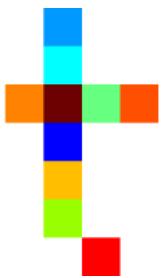




LEVEL^{UP} WORKSHOP III

Part B

0303 - Influence of Settings



- 1 Navigate to your workshop project and **clone** the **first workflow**.
Save it to your project and give it the name "**3. Influence of Settings**".

The screenshot shows a project named "agouy" with the following details:

- Project** tab is selected.
- Activities** tab is visible.
- Clustering and Dimension Reduction** section is shown with no description provided.
- Workflow History:**
 - agouy updated workflow **2. Cluster exploration** 17 minutes ago
 - agouy README.md 23 hours ago
 - Melanoma_AntiPD1_Sample_annotation.csv 5 hours ago
 - Melanoma_AntiPD1_Sample_annotation.csv 2 hours ago
 - Melanoma_AntiPD1_Baseline_FCS.csv 23 hours ago
 - dev 2 hours ago
 - 2. Cluster exploration 17 minutes ago
 - 1. Clustering and dimension workflow 5 hours ago
- File List:**
 - README.md

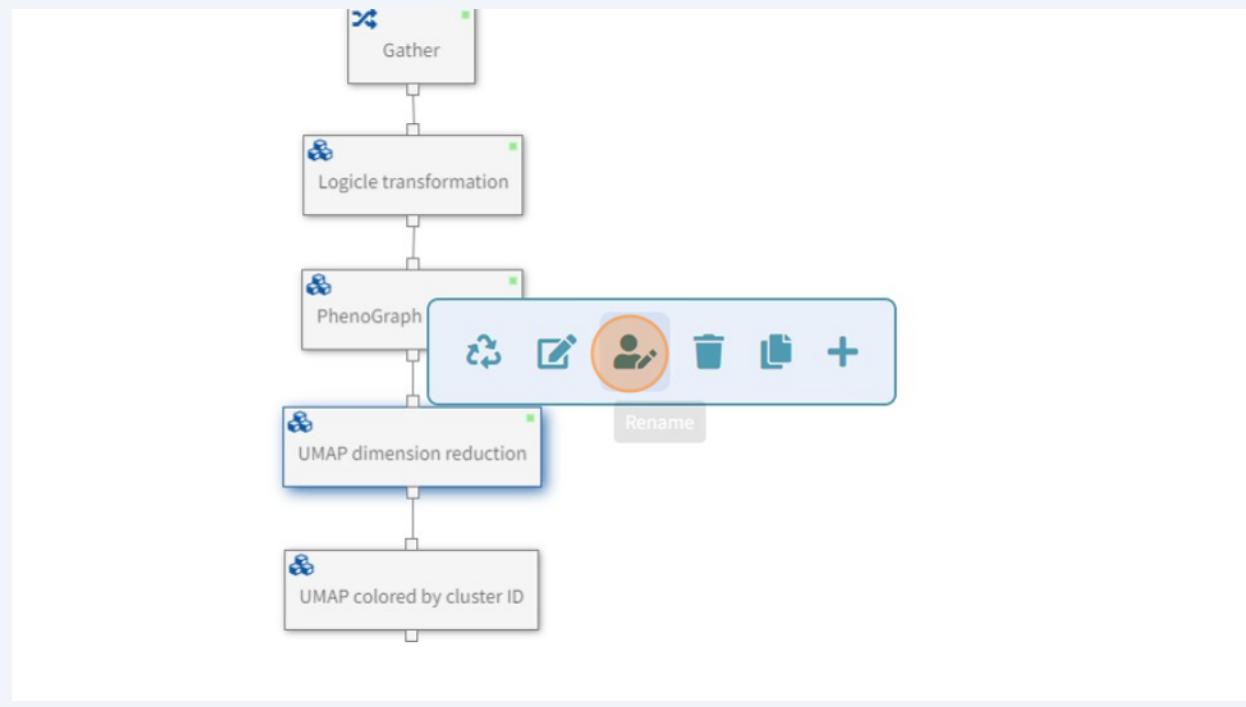
How do settings influence UMAP results?

We will see how the **min_dist** UMAP setting affects the results. To do so, we will duplicate the UMAP and visualisation steps and tweak settings.

This setting is the **minimum distance apart** that data points are allowed to be in the lower dimensional representation. Therefore it controls how tightly UMAP will group points together.

- 2 We recommend naming steps with more detail to help organisation.

Rename this step to **UMAP - Default min_dist = 0.5**

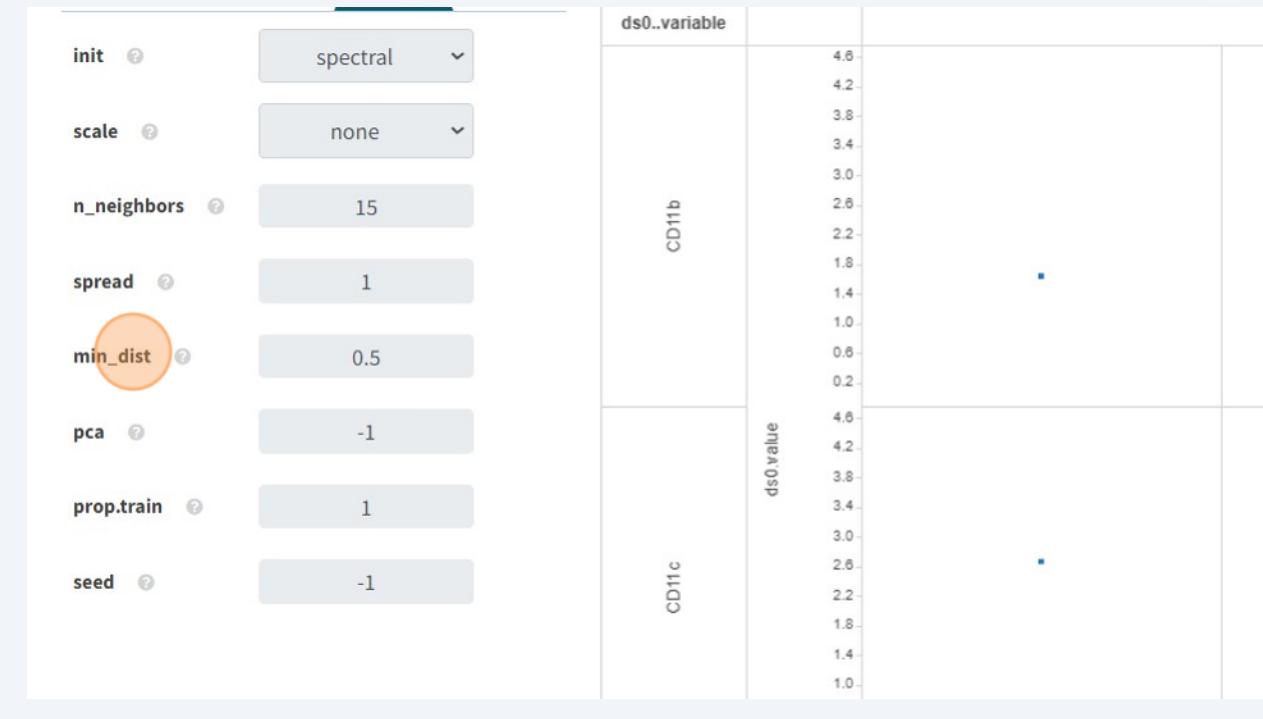


- 3 Click the Data Step and select **Edit**

Navigate to the **Settings** tab.

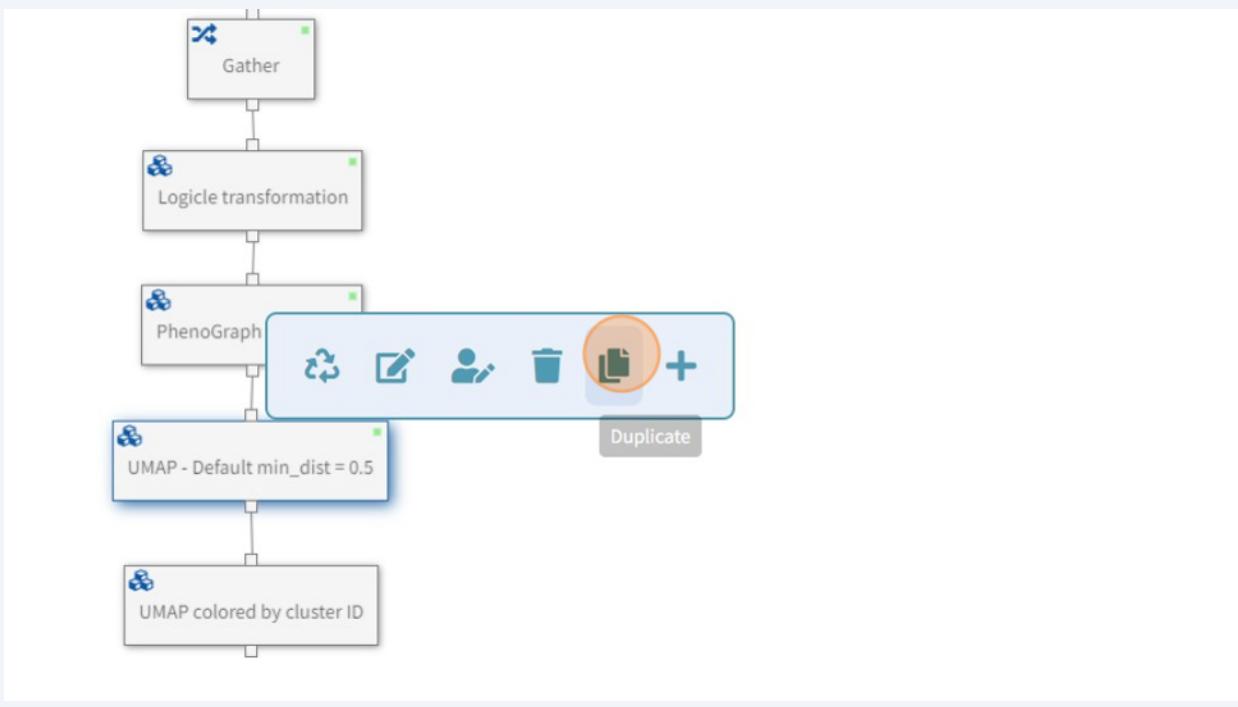
You will see the **default values** for the operator.

min_dist has a default setting of 0.5

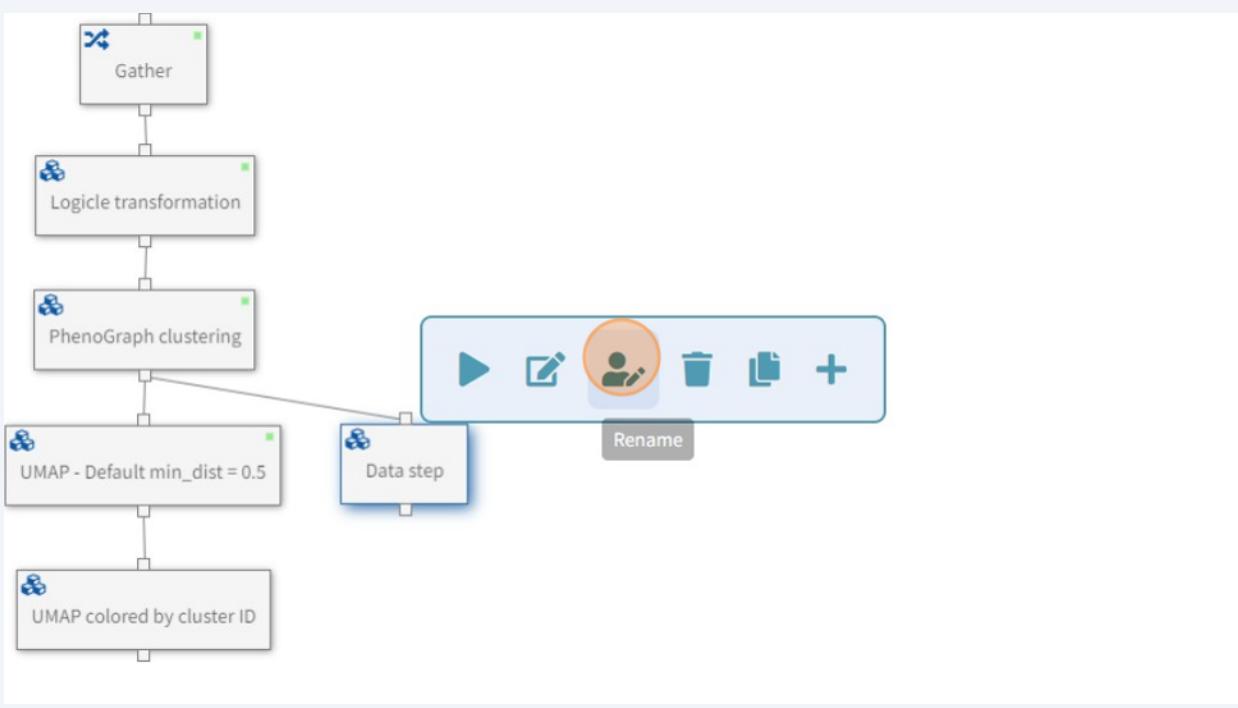


- 4 Go back to the workflow canvas and **Duplicate** this step.

That way, we do not have to re-create the same input projection as before.



- 5 **Rename** the step to reflect this change (**min_dist = 0.0001**).



