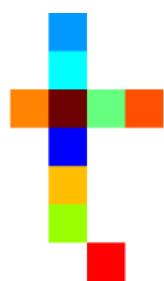




LEVEL ^{UP} WORKSHOP III

Part A

0301 - How to perform Clustering and Dimension Reduction in Tercen



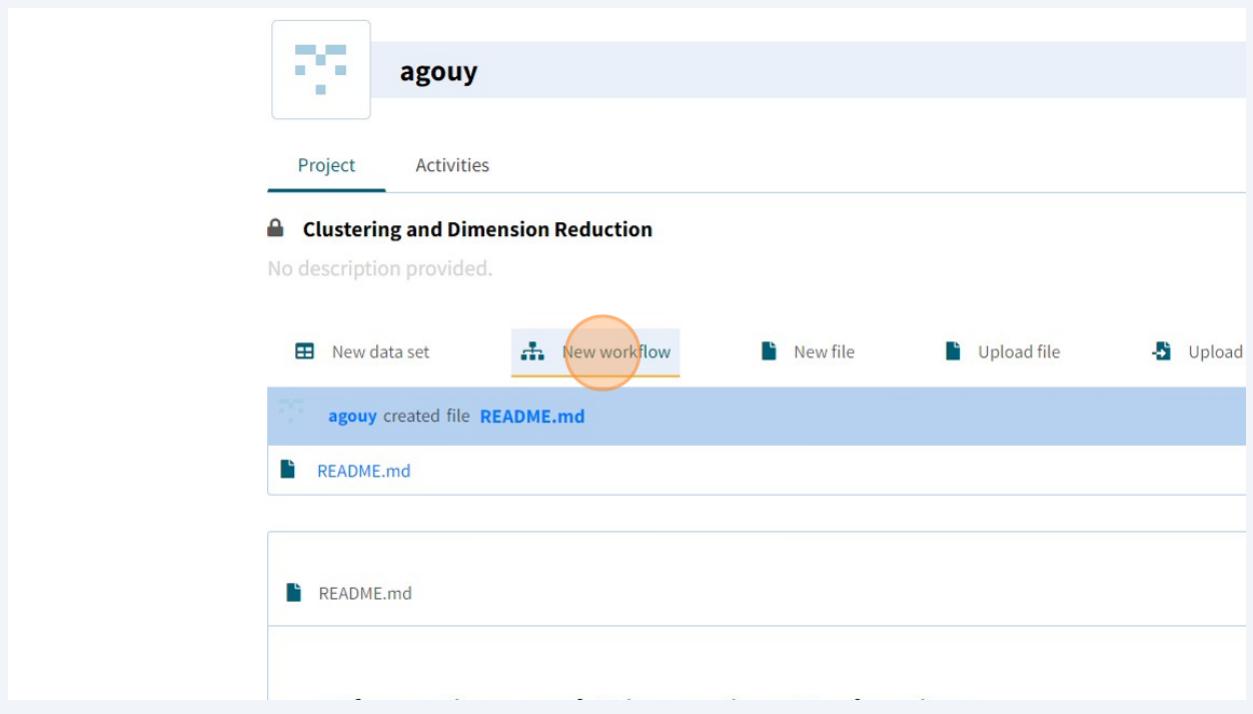
- 1 Navigate to the Workflow III project and clone it.

The screenshot shows the Tercen platform interface. At the top, there's a navigation bar with the Tercen logo, a search bar labeled "Search Tercen", and a "Learn" button. Below the navigation is a breadcrumb trail: Home > LevelUpWorkshopsTeam > Workshop III. The main area is titled "LevelUpWorkshopsTeam" and has tabs for "Project" (which is selected) and "Activities". Under the "Project" tab, there's a section for "Workshop III" which is described as "Clustering and Dimension Reduction". Below this, there's a list of recent activity items:

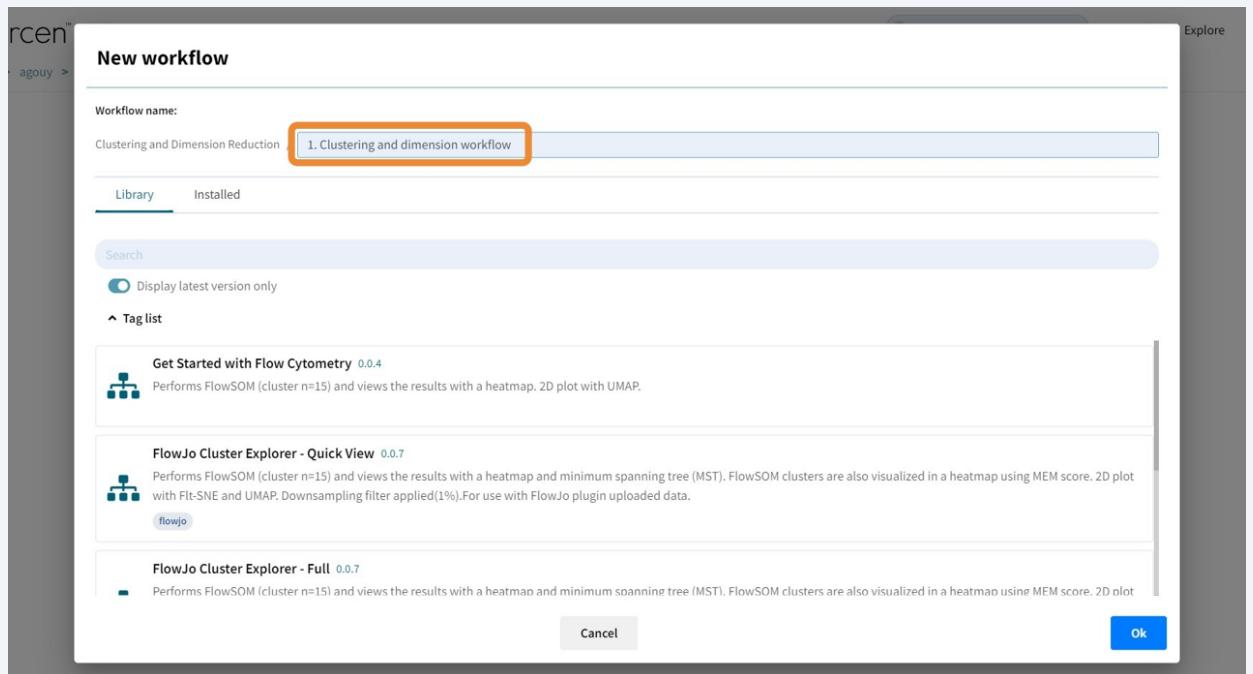
Action	Time Ago
agouy updated file A1 - Setup, Running Phenograph and UMAP.md	30 seconds ago
README.md	1 hours ago
pdfs	1 hours ago
Guides	1 hours ago
Example Data	1 hours ago

At the bottom of the project page, there's a sidebar with a "README.md" file listed. The "Clone project" button in the top right of the main content area is highlighted with an orange box.

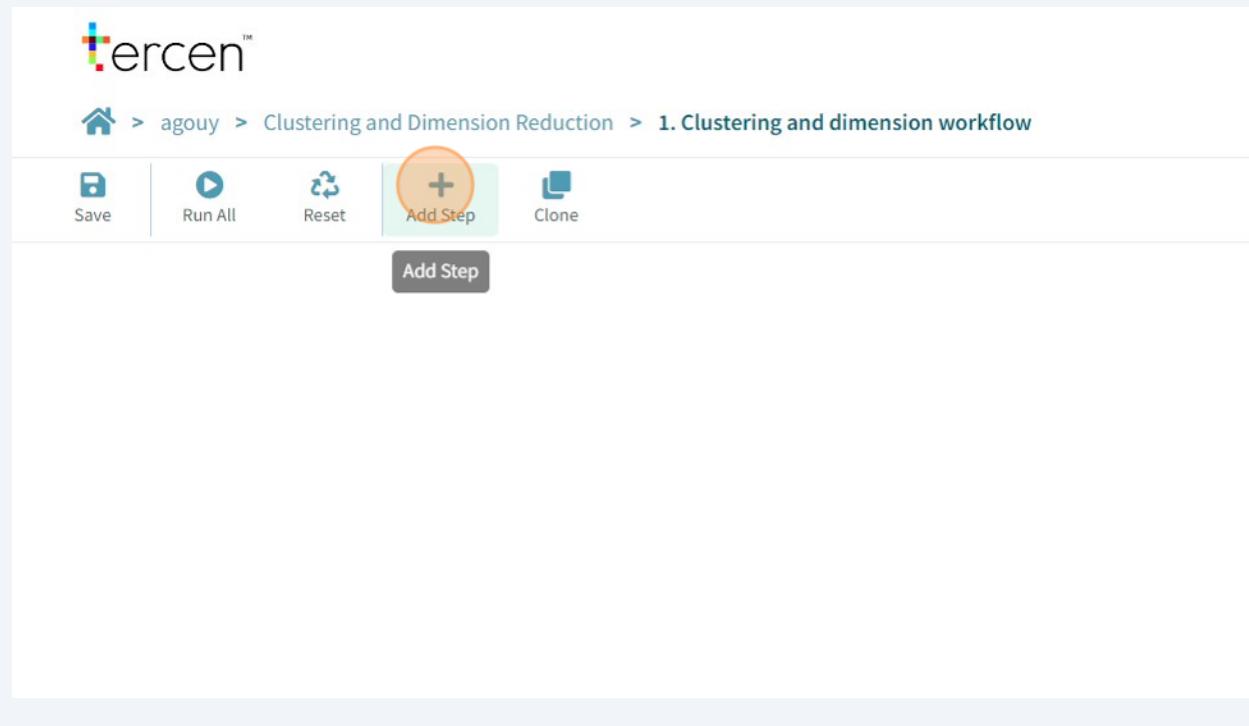
2 Create a new workflow.



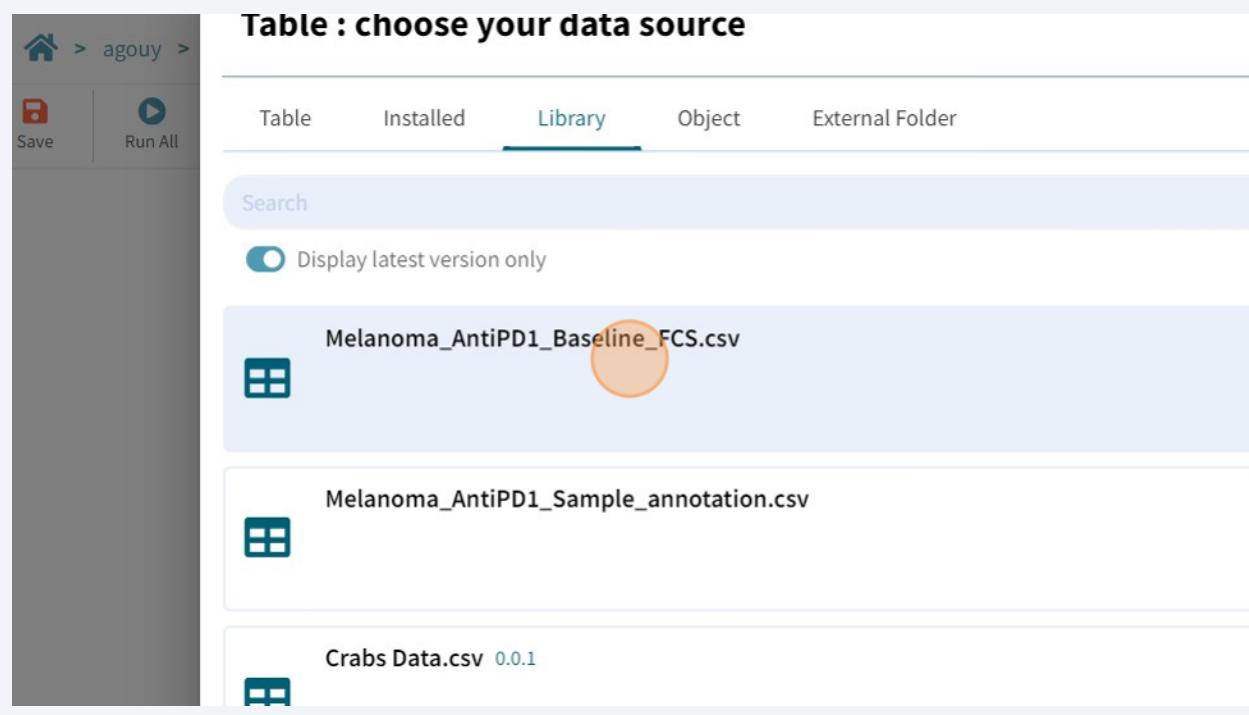
3 Give a relevant **name** to your workflow, such as "1. Clustering and dimension reduction" and click on **OK**.



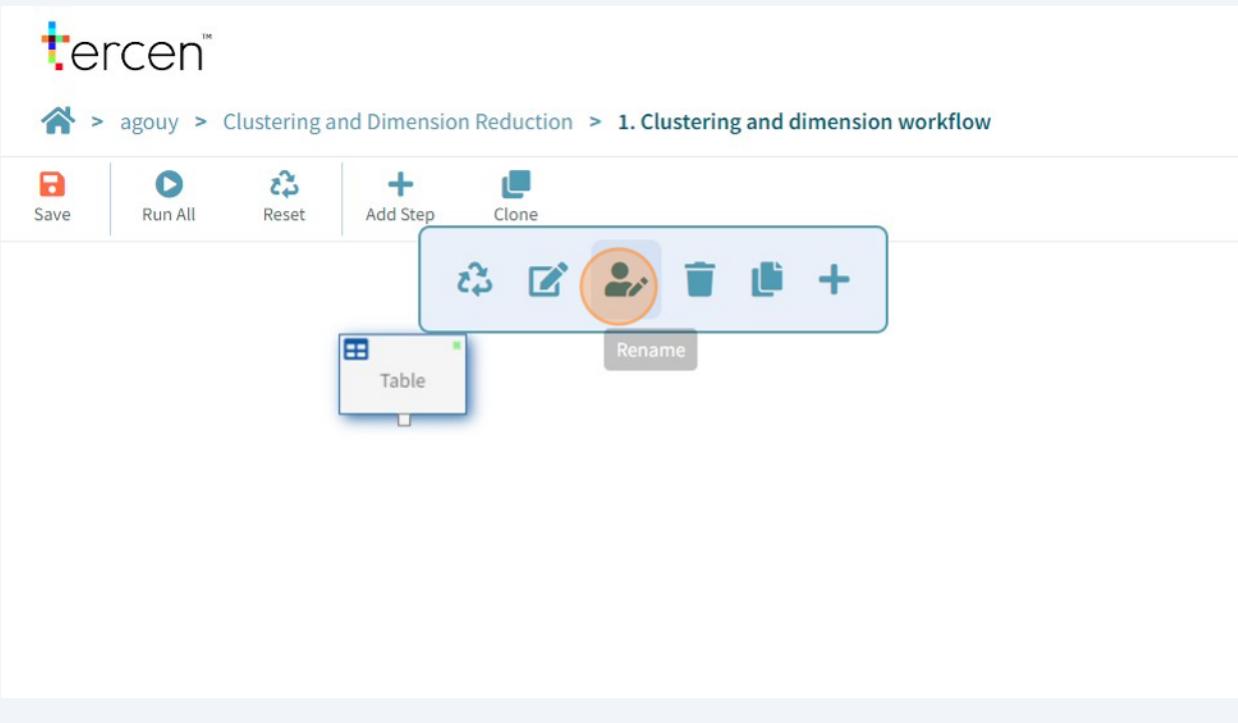
- 4 On the newly created workflow canvas, we'll start by adding some **flow cytometry data**.
Click on "Add Step" and select **Table**



- 5 Navigate to the **Library** tab and select the **Melanoma_AntiPD1_Baseline_FCS** dataset, then click **OK**.



- 6 You can **rename** the newly created **Table** to **FCS Data** for clarity.

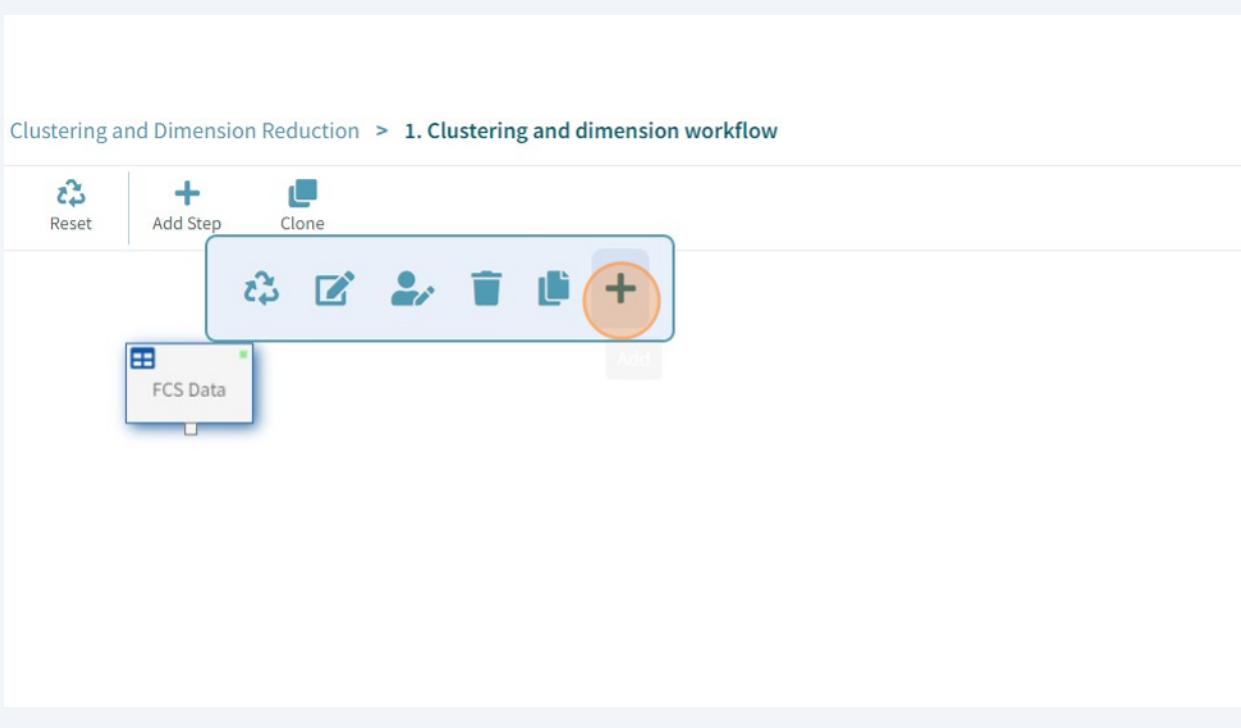


About this dataset

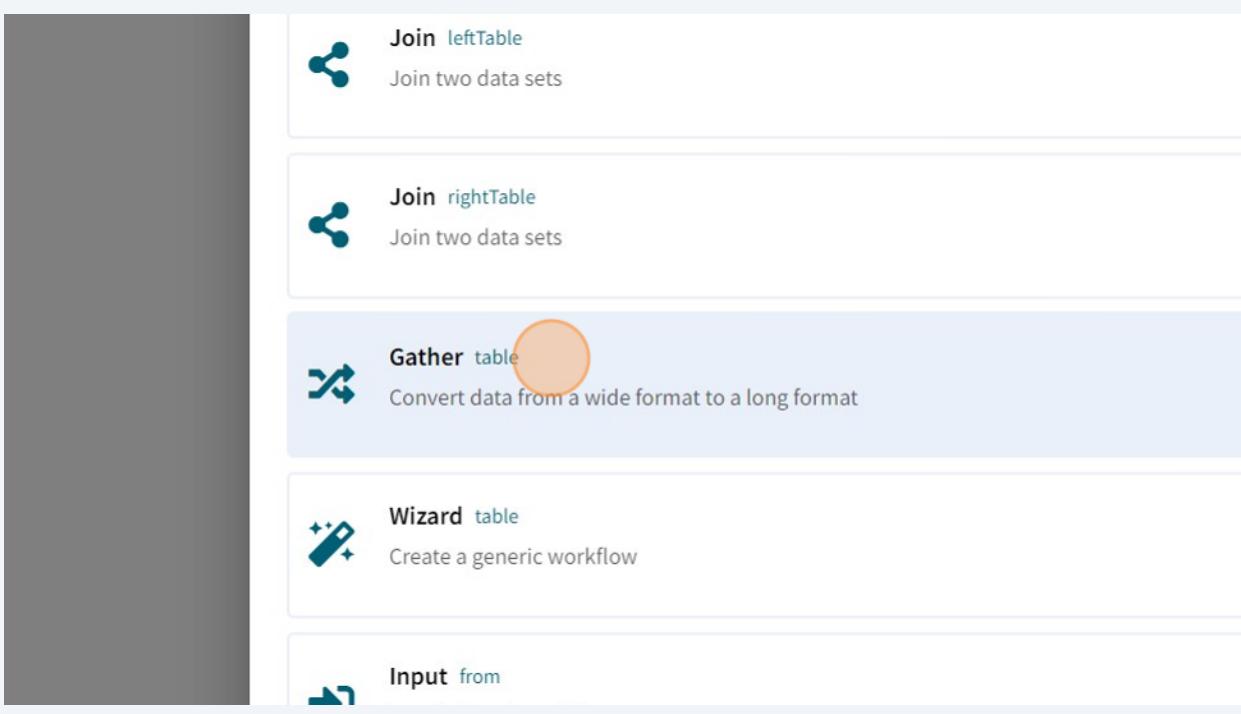
The data comes from a melanoma anti-PD-1 treatment response study (Krieg et al. 2018). It is a subset of mass cytometry data at baseline - for 20 patients (peripheral blood). Among them, we have 12 Responders and 8 Non-Responders to the treatment.

Reference: Krieg et al. High-dimensional single-cell analysis predicts response to anti-PD-1 immunotherapy. Nat Med 24, 144–153 (2018).

- 7 Add a **new step** after the dataset.



- 8 Add a **Gather step**.



- 9 On the local toolbar **Edit** the **Gather** step.

