t-SNE

Two-dimensional data projection with t-SNE.

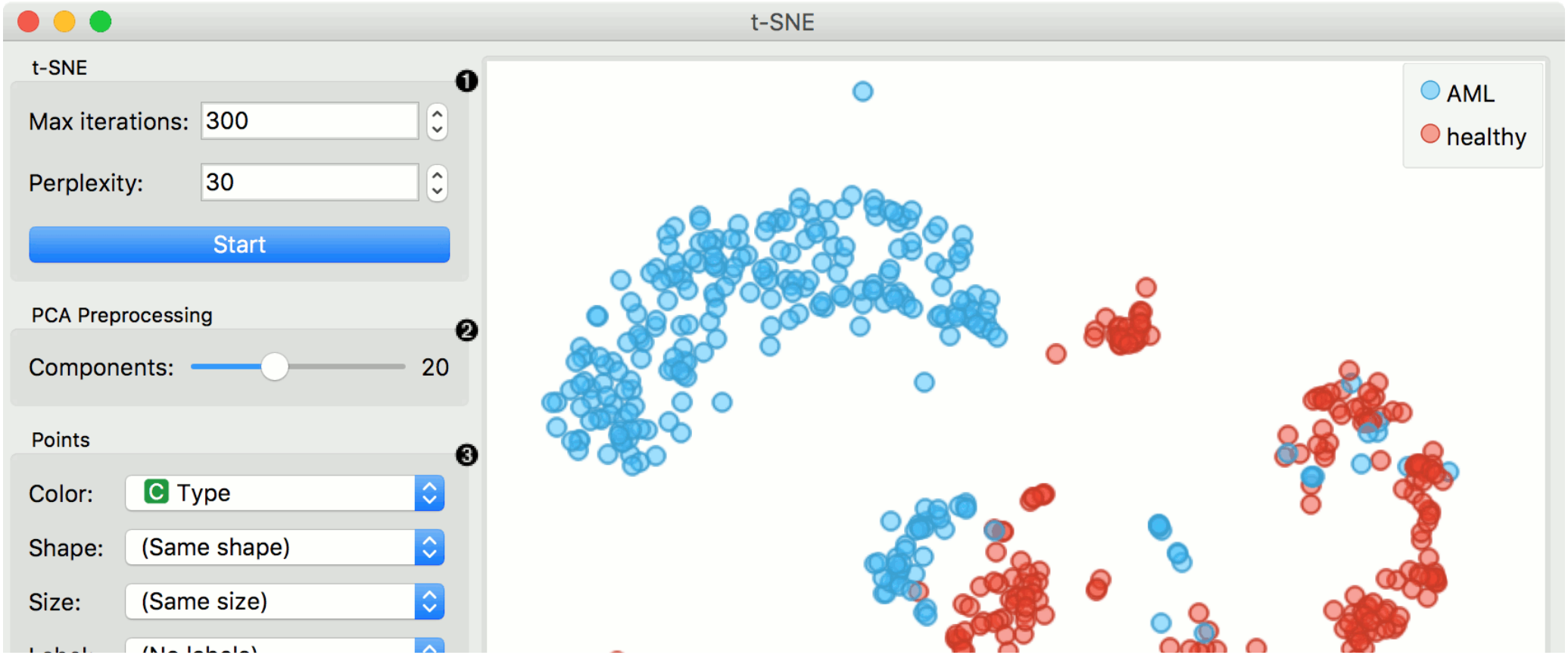
Inputs

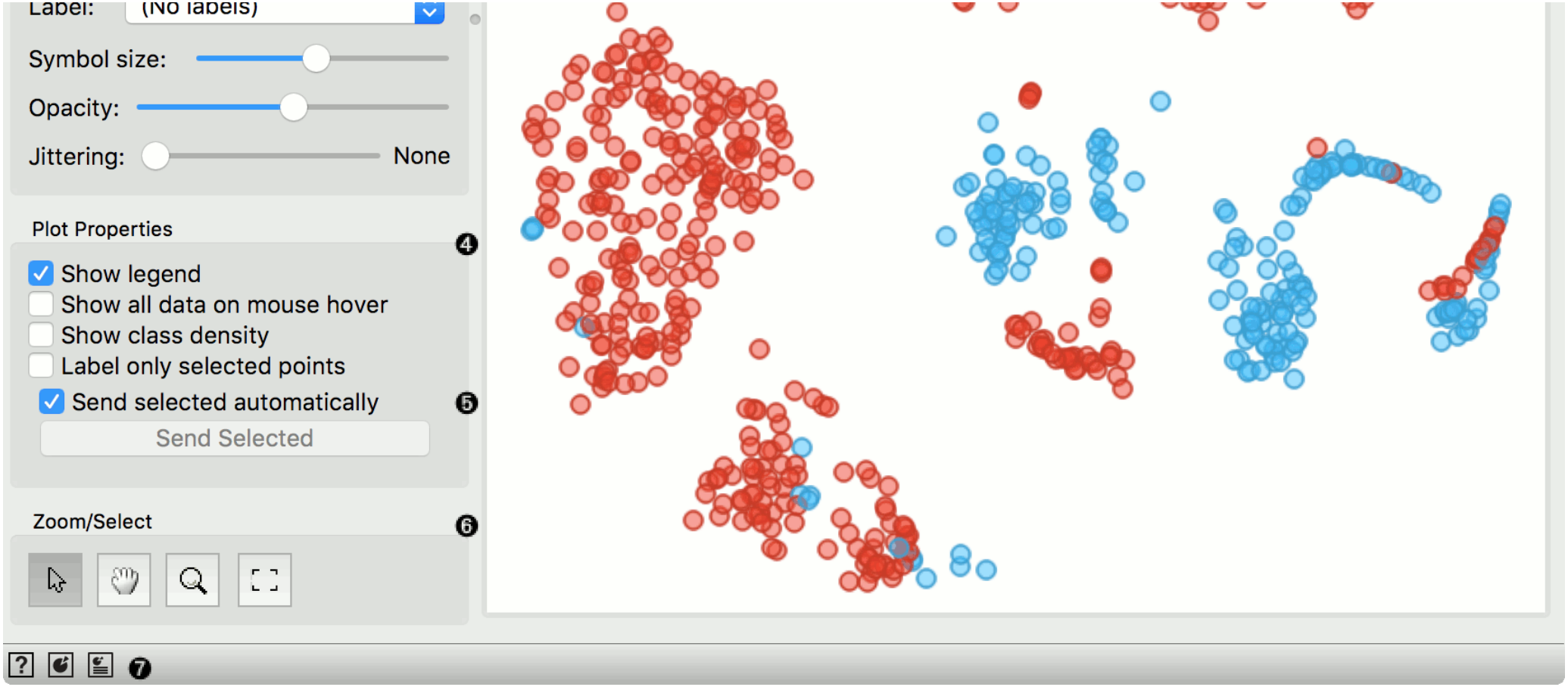
- Data: input dataset
- Data Subset: subset of instances

Outputs

- Selected Data: instances selected from the plot
- Data: data with an additional column showing whether a point is selected

The **t-SNE** widget plots the data with a t-distributed stochastic neighbor embedding method. t-SNE is a dimensionality reduction technique, similar to MDS, where points are mapped to 2-D space by their probability distribution.





1. Number of iterations for optimization and the measure of perplexity. Press Start to (re-)run the optimization.
2. Select the number of PCA components used for projection.
3. Set the color of the displayed points (you will get colors for discrete values and grey-scale points for continuous). Set shape, size and label to differentiate between points. Set symbol size and opacity for all data points. Set jittering to randomly disperse data points.
4. Adjust *plot properties*:
    - *Show legend* displays a legend on the right. Click and drag the legend to move it.
    - *Show all data on mouse hover* enables information bubbles if the cursor is placed on a dot.
    - *Show class density* colors the graph by class.
    - *Label only selected points* allows you to select individual data instances and label them.
5. If *Send selected automatically* is ticked, changes are communicated automatically. Alternatively, press *Send Selected*.
6. *Select, zoom, pan and zoom to fit* are the options for exploring the graph. The manual selection of data instances works as an angular/square selection tool. Double click to move the projection. Scroll in or out for zoom.
7. Access help, save image or produce a report.

