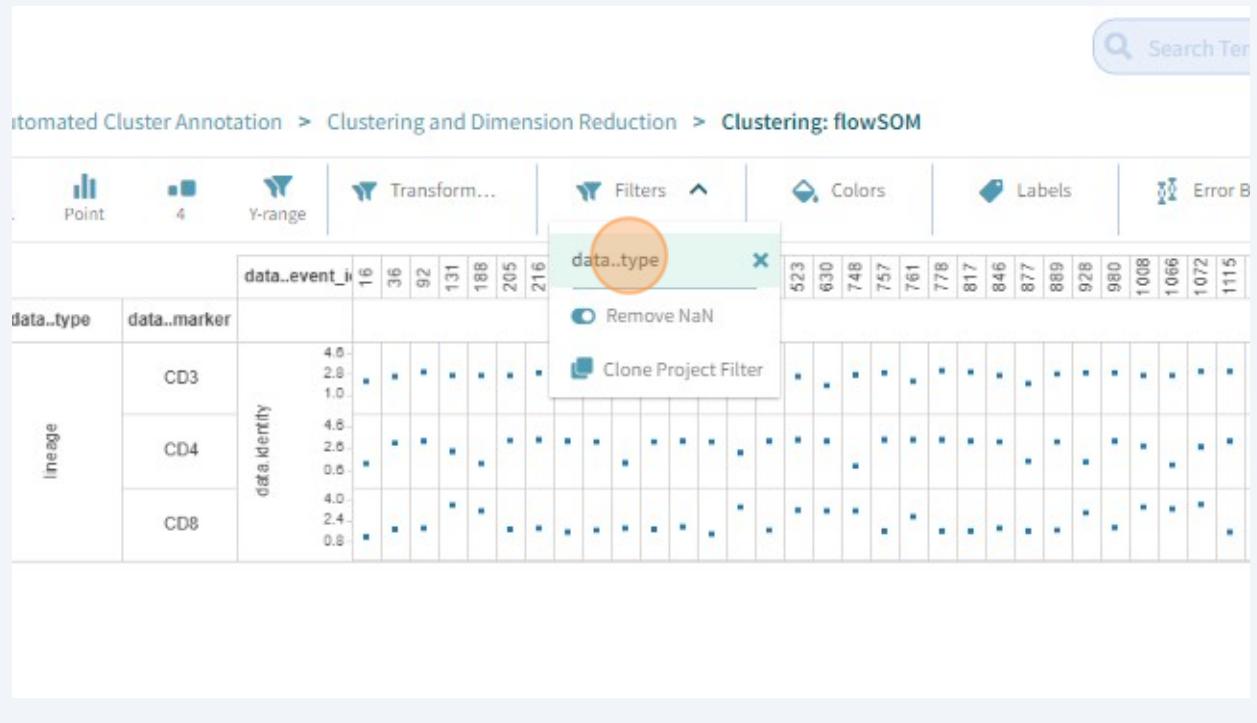
24 Reset the step



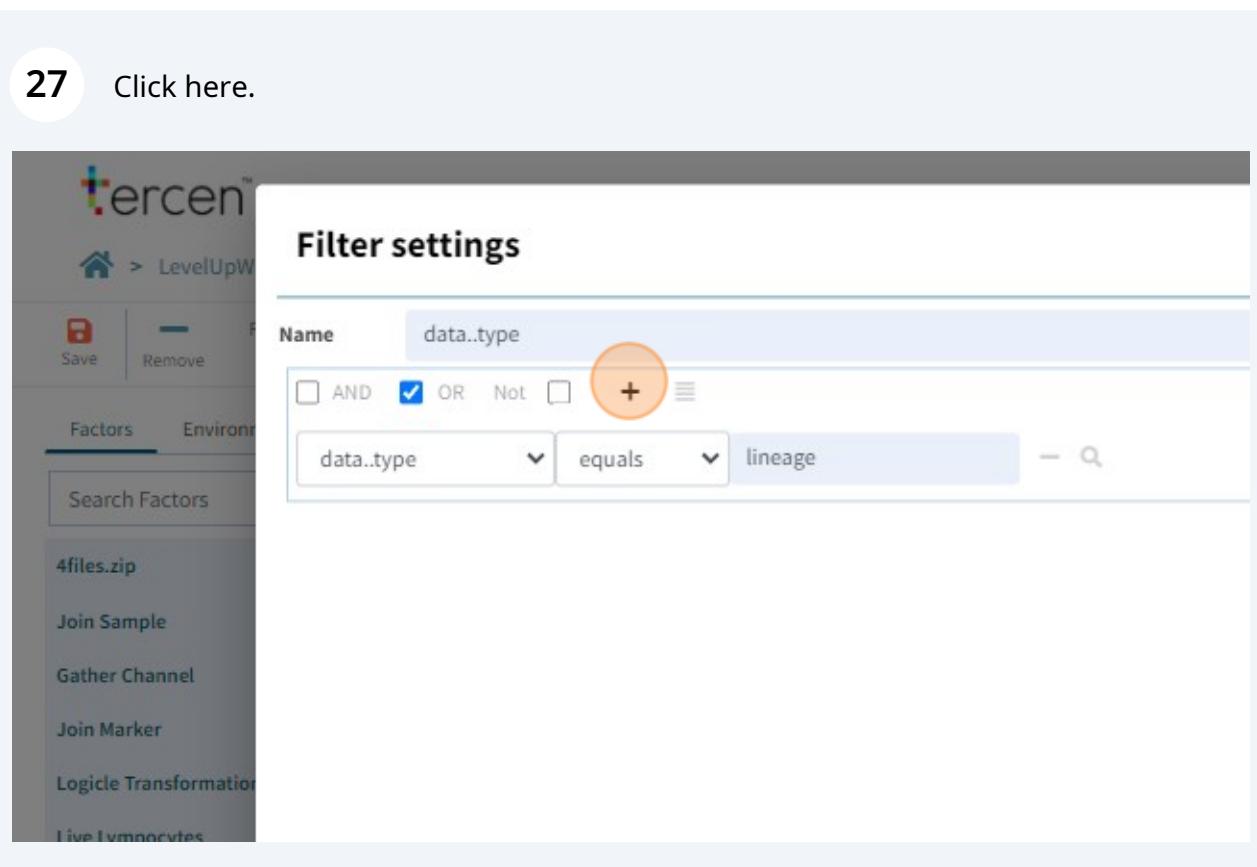
25 Double-click here.



26 Go to the filters and click "data..type"



27 Click here.



28 Click this dropdown.

The screenshot shows the Tercen software interface. On the left, there's a sidebar with options like 'Save', 'Remove', 'Factors', 'Environ', 'Search Factors', '4files.zip', 'Join Sample', 'Gather Channel', 'Join Marker', 'Logicle Transformation', and 'Live Lymphocytes'. The main area is titled 'Filter settings' with a name 'data..type'. It has an OR operator selected. There are two filter conditions: 'data..type equals lineage' and 'FSC-A equals NaN'. The dropdown menu in the second condition is highlighted with an orange circle.

29 Click this text field and type "signal"

The screenshot shows the Tercen software interface. The 'Factors' section is visible on the left. The main area is titled 'Filter settings' with a name 'data..type'. It has an OR operator selected. There are two filter conditions: 'data..type equals lineage' and 'data..type equals NaN'. The text input field for the second condition ('data..type') is highlighted with a blue selection, and the dropdown menu next to it is highlighted with an orange circle.