Change the Namespace to **channel**,

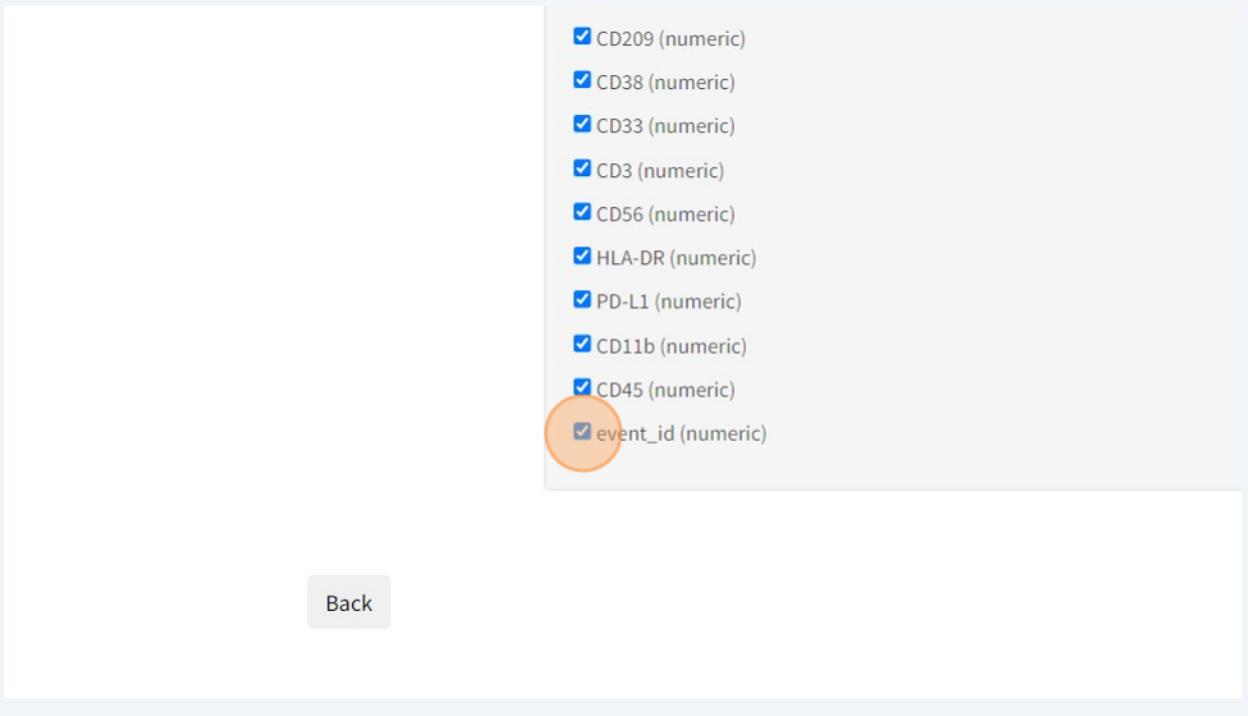
Then **Select all** the channels.

The screenshot shows the Tercen software interface with the following details:

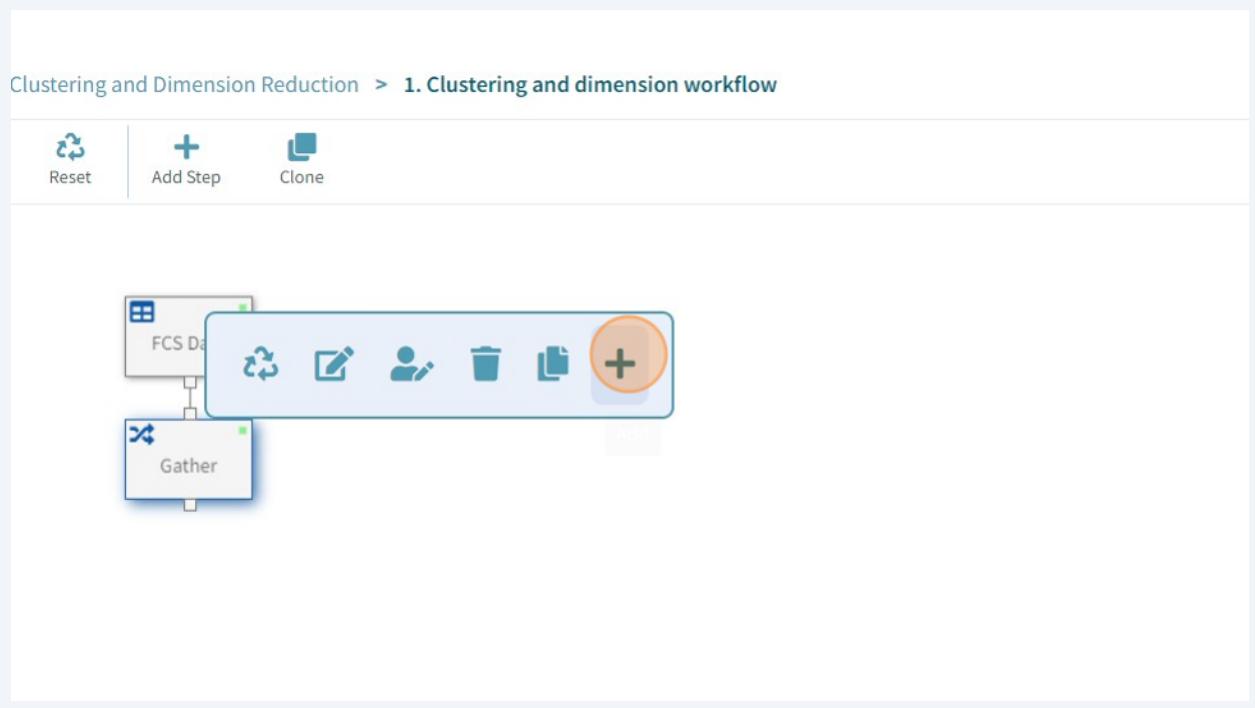
- Namespace:** channel
- Selection pattern:** (empty)
- Factor type:**
 - Numeric
 - Character
- Factors:** A list of channels:
 - CD19 (numeric)
 - CD64 (numeric)
 - CD303 (numeric)
 - CD34 (numeric)
 - CD141 (numeric)
 - CD61 (numeric)
 - CD123 (numeric)
 - CD66b (numeric)
 - CD62L (numeric)
 - ICAM-1 (numeric)
 - CD1c (numeric)

- 10** Don't forget to remove non-relevant variables, such as the **event_id**.

Then click the **Save & Run Step** button.



- 11** Go back to the **workflow** canvas and add a new **Data Step** after the Gather step.



12 Add a **Data step**.

Add

Step Operator Operator Library Installed Apps App Library

Search

 **Data step** `data`
Perform computation on user defined projection

 **Multi data step** `data`
Perform computation on user defined projection

 **Join** `leftTable`
Join two data sets

13

Now we will **prepare the input projection** required for data transformation.

Drag and drop:

- the "**channel.value**" factor to the **y axis**
- the "**channel.variable**" factor to the **rows**
- the "**event_id**" factor to the **columns**

The screenshot shows the Tercen software interface with the 'Factors' tab selected. At the top, there are tabs for 'Factors', 'Environment', and 'Settings'. Below the tabs is a search bar labeled 'Search Factors'. Under the 'Factors' tab, there is a section titled 'FCS Data' with a dropdown arrow, and another section titled 'Gather' with an upward arrow. Inside 'Gather', there are two entries: 'channel.value' and 'channel.variable'. The 'channel.variable' entry is highlighted with a green background and has an orange circle drawn around it, indicating it is the target for dragging and dropping.



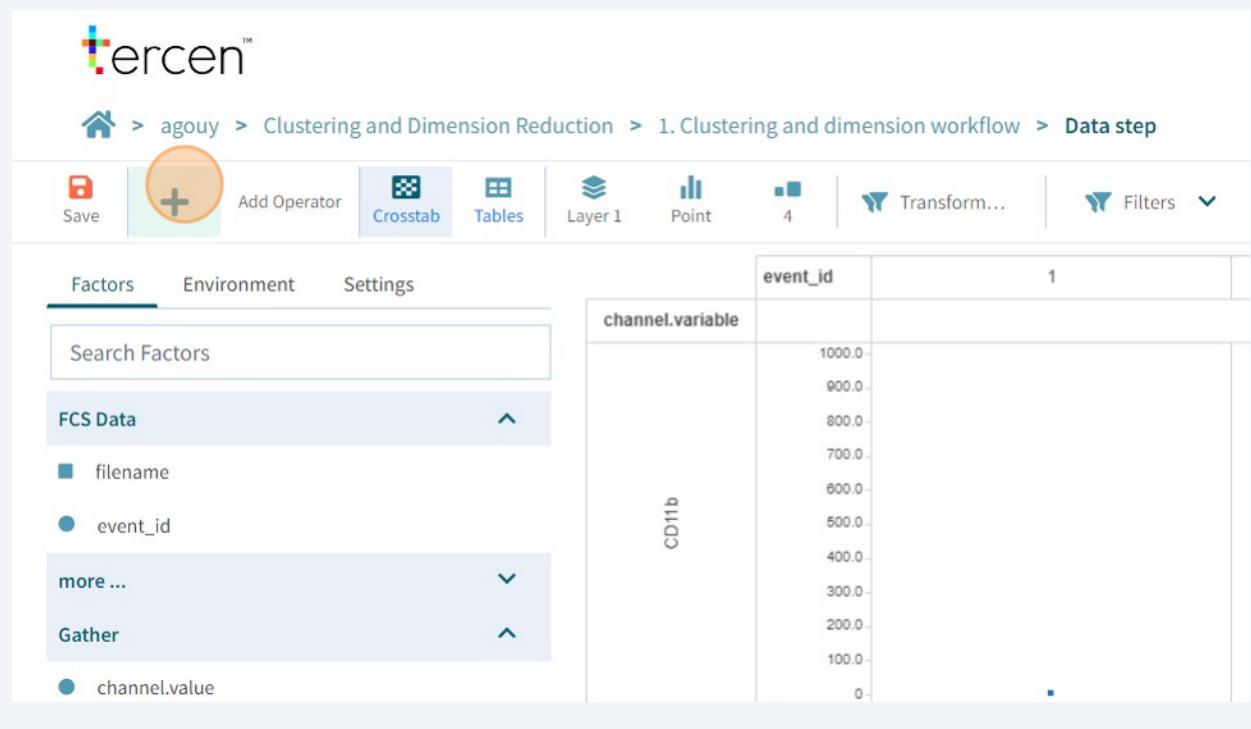
About this input projection

This **projection** is very common in Tercen. We project the **measurement** in the **y axis**, the **variable** in the **Rows**, and **observations** identifiers in the **Columns**.

Almost every clustering and dimension reduction operator uses this projection.

14

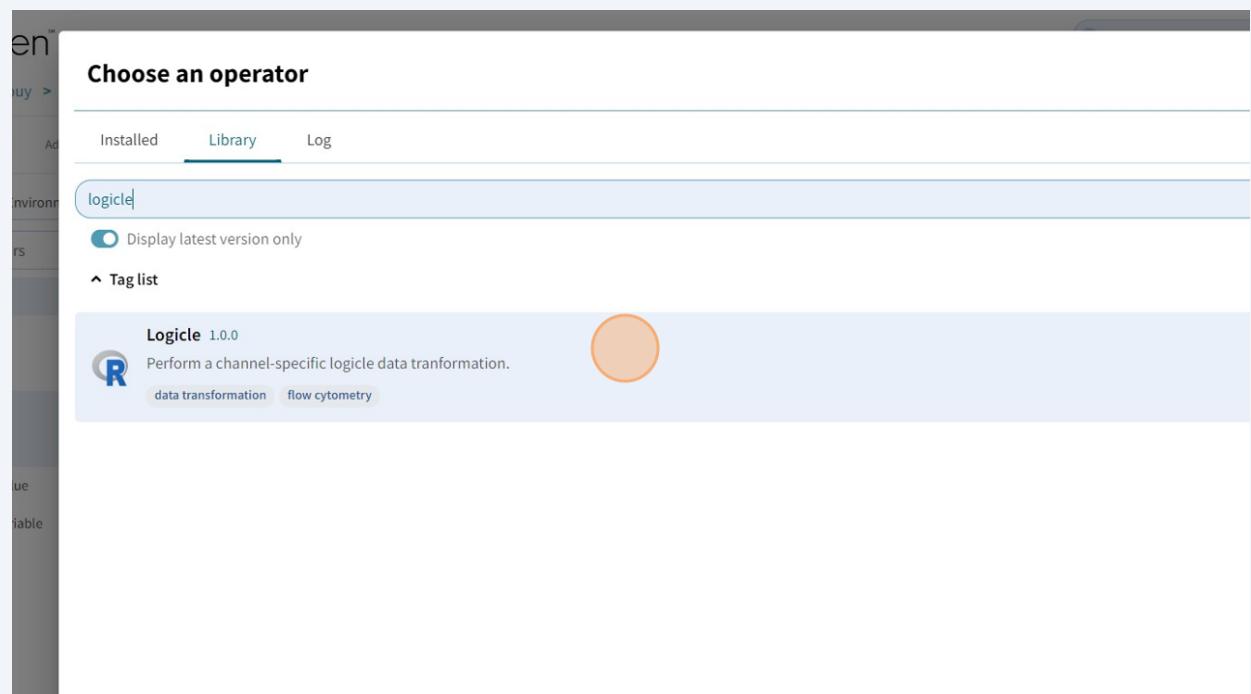
You should obtain a crosstab projection as below.
Now add an operator to perform a **logicle transformation**.



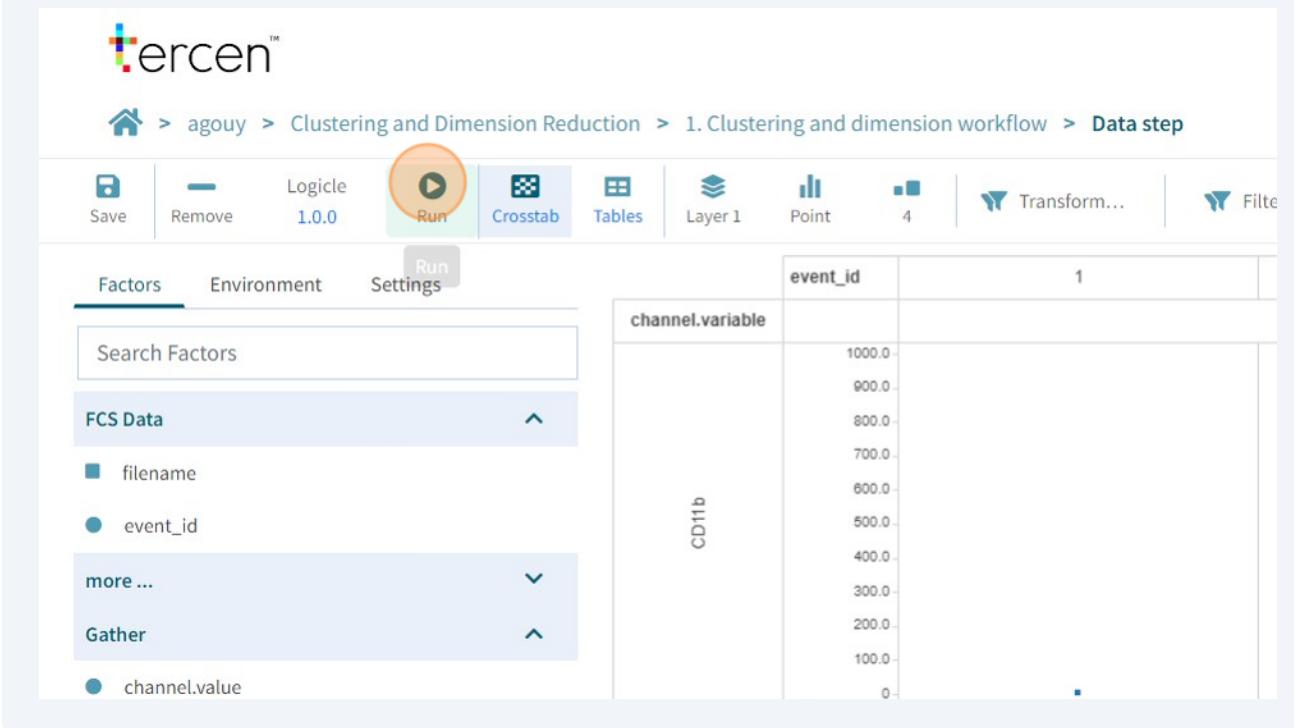
15

Search for the **Logicle** operator and add it to the data step.

If the operator is not part of the **Installed** operators, you will find it in the **Library** tab.



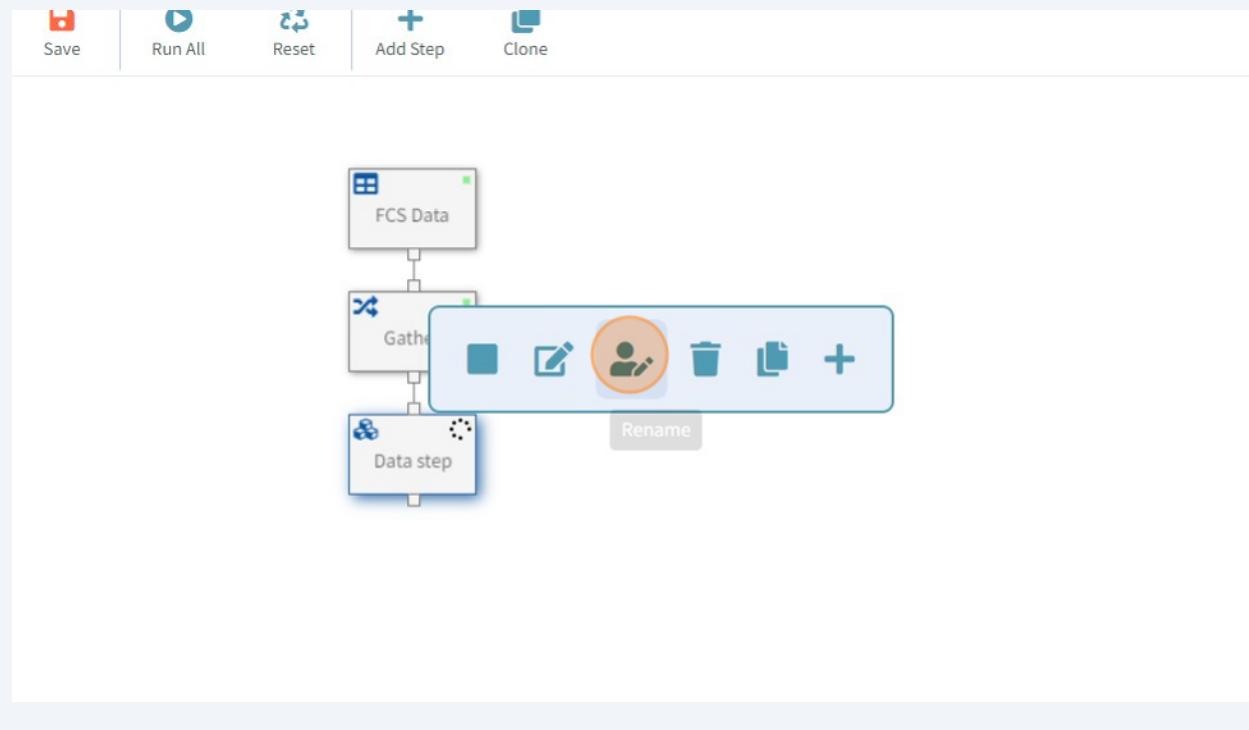
16 Run the operator.



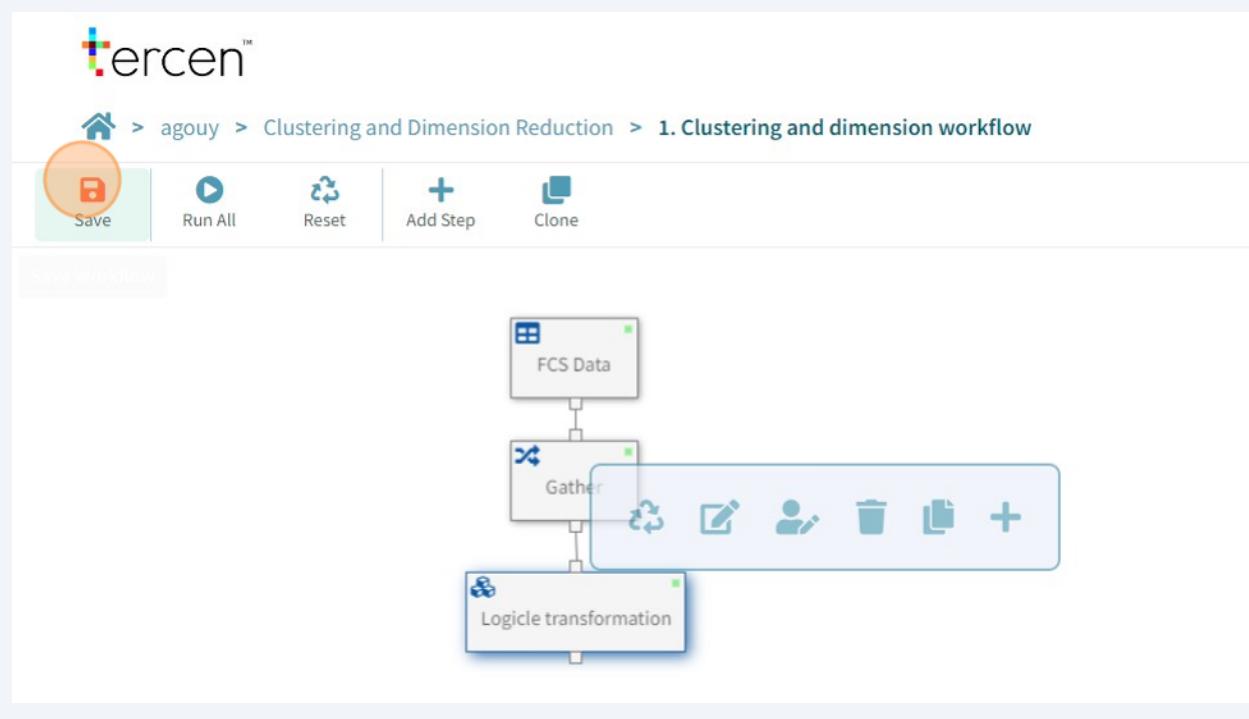
17 Click back to **1.Clustering and dimension workflow**

Operators will run in the background while you continue working with Tercen.

