

How to UMAP

Introduction

Uniform Manifold Approximation and Projection (UMAP) is a machine learning algorithm used for dimensionality reduction to visualize high parameter datasets in fewer dimensions. UMAP is capable of maintaining both global and local structures between populations within a data matrix, and while UMAP is primarily used for data visualization, it can also aid in identifying underlying structures in high-dimensional space and preserve them in a lower-dimensional embedding of two (or more) parameters (1). UMAP is most often used for unsupervised embedding. However, UMAP Plugin versions 3.3+ can also perform supervised dimensionality reduction, where a derived parameter created from a separate clustering algorithm plugin can then be specified within the UMAP initiation window to help guide the UMAP embedding as it calculates, potentially resulting in improved separation of individual cluster populations. UMAP has two main stages: First it finds the nearest neighbors for each input event and learns the manifold structure of the high-dimensional data, next it uses this manifold to embed events into a lower-dimensional space.

More information about the tool can be read here:

[McInnes, Leland, John Healy, and James Melville. "Umap: Uniform manifold approximation and projection for dimension reduction." *arXiv preprint arXiv:1802.03426* \(2018\).](#)

This version is an implementation of the [UWOT package](#) developed for R by James Melvill, Aaron Lun, Mohamed Nadhir Djekidel, and Yuhan Hao.

Please review our [video](#) and [documentation](#) for installing FlowJo Plugins.

Download and Installation

1. Place the plugin .jar file in your Plugins folder, and direct FlowJo to that folder using the Diagnostics section of the Preferences. Restart FlowJo.
2. Make sure you have [R installed](#) and the R path is specified in the R Path field of the Diagnostics section of the Preferences.

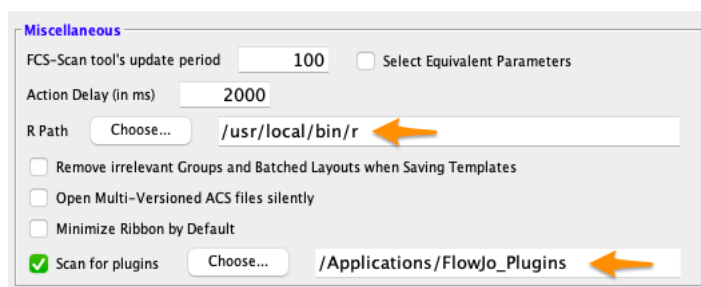


Figure 1 - The R path shown above is for MacOS. A typical Windows R path is C:\Program Files\R\R-4.2.2\bin\x64 for R version 4.2.2. A path to the plugins folder can also be specified.

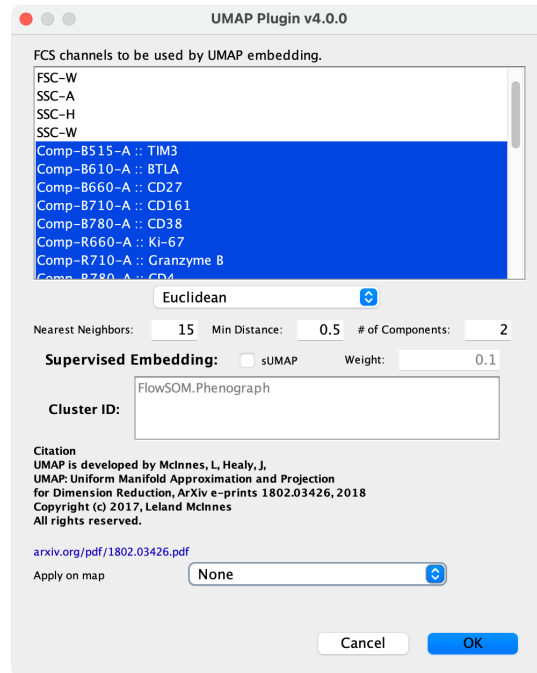
3. Running the plugin for the first time will install the needed R packages to allow the calculation to run in the R environment. Sometimes these installations can fail and you may need to manually install the R packages. To install the required R packages, use the following commands in R:

```
install.packages(c('ggplot2', 'FNN', 'igraph', 'Matrix', 'cowsay', 'umap', 'uwot', 'utils', 'devtools', 'data.table', 'dplyr'));
```

Note This plugin was tested in R versions 4.3 and UWOT R package version 0.1.14.

Usage

To run the UMAP plugin on your FCS file, select the population of interest within the workspace. Go to the Workspace tab and select the UMAP option from within the Plugins drop-down menu. This will bring up a dialog where researchers can choose which compensated parameters to use for dimensionality reduction. There are a variety of other options available to customize the embedding. Mousing over the fields will bring up a tool tip which describes the field's function.



Notes

- * **Parameters:** All Comp- prefixed parameters will be selected initially. Choose any subset of parameters that you want to be considered for dimensionality reduction.
 - * **Nearest Neighbors:** The number of nearest neighbors according to the distance metric.
 - * **Min Distance:** The minimum distance of embedded points.
 - * **Number of Components:** The number of dimensions to return.
 - * **sUMAP:** Select this option to perform supervised embedding then select a 'Cluster ID' parameter.
 - * **Weight:** The relative weighting of a clustering parameter (Cluster ID) to be used in the calculation.
 - * **Cluster ID:** The previously derived clustering algorithm output parameter to be used for supervision.
- By running the tool, a new set of UMAP parameters should be generated as a CSV file within the plugin derivatives folder, and that information will be merged into the current workspace automatically.