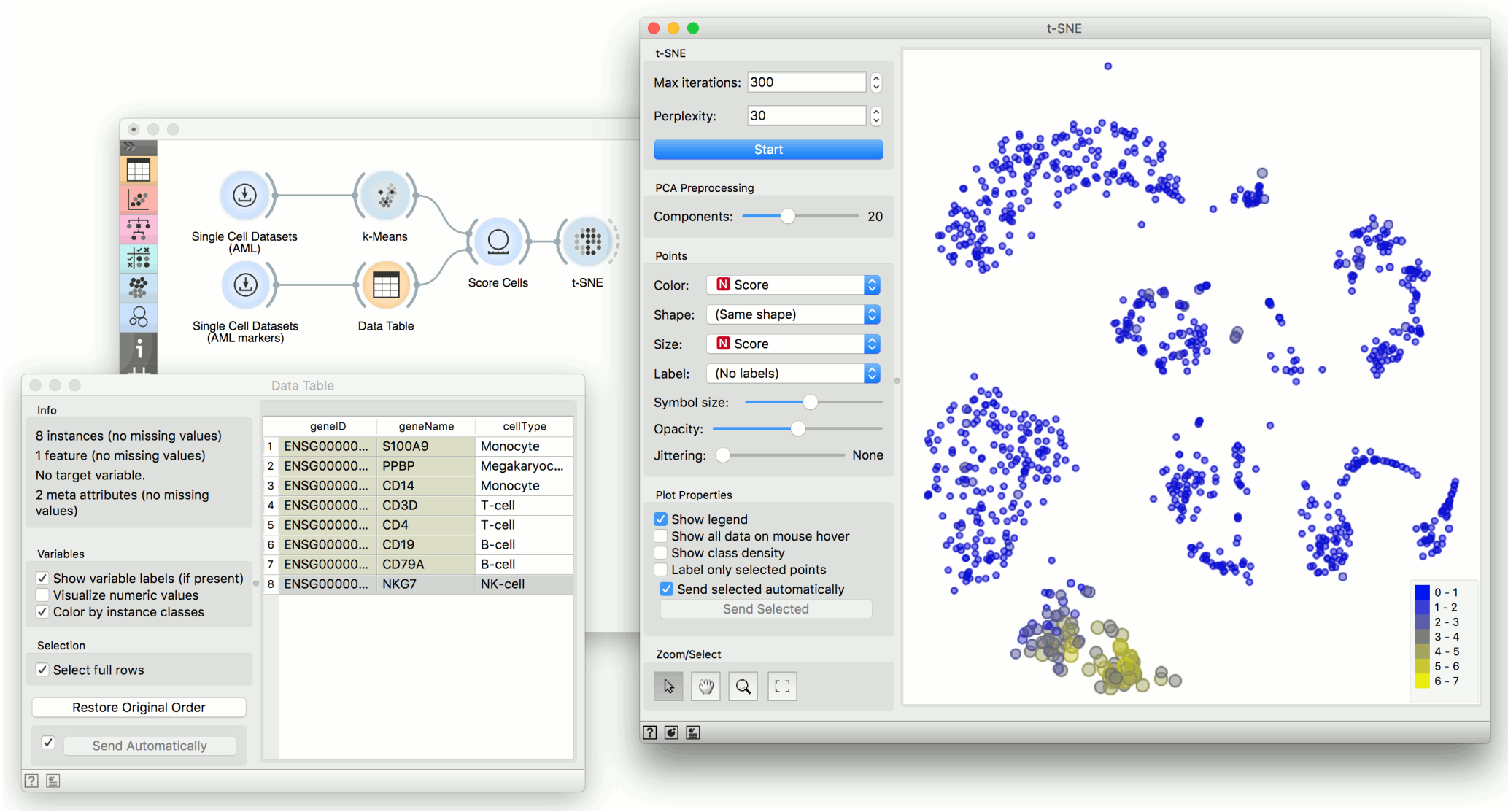
Example

We will use **Single Cell Datasets** widget to load *Bone marrow mononuclear cells with AML (sample)* data. Then we will pass it through **k-Means** and select 2 clusters from Silhouette Scores. Ok, it looks like there might be two distinct clusters here.

But can we find subpopulations in these cells? Let us load *Bone marrow mononuclear cells with AML (markers)* with **Single Cell Datasets**. Now, pass the marker genes to **Data Table** and select, for example, natural killer cells from the list (NKG7).

Pass the markers and k-Means results to **Score Cells** widget and select *geneName* to match markers with genes. Finally, add **t-SNE** to visualize the results.

In **t-SNE**, use *Scores* attribute to color the points and set their size. We see that killer cells are nicely clustered together and that t-SNE indeed found subpopulations.


Single Cell Datasets (AML)
k-Means
Single Cell Datasets (AML markers)
Data Table
Score Cells
t-SNE
Data Table
Info
8 instances (no missing values)
1 feature (no missing values)
No target variable.
2 meta attributes (no missing values)
Variables
Show variable labels (if present)
Visualize numeric values
Color by instance classes
Selection
Select full rows
Restore Original Order
Send Automatically
geneID geneName cellType
1 ENSG00000... S100A9 Monocyte
2 ENSG00000... PPBP Megakaryoc...
3 ENSG00000... CD14 Monocyte
4 ENSG00000... CD3D T-cell
5 ENSG00000... CD4 T-cell
6 ENSG00000... CD19 B-cell
7 ENSG00000... CD79A B-cell
8 ENSG00000... NKG7 NK-cell
t-SNE
t-SNE
Max iterations: 300
Perplexity: 30
Start
PCA Preprocessing
Components: 20
Points
Color: Score
Shape: (Same shape)
Size: Score
Label: (No labels)
Symbol size:
Opacity:
Jittering: None
Plot Properties
Show legend
Show all data on mouse hover
Show class density
Label only selected points
Send selected automatically
Send Selected
Zoom/Select
0 - 1
1 - 2
2 - 3
3 - 4
4 - 5
5 - 6
6 - 7