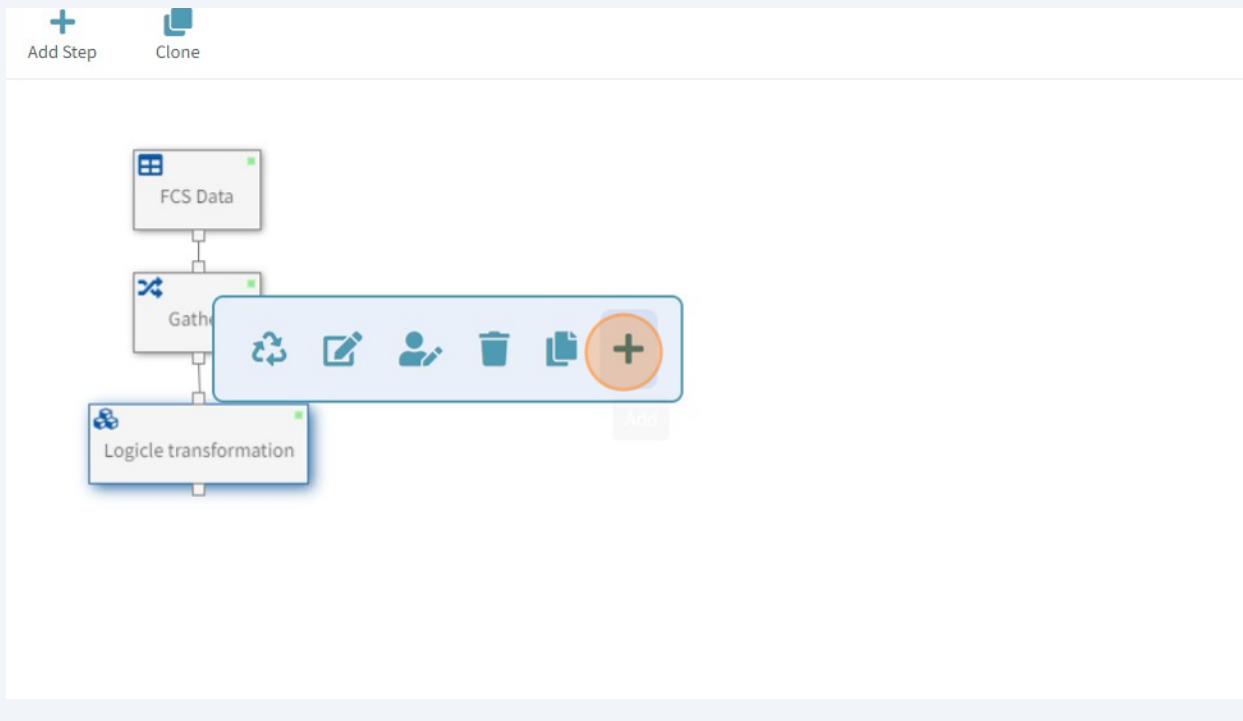
Rename the Data Step to **Logicle transformation** while it is running.



18 Do not forget to **regularly save** the workflow!



- 19 We will now perform **clustering** on our data.
Add a new **Data step** to the Logicle transformation.



20 Tercen will attempt to project your new data automatically.

- the **newly transformed value** in the **y axis**.
- **event ID** in the **columns**,
- **variable** in the **rows**

Add an **operator**.

The screenshot shows the Tercen software interface for a 'Data step'. The top navigation bar includes a home icon, user 'agouy', 'Clustering and Dimension Reduction', '1. Clustering and dimension workflow', and 'Data step'. The toolbar features icons for 'Save' (red folder), 'Add Operator' (orange plus sign), 'Crosstab' (grid), 'Tables' (grid with lines), 'Layer 1' (stacked bars), 'Point' (bar chart), a count of '4', 'Transform...' (blue funnel), and 'Filters' (blue funnel with dropdown). The main workspace has tabs for 'Factors' (selected), 'Environment', and 'Settings'. Under 'Factors', there's a search bar and a dropdown menu showing 'FCS Data', 'Gather', and 'Logicle transformation'. Below these are three items: 'ds0.value' (radio button), 'ds0.event_id' (radio button), and 'ds0.variable' (checkbox). To the right, a table structure is shown with columns 'ds0..event_id' and 'ds0..variable'. The 'ds0..event_id' column has four rows, and the 'ds0..variable' column has one row labeled 'ds0.value'.

21 Search for the **PhenoGraph** operator and add it.

The screenshot shows a search interface for operators. At the top, there are tabs for 'Installed', 'Library' (which is selected), and 'Log'. A search bar contains the text 'pheno'. Below the search bar is a checkbox labeled 'Display latest version only'. Underneath, there is a section titled '^ Tag list'. The main result is 'PhenoGraph 2.0.0', which includes an R logo, a brief description 'A graph-based method for flow cytometry data clustering.', and two tags: 'clustering' and 'flow cytometry'. To the right of the description is a large orange circular icon.

22 Click the **Environment** tab

This brings up the settings of an operator.

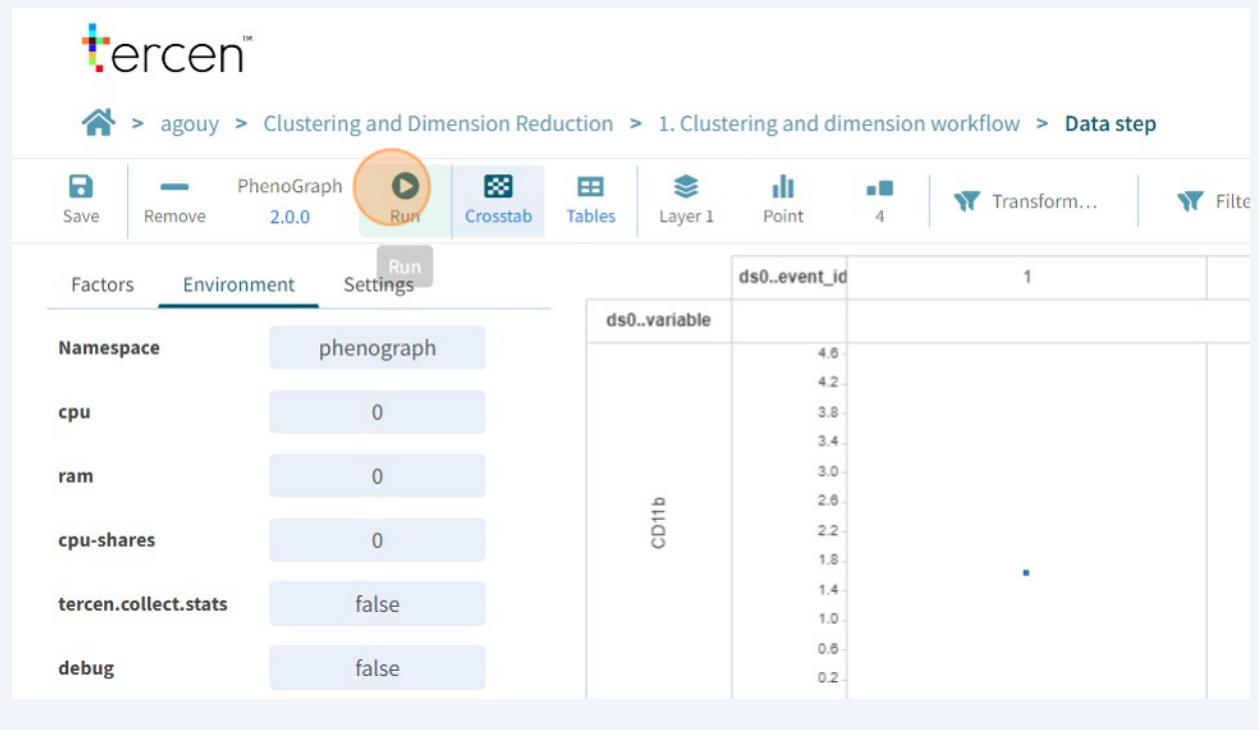
Change the **namespace** to phonograph to give it a more meaningful prefix.

The screenshot shows the tercen™ software interface. At the top, there is a navigation bar with steps: Home > agouy > Clustering and Dimension Reduction > 1. Clustering and dimension workflow > Data step. Below the navigation is a toolbar with icons for Save, Remove, PhenoGraph 2.0.0 (selected), Run, Crosstab, Tables, Layer 1, Point, a count of 4, Transform..., and Filter. The main area has tabs for Factors, Environment (which is selected and highlighted with a blue border), and Settings. Under the Environment tab, there is a 'Namespace' input field containing 'ds1'. To the right, there is a data preview table:

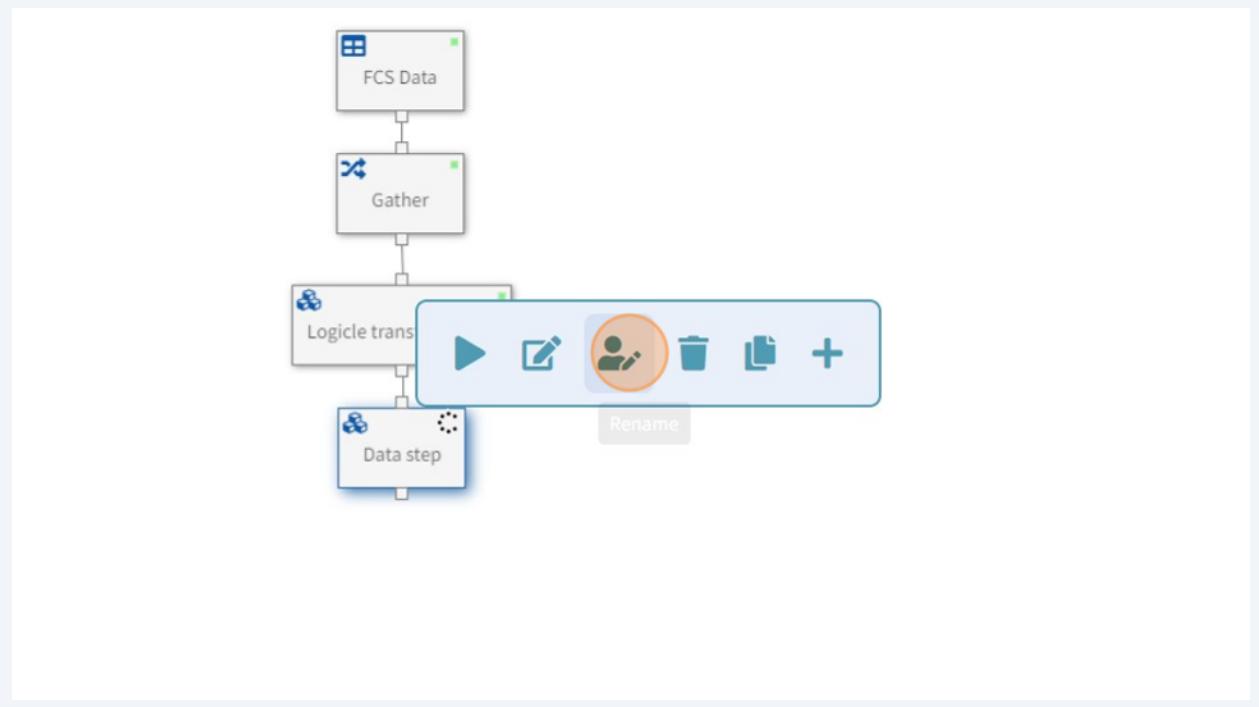
	ds0..event_id	1
ds0..variable		
cpu	4.8	
ram	4.2	
cpu-shares	3.8	
tercen.collect.stats	3.4	
debug	3.0	
	2.6	
	2.2	
	1.8	
	1.4	
	1.0	
	0.6	
	0.2	

The table shows data for the 'CD11b' cluster, with 'ds0..event_id' set to 1. The 'ds0..variable' column lists various system metrics like CPU usage, RAM, and memory shares, each with a value between 0.2 and 4.8.

23 Run the operator.



24 While the operator is computing, rename the data step on the workflow canvas to **Phenograph Clustering**.





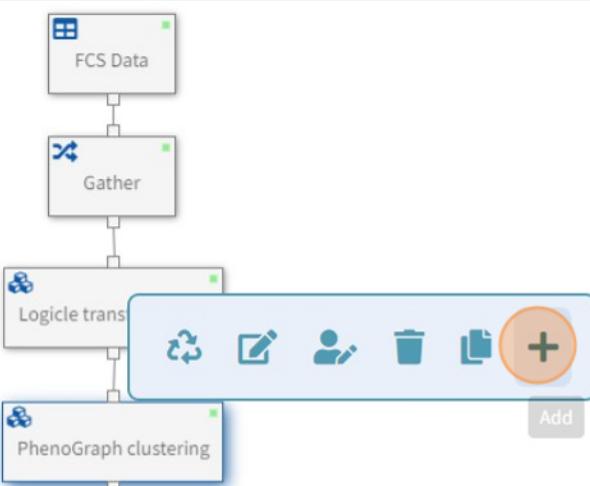
We have applied a **clustering** algorithm (PhenoGraph) to our data.

In order to better **visualise** those results, we first need to apply another operator to do **dimension reduction** (UMAP).

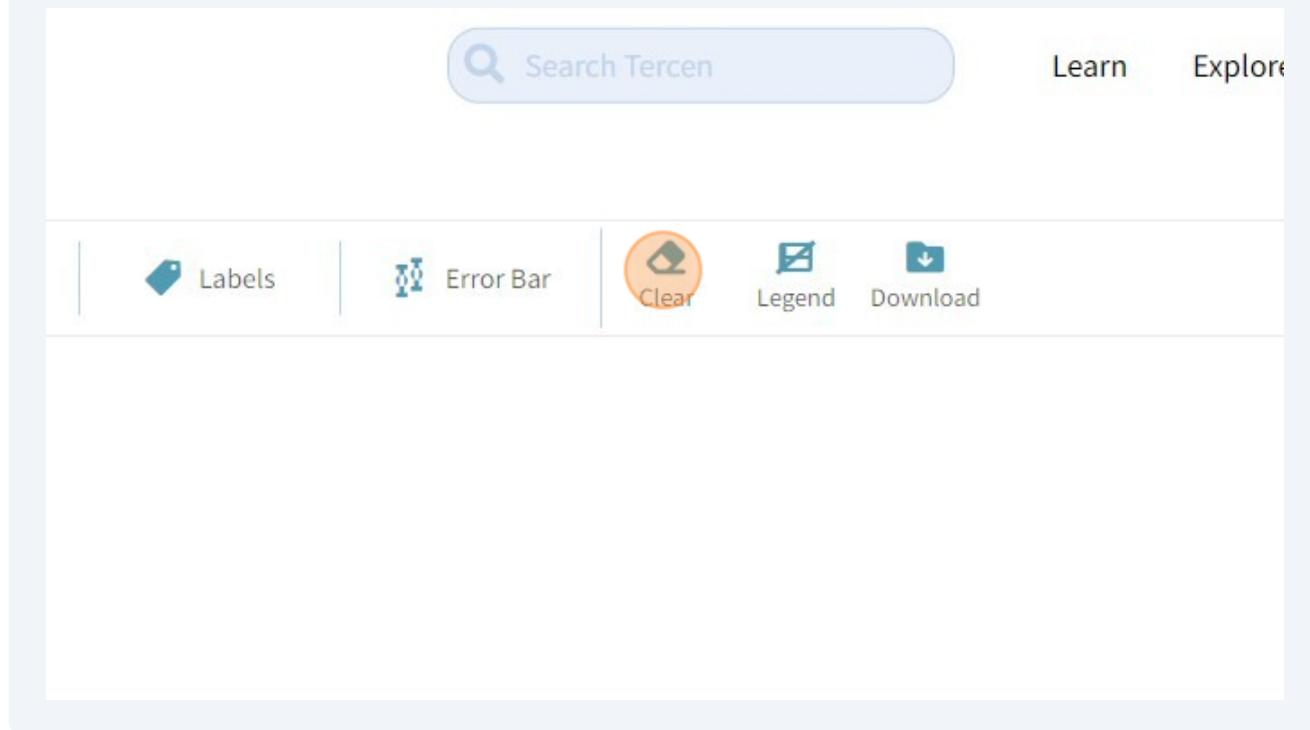


25 Add a new **Data step** to Phenograph Clustering

We will perform **dimension reduction** on our data.



26 **Clear** all the existing factors from the crosstab.



27 Prepare the following input projection:

- the transformed value (**ds0.value**) in the **y axis**,
- the event ID (**ds0.event_id**) in the **columns**,
- the variable (**ds0.variable**) in the **rows**,

The screenshot shows a software interface for preparing input projections. On the left, there's a sidebar with tabs for 'Factors', 'Environment', and 'Settings'. The 'Factors' tab is selected, showing a search bar labeled 'Search Factors' and a list of factor types: 'FCS Data', 'Gather', 'Logic transformation', 'PhenoGraph clustering', and 'ds0..variable'. The 'Logic transformation' section is expanded, showing three items: 'ds0.value' (which is highlighted with an orange circle), 'ds0..event_id', and 'ds0..variable'. On the right, there's a large grid titled 'ds0..variable' with columns labeled 'Row', 'Y-Axis', and 'Column'. The 'Column' column has a sub-label 'X-Axis'. The grid is currently empty.

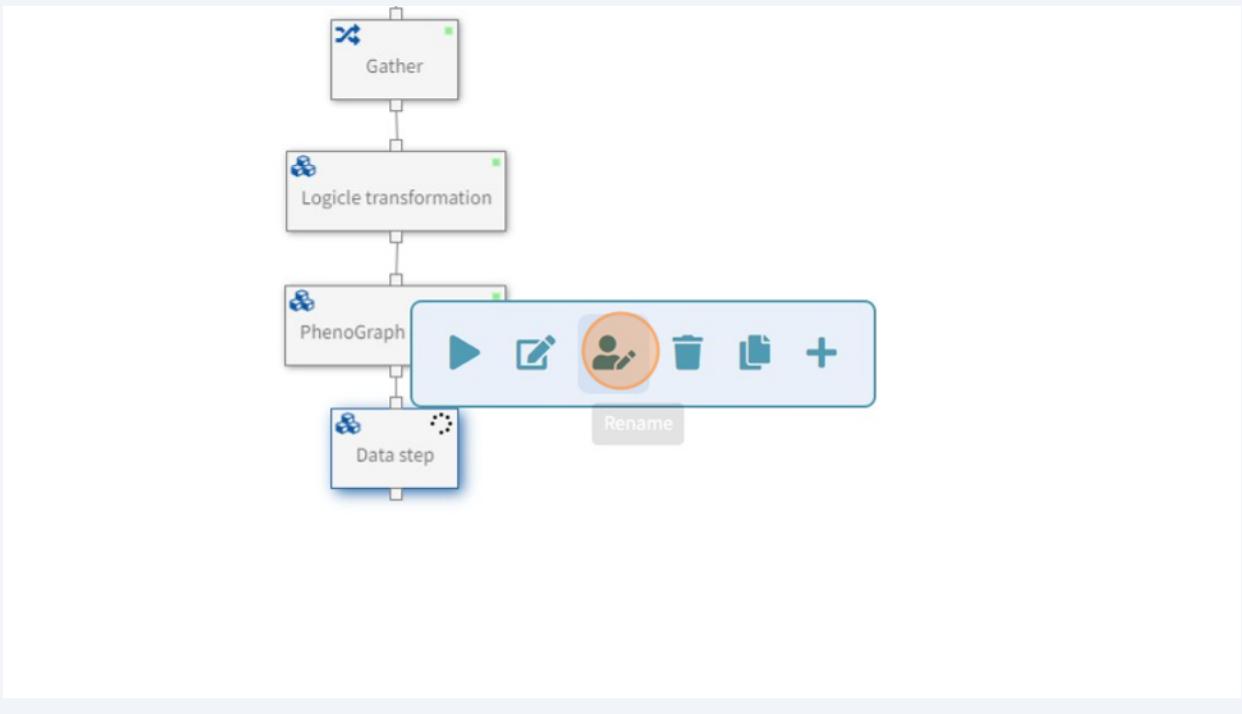
28 Search for the **UMAP** operator and add it.

The screenshot shows the Cellxenon Platform's search results for 'umap'. The 'Installed' tab is active, showing the UMAP 1.2.3 operator. The operator card provides a brief description: 'Uniform Manifold Approximation and Projection, a dimensionality reduction method.' and includes tags: 'dimensionality reduction', 'flow cytometry', 'omics', and 'scRNASeq'. A large orange circle highlights the 'Run' button on the operator card.

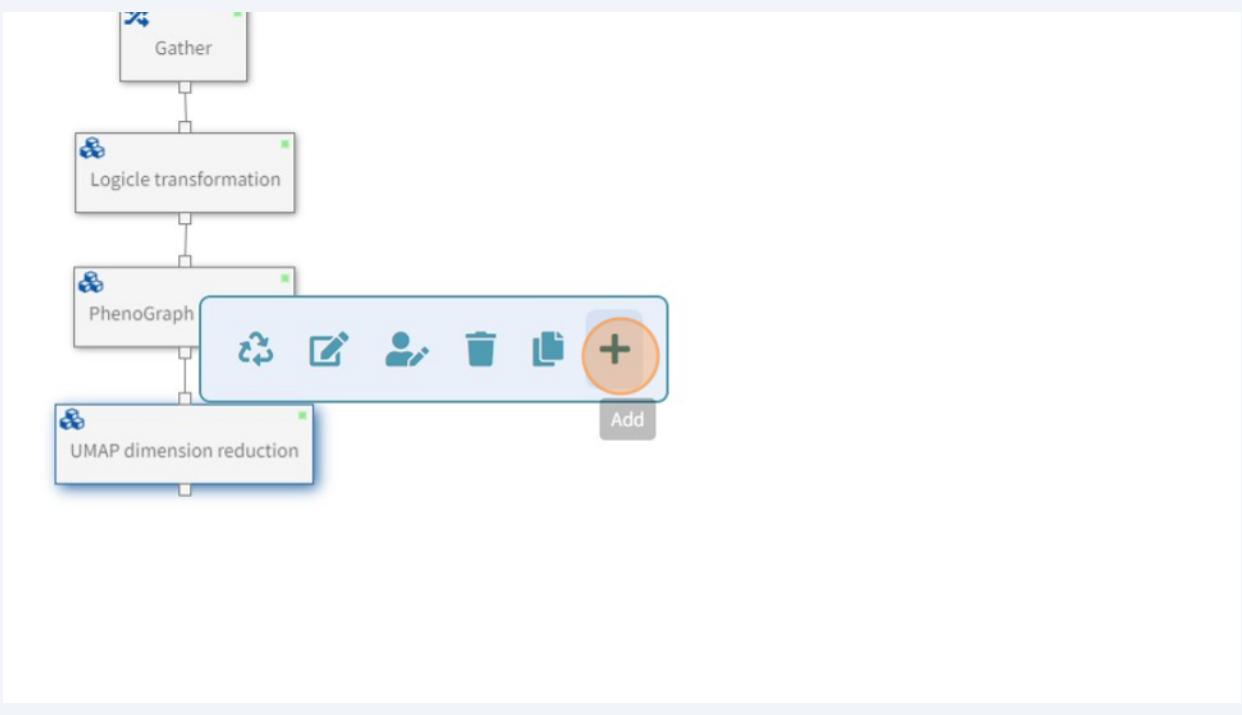
29 Run the operator.

The screenshot shows the Cellxenon Platform interface with the UMAP operator running. The 'Run' button is highlighted with an orange circle. The right panel displays a scatter plot of 'ds0.event_id' vs 'ds0.variable'. The plot shows a single point at event_id 1 and variable CD11b, with a color scale ranging from 0.2 to 4.6.