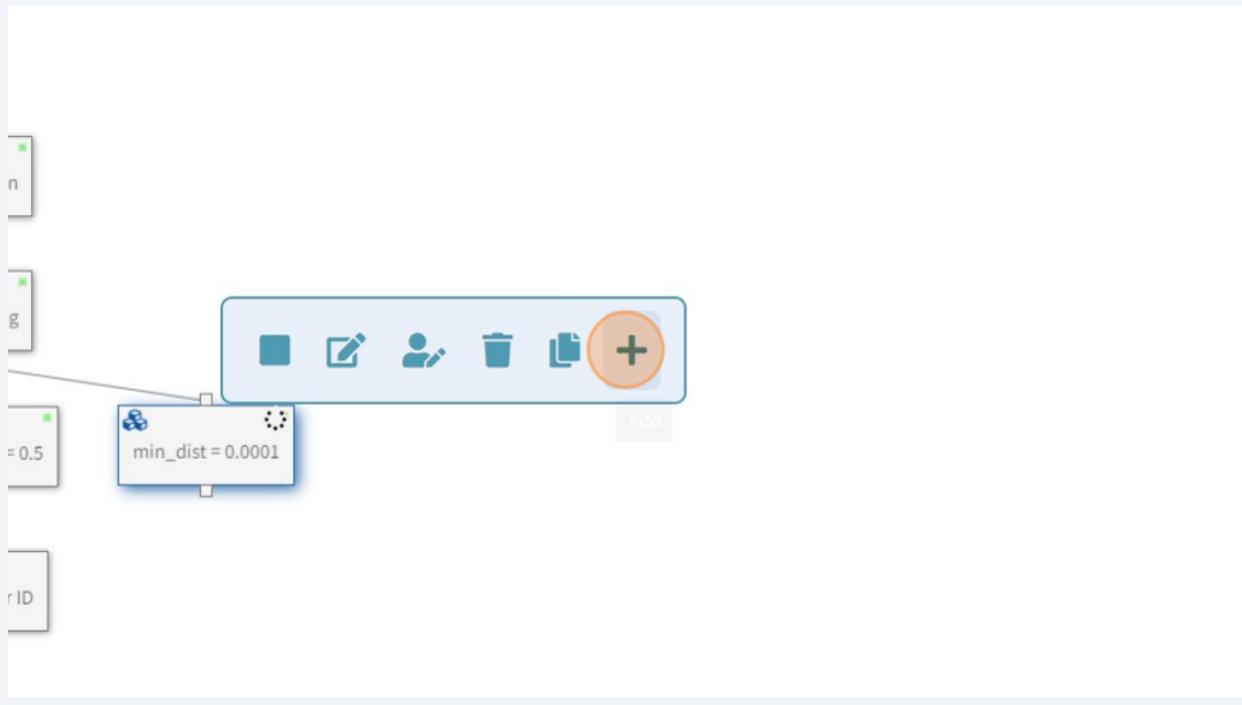
6 Edit the Data step

Set the **min_dist** value to **0.0001**.

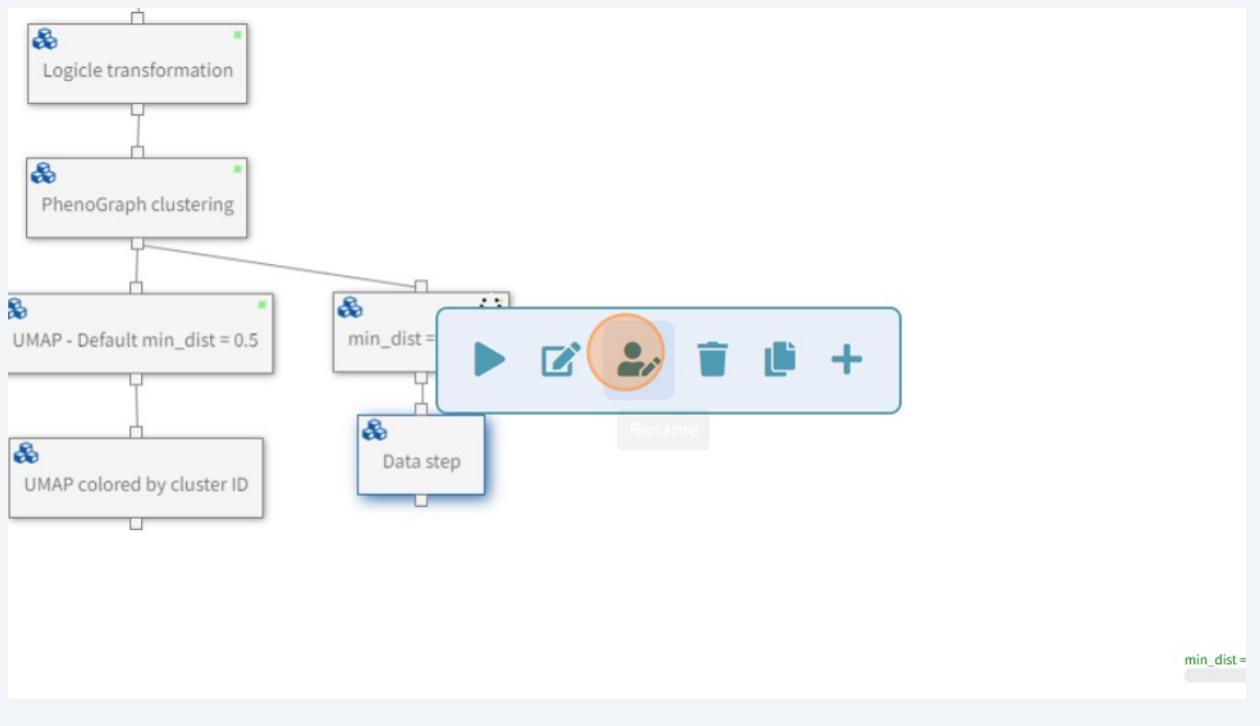
Then **Run** the operator.



7 Add a Data step to visualise the results.



8 Rename the Data Step to **View Results**

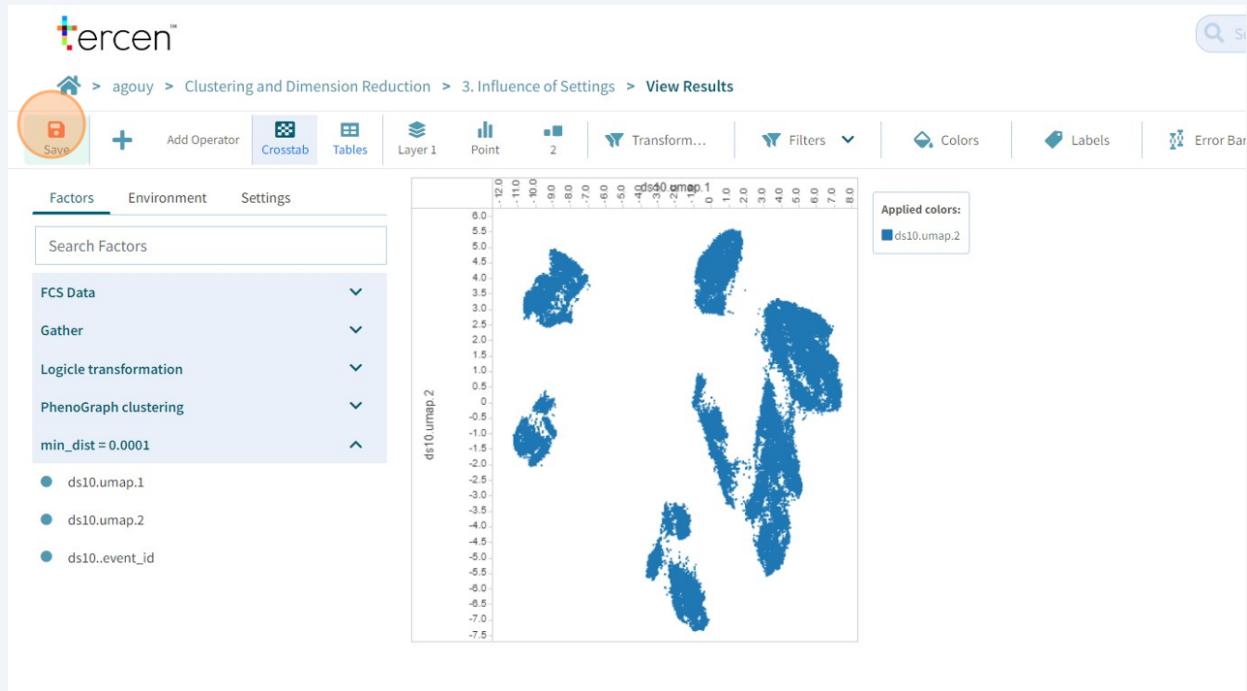


9 Remove the automatic factors with the **Clear Button**

Make this projection in the crosstab.

- **ds.10.umap.2** to the **y-axis**
- **ds10.umap.1** to the **x-axis**

What do you observe?



We observe a different results when we lower the **min_dist** setting.

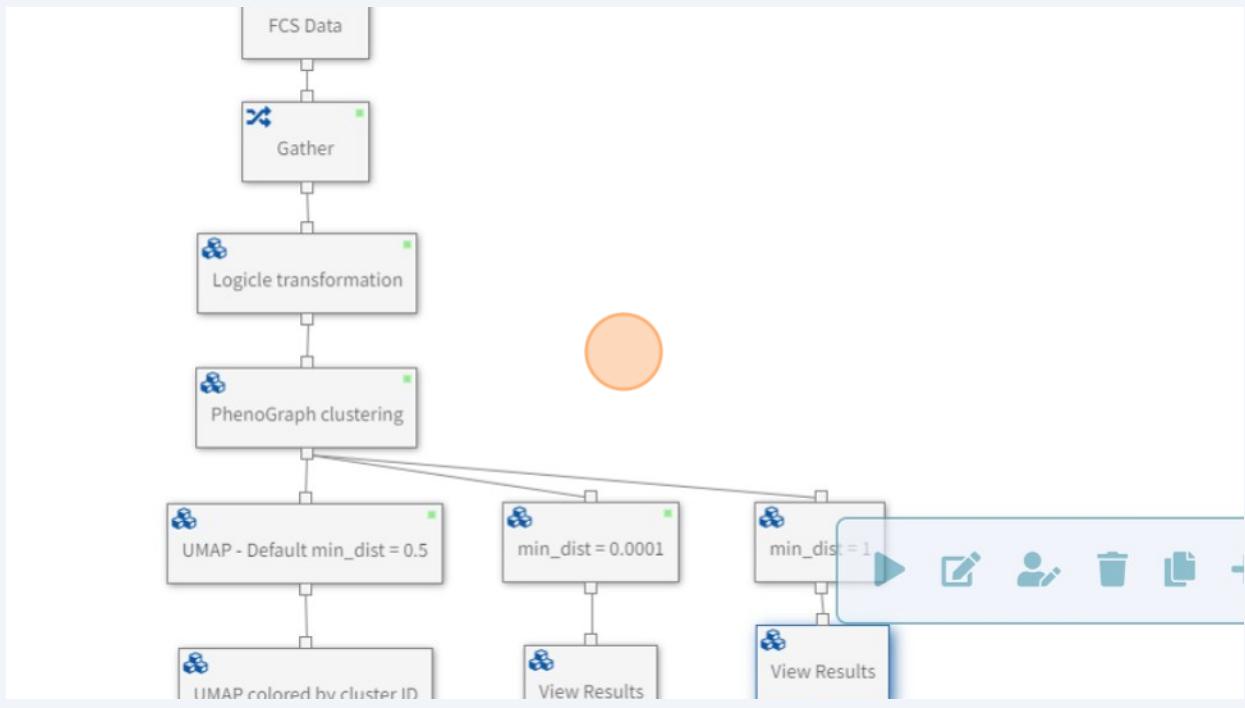
Similar data points are more **tightly packed** together.

10 Let's repeat the exercise, this time with a **min_dist** value of **1**.

Duplicate a Data Step, rename it, and modify the settings.

You should obtain a workflow as below.

What do you observe?



The **min_dist setting** has a strong influence on the UMAP results.

The **higher** it is, the more **stretchy** the data is on the UMAP.

This parameter is the **most important one** to know if the default settings do not produce a satisfying output.

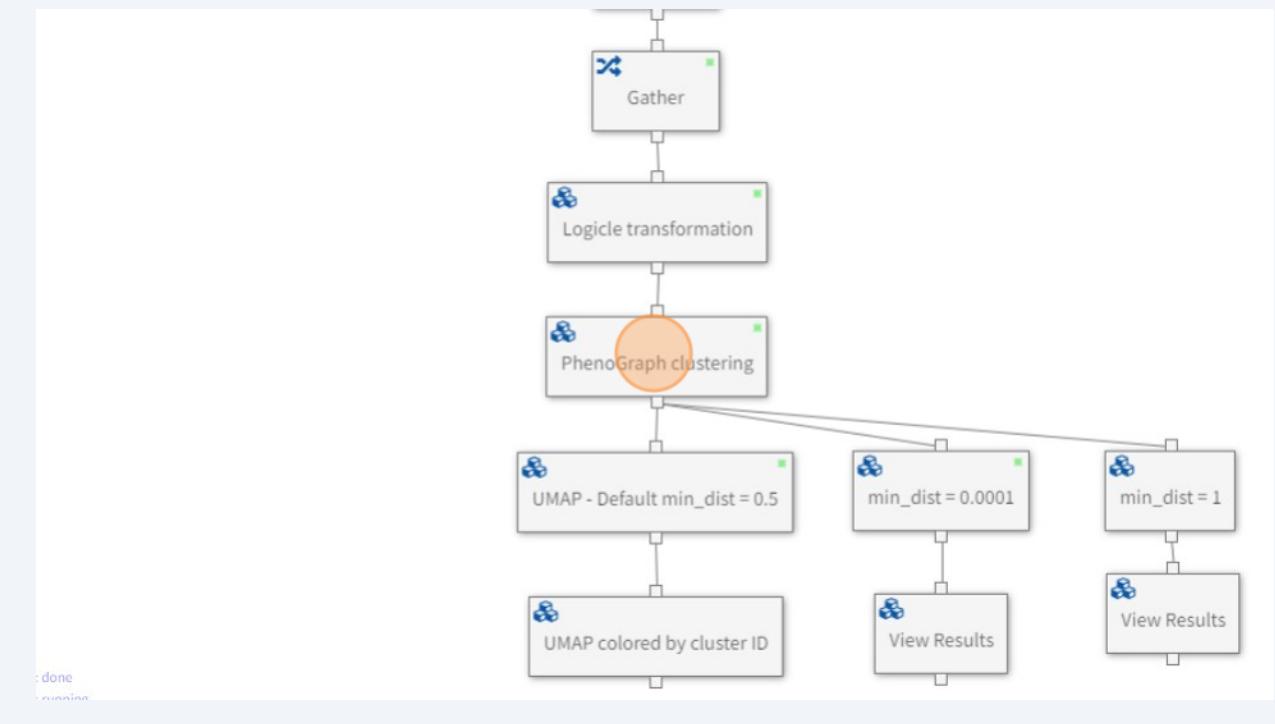
11

We will now see the influence of an important **PhenoGraph** setting called **k**

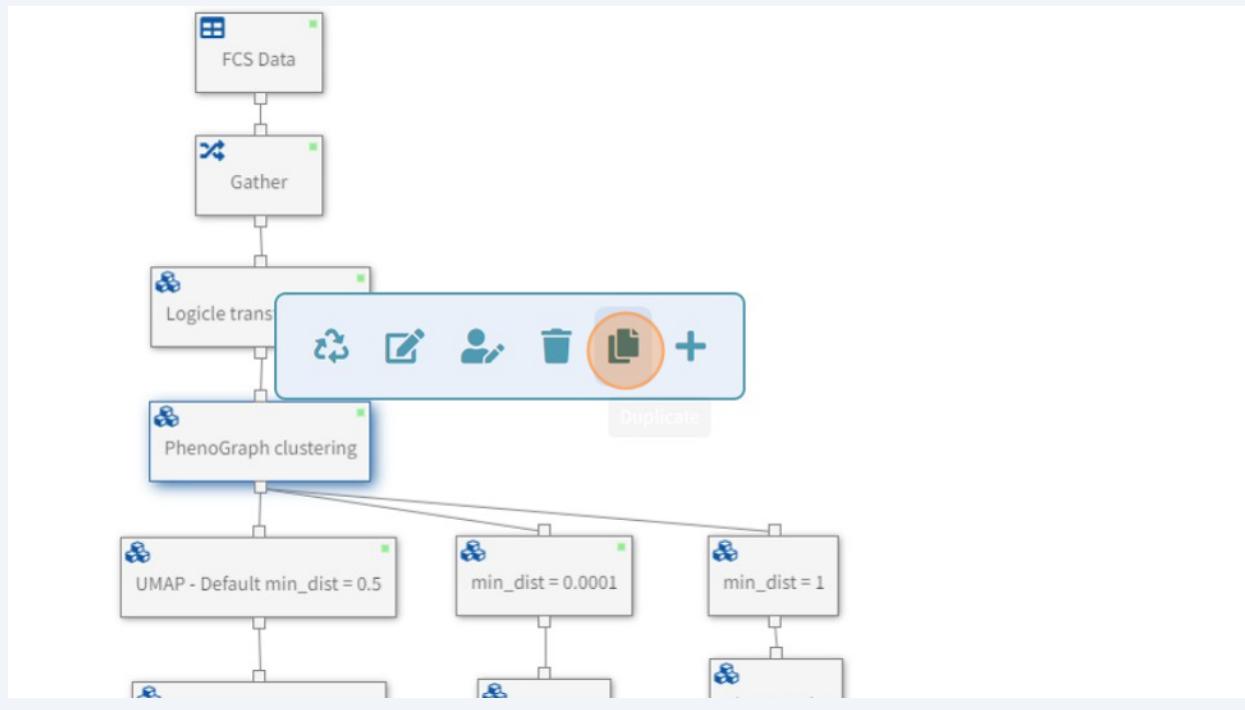
($k =$ the number of nearest neighbours).