30 Click here.

The screenshot shows a search interface with two filters applied:

- data.type equals lineage
- data.type equals signal

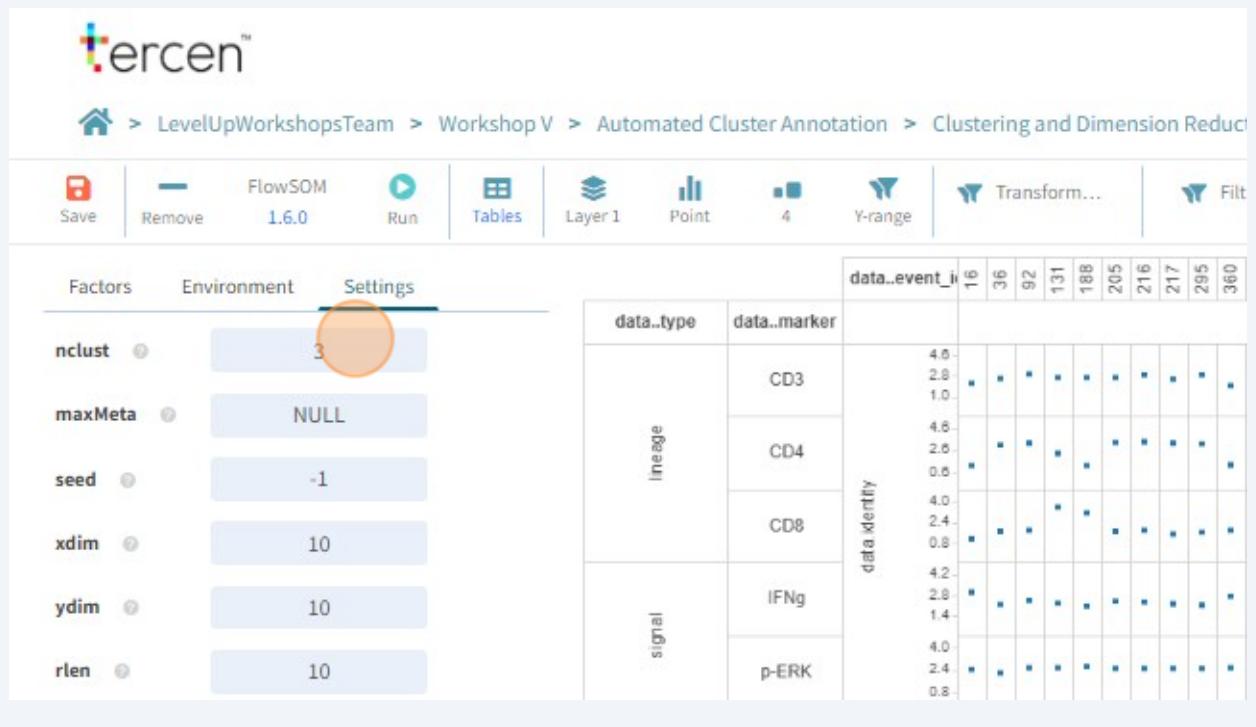
An orange circle highlights the search icon (magnifying glass) next to the "lineage" filter.