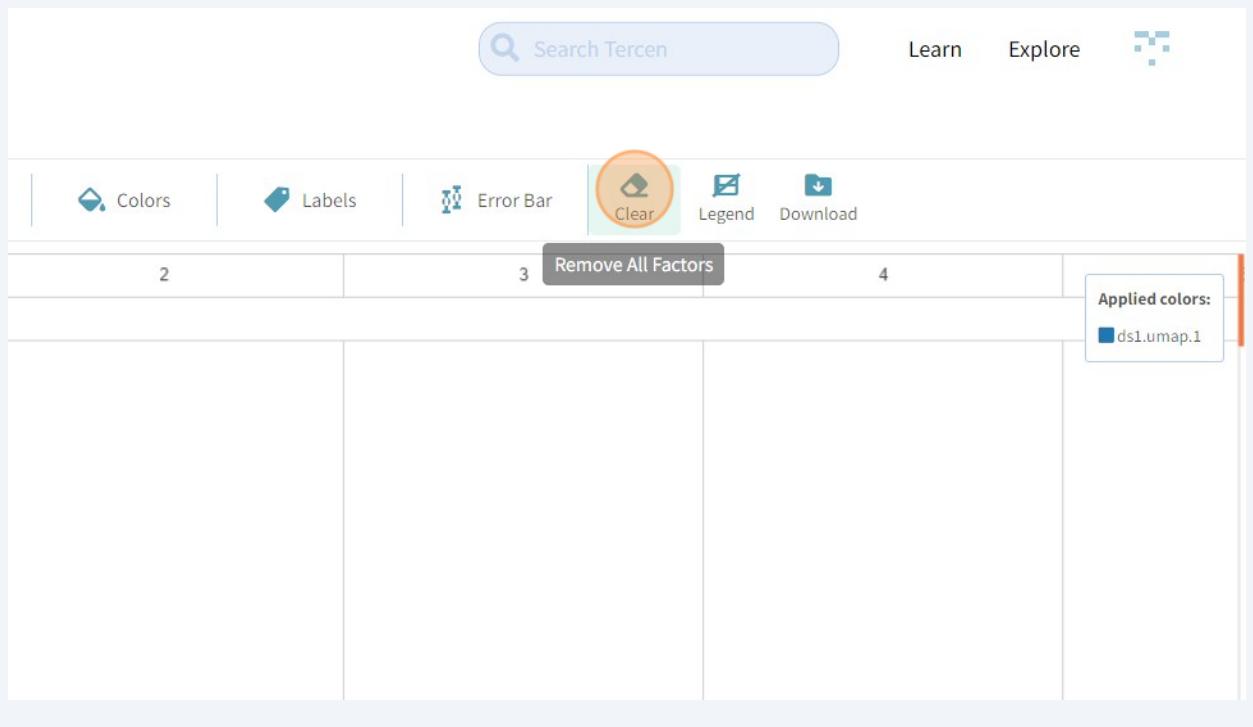
30 While the operator runs rename your data step to **UMAP dimension reduction**



31 We will now add a new **Data step** to **visualise our results**.



32 Clear all factors.



33 UMAP has computed two new factors. Let's project them on the crosstab.

- **ds1.umap.1** to X-Axis
- **ds1.umap.2** to Y-Axis



34 Press the **Point Size** button and reduce to **1**.

Drag the black lines to resize your chart



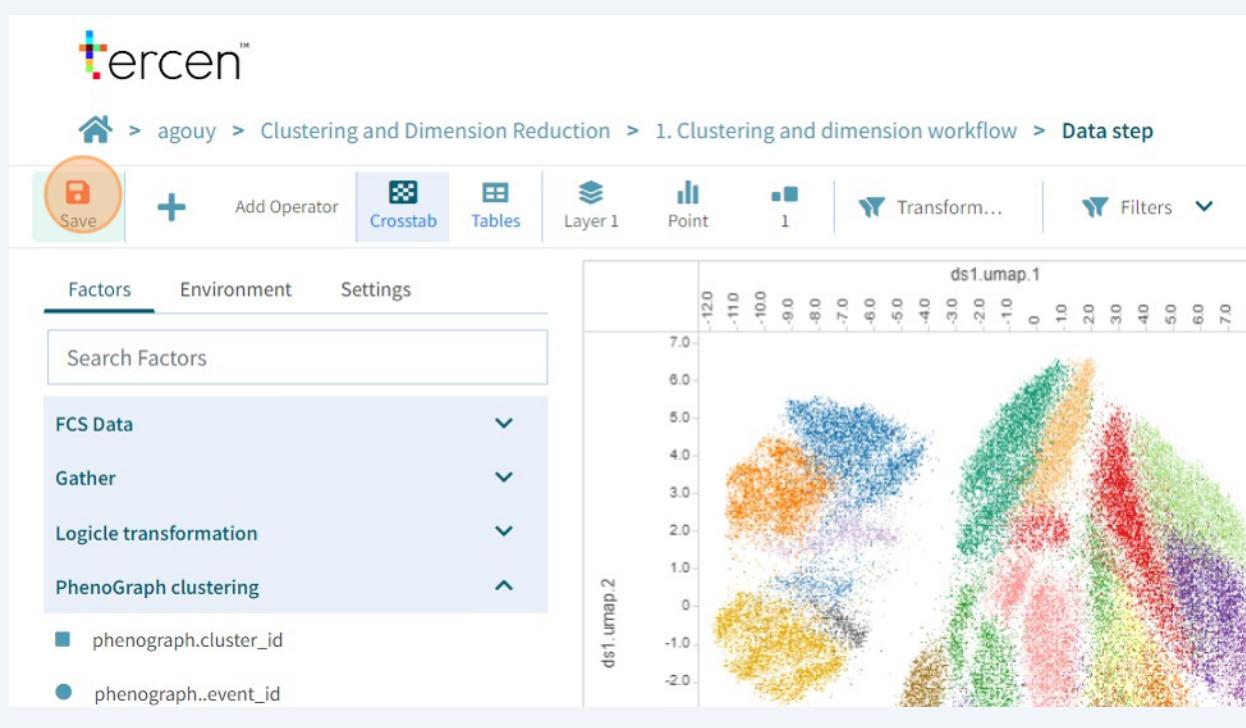
35 Open the **PhenoGraph clustering** factors.

Drag **phenograph.cluster_id** to **colors**.



36 Do not forget to **Save** the workflow!

You can also rename this data step for clarity.

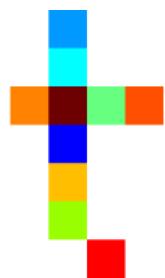




Et voilà! You have performed clustering and dimension reduction on a flow cytometry dataset.

We will next see how interpret those clusters.

0302 - Cluster Exploration



- 1 Navigate to your workshop project and **clone** the **first workflow**.
Save it to your project name it **2. Cluster Exploration**.

The screenshot shows a workshop project interface with the following details:

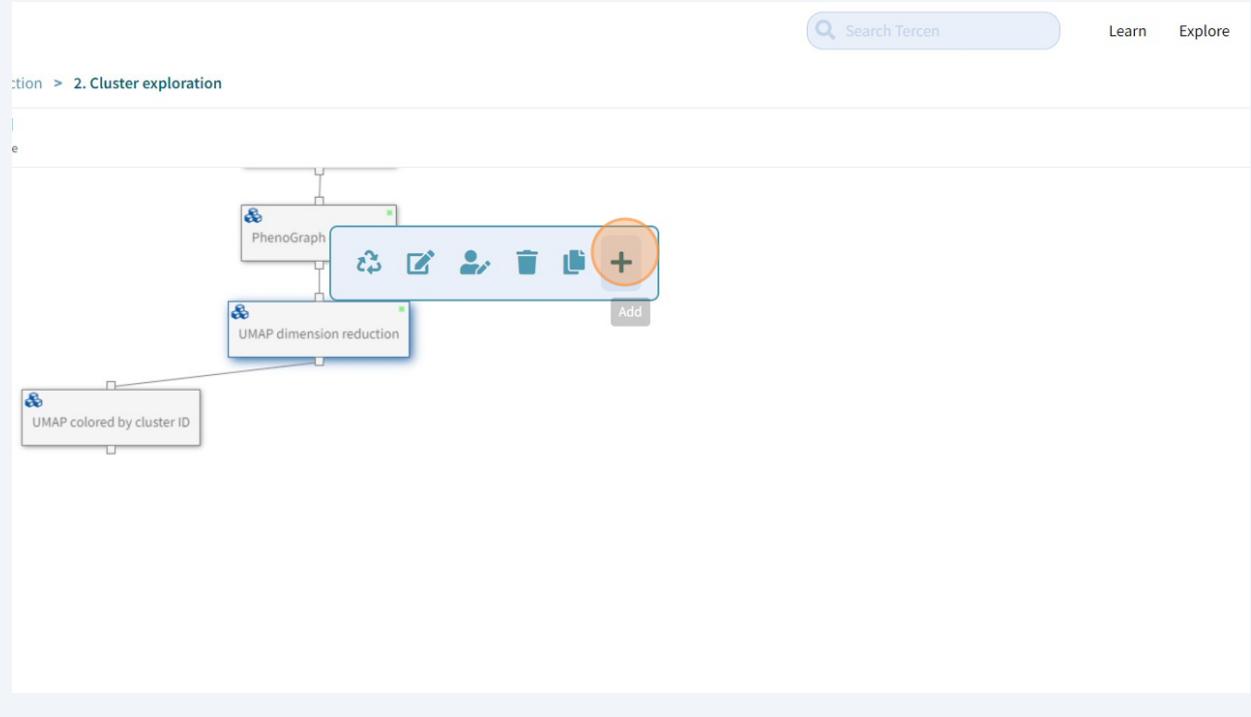
- Project:** agouy
- Activities:** Clustering and Dimension Reduction
- Files:**
 - README.md (6 minutes ago)
 - Melanoma_AntiPD1_Sample_annotation.csv (21 hours ago)
 - Melanoma_AntiPD1_Baseline_FCS.csv (3 hours ago)
 - dev (6 minutes ago)
 - 1. Clustering and dimension workflow (3 hours ago)
- Actions:** New data set, New workflow, New file, Upload file, Upload workflow, Project settings, Clone project
- Cloned Workflow:** A cloned workflow titled "2. Exploring clustering results" is shown below the main project area. It contains:
 - README.md
- Workflow Details:** The workflow is named "Clustering and Dimension Reduction".
- Icons:** There are several small icons at the bottom right of the workflow area, including a red circle with a white icon.

2 Open the workflow canvas.

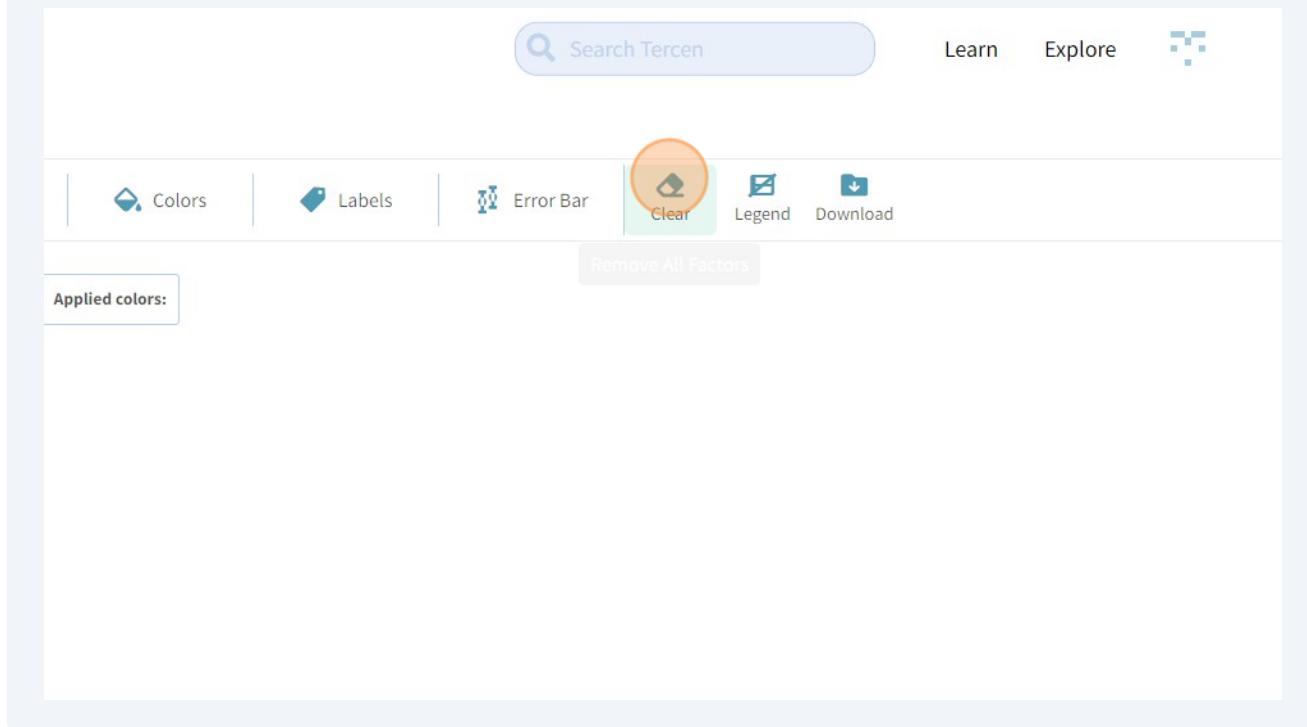
In this tutorial we will produce visualisations to explore clustering results.

Starting with a visualisation of marker expression.

Select UMAP dimension reduction and Add a new data step.



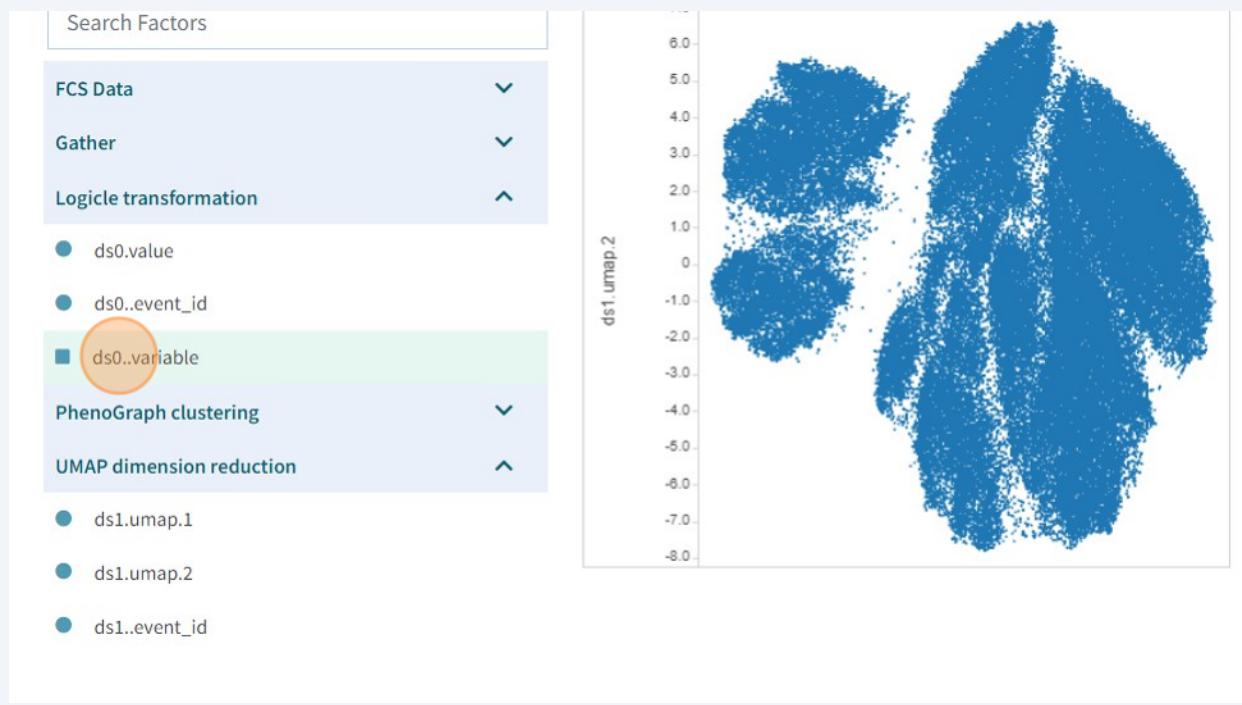
- 3** **Clear** the crosstab.



4 Prepare the classical UMAP visualisation with:

- **ds1.umap.2 to y axis**
- **ds1.umap.1 to x axis**
- **ds0.variable factor to Row**

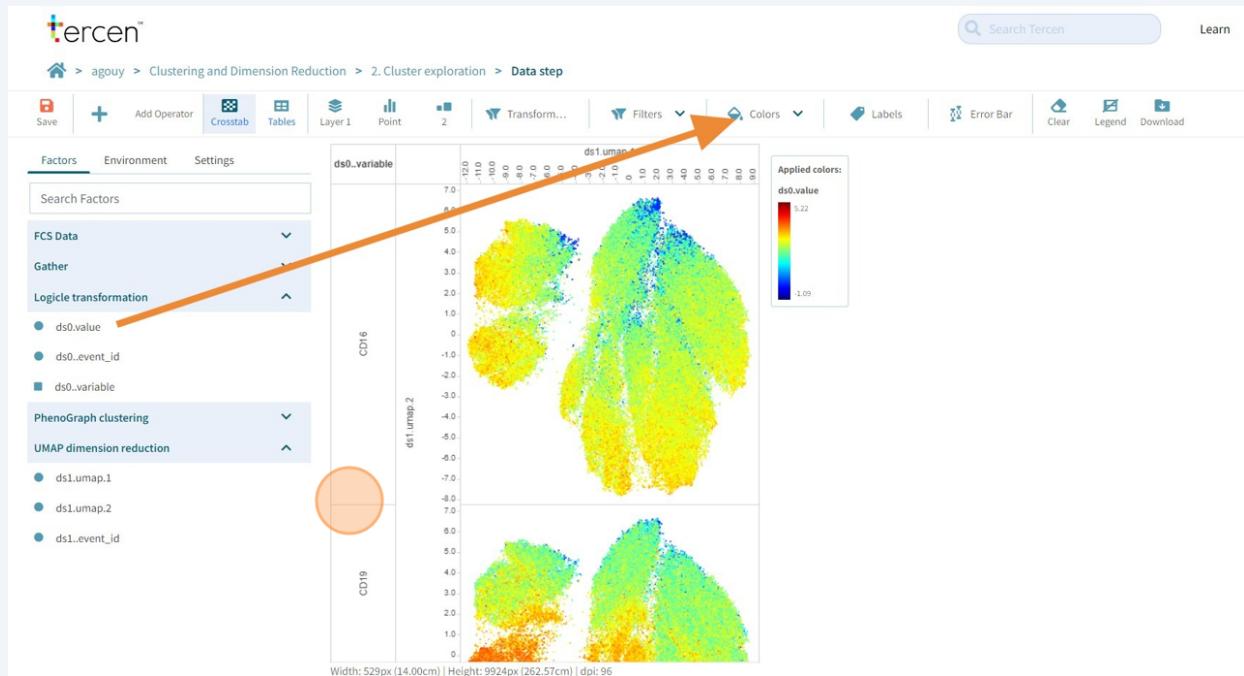
This will stratify the visualisation by marker.



5 From the Logicle transformation group

Drag ds0.value to Colors.

What do you observe?

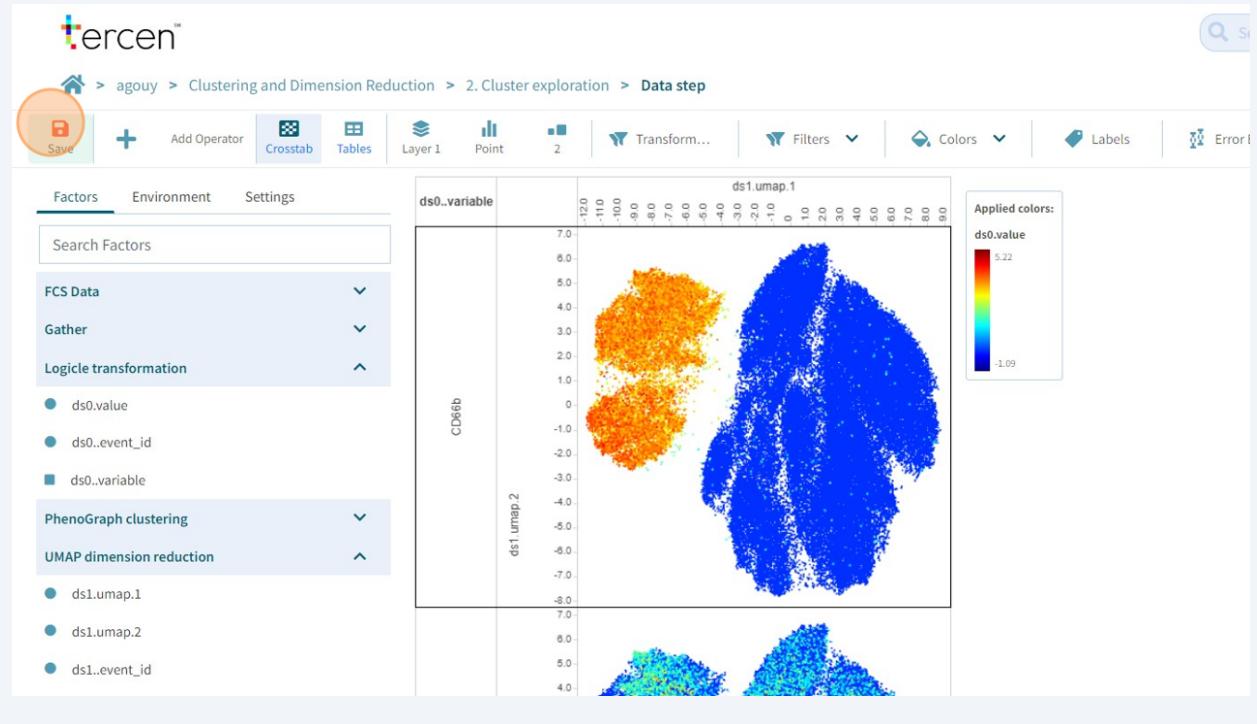


6

You now see the **expression values** on the **UMAP**.

Check the **CD66b marker** for example.