Click and Edit the **PhenoGraph clustering** data step,

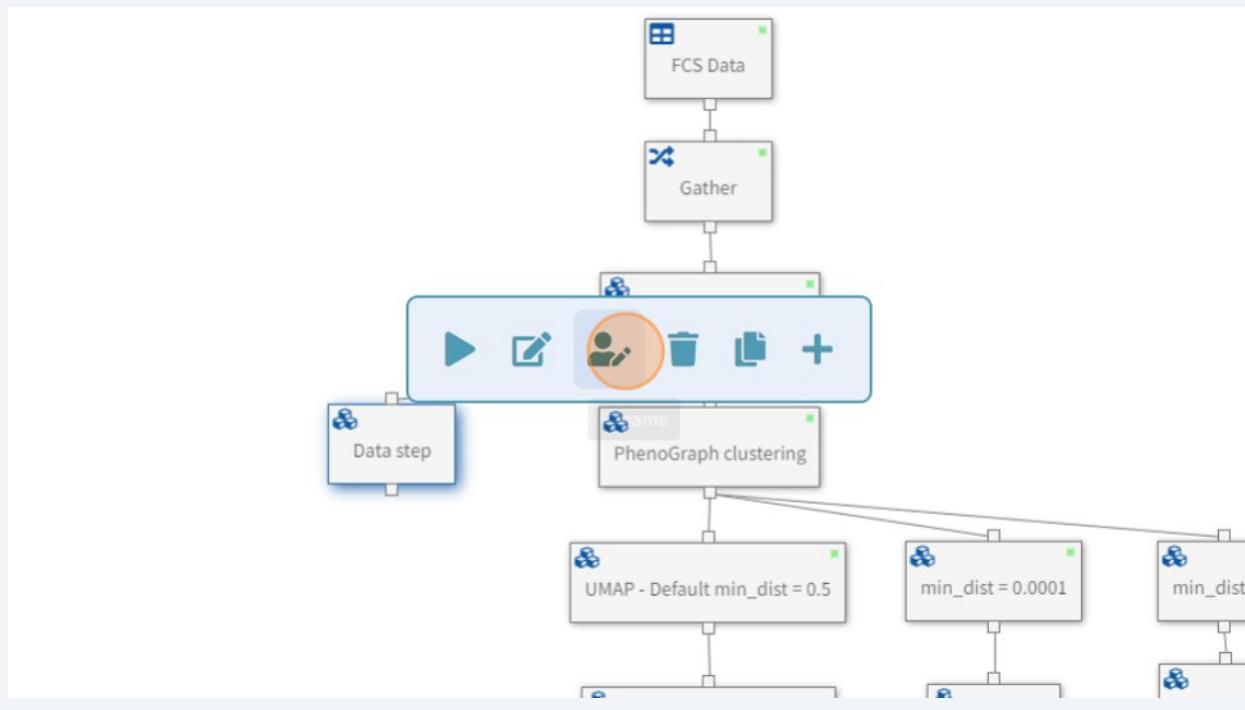
In the setting tab you can see that the **default k** value is set to **30**.



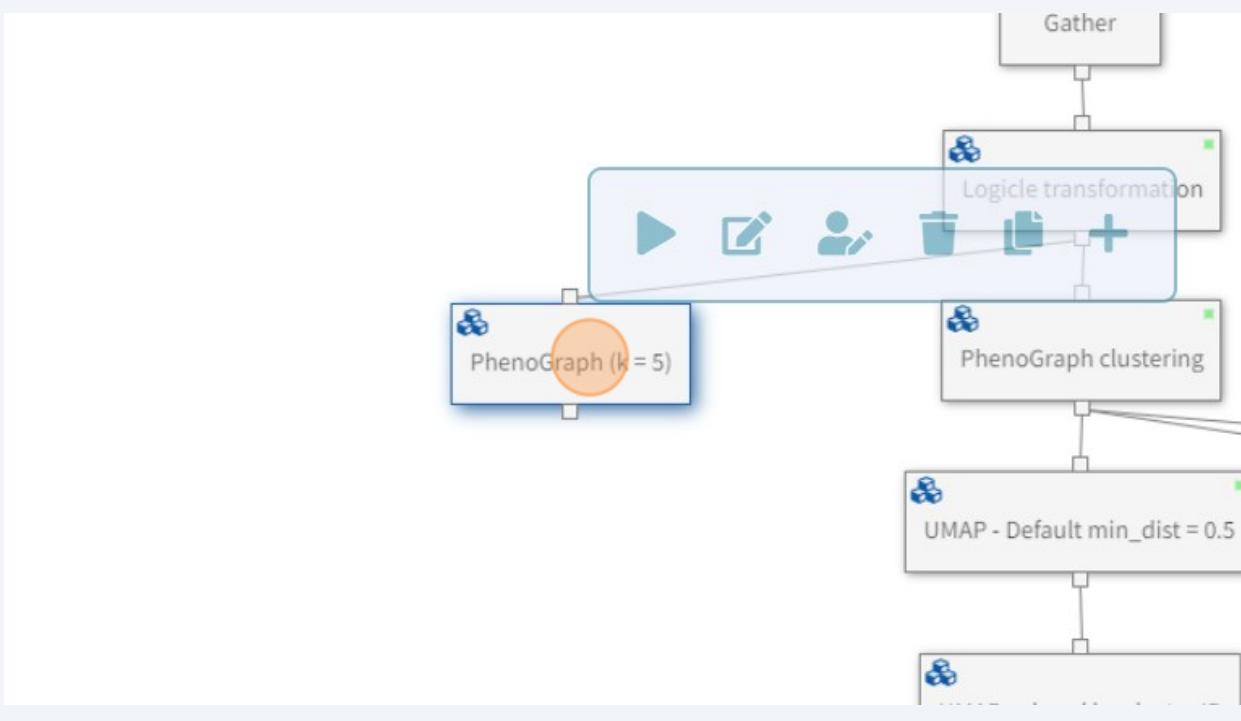
- 12 Return to the workflow canvas and **duplicate** this step to tweak this setting.



- 13 Rename the new step to **Phenograph (k = 5)**.



- 14 Edit the step to set the **k** value to **5** in the **Settings** tab.



- 15 Run the operator.

The screenshot shows the tercen™ software interface with the following details:

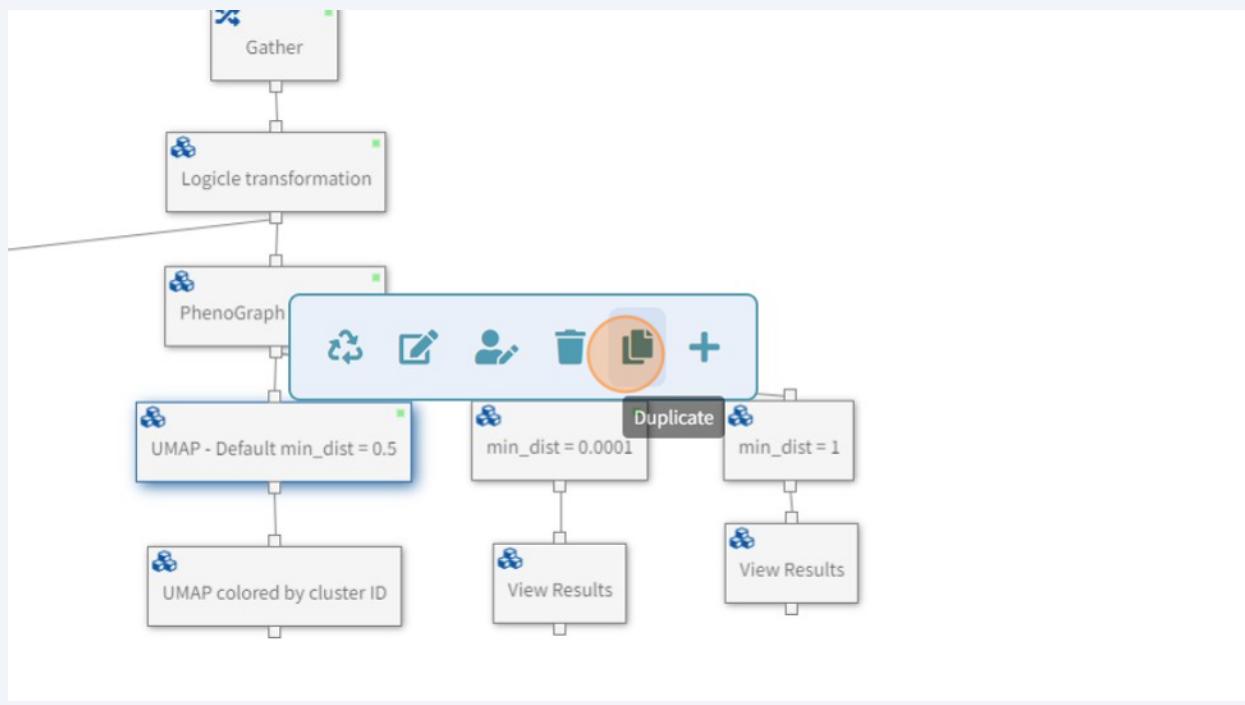
- Header:** tercen™, agouy > Clustering and Dimension Reduction > 3. Influence of Settings > PhenoGraph (k = 5)
- Toolbar:** Save, Remove, PhenoGraph 2.0.0, Run (highlighted with an orange circle), Crosstab, Tables, Layer 1, Point, Transform..., Filter.
- Settings Tab:** Factors, Environment, **Settings** (highlighted with a grey bar).
 - k:** 5
 - seed:** -1
 - implementation:** FastPG
- Scatter Plot:** A 2D scatter plot showing data points for 'ds0..event_id' (1) and 'ds0..variable' (CD11b). The y-axis ranges from 0.2 to 4.8.

16

We will also run **UMAP** after the newly-created Phenograph step.

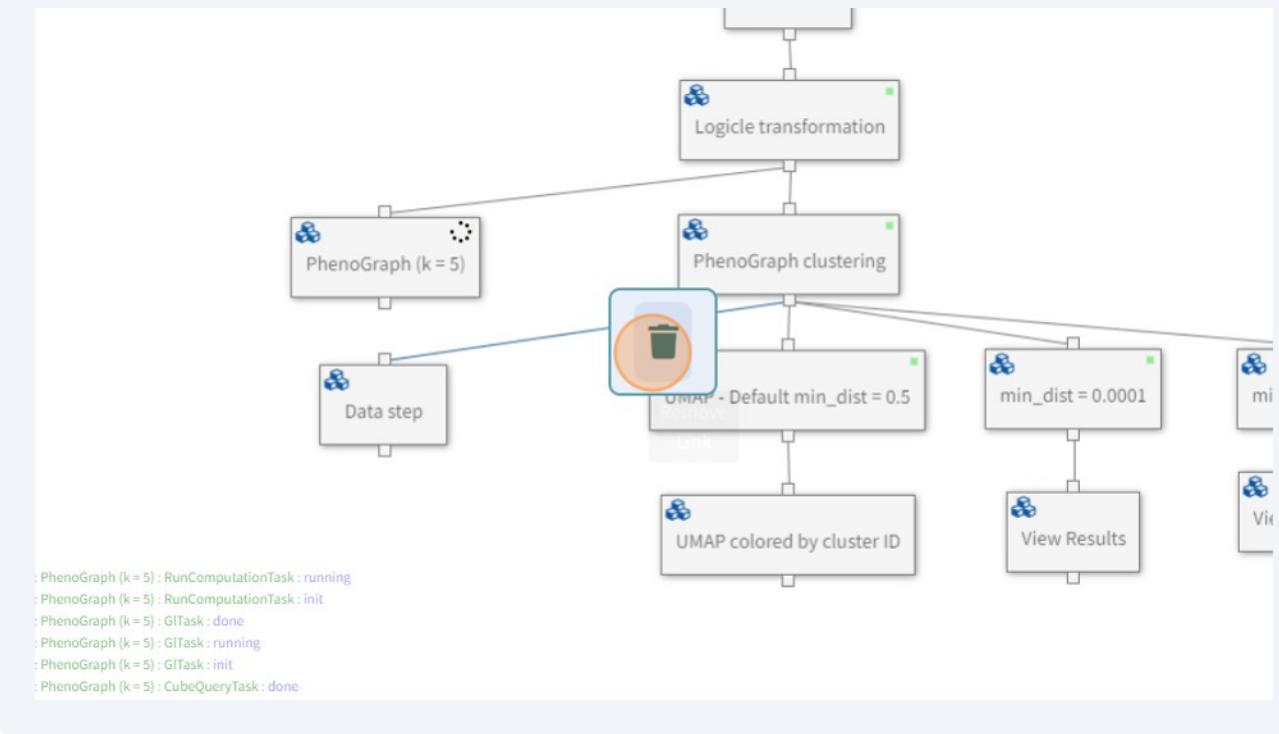
To do so, you can **Duplicate** the an existing UMAP step.