31 Click "Settings"

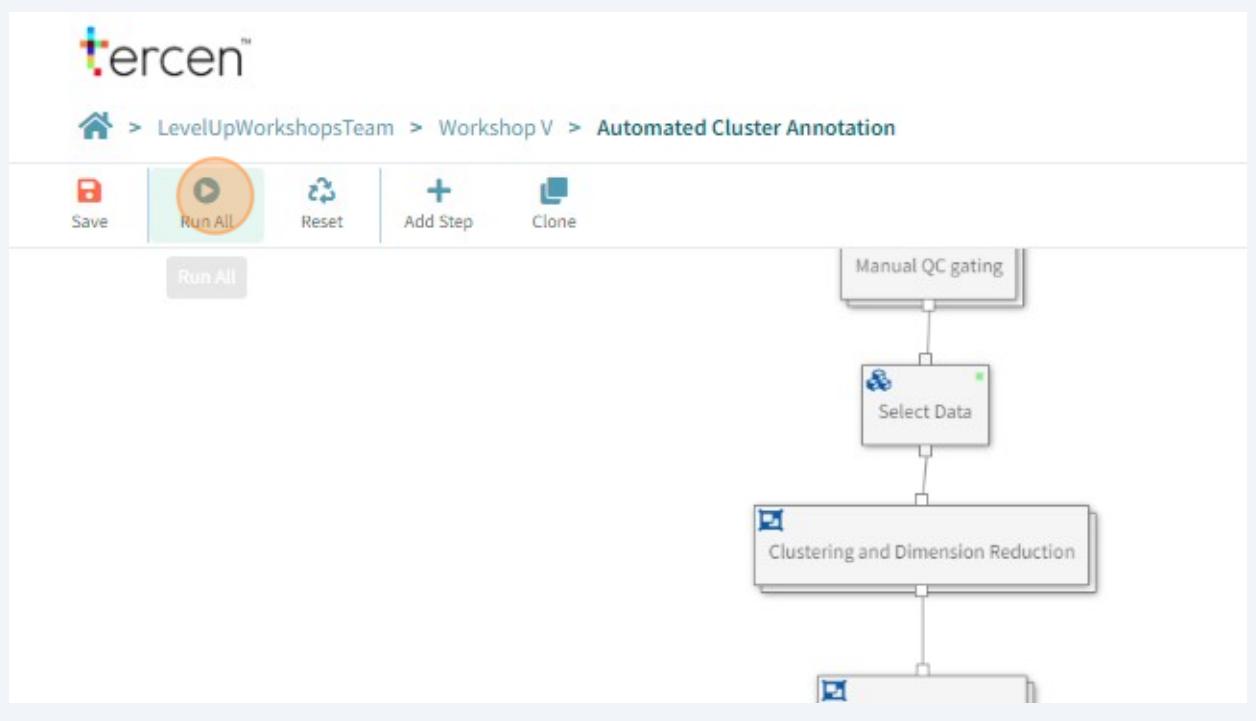
The screenshot shows the Tercen software interface with the "Settings" tab selected. On the left, there is a sidebar with various options like "4files.zip", "Join Sample", "Gather Channel", "Join Marker", "Logic Transformation", and "Live Lymphocytes". On the right, there is a heatmap visualization with columns labeled "data.event_id" and rows labeled "data.identity". The legend on the right maps marker names to colors:

Marker	Color
CD3	Blue
CD4	Orange
CD8	Green
IFNg	Red
p-ERK	Purple

32 Put 20 as the number of clusters



33 Lets click on the Run all button to use the new clustering and annotations.



34

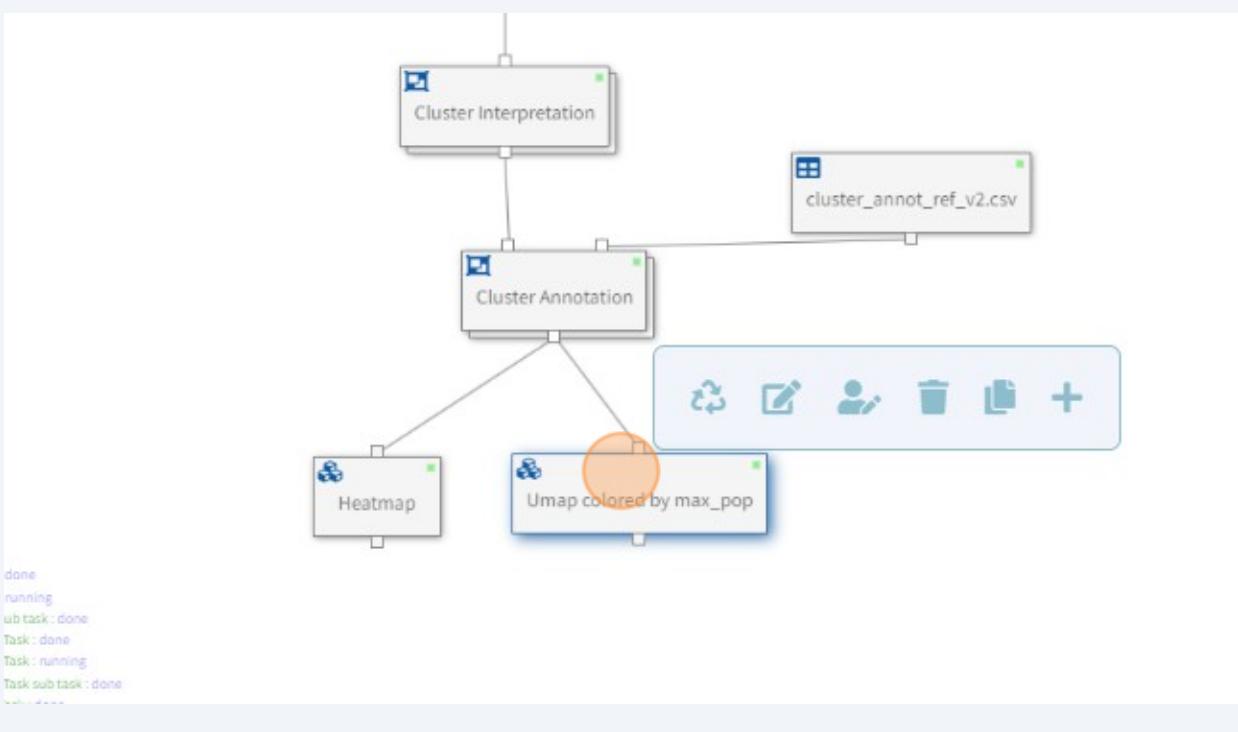
Wait till everything is green (it will take 5 mins)