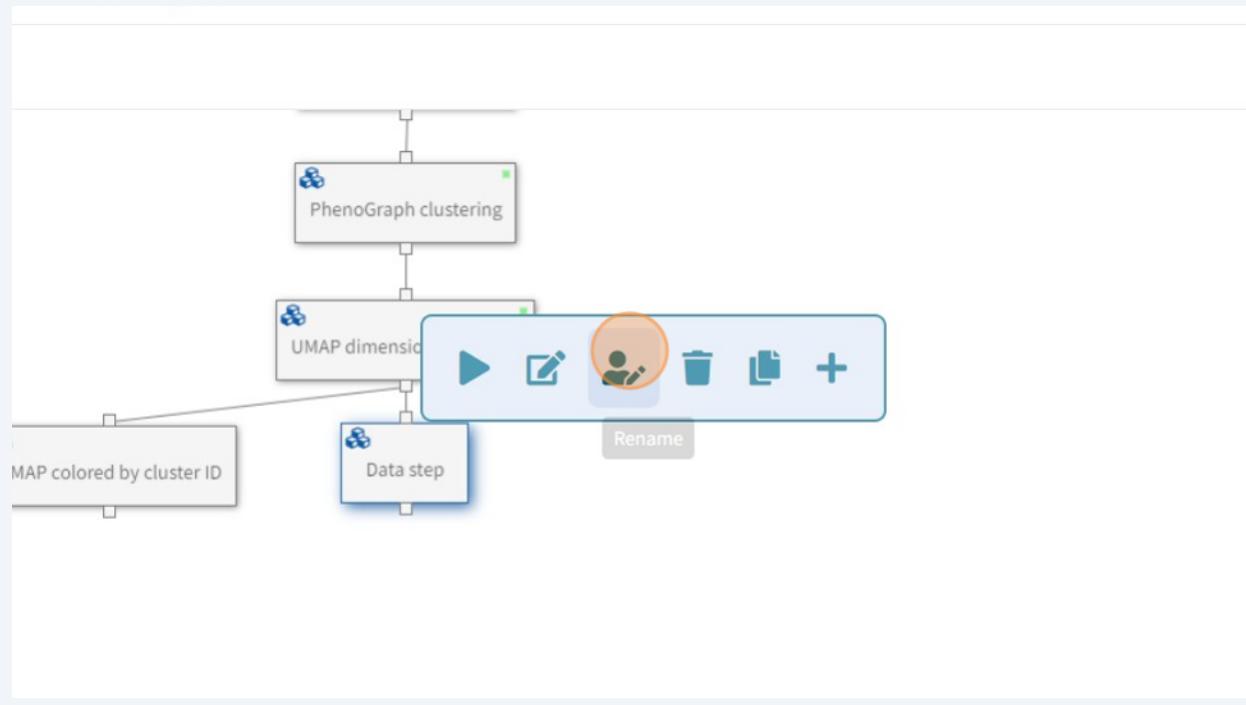
We can clearly see a difference in expression between the clusters on the left and on the right.



7

Save the step.

Go back to the workflow canvas and **Rename** this step to **UMAP vs. channel intensity**.



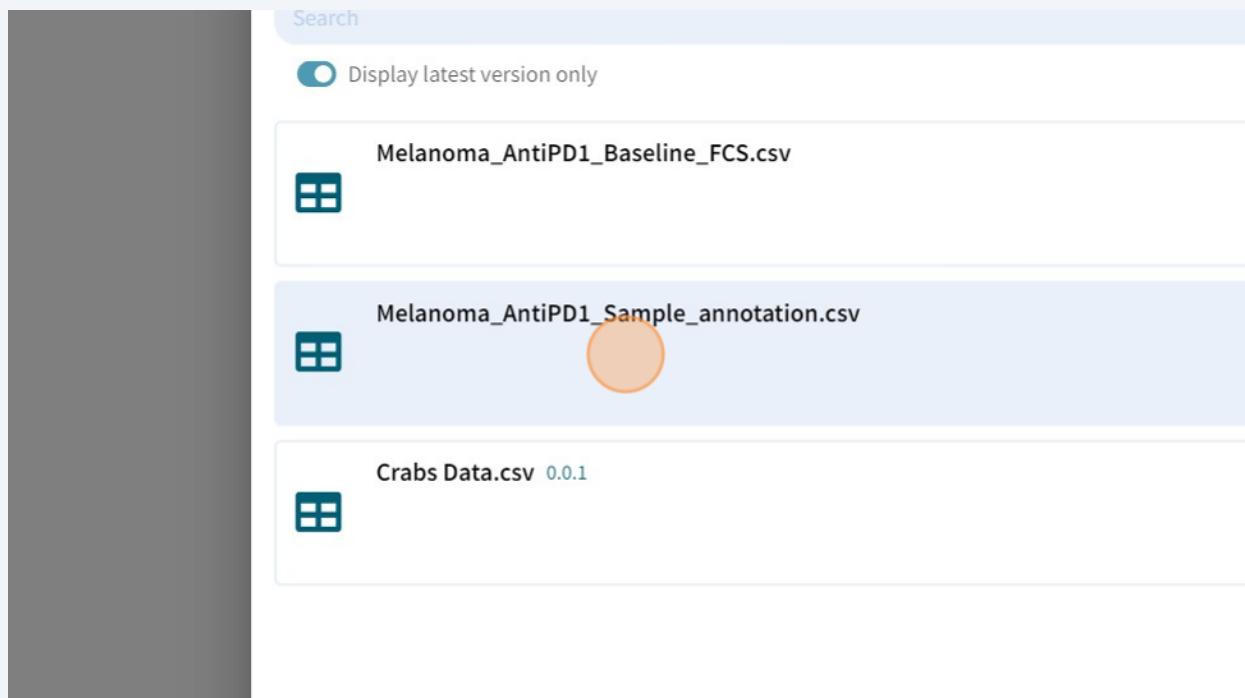
8

We will now add some biological information to our visualisations using a Sample Annotation file.

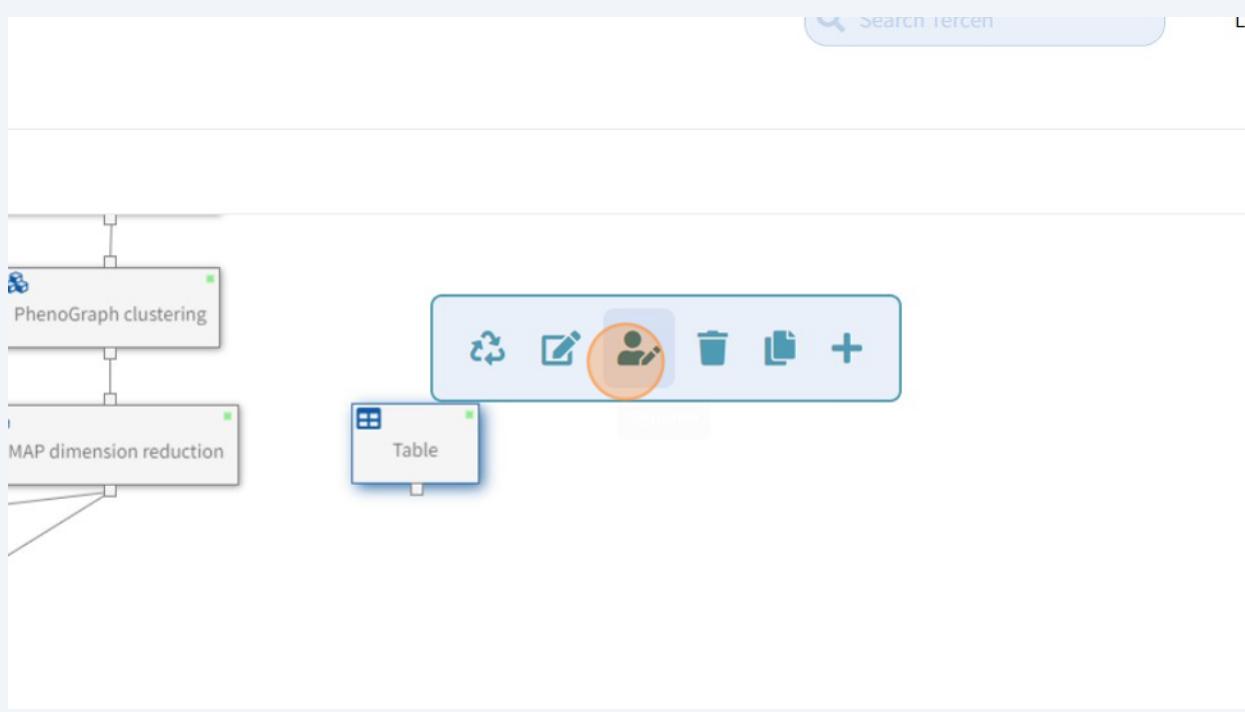
Click "**Add Step**".



- 9 Find the **Melanoma_AntiPD1_Sample_annotation.csv** dataset in the **Library**.



- 10 Rename it **Sample annotation**.



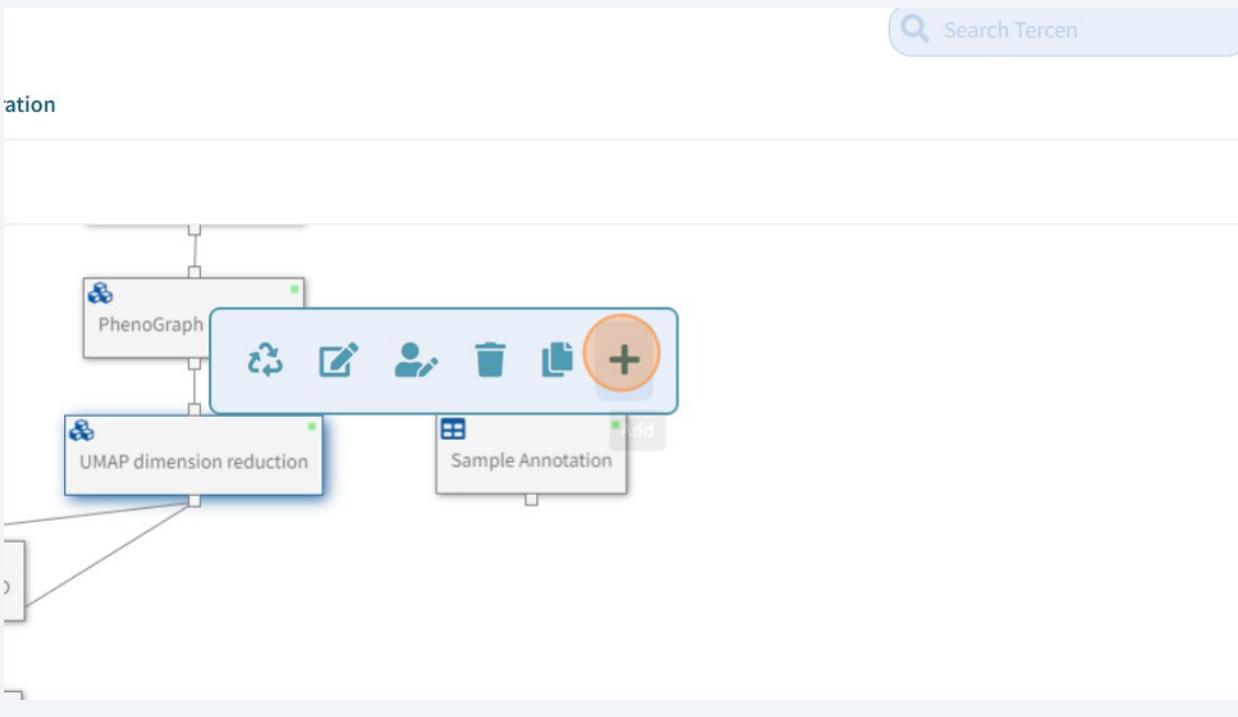
11 You can **open this dataset** to see the information it contains about our samples.

The most important variable is **Response**, which tells us if the patient is a **Responder (R)** or **Non-Responder (NR)** to the anti-PD1 treatment.

#	filename	Patient ID	Response	Patient number	Treatment
1	BASE_CK_2016-06-23_03_NR1.fcs	NR01	NR	1.0	Pre
2	BASE_CK_2016-06-23_03_NR2.fcs	NR02	NR	2.0	Pre
3	BASE_CK_2016-06-23_03_NR3.fcs	NR03	NR	3.0	Pre
4	BASE_CK_2016-06-23_03_NR4.fcs	NR04	NR	4.0	Pre
5	BASE_CK_2016-06-23_03_NR5.fcs	NR05	NR	5.0	Pre
6	BASE_CK_2016-06-23_03_R1.fcs	R01	R	9.0	Pre
7	BASE_CK_2016-06-23_03_R2.fcs	R02	R	10.0	Pre
8	BASE_CK_2016-06-23_03_R3.fcs	R03	R	11.0	Pre
9	BASE_CK_2016-06-23_03_R4.fcs	R04	R	12.0	Pre
10	BASE_CK_2016-06-23_03_R5.fcs	R05	R	13.0	Pre

12 Go back to the workflow canvas.

Join the annotation by adding a **Join step**.



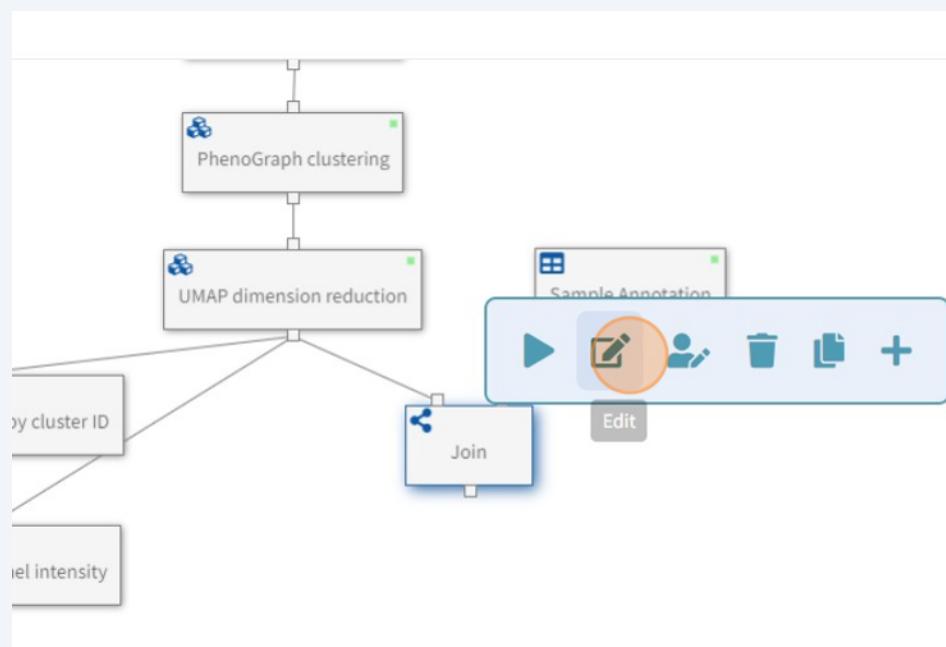
13 Click "Join leftTable".

The screenshot shows a list of data steps in a pipeline editor:

- Data step data**: Perform computation on user defined projection
- Multi data step data**: Perform computation on user defined projection
- Join leftTable**: Join two data sets (highlighted with an orange circle)
- Join rightTable**: Join two data sets
- Gather table**: Gather data from multiple sources into a single table

14 Link the Join Step to the Sample Annotation table.

Then **Edit** the Join step.



15 The key to this join is **filename (character)**

Check the filename factor on both lists.

Save and run the step.

The screenshot shows a software interface with a list of variables on the left and a dropdown menu on the right. The list includes:

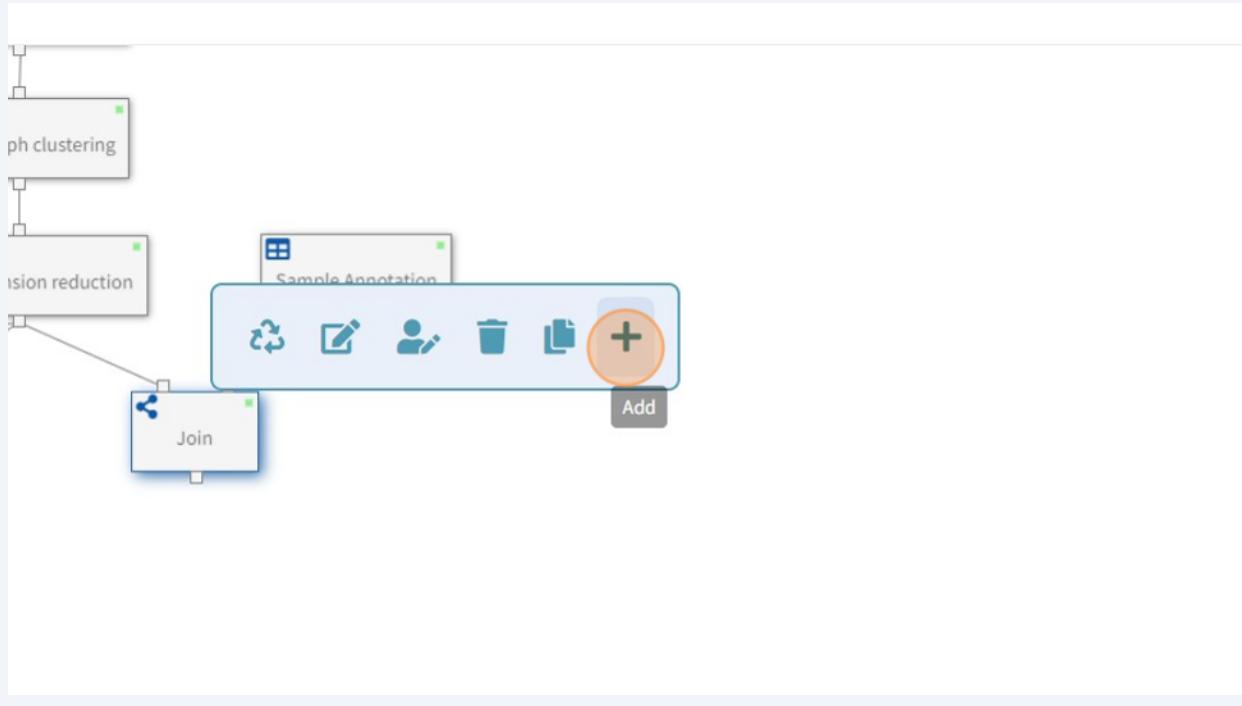
- 19 (numeric)
- 54 (numeric)
- 303 (numeric)
- 34 (numeric)
- 141 (numeric)
- 51 (numeric)
- 123 (numeric)
- 56b (numeric)

To the right, a dropdown menu displays several options:

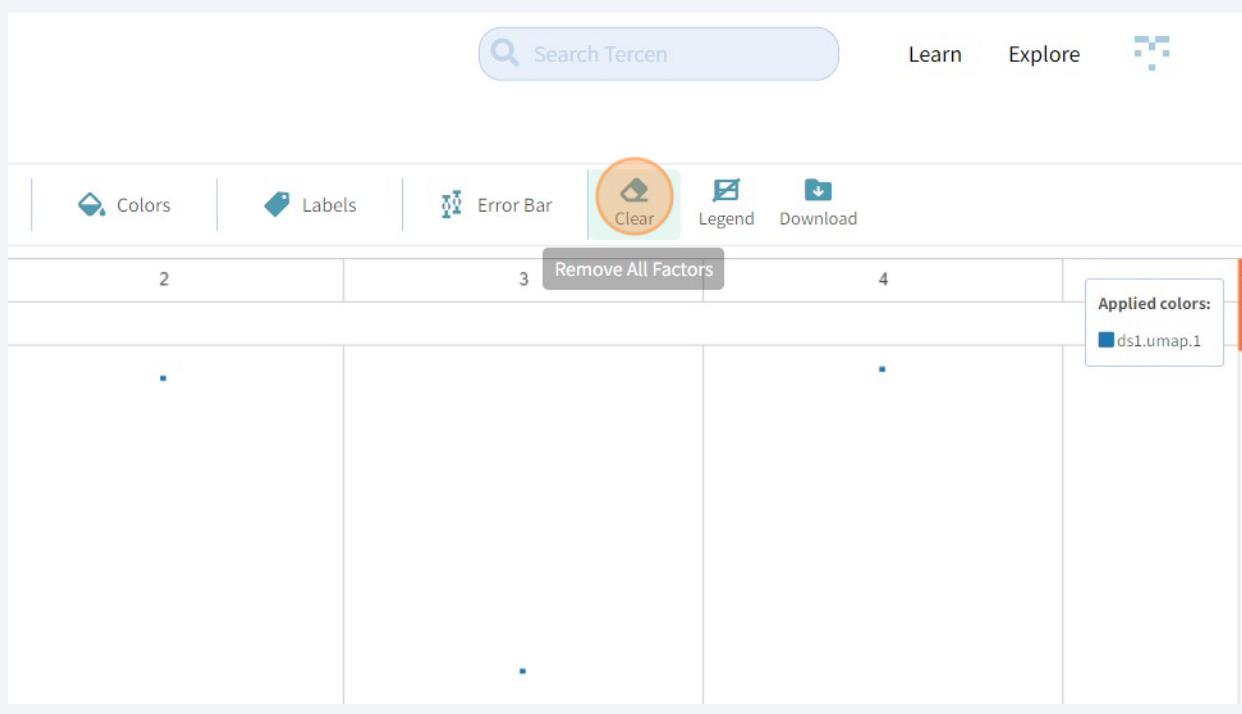
- Select all
- filename (character) (highlighted with an orange circle)
- Patient ID (character)
- Response (character)
- Patient number (numeric)
- Treatment (character)
- 66228a55cfc88e19174264f1fa15657b._rids (numeric)
- 66228a55cfc88e19174264f1fa15657b.tlbd (character)

16 The next visualisation will use the joined data to stratify our UMAP results by the response status of the patients.

Add a data step to the Join.