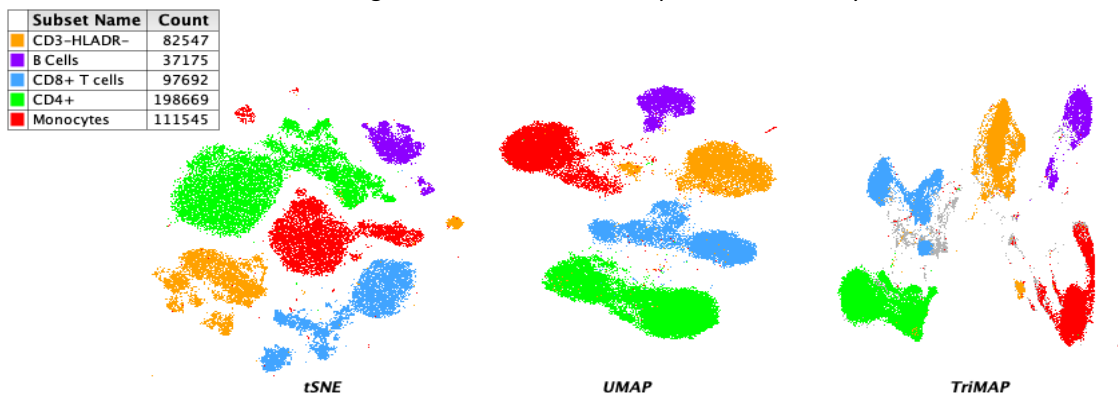


Figure 2 - Comparisons of the different embedding options run on CD45+ gated events.

Supervised UMAP

Supervised Embedding options in UMAP. When selecting the **sUMAP** checkbox, the relevant options will become actionable in the plugin interface. You have control over the weighting of your labels and which derived parameter to use as labels and drive the supervision.

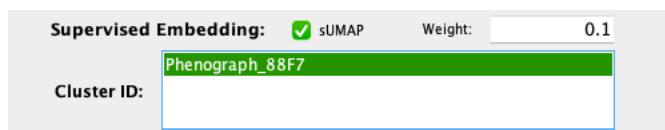
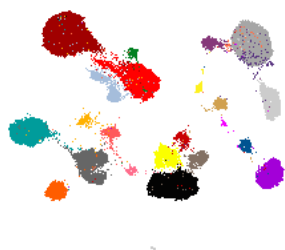


Figure 3 - The Supervised UMAP section of the plugin interface. You must select a clustering id or derived parameter here.

Enter a value in the **Weight** field or leave the default setting of 0.1. This value is the relative weight of the supervised parameter. A higher value will use more supervision and drive the events into tighter groups. A lower value will deemphasize the effect of cluster input on the resulting low dimensional embedding. Please note that setting this value too high can sometimes have the opposite effect and the groups become spread out and do not represent the labels well. In most cases, the default value is sufficient to drive supervision and separate the classes of cells. Next, select a Cluster ID parameter from a previous clustering analysis in FlowJo. If you do not select the Cluster ID before clicking OK, you will be prompted again to select this derived parameter before you can continue.



Supervised UMAP

Figure 4 - UMAP embedding using Phenograph clusters as labels. The embedding of the same CD45+ population as shown in Figure 3 now shows cells

Use Previous UMAP Maps

The **Apply on Map** option can be used to fit a new data set to a map previously generated by the UMAP plugin (version 4.0 or newer). This will generate UMAP parameters for new fcs files that have the same shape as previous UMAP runs, allowing users to compare across samples and conditions. For best results, the initial map should be created using a population that contains all the subpopulations present in the populations that will later be mapped using the “Apply on Map” feature.

To use this feature the user needs to select a R data object (RDS file) generate by a previous run from the “Apply on Map” dropdown menu. Any UMAP runs performed previously on this workspace should be listed in the dropdown menu.

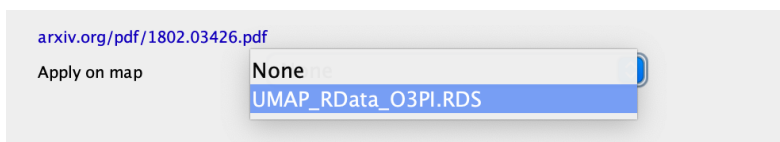


Figure 5 – A previous UMAP run can be selected from the Apply on Map dropdown. It will only list run IDs occurring in the same workspace.

The resulting UMAP image should retain the same shape, allowing us to compare across samples and conditions.

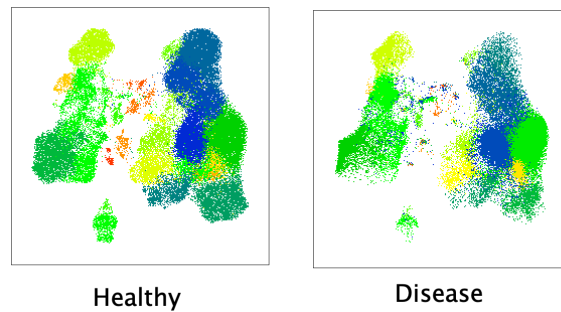


Figure 6 – Using the “Apply on map” option the same UMAP shape can be applied to multiple samples.

Leave us your feedback

Please write to flowjo@bd.com or seggeq@bd.com with any questions or concerns.