Pick the UMAP min_dist = 0.5

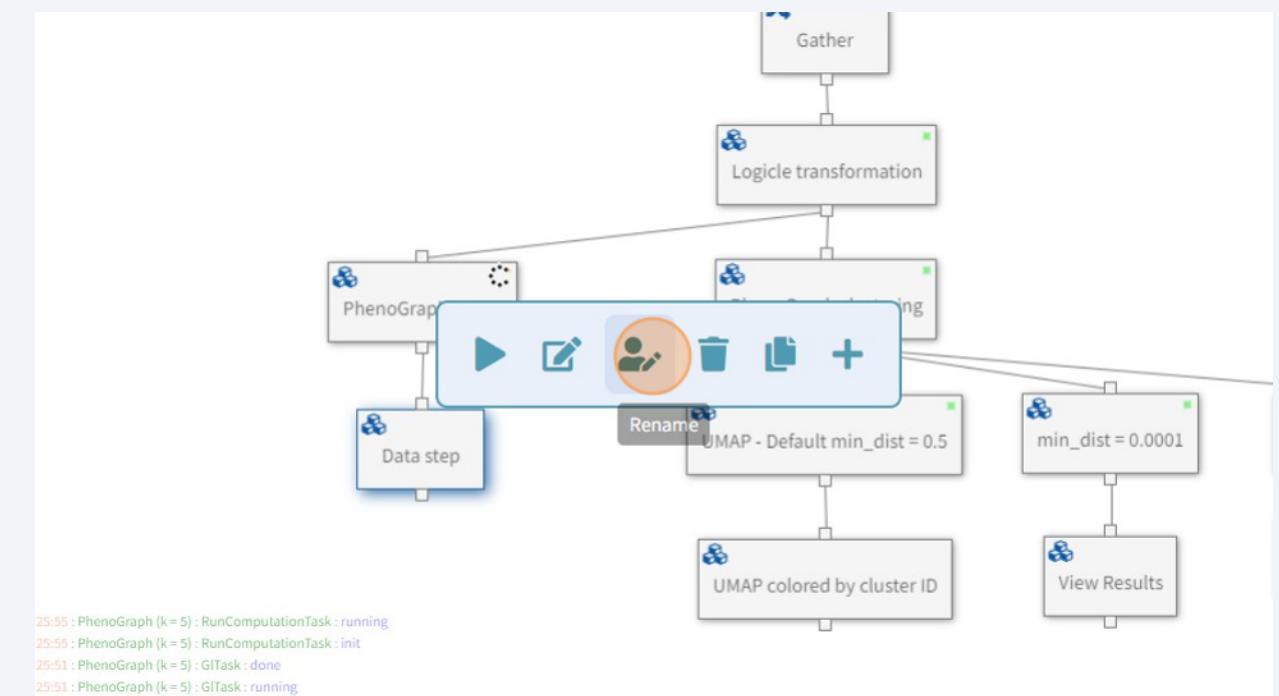


17 Click on the black line and Delete the existing link

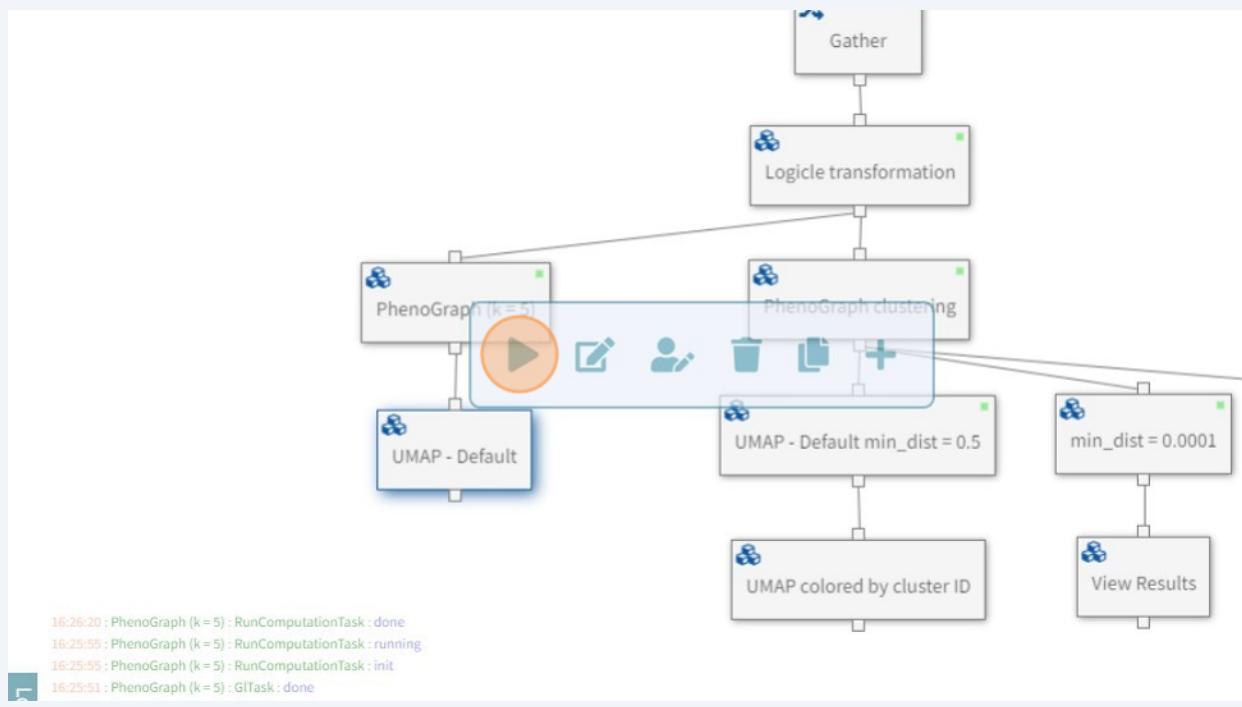
Build a new link by clicking the **node** at the bottom of the Phenograph (k=5) and then the node at the top of the new Data Step.



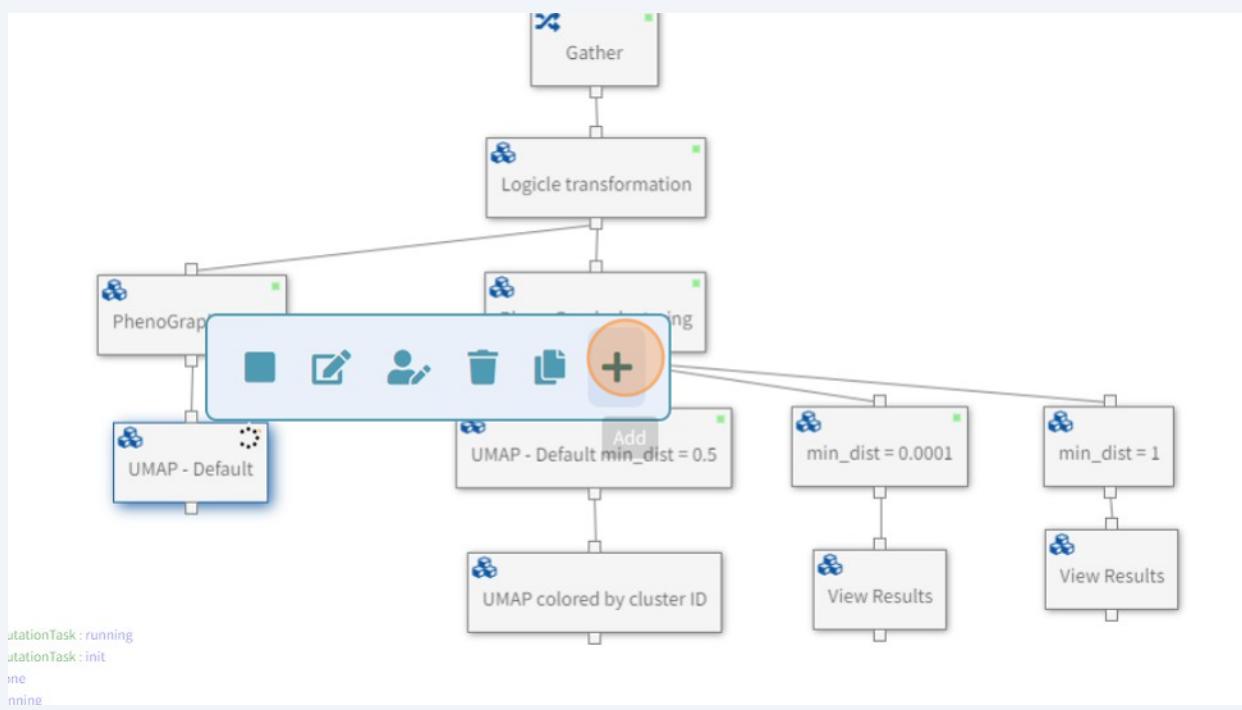
18 Name this step **UMAP - Default**



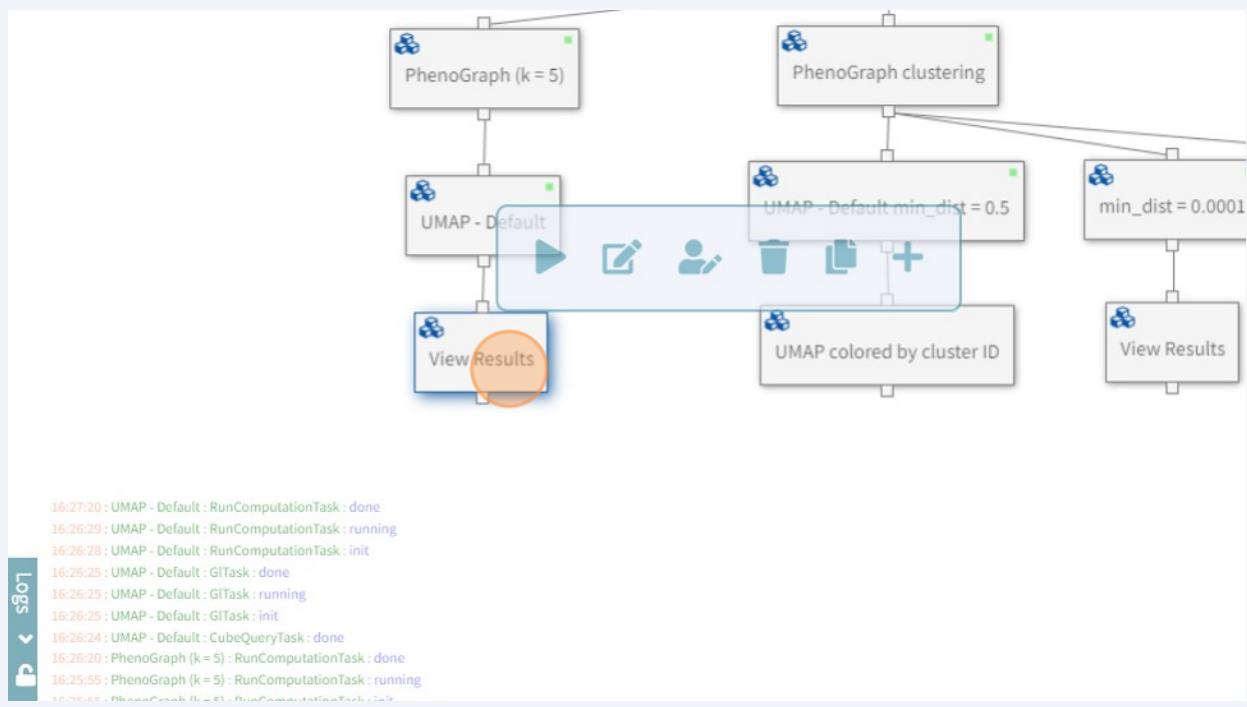
19 Run the step.



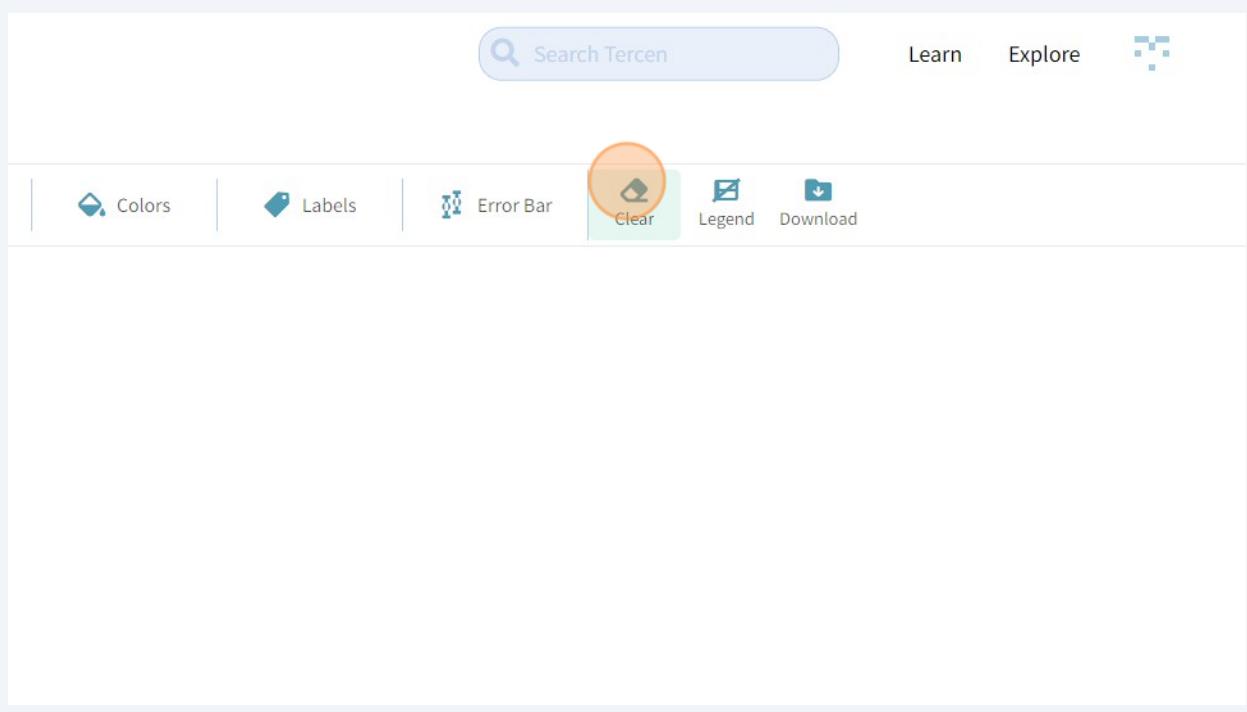
20 Once the computation has finished, you can **Add** a new **Data Step** to visualise the results.