While it's running, discuss with your colleague what outcome you expect from the change.

Once finished, we will now look at the UMAP colored by max_pop and the heatmap, and the diagnostic plots.

35

Double-click here.

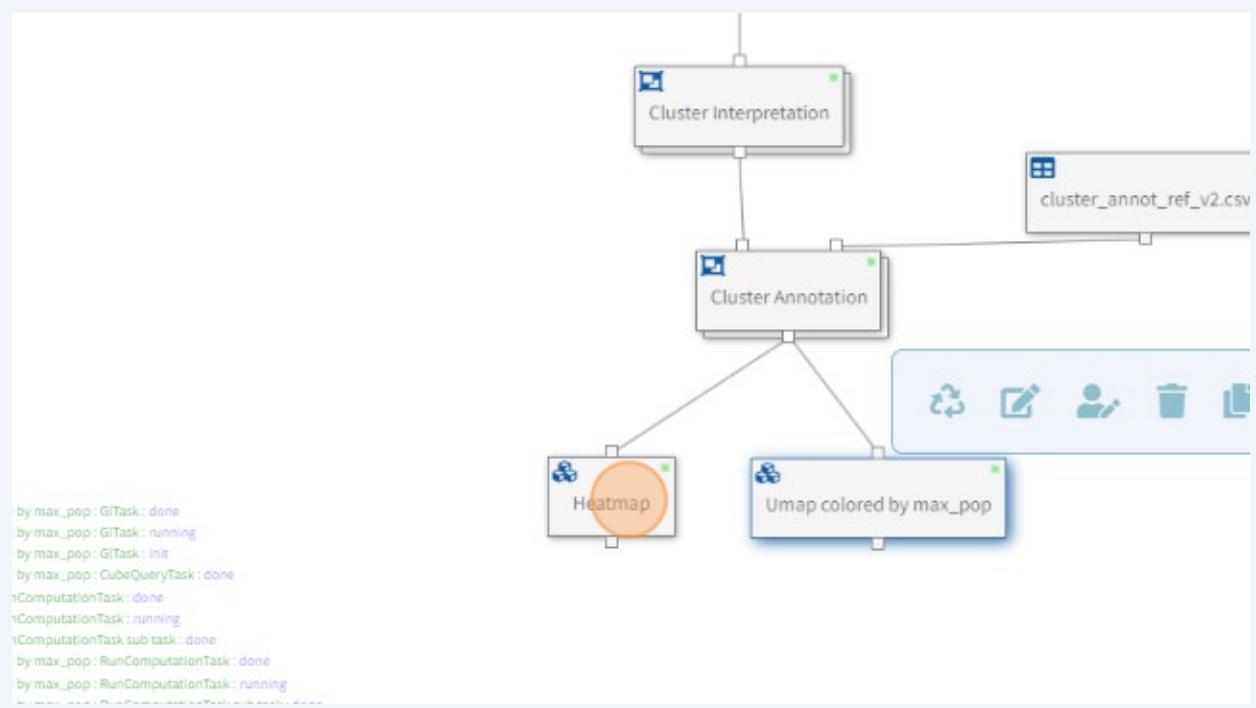


36 This open the umap with the new annotated clusters.