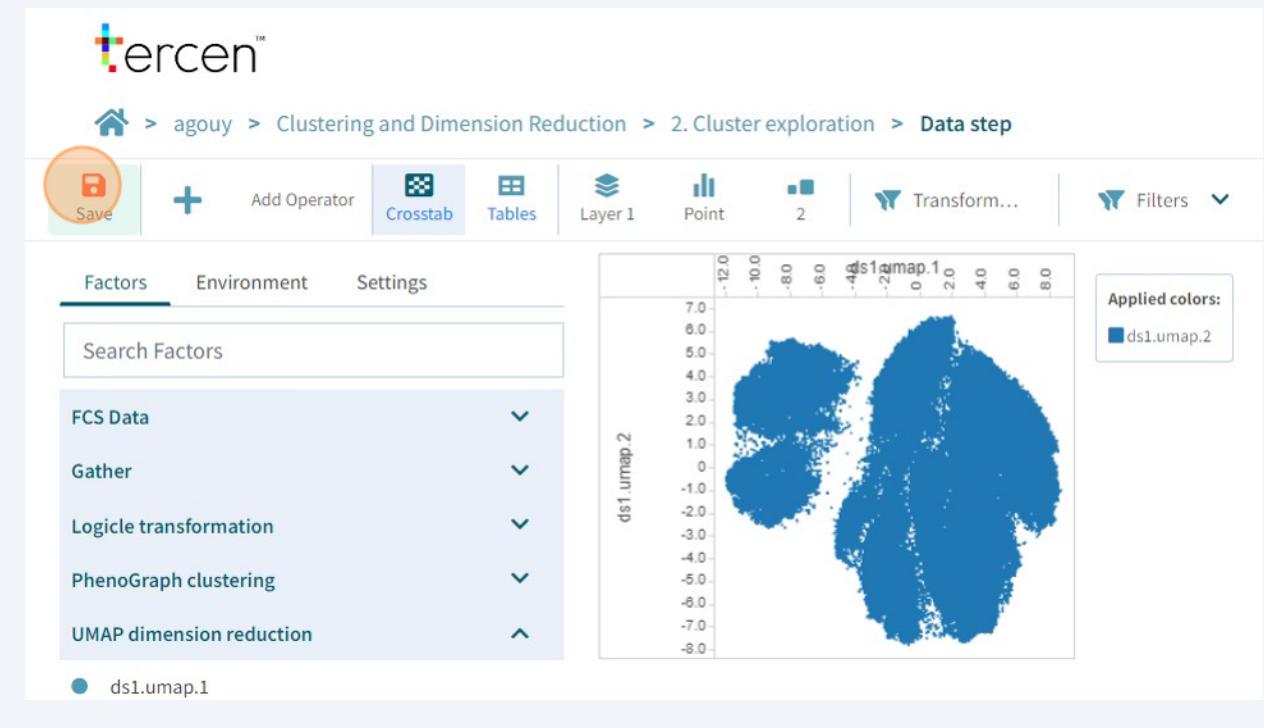


17 Clear the crosstab.

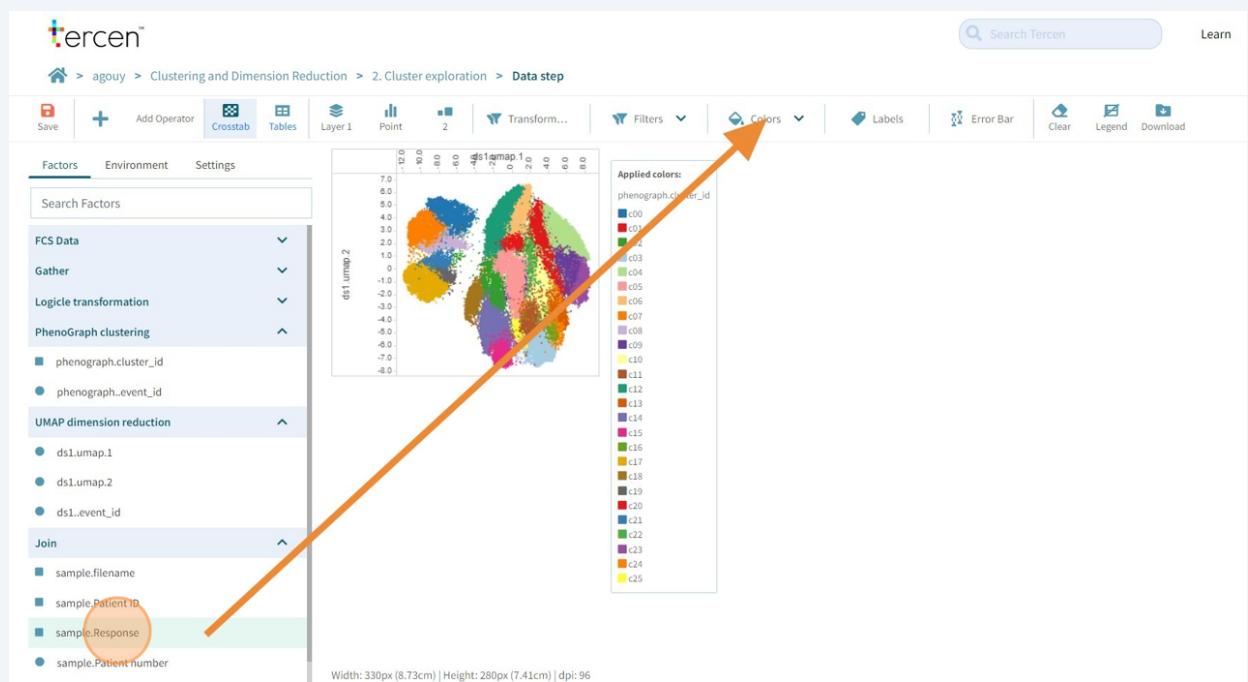


18 Prepare the standard UMAP results projection:

- **umap.2** in the y axis
- **umap.1** in the x axis
- **cluster_id** in the colors



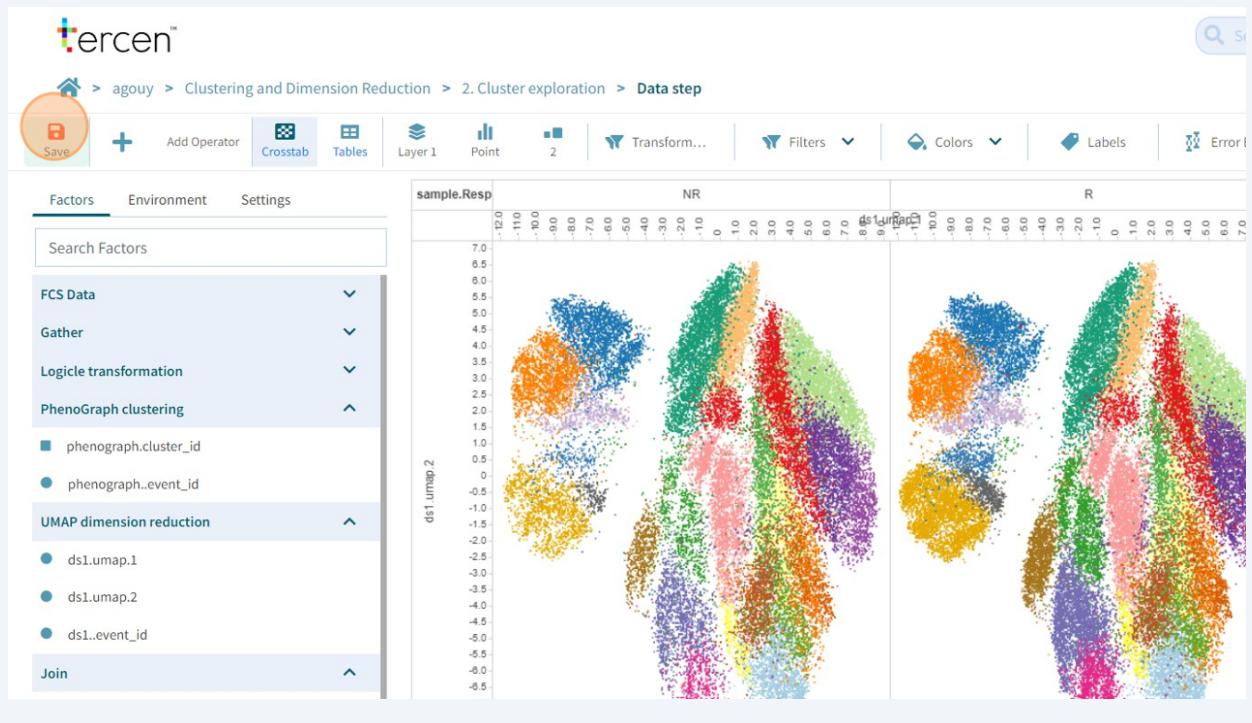
19 Now you can drag and drop the "**sample.Response**" factor to the **Columns**.



20

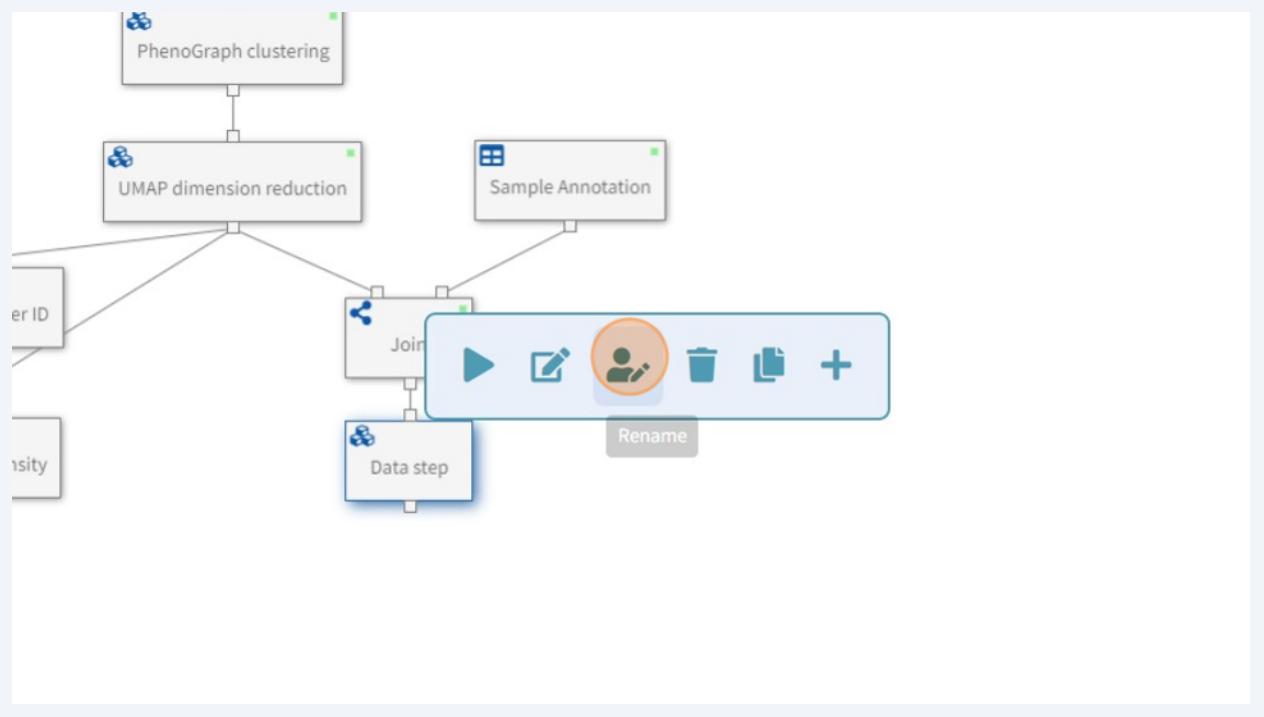
Our data points are now separated according to the **response status** of the patient.

We can get a sense of the broad differences in cluster abundances between the conditions.



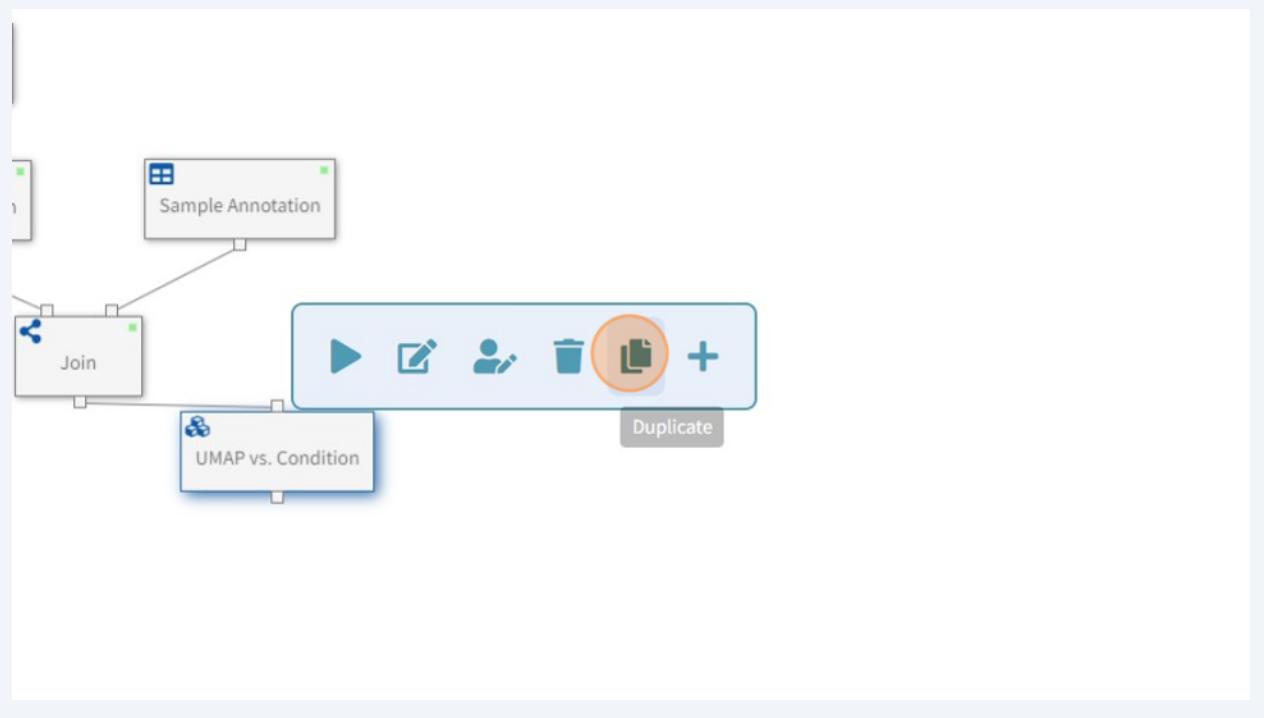
21 Save the step and go back to the workflow canvas.

Rename the step "**UMAP vs. conditions**".

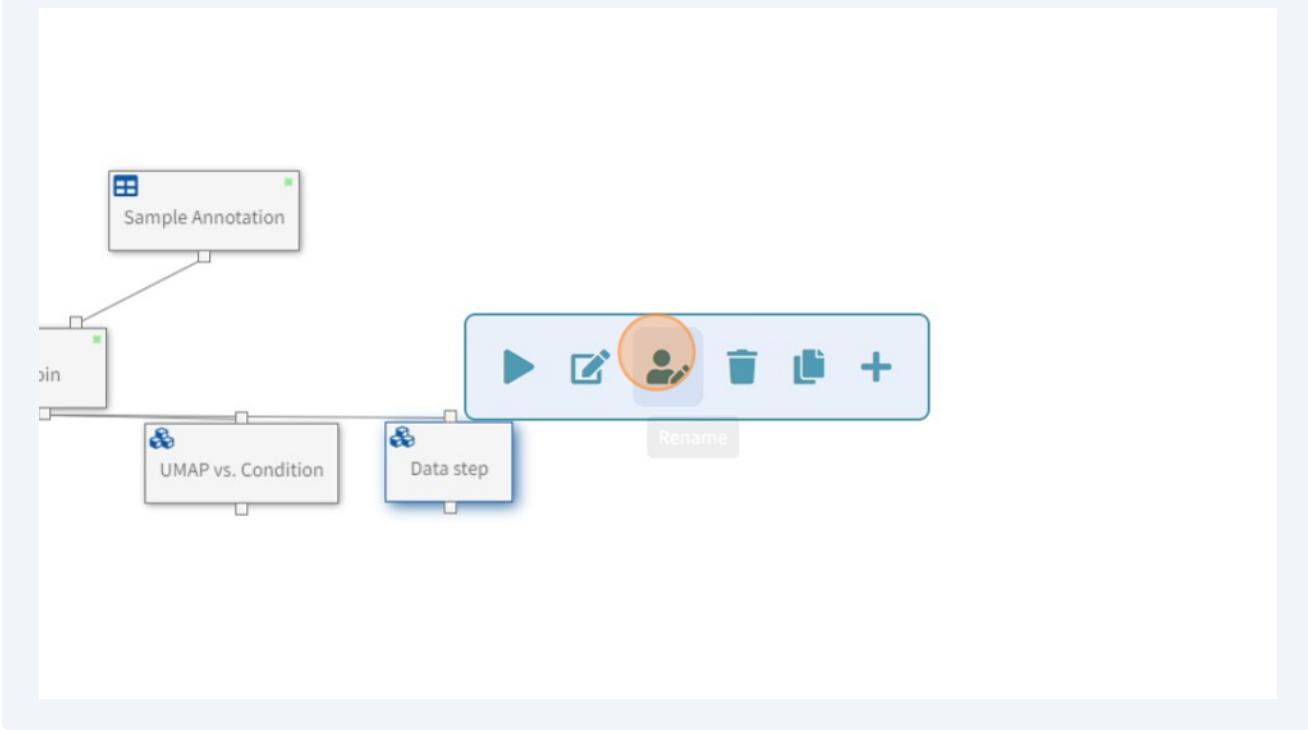


22 Now we will stratify the data with the Patient ID.

Duplicate the UMAP vs. Condition step.



23 Rename it "UMAP vs. sample"



24 Edit the step

Drop the sample identifier (**sample.Patient ID**) to the Columns.

Now we see one UMAP per patient.

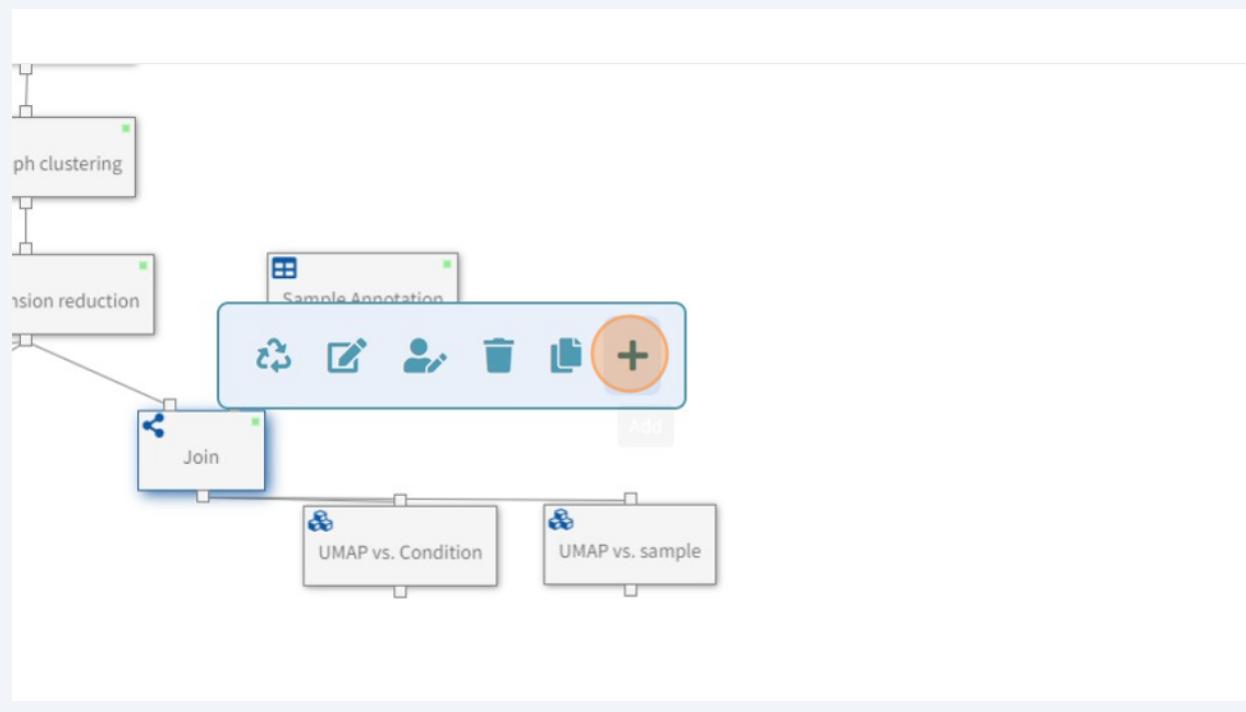
This can be helpful to identify an outlier sample.



25

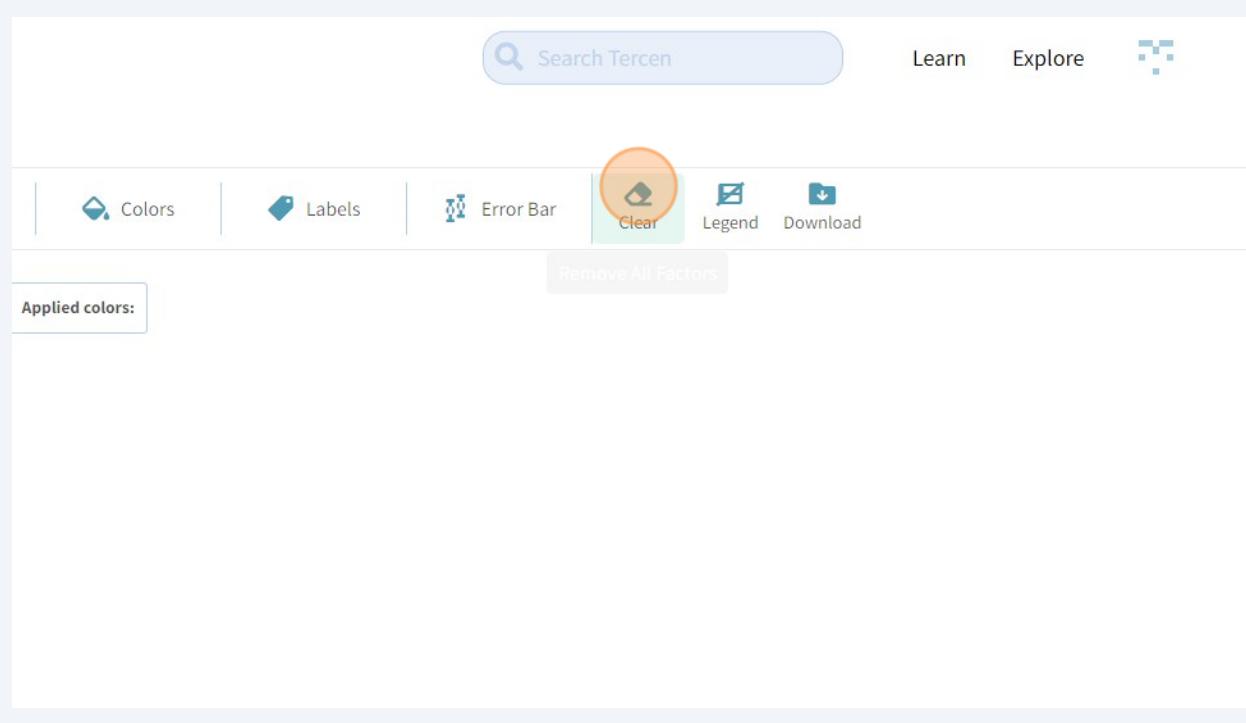
Now we will characterise the clusters in terms of **marker expression** by computing the median expression per cluster.

Add a data step after the **Join** step.



26

Clear the crosstab.