21 Edit the data step.

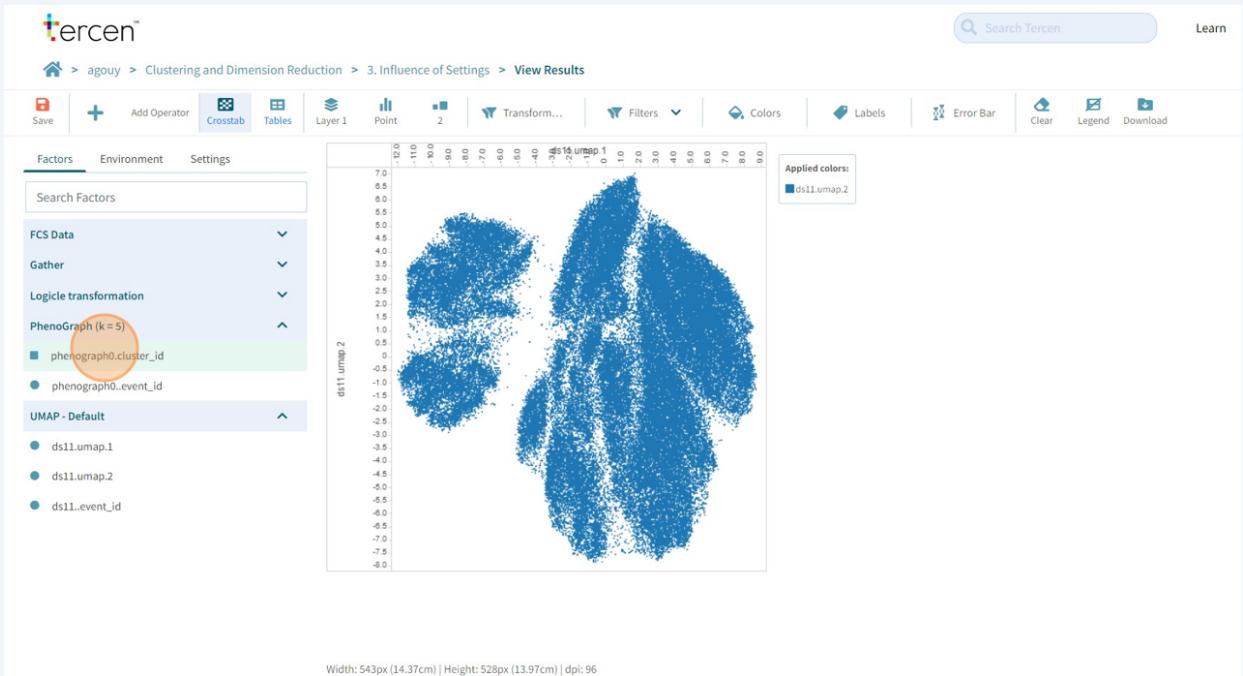


22 Clear the data step.



23 Project the UMAP dimensions

- **ds11.umap.2 to y-axis**
- **ds11.umap.1 to x-axis**
- **phenograph0.cluster_id to Colors.**

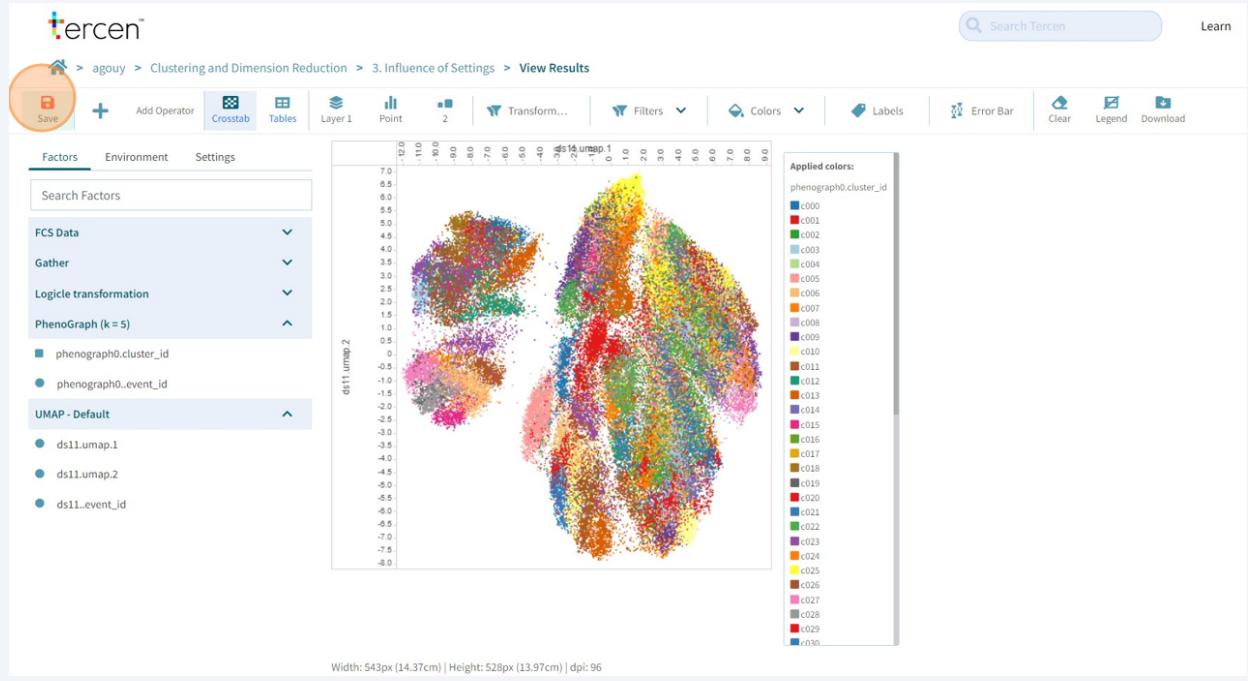


24

You can now see the PhenoGraph **clusters** that were generated with a **smaller k** value.

What do you observe?

Do not forget to **Save** your workflow!

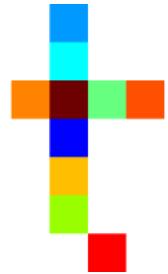


The **k** (number of nearest neighbours) setting value of the PhenoGraph algorithm has an influence on the size and number of clusters.

The **smaller k** is, the **smaller** the **clusters** will be.

It can be useful to tweak this setting if you think the algorithm is **over-** or **under-clustering** your data.

0304 - Influence of Algorithms



- 1 Navigate to your workshop project.

Clone the 1. Clustering and dimension workflow

Save it to your project and name it **4. Influence of Algorithms**.

The screenshot shows the Agouy workshop interface. At the top, there's a header with a user icon and the text "agouy". Below the header, there are two tabs: "Project" (which is selected) and "Activities". Under the "Project" tab, there's a section titled "Clustering and Dimension Reduction" with a note: "No description provided.". Below this, there's a list of files and workflows:

File/Workflow	Last Modified
agouy updated workflow 4. Influence of Algorithms	22 seconds ago
README.md	1 days ago
Melanoma_AntiPD1_Sample_annotation.csv	6 hours ago
Melanoma_AntiPD1_Sample_annotation.csv	3 hours ago
Melanoma_AntiPD1_Baseline_FCS.csv	1 days ago
dev	3 hours ago
3. Influence of Settings	51 minutes ago
2. Cluster exploration	1 hours ago
1. Clustering and dimension workflow	6 hours ago

In the center of the dashboard, there's a circular orange button with three small icons: a cloud, a gear, and a trash can. At the bottom right, there's a red circular button with a white icon.

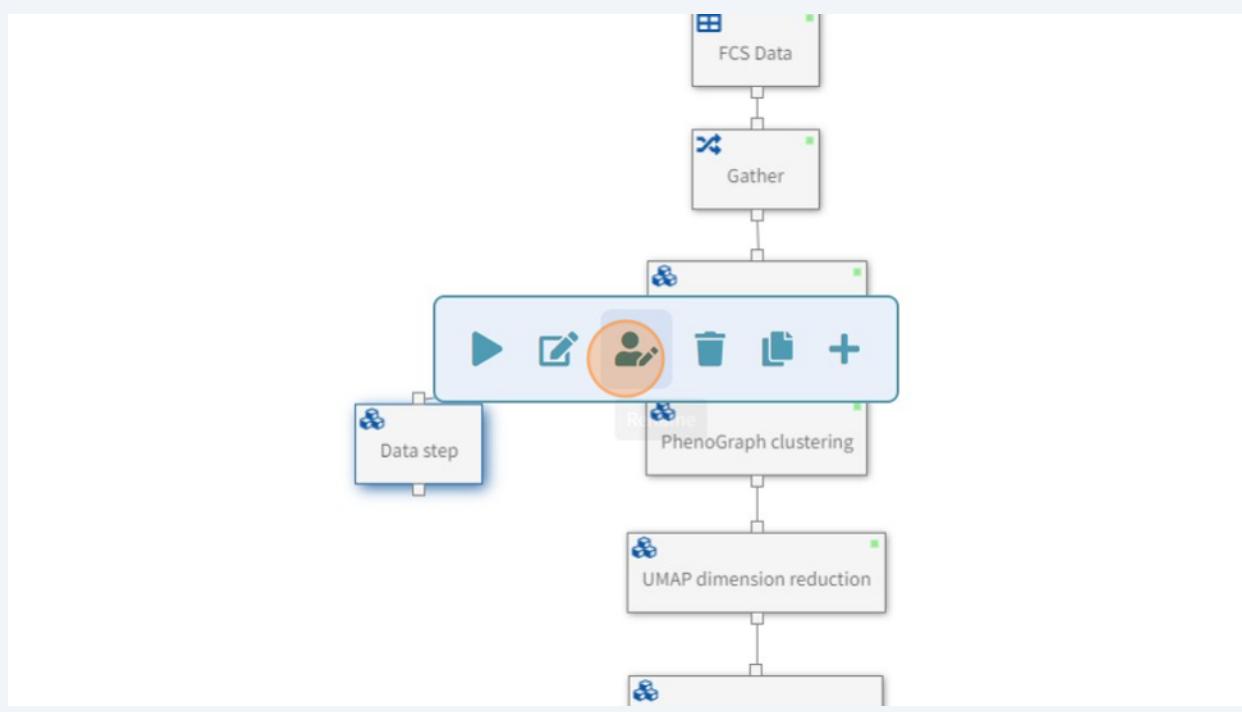
- 2 We will try an **alternative** to the **PhenoGraph** clustering algorithm.

FlowSOM is a popular algorithm in Flow Cytometry analysis.

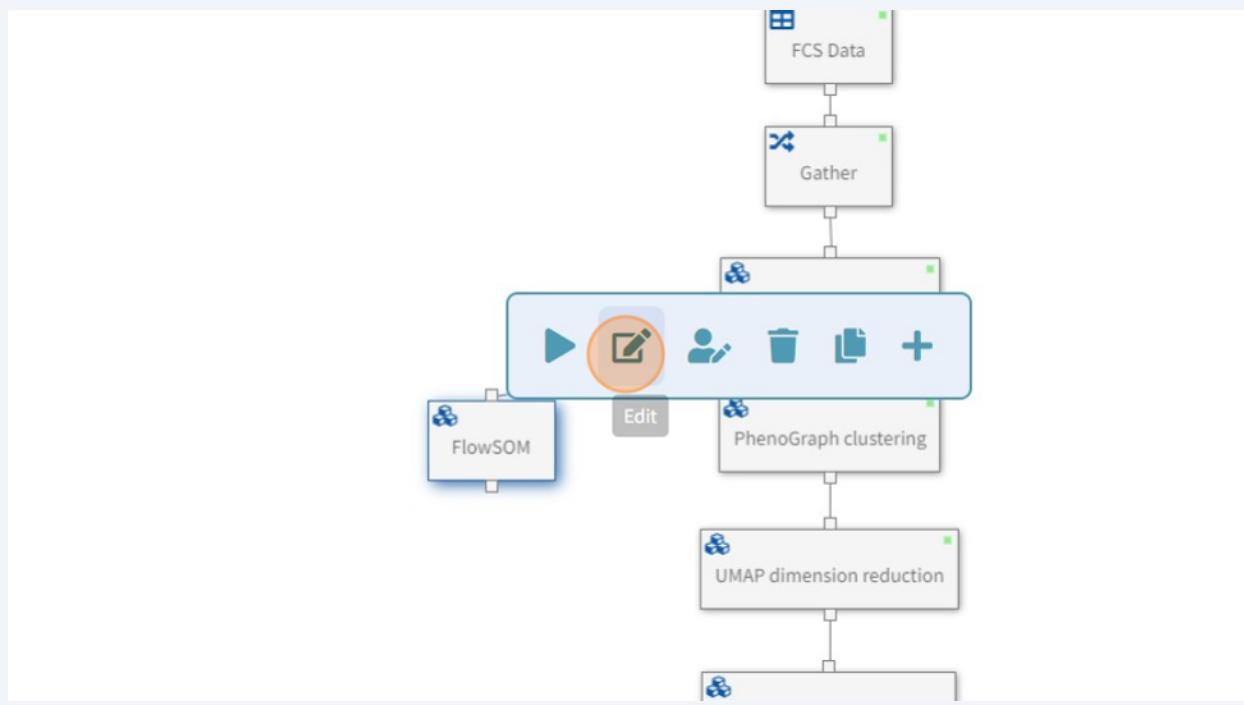
Duplicate the Phenograph clustering step.



- 3 Name the new step **FlowSOM**.



4 Let's **Edit** the data step.



5 Remove the PhenoGraph **operator**.

The screenshot shows the tercen™ software interface. At the top, there is a navigation bar with a home icon, the user name 'agouy', and several menu items: 'Clustering and Dimension Reduction', '4. Influence of Algorithms', and 'FlowSOM'. Below the navigation bar is a toolbar with various icons: Save, Remove (highlighted with an orange circle), PhenoGraph 2.0.0, Run, Crosstab, Tables, Layer 1, Point, Transform..., and Filter. The main area is divided into two sections: 'Factors' and 'Environment'. The 'Environment' tab is selected. It contains several configuration parameters:

Namespace	Value
flowsom	flowsom
cpu	0
ram	0
cpu-shares	0
tercen.collect.stats	false
debug	false

To the right of the environment settings is a data visualization section. It features a scatter plot with axes labeled 'ds0..event_id' (ranging from 1 to 4) and 'ds0..variable' (ranging from 0.2 to 4.6). The plot shows a single data point at approximately (1, 1.4) labeled 'CD11b'.