What do you think has changed with the clusters and their annotation?



37 Double-click here.

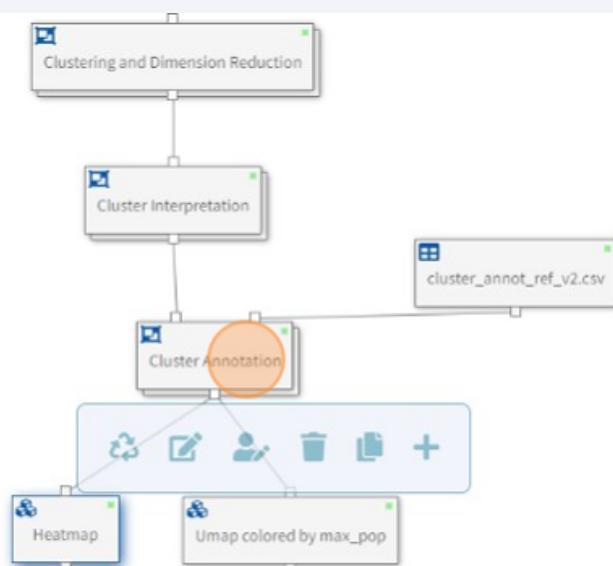


38 This open the heatmap.

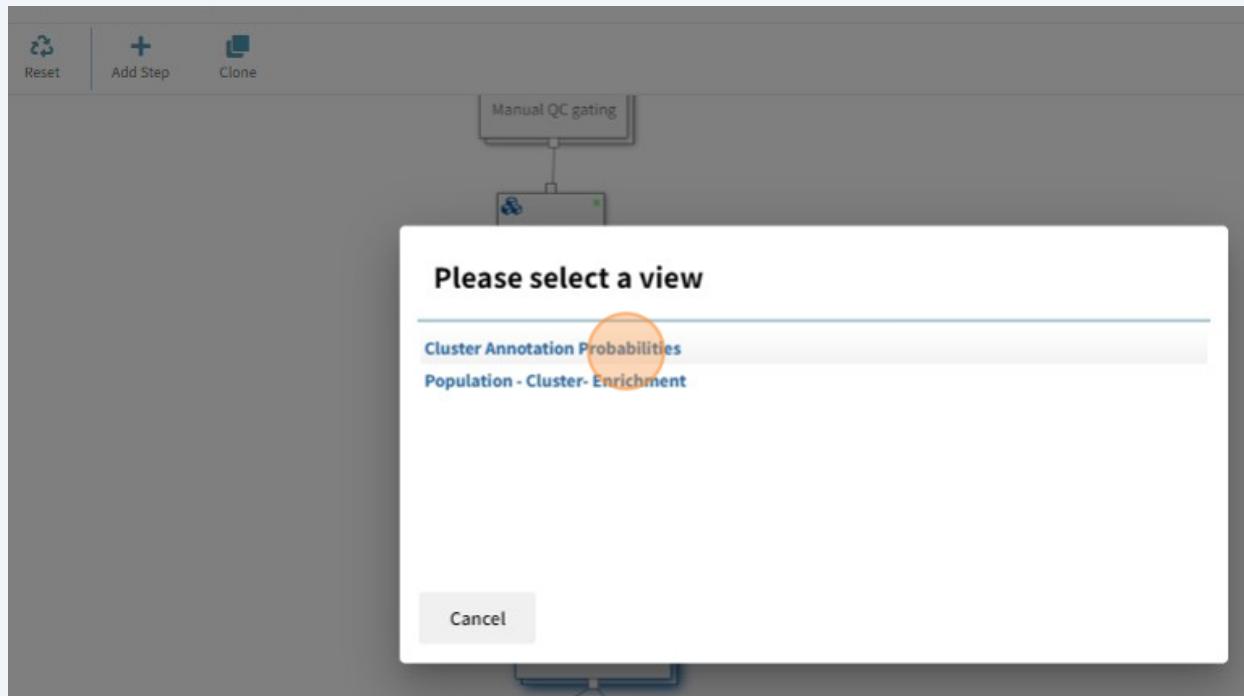
Observe what has changed to how the clusters align with the signalling markers.