27

Prepare the input projection to compute the median expression per cluster:

From Logicle transformation

- **ds0.value to y axis**
- **ds0.variable to Rows**

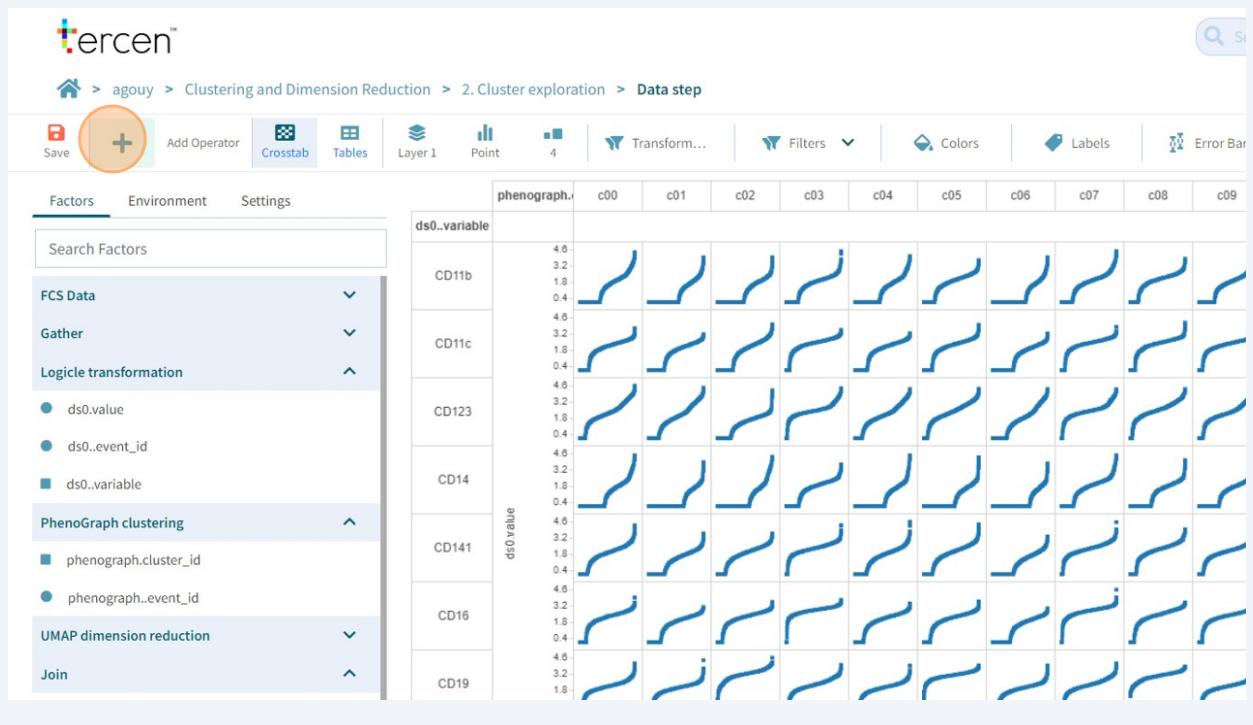
From PhenoGraph clustering

- **phenograph.cluster_id to Columns**

The screenshot shows a software interface with a sidebar on the left and a main workspace on the right. The sidebar has three tabs: 'Factors' (which is selected and highlighted in blue), 'Environment', and 'Settings'. Below the tabs is a search bar labeled 'Search Factors'. Under the 'Factors' tab, there is a dropdown menu with the following options:

- FCS Data
- Gather
- Logicle transformation
- PhenoGraph clustering
 - phenograph.cluster_id (highlighted and circled in orange)
 - phenograph.event_id
- UMAP dimension reduction
- Join
 - sample.filename
 - sample.Patient ID
 - sample.Response
 - sample.Patient number

28 Add a new operator.



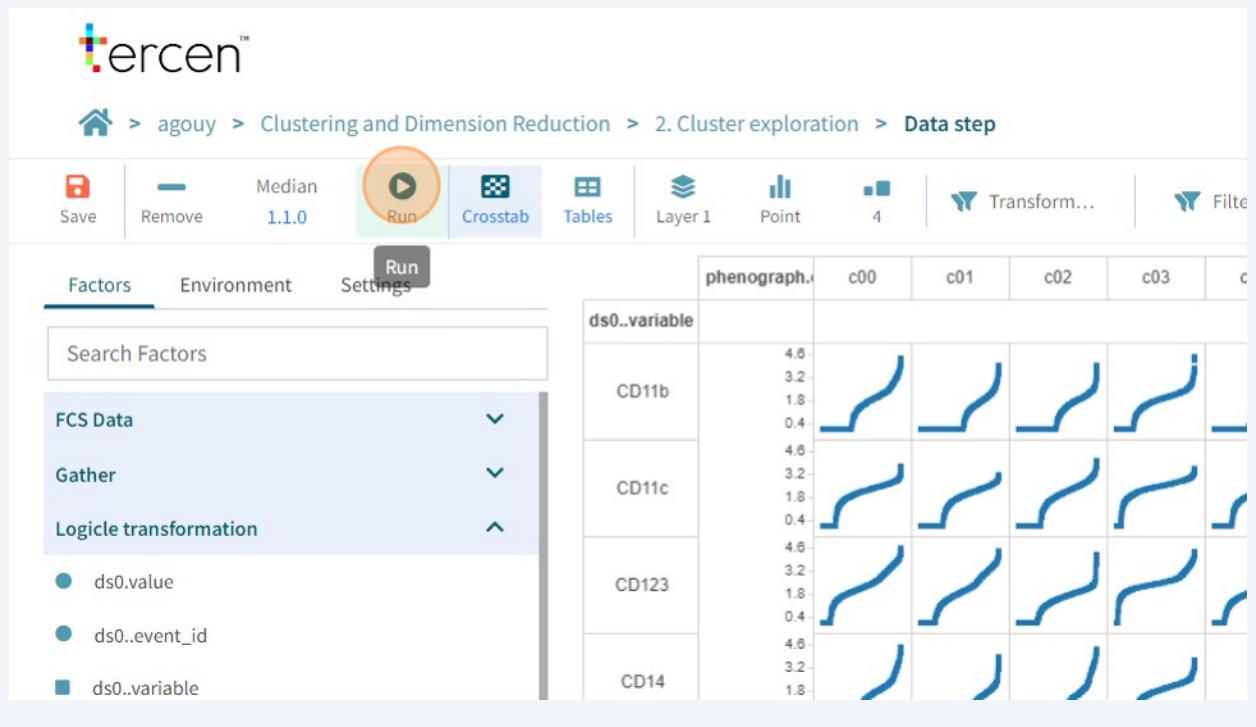
29 Find the Median operator in the Library.

The screenshot shows the RStudio package library interface. At the top, there are tabs for "Installed", "Library" (which is selected and highlighted in blue), and "Log". Below the tabs, there is a search bar containing the text "median". There is also a checkbox for "Display latest version only" which is checked. A "Tag list" button is visible. The main area displays two packages: "Median 1.1.0" and "Median and MAD 0.0.1".

Median 1.1.0
R Calculates the middle value of a range. The value that has an equal number of values above and below it.
basic statistics

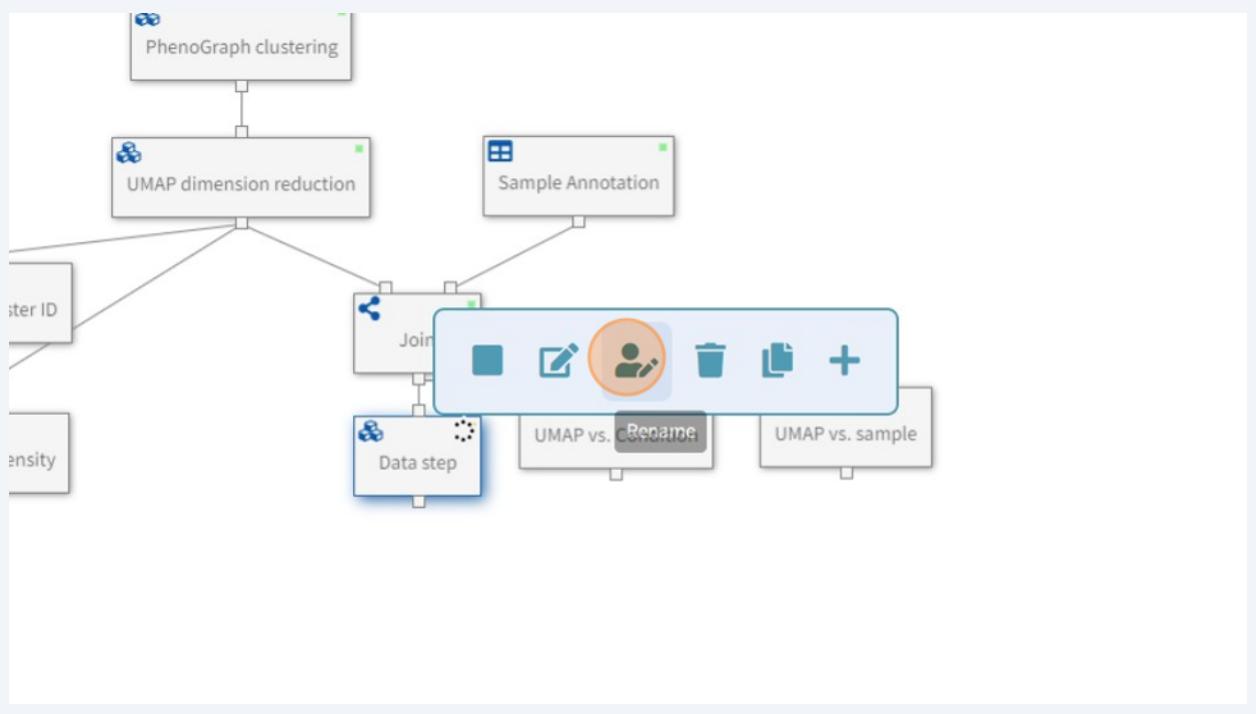
Median and MAD 0.0.1
R Compute the median and Median Absolute Deviation (MAD) of the data points in a cell.
basic statistics

30 Run the operator.

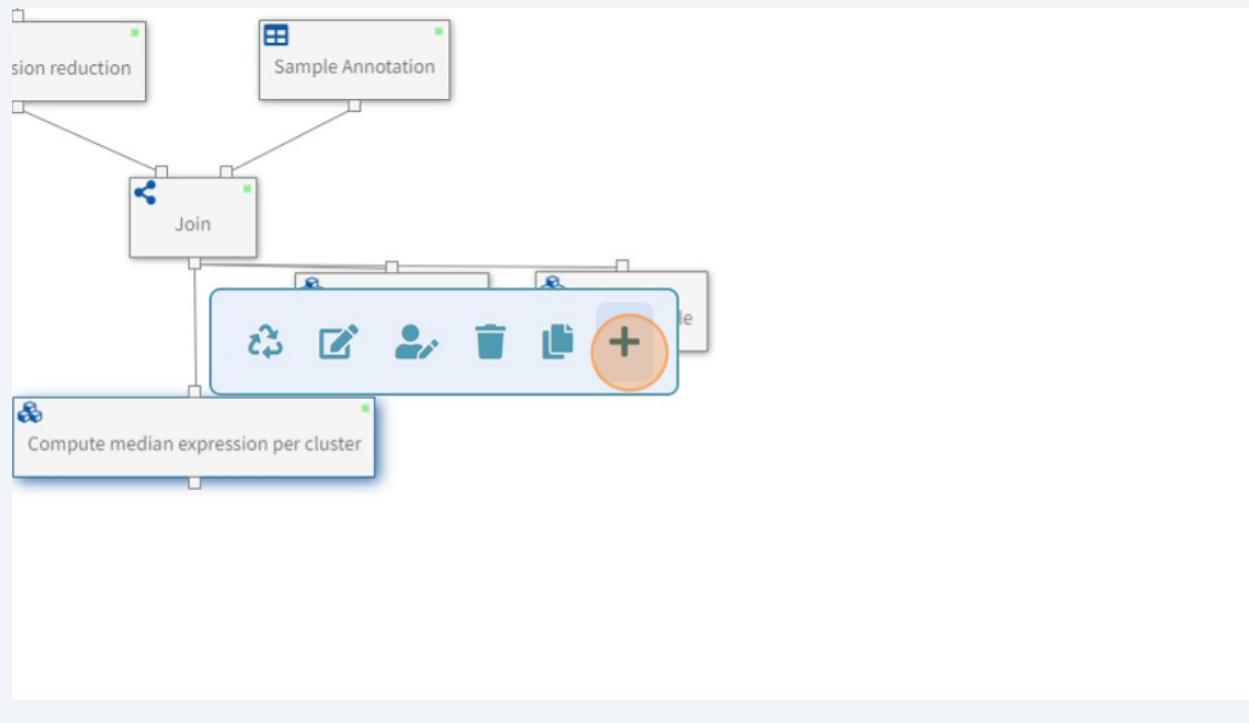


31 Navigate back to the workflow canvas

Rename the step to "**Median expression per cluster**".



32 Add a **new data step** to visualise the results.



33

We now see the **median expression value** per marker, per cluster.

Let's improve this view.

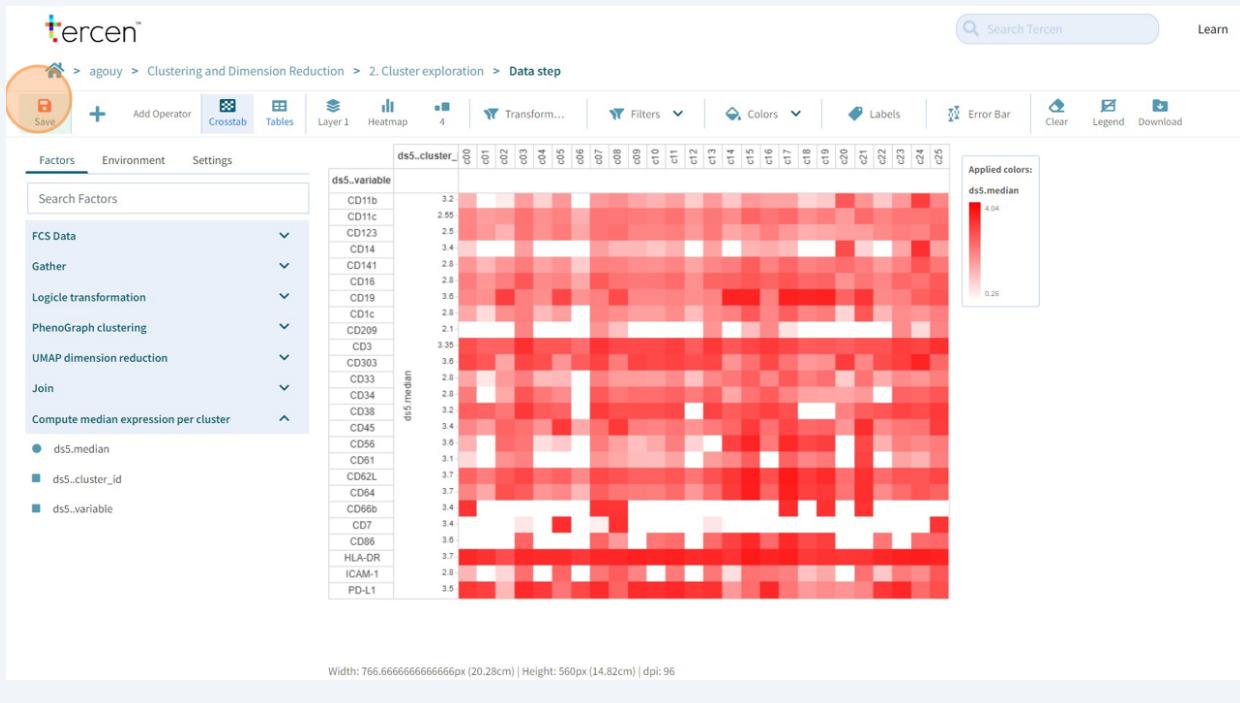
Drag and drop the **median** factor to the **Colors**.

Change the chart type to "**heatmap**".



34

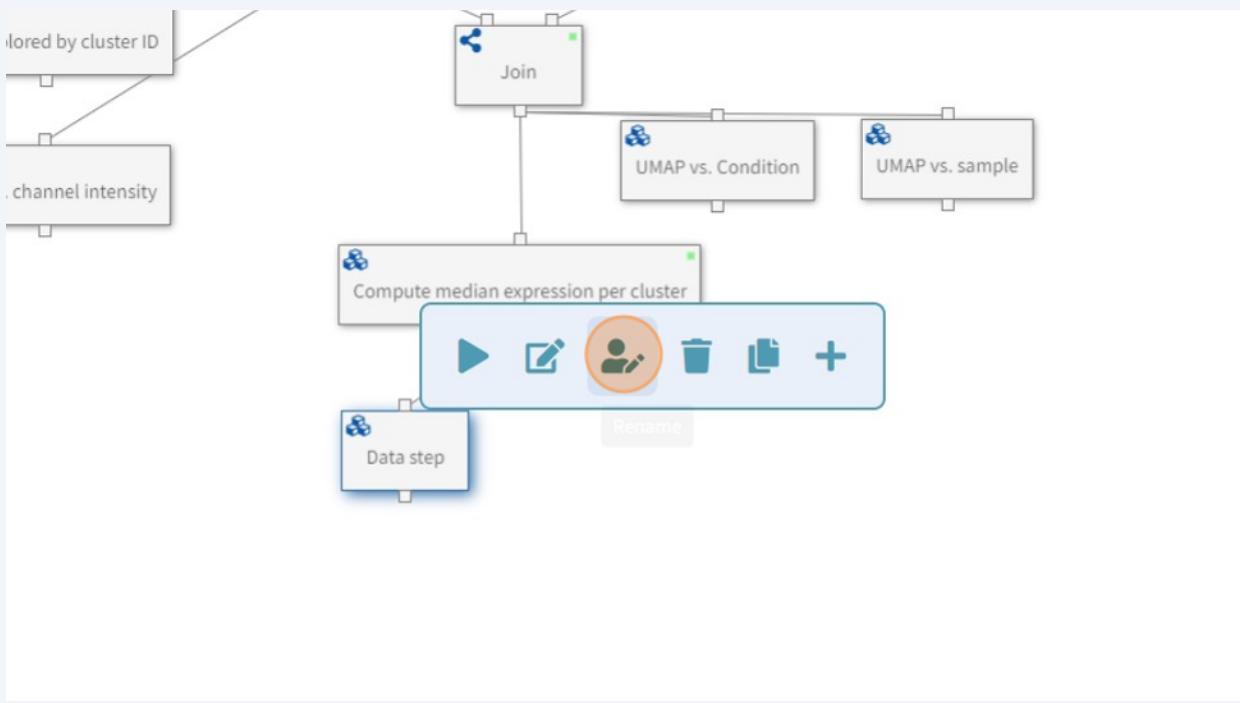
You can now see on the **heatmap** the median expression values of each cluster for each marker.



35

Return to the workflow canvas.

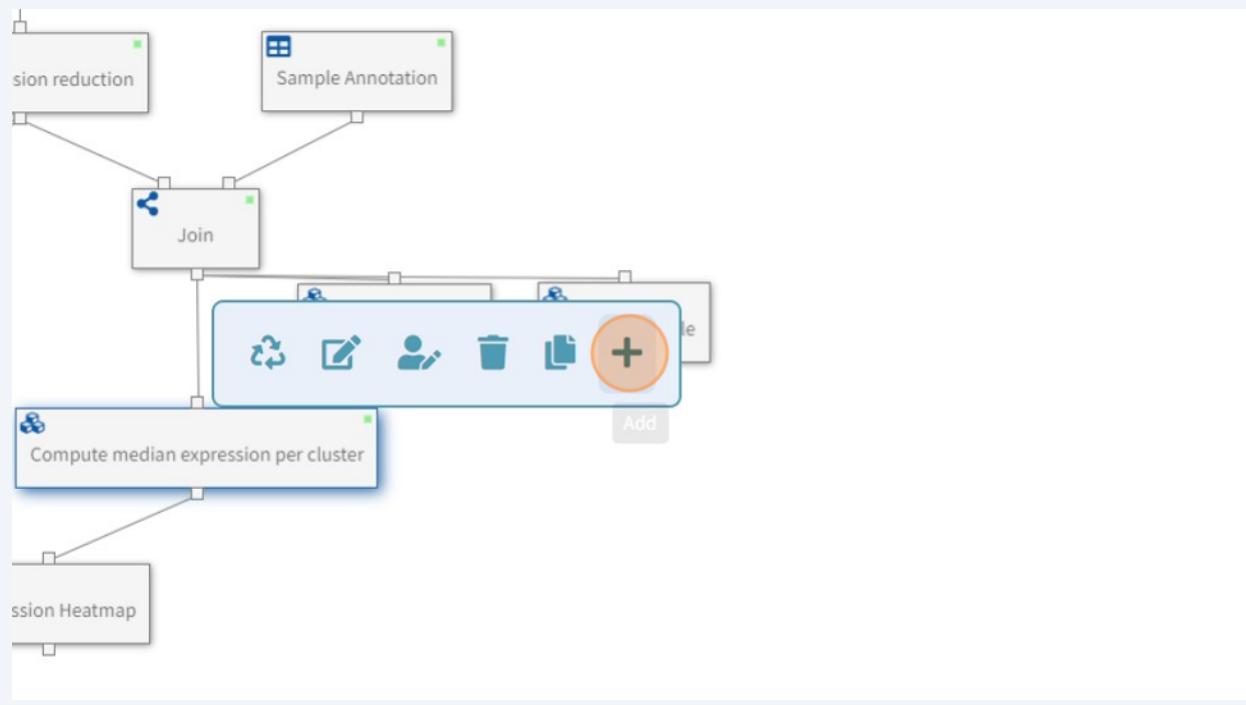
Rename the step to **Expression Heatmap**.



36

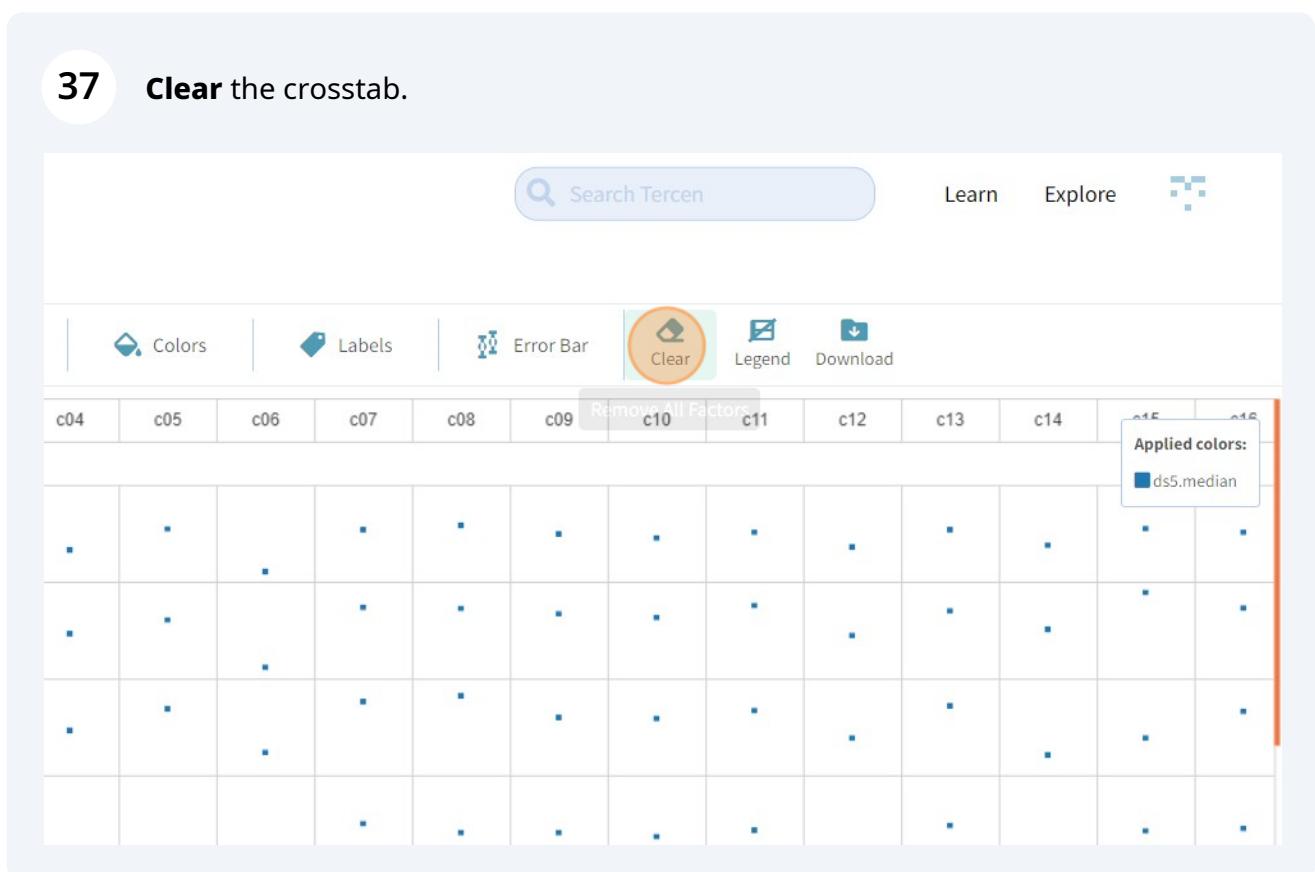
We will now produce **another visualisation**.