6 Press the plus button

Replace it by the **FlowSOM operator** you will find in the **Library**.

Installed Library Log

flowsom

Display latest version only

▲ Tag list

FlowSOM 1.6.0
Flow Cytometry Self Organising Maps: An algorithm that clusters cells based on chosen channels and maps them into a 2D space.
flow cytometry | clustering

FlowSOM MST 0.0.3
Generate plots of FlowSOM Minimum Spanning Trees.
data visualisation

cytonorm 1.2.0

- 7 Click the **Environment** tab.

Edit the **namespace** setting to **flowsom**

The screenshot shows the tercen™ software interface with the following details:

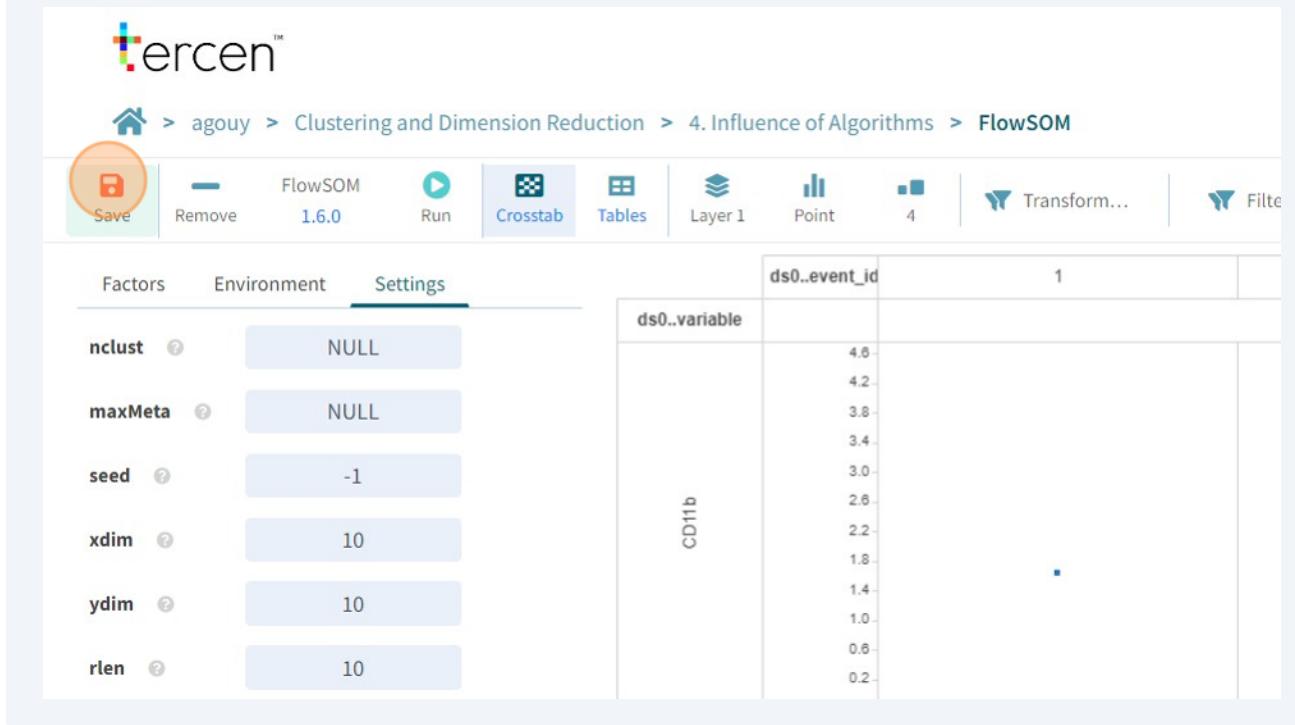
- Top Bar:** tercen™ logo, navigation path: Home > agouy > Clustering and Dimension Reduction > 4. Influence of Algorithms > FlowSOM, toolbar with Save, Remove, Run, Crosstab (selected), Tables, Layer 1, Point, 4, Transform..., and Filter.
- Environment Tab:** Active tab, showing configuration settings:
 - Namespace:** flowsom (highlighted with an orange circle)
 - cpu:** 0
 - ram:** 0
 - cpu-shares:** 0
 - tercen.collect.stats:** false
 - debug:** false
- Table View:** A data table titled "ds0..variable" with one row. The row contains "ds0..event_id" and the value "1". To the left of the table, the identifier "CD11b" is visible.

8

Click the **default settings**.

FlowSOM has more settings than PhenoGraph

The one to note is the **number of clusters (nclust)**. By default, FlowSOM will estimate the "best" number of clusters in the data. If the results are not satisfying, the number of clusters to estimate can be fixed by tweaking this setting.

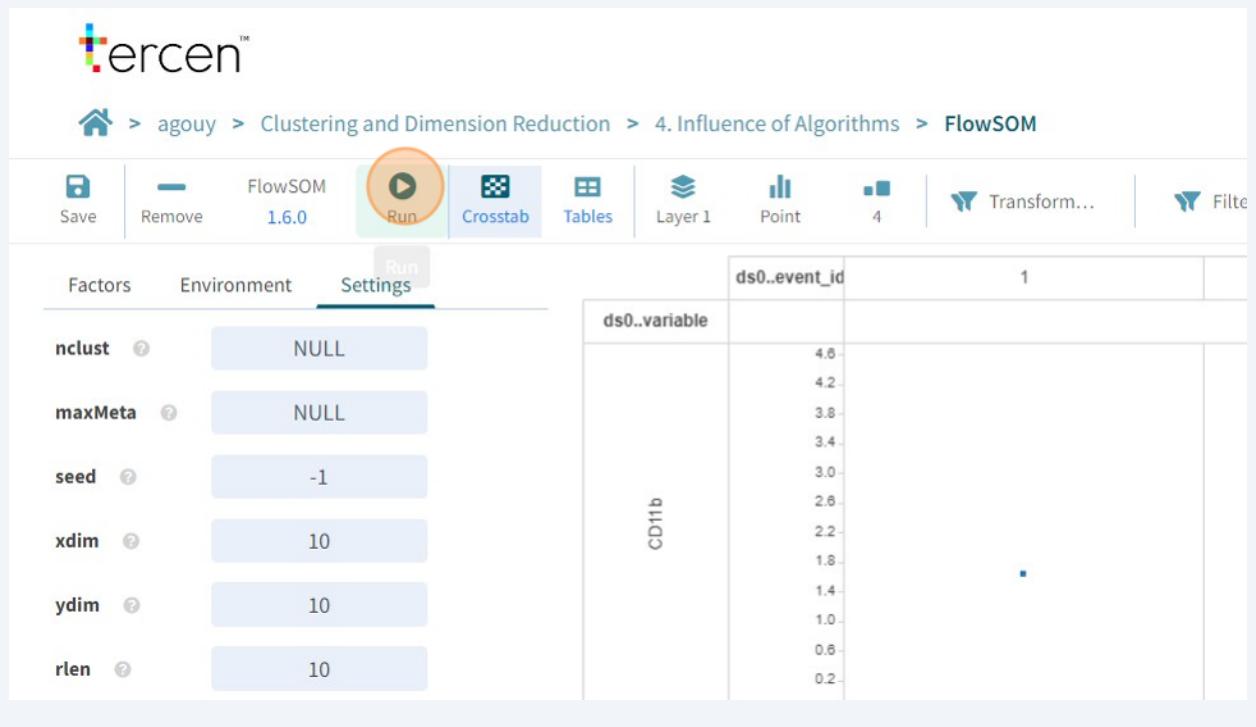


FlowSOM's nclust setting

Estimating the **right number** of clusters when performing a FlowSOM analysis is an **iterative** process.

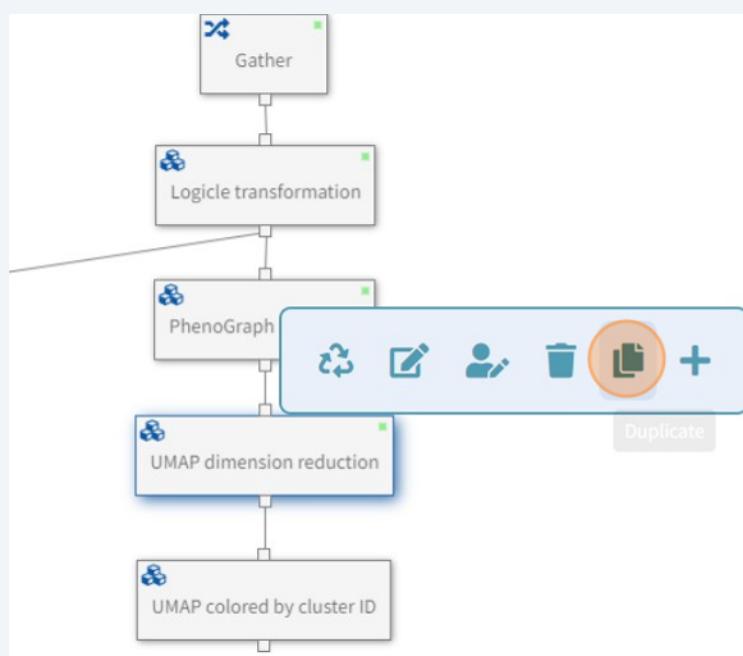
One would typically tweak the setting and **check the quality** of the estimated clusters. In particular, we want to avoid **underclustering** our results, which would mean that we have grouped different cell populations together.

9 Run the operator.



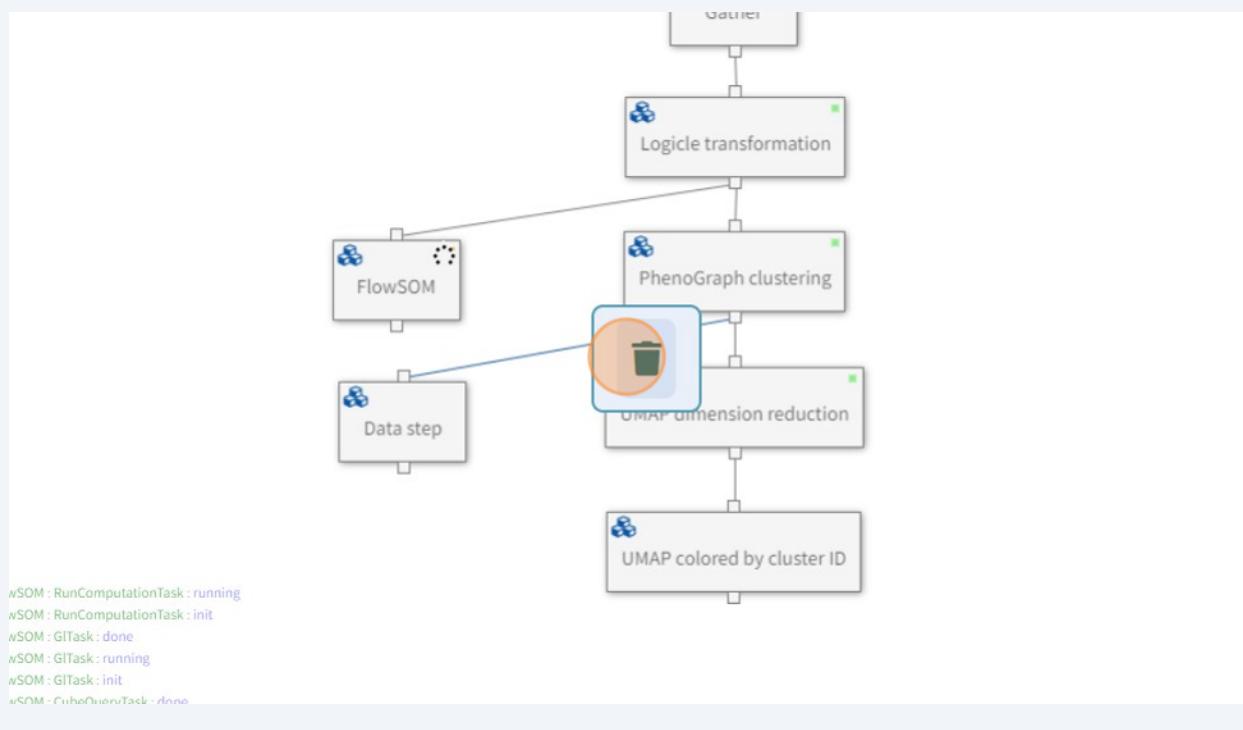
10 Go back to the workflow canvas.