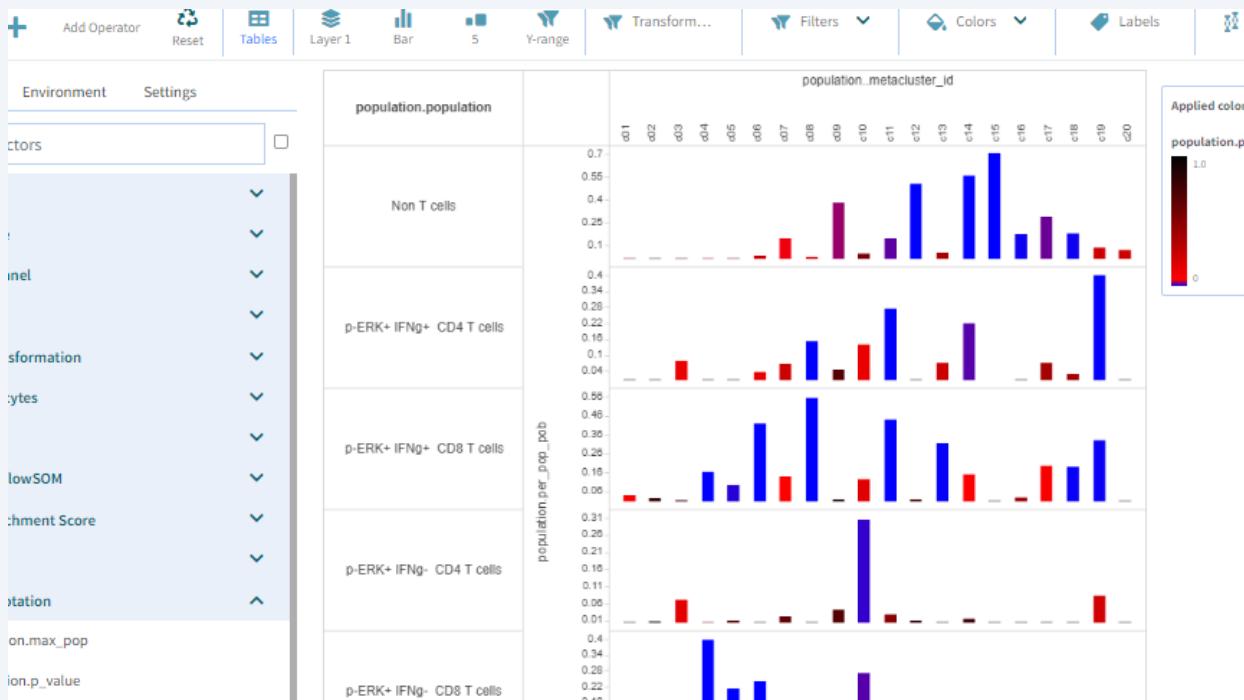
39 Double-click here.



40 Click "Cluster Annotation Probabilities"



41 Here is the diagnostic plot for probabilities





Discuss!

Scan the highest probability bars for each cluster in each population group.

Is there any cluster having many high probabilities of belonging to several population names?



42 Look at the enrichment plot



43 Save this workflow



Discuss!

What was achieved by using the new reference file and cluster settings?

When would this approach be useful?