Add a **new data step** after the median computation step.



37

Clear the crosstab.



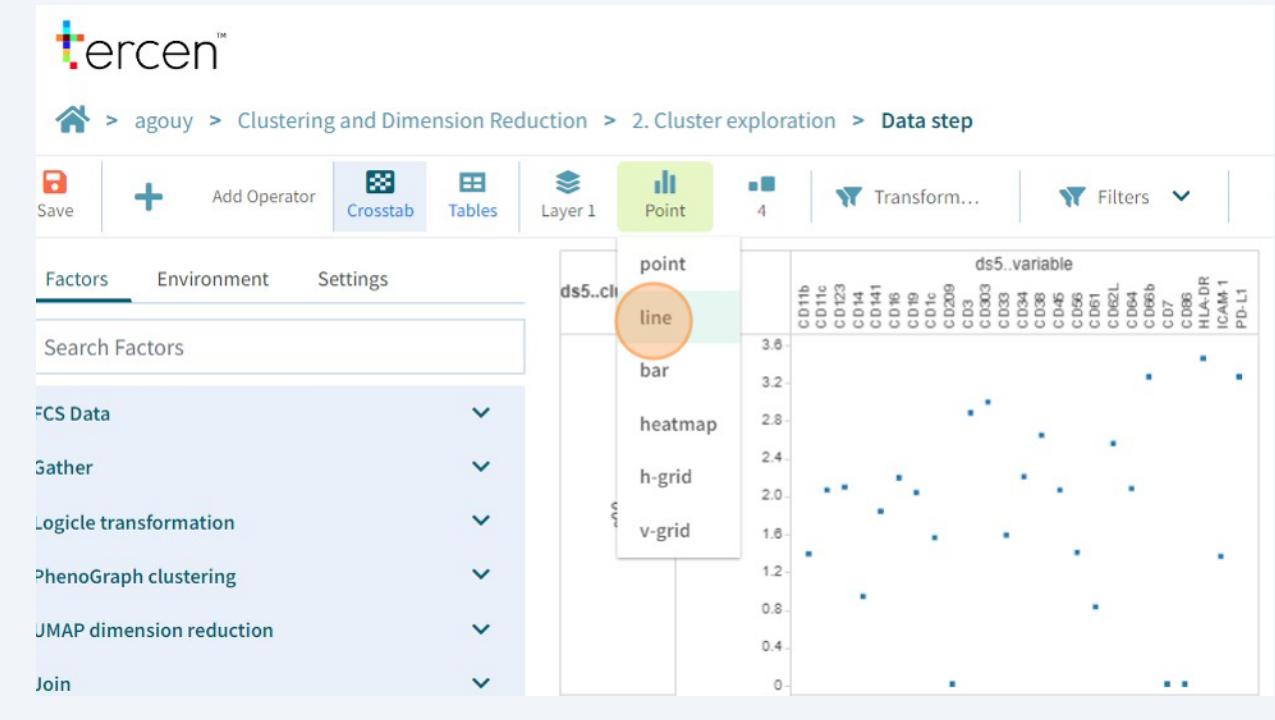
38

We will display the marker **expression profile** of each **cluster**.

To do so, you can **drag and drop**:

- **median** to the **y axis**
- **variable** to the **x axis**
- **cluster_id** to the **rows**

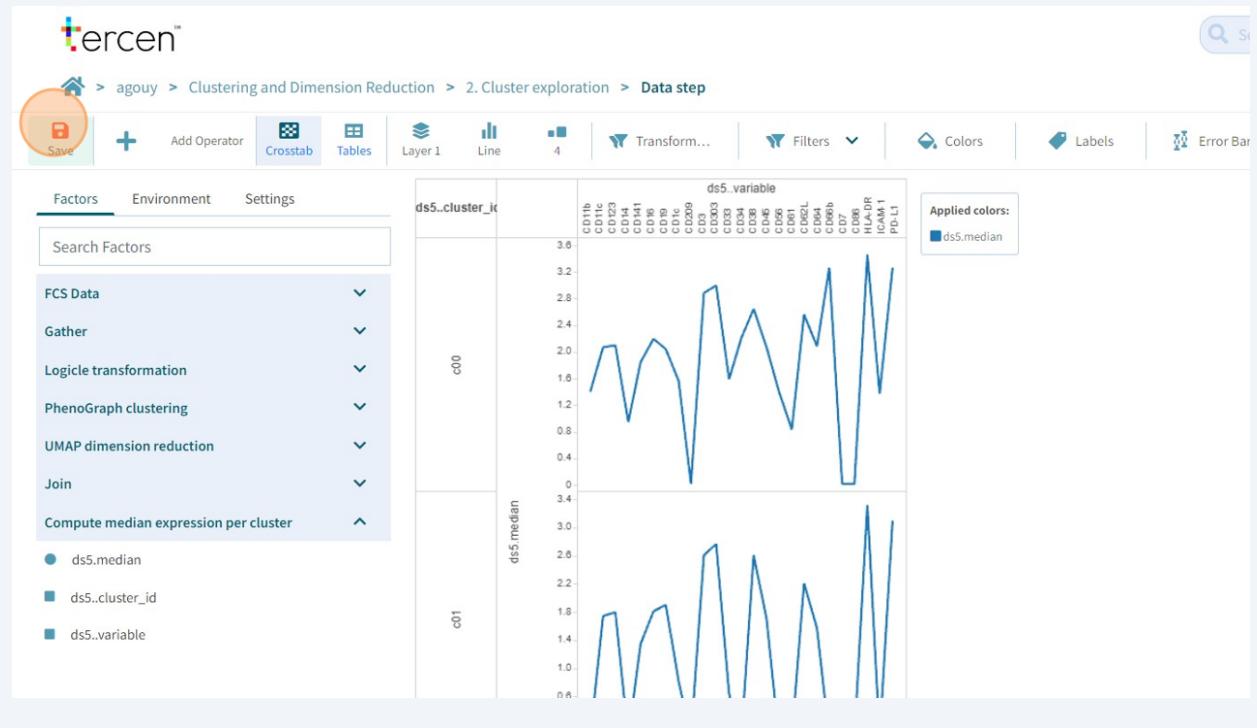
Change the chart type to "**line**".



39

You can now see the **expression profiles** of each clusters. It shows the average expression level of each marker within each cluster, i.e. the phenotype of the cell population underlying it.

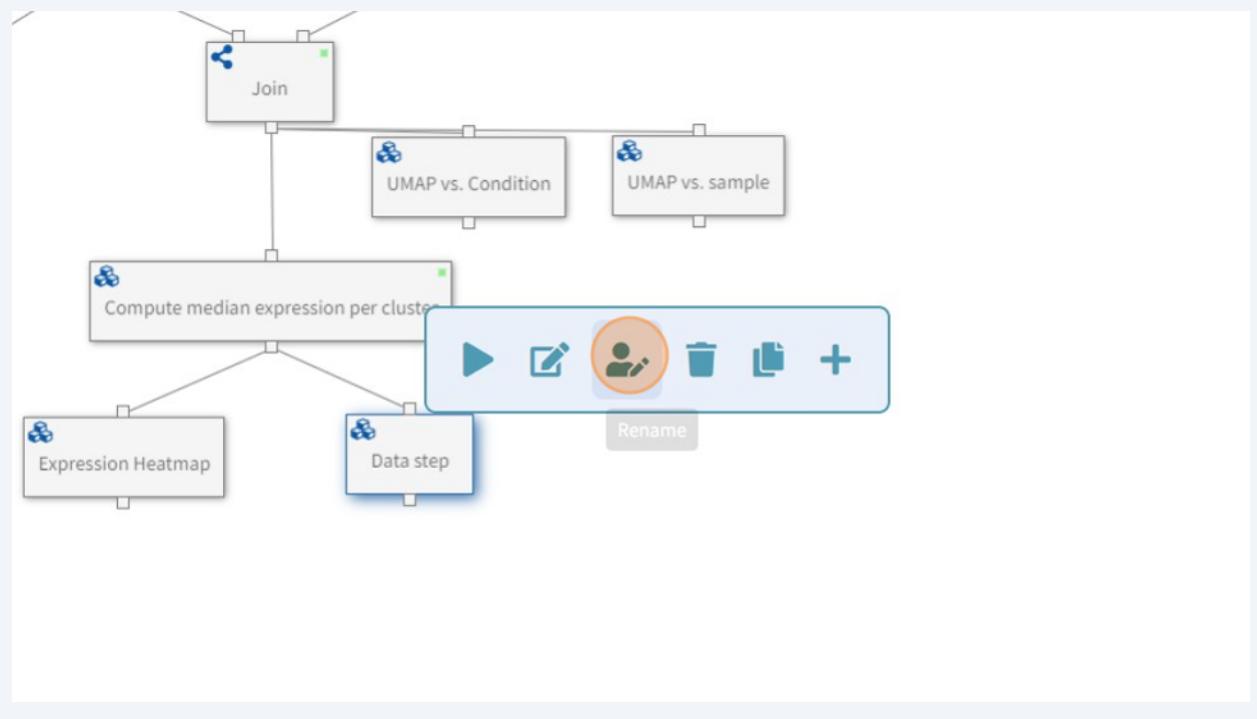
Do not forget to **Save** the workflow!



40

Back to the workflow canvas

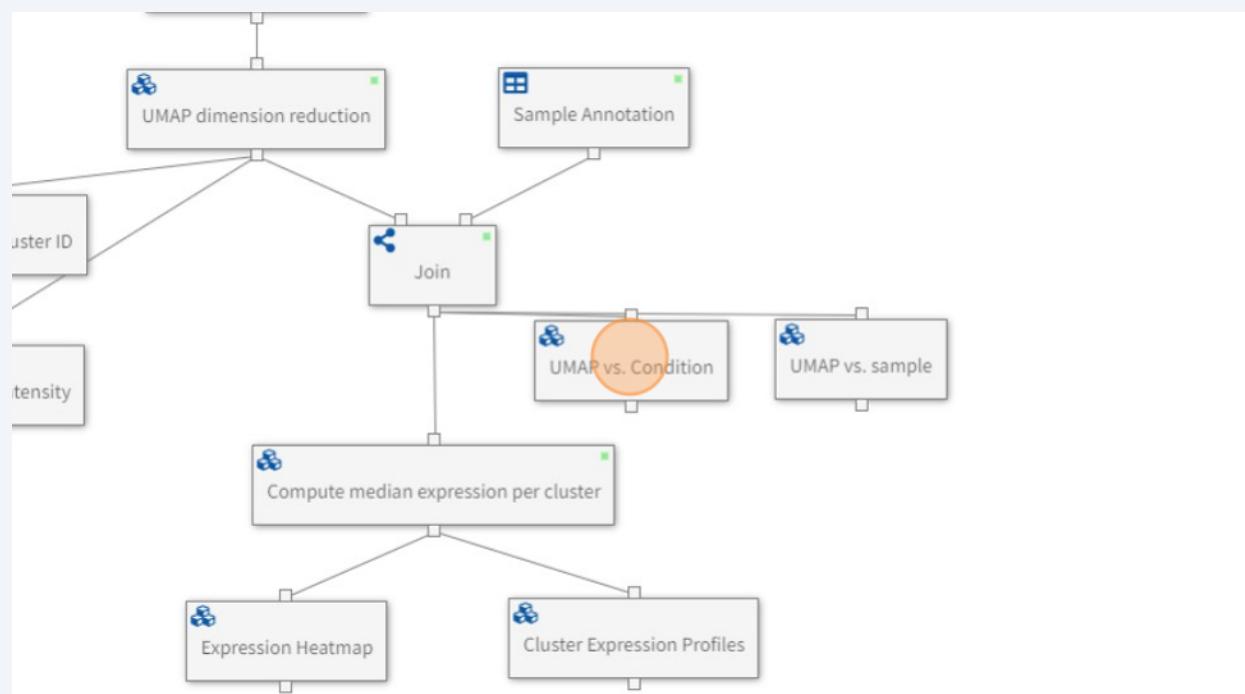
Rename the data step to "**Cluster expression profiles**".



41

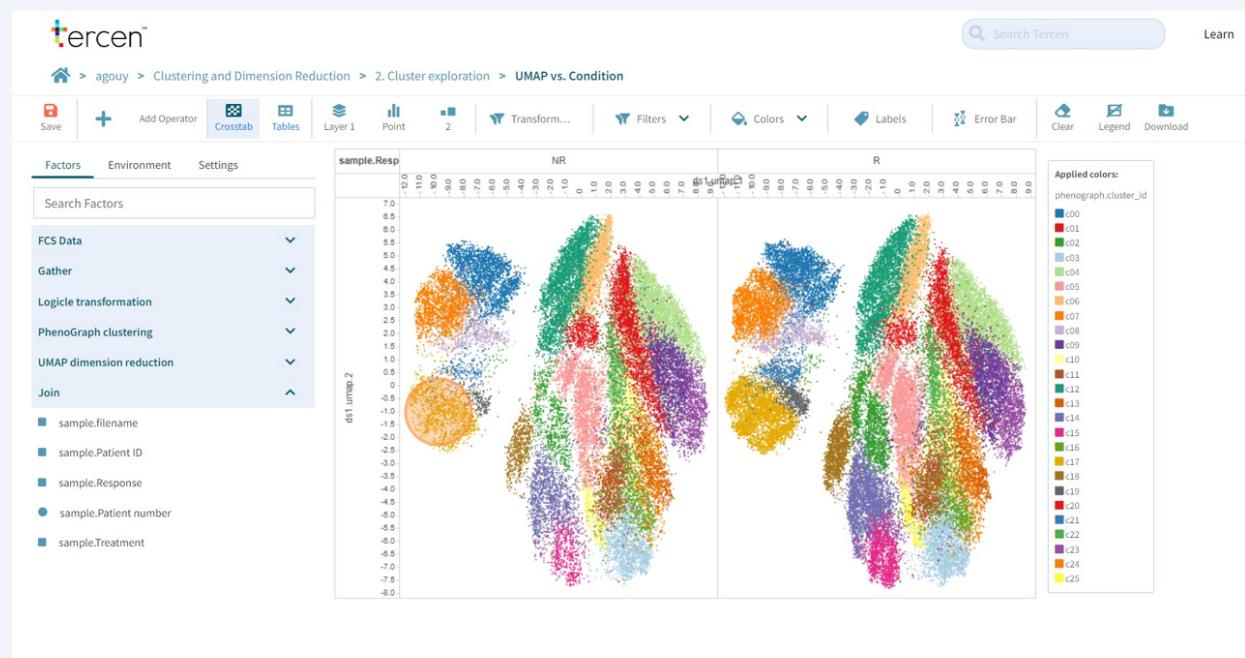
Let's explore further our dataset.

Open the **UMAP vs. Condition** step.



42

We will investigate a specific cluster, **c17**, and try to understand which cell type underlies it.





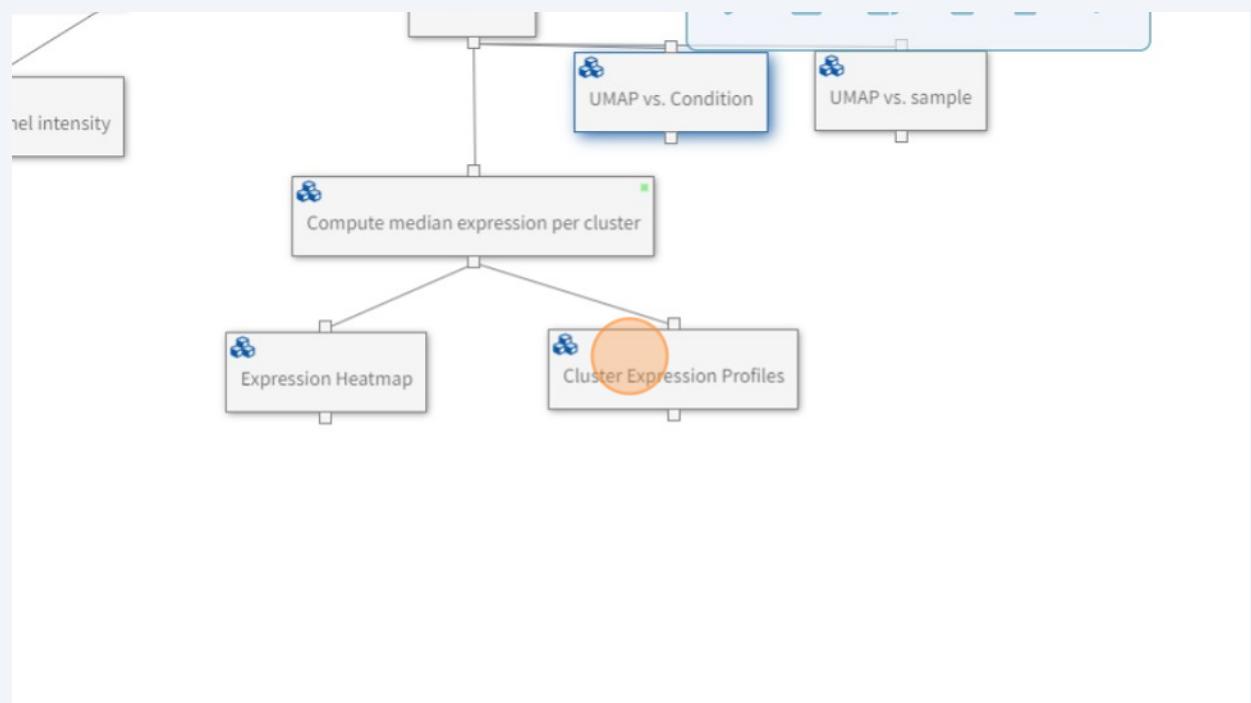
Here, we are **visually exploring** our dataset.

Even though we can speculate on differences between conditions, results should be interpreted with caution. To draw **formal conclusions** on the differences between conditions, a **more robust** analysis should be performed.

Differential analysis will be addressed this in the next workshops!

43

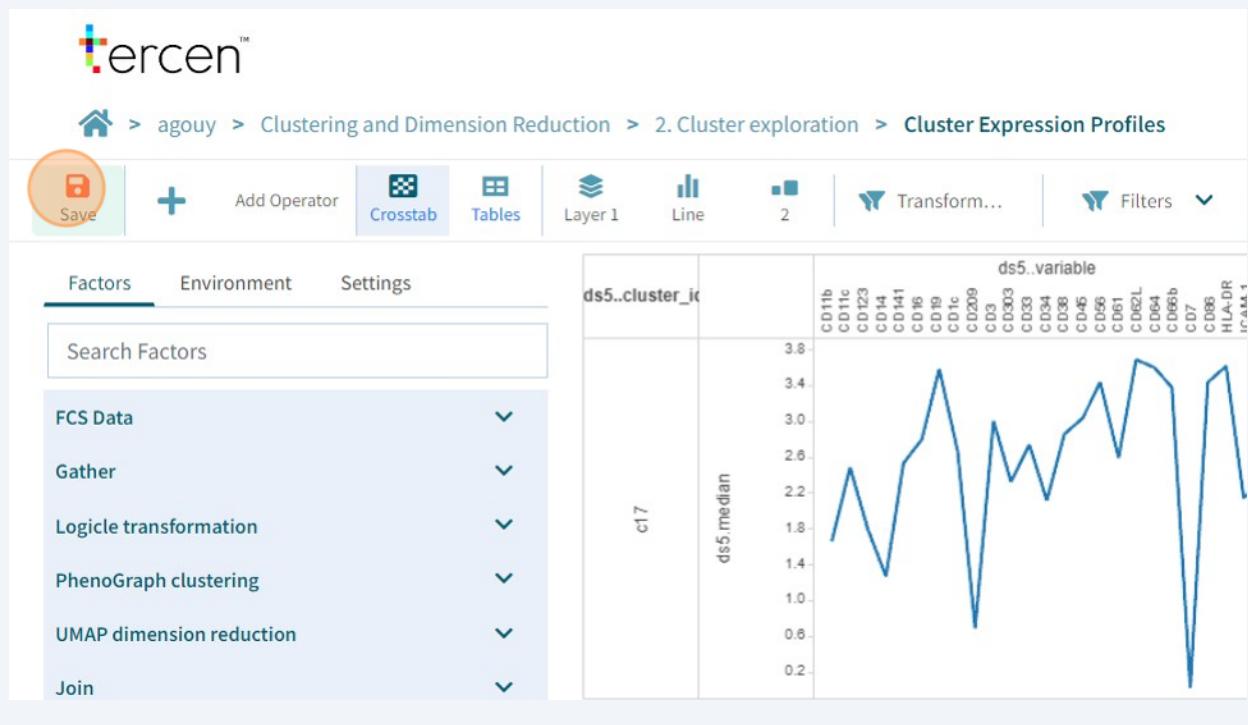
Let's check c17 on the **Cluster Expression Profiles** step.



44

Examine the Cluster 17 expression profile.

Can you guess the cell type?



These are the kind of signals we can observe through exploratory analysis.

This type of analysis confirms the presence of interesting patterns in the data but further analyses are required to formally assess differences between groups.

Future workshops will introduce more structured methods to perform differential analysis of patient groups and better characterise cell types (which is not easy to do for high-dimensional data).