Duplicate the UMAP step

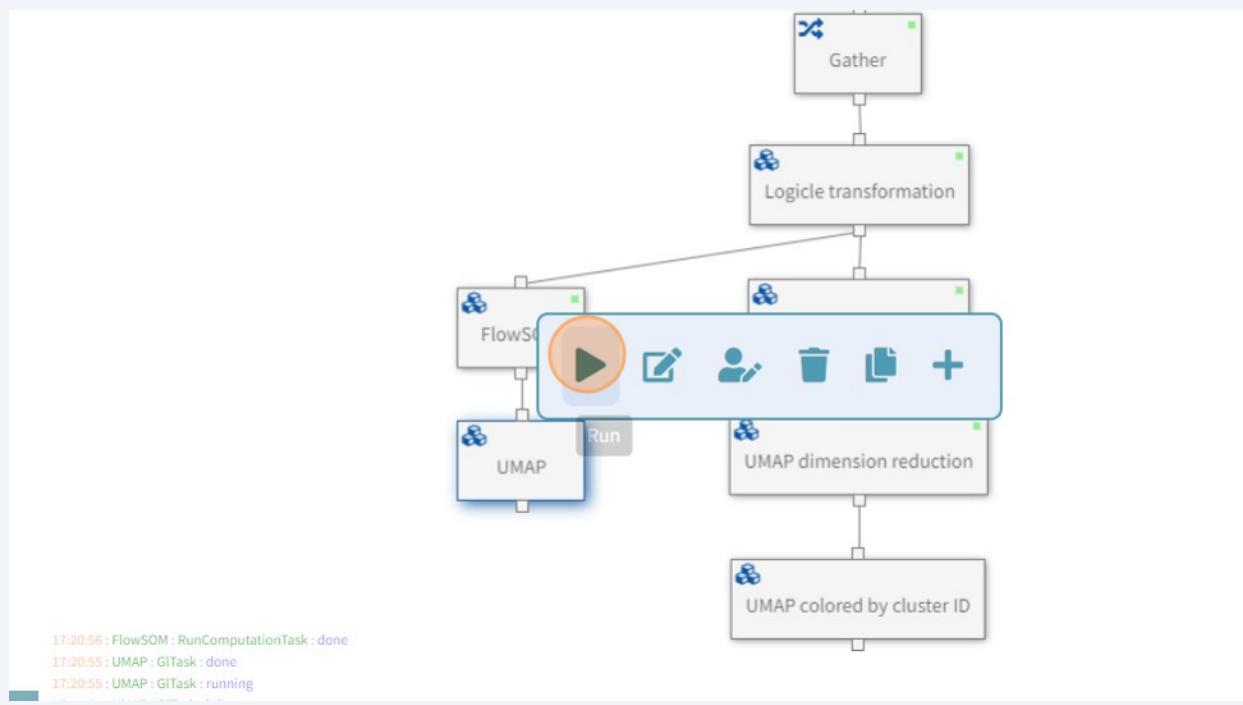


11 Delete the existing link to Phenograph

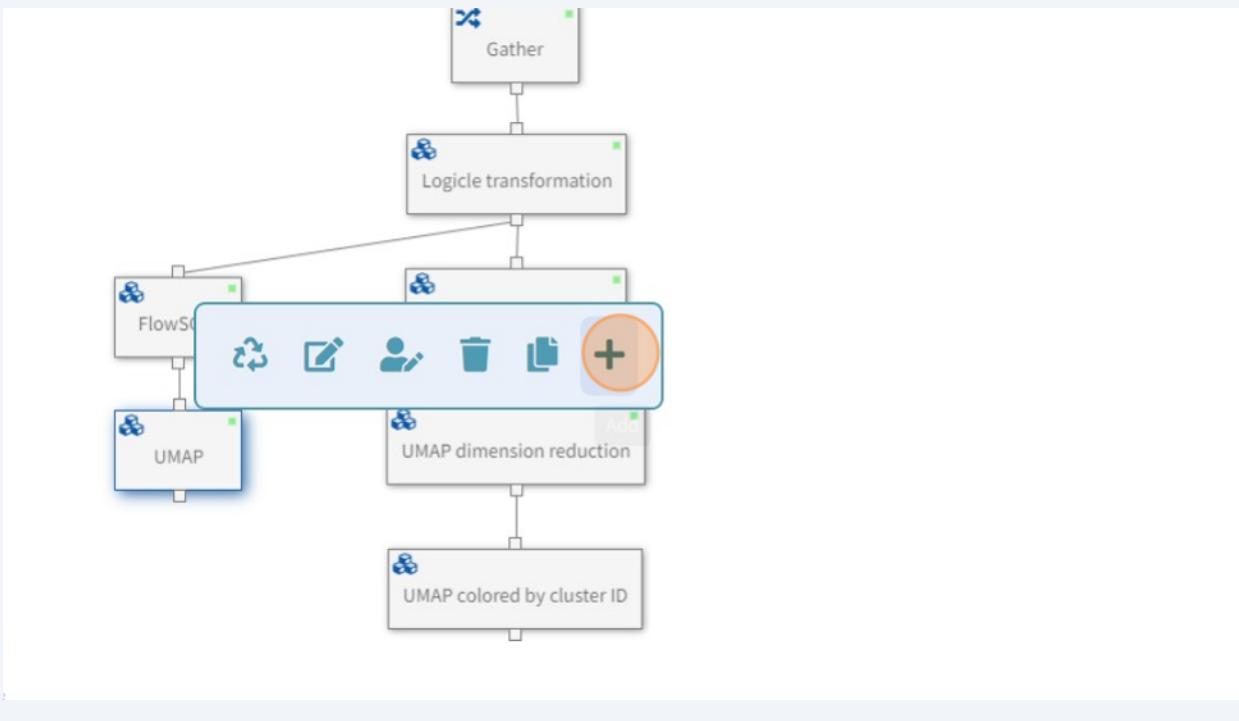
And **create a new one** to FlowSOM.



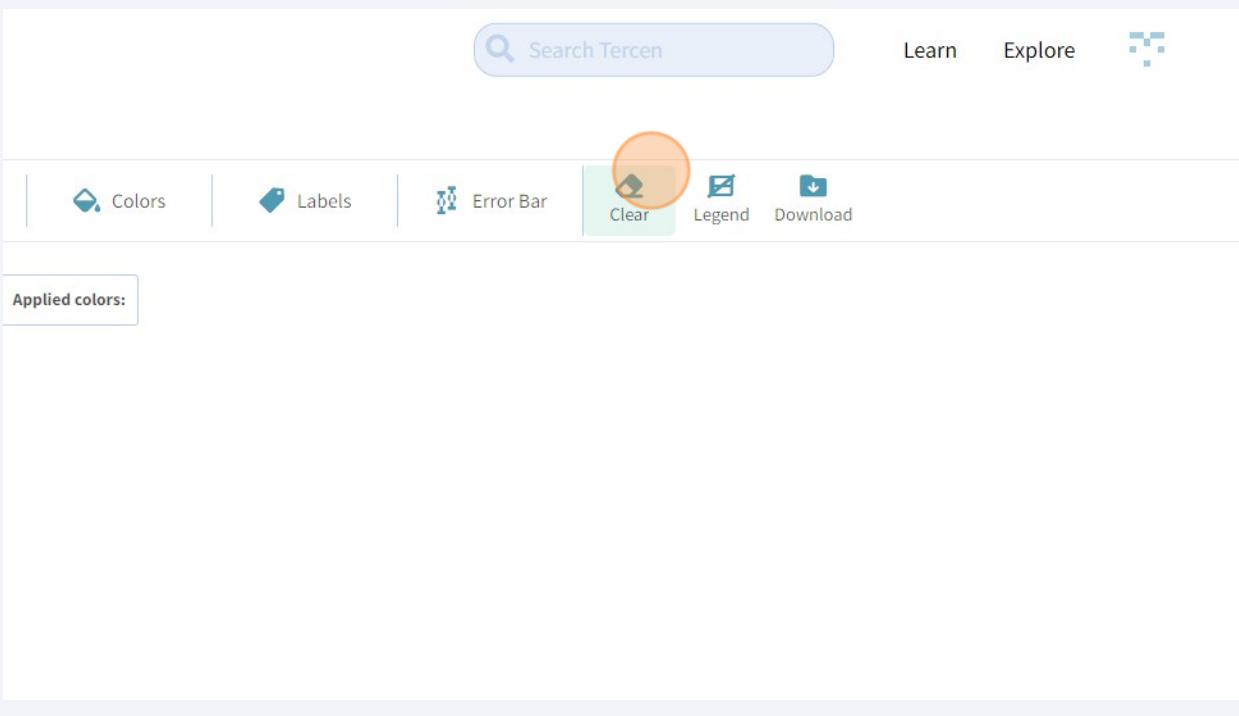
12 Run the UMAP.



13 After the computation has finished, **add a data step** to visualise the results.



14 **Clear** the crosstab.



15 Project the two UMAP dimensions

- **ds10.umap.1 to x-axis**
- **ds10.umap.2) to y-axis**
- **flowsom.metacluster_id" to Colors**

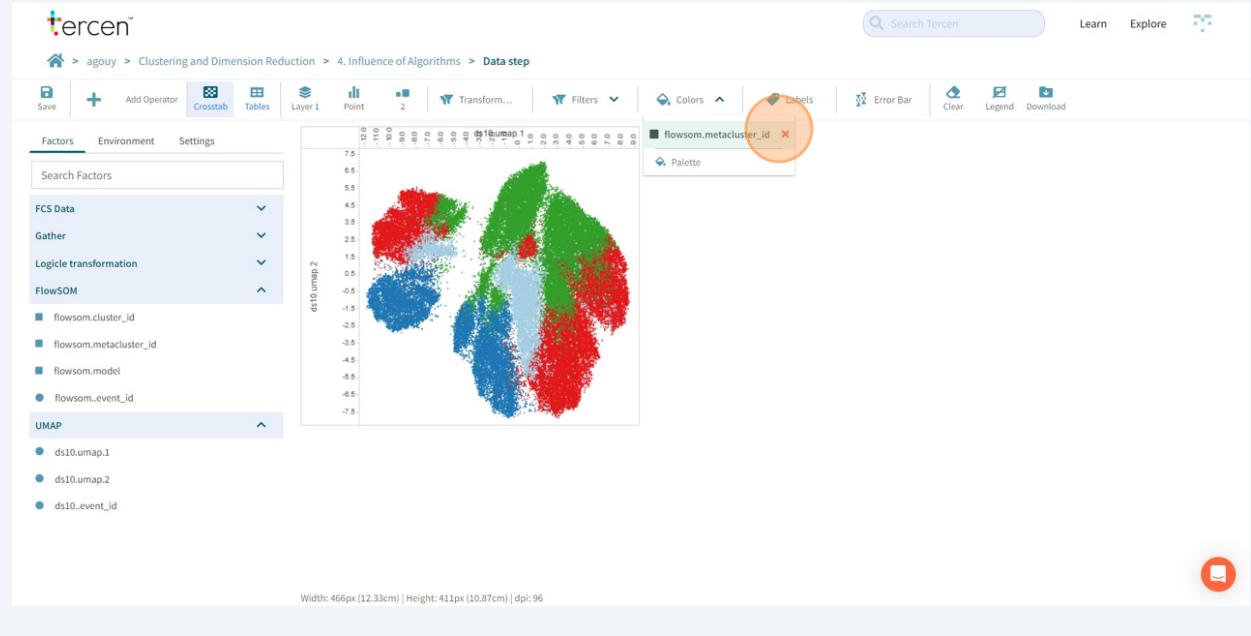
What do you observe?



16 Remove **flowsom.metacluster_id** from colors

Replace it with **flowsom.cluster_id**.

What do you observe?



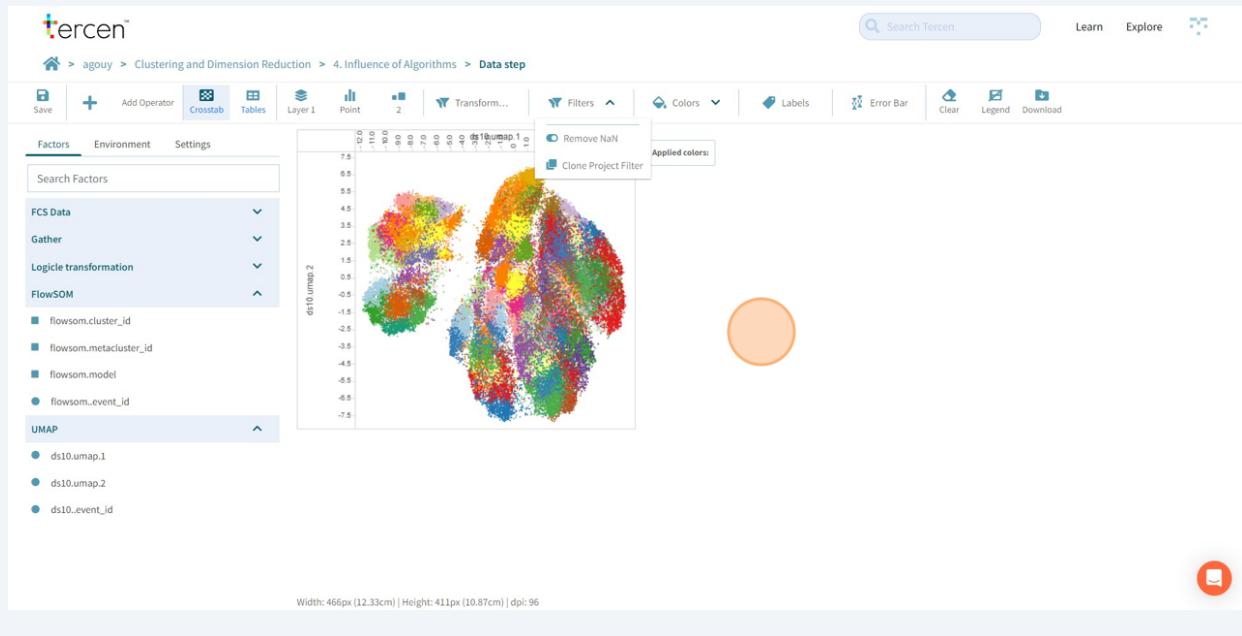
FlowSOM estimates **clusters**, and then group them together based on their similarity (**metaclusters**).

Here, we see that FlowSOM has produced some **very large metaclusters** by default (underclustering). On the other hand, the **clusters** it estimates are **quite small**.

In a real-life situation, it would be worth increasing the number of estimated metaclusters until it provides satisfying results.

17

We have seen that different clustering algorithms (FlowSOM and Phenograph) can produce different results on the same flow cytometry data.

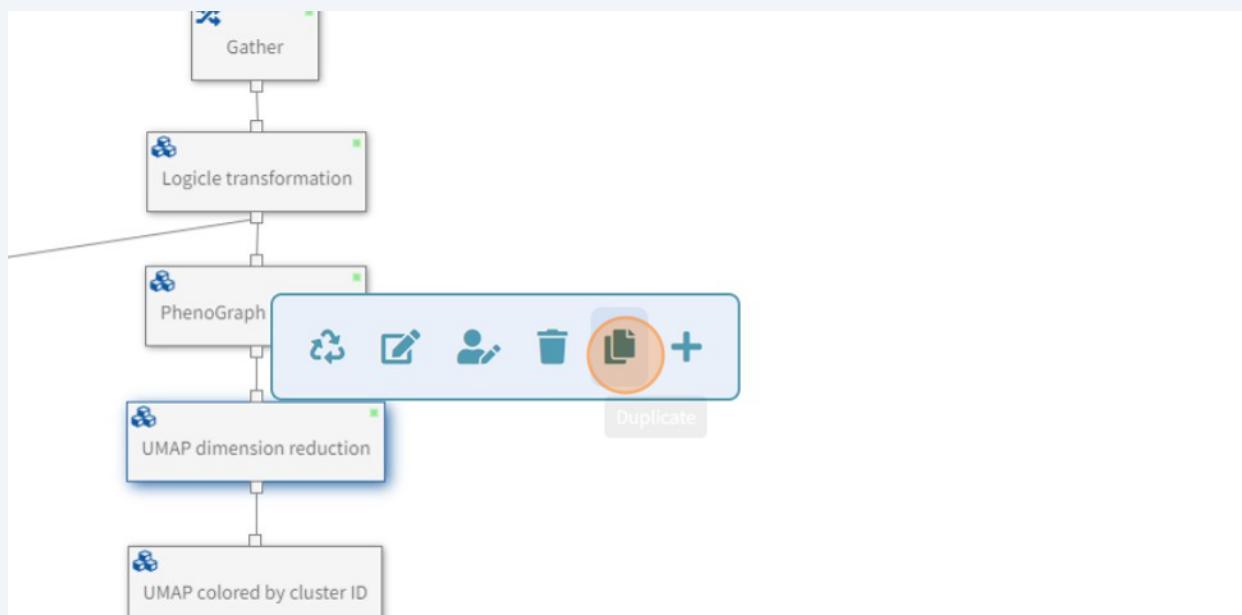


18

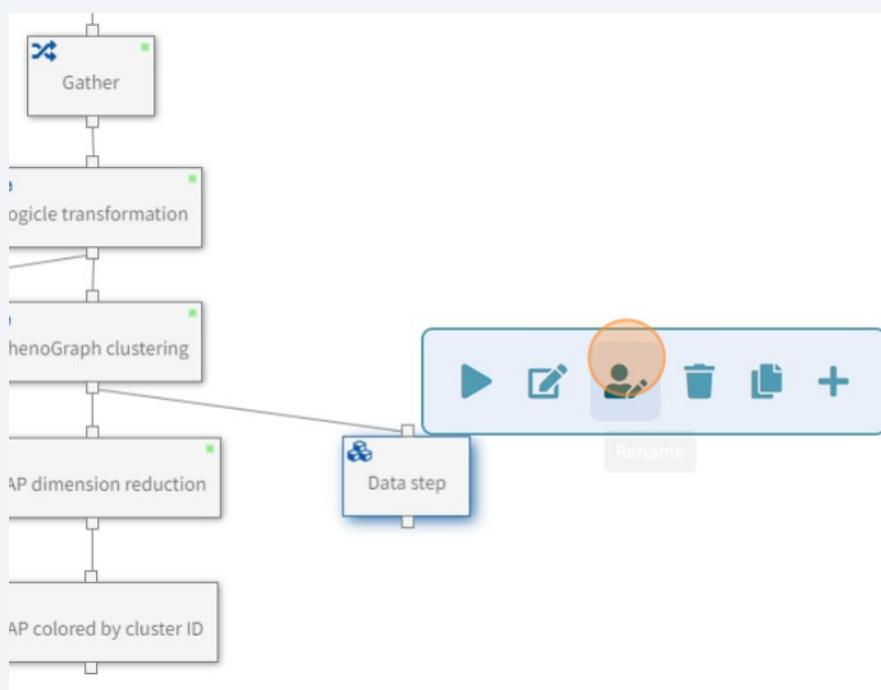
Now lets explore how different dimension reduction algorithms can represent the same data.

t-SNE is a commonly used alternative to UMAP.

Duplicate the original UMAP step.



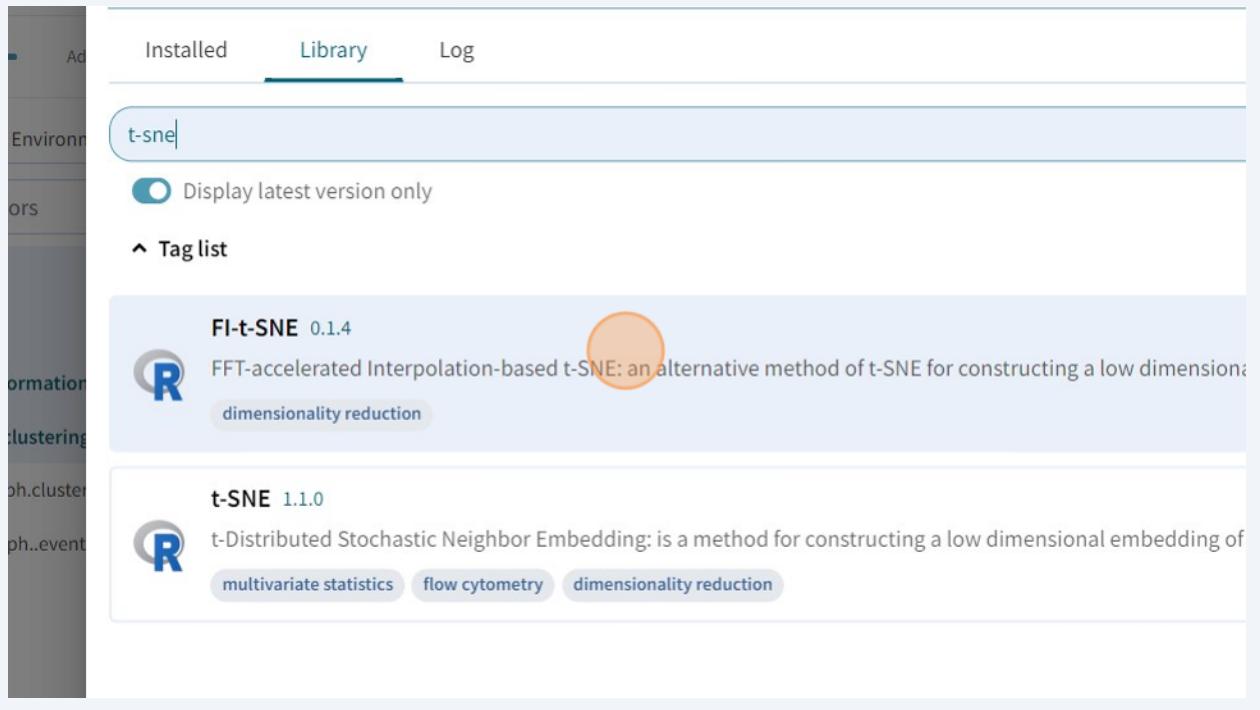
19 Name it t-SNE.



20 Remove the existing operator.

ds0..event_id	1
ds0..variable	CD11b 4.8 4.2 3.8 3.4 3.0 2.6 2.2 1.8 1.4 1.0 0.6 0.2

21 Look for the **FI-t-SNE operator**.

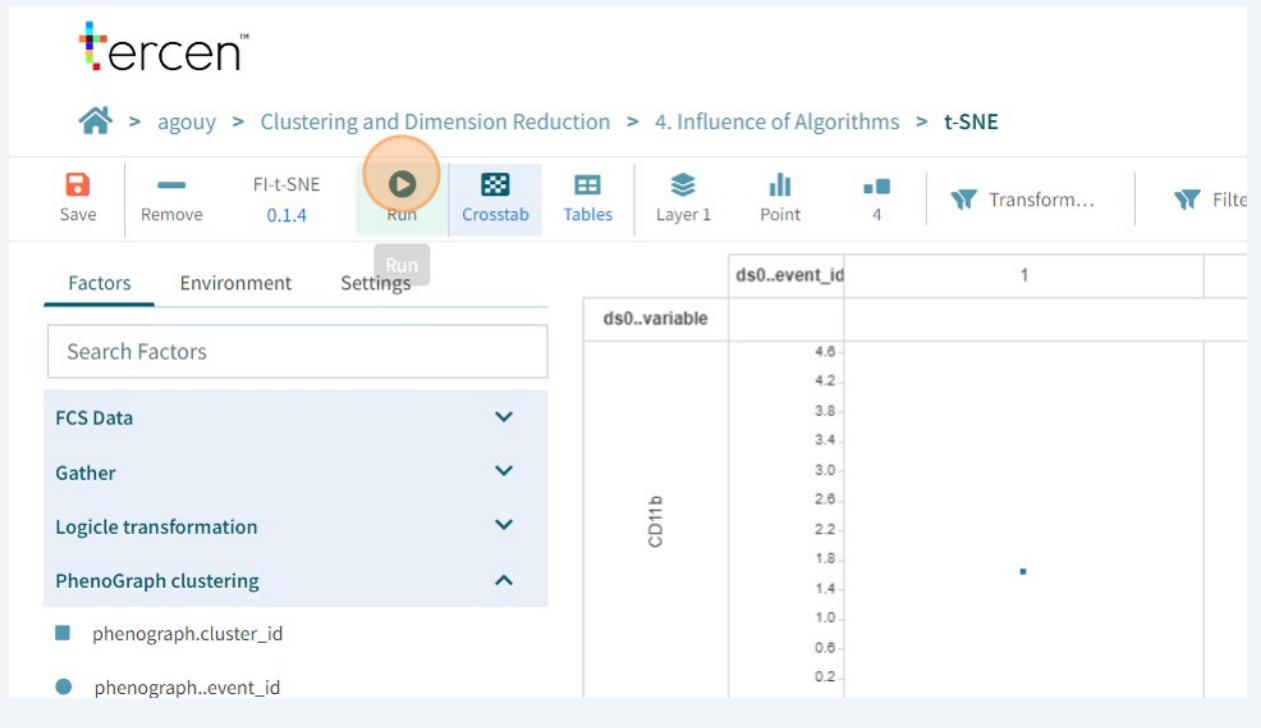


You might have noticed there were **two t-SNE operators** in Tercen.

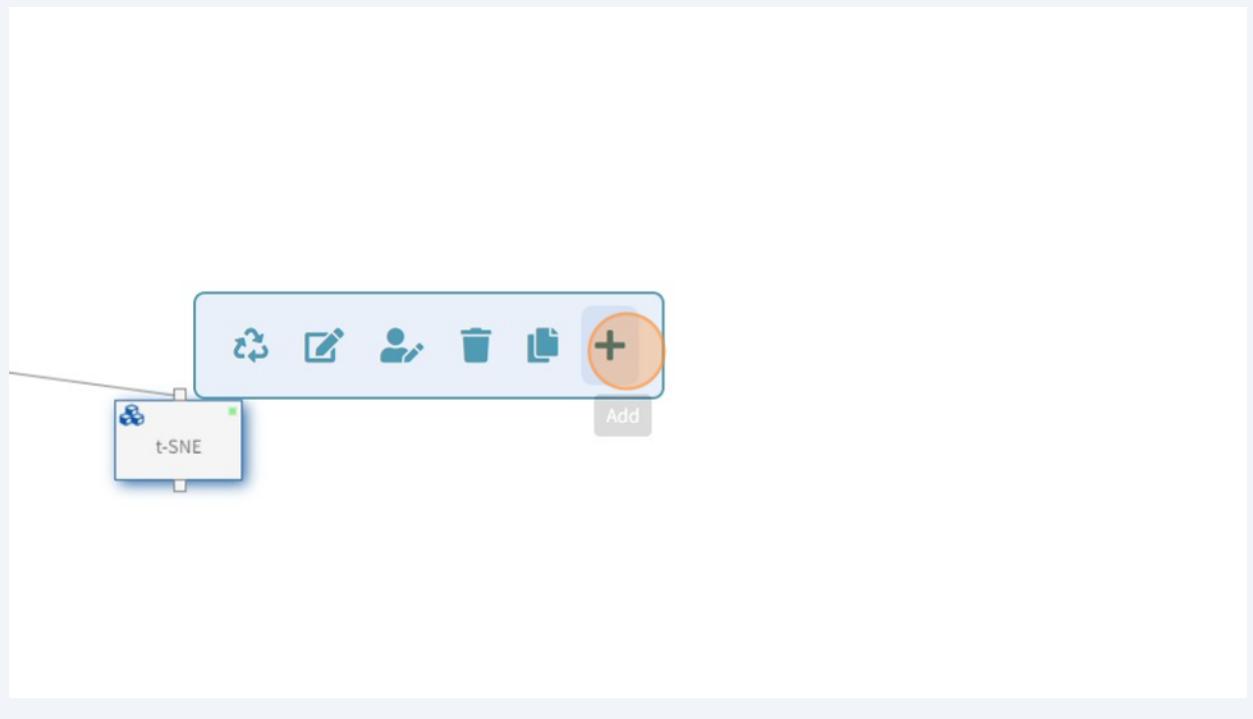
They correspond to **two different implementations** of the algorithms.

The **FI-t-SNE** is a **faster** approximation to the t-SNE that gives very similar results to the original one.

22 Run the algorithm.

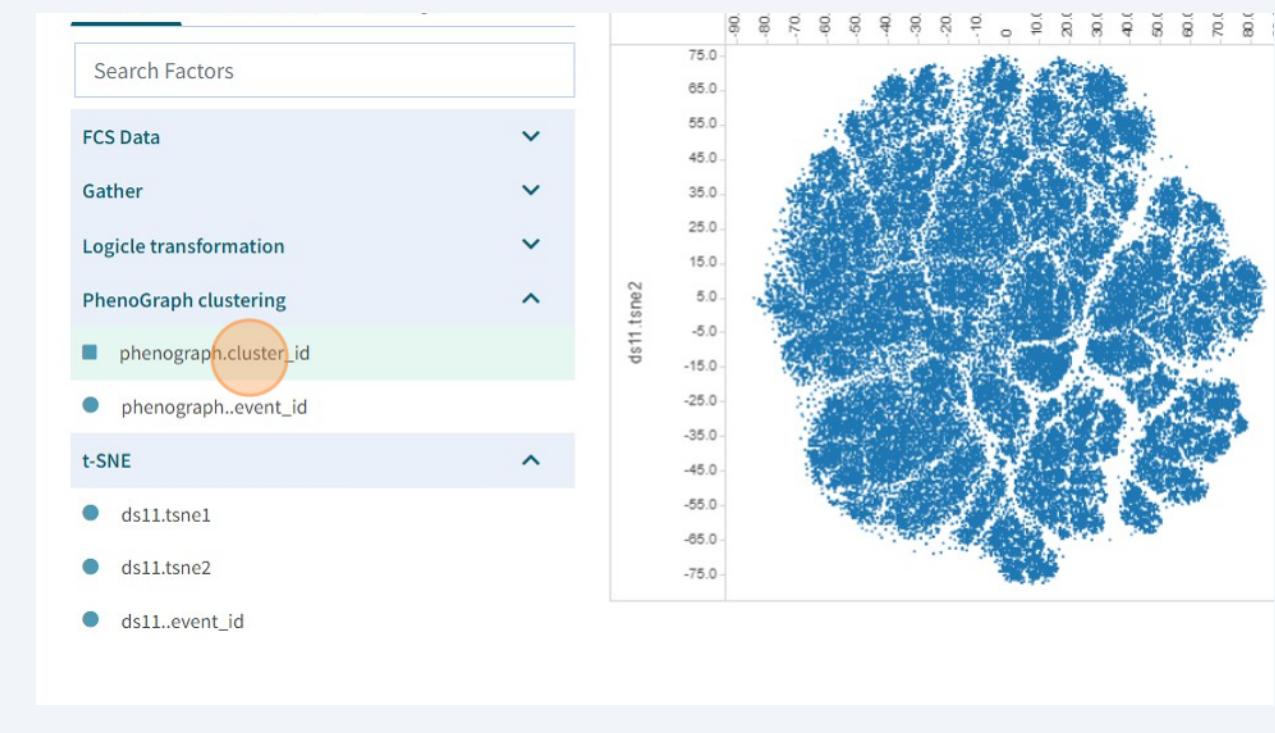


23 After the computation, **add a data step** to visualise the results.



24 Clear the crosstab and project

- **ds11.tsne1 to x-axis**
- **ds11.tsne2 to y-axis**
- **phenograph.cluster_id to Colors**



25 What do you observe?

