

“How To CyCombine”

CyCombine plugin for FlowJo

CyCombine is a R package that allows for batch correction or integration of single-cell data sets and has been implemented in FlowJo as a plugin. Batch effects can be thought of as non-biological variability between groups or batches of samples processed at different times. This variability can come from using different lots of reagents, or from different technicians, and from using different instruments to process samples.

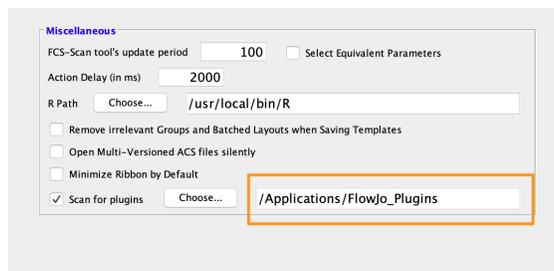
To increase the statistical power of our analyses, we often perform a concatenation step in FlowJo to create one file where we can perform gating on populations of cells, or using clustering and dimensionality reduction algorithms. When there are batch effects in the data, we may not be able to take full advantage of a concatenated file. This technical variability can lead to additional clusters from clustering algorithms or additional islands seen in tSNE or UMAP space, reducing the statistical power we get from the concatenated data. Batch correction aims to reduce this technical variability across batches to help us make better comparisons with the data.

The cyCombine algorithm has been developed by Lars Ronn Olsen, Christina Bligaard Pedersen, and Soren Helweg Dam at the Technical University of Denmark and the source code is available here: <https://github.com/biosurf/cyCombine>. You may read more about cyCombine in the nature communications paper here:

Pedersen, C.B., Dam, S.H., Barnkob, M.B. et al. “cyCombine allows for robust integration of single-cell cytometry datasets within and across technologies.” Nat Commun 13, 1698 (2022) <https://doi.org/10.1038/s41467-022-29383-5>

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.
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.



3. Running the plugin for the first time will auto-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('dplyr','readr','BiocManager','tibble','cowsay',
                  'data.table','cowplot','devtools','tidyr','ggridges',
                  'emdist'))
```

```
BiocManager::install(c('flowCore', 'sva'))
```

```
devtools::install_github('biosurf/cyCombine')
```

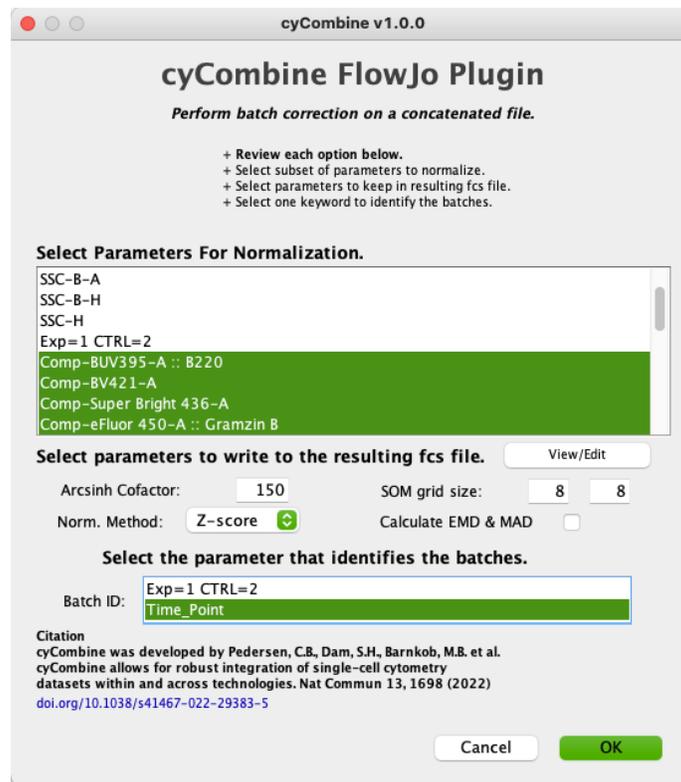
Note This plugin was tested in R versions 4.1 & 4.2 and cyCombine version 0.2.13.

Usage

Before running cyCombine, you will need to perform a concatenation step in FlowJo to create a categorical keyword (**with no spread**) to identify the different batches in your data set. Please review our video on FlowJo University on how to include keywords during concatenation [HERE](#).

When performing batch correction with cyCombine, it is best to pass in a “cleaned up” population of cells into the algorithm. To clean up the data, perform gating steps to remove dead cells, doublets, debris, or even gate down further to a population of interest. This can be done before or after concatenation in FlowJo.

Select the population of interest in FlowJo and then start the CyCombine plugin by selecting the plugin from the plugin drop down menu in FlowJo. Once started, you will be presented with the plugin UI shown below.



In the top parameter selector, choose the parameters that you would like to normalize in the resulting .fcs file. We recommend selecting all of your compensated parameters for normalization. Below the parameter selector you can click the **View/Edit** button to select the parameters to export to the new .fcs file. By default, all of your parameters will be selected here but you can choose to deselect parameters that you do not want

exported. You can also choose to export more than one of your categorical parameters at this step so they are available for making comparisons in the normalized file. The categorical parameter selector at the bottom is where you will select the keyword that identifies your batches in the concatenated file, choose only one.

Prior to performing the batch correction, the data are arcsinh transformed using the cofactor entered in the **Arcsinh Cofactor** field. For cyTOF data, it is recommended to use a cofactor of 5, for traditional flow data a cofactor of 150, and for spectral flow data a cofactor of 6000. Choose a normalization method from the **Norm. Method** drop down menu. The “Z-Score” method is recommended for similar data sets (e.g. multiple batches from same instrument). The “Rank” method is recommended for less similar datasets (e.g. multiple batches from different instruments).

Use the **SOM grid size** to change the cluster number used to partition the cells. The default 8x8 grid size is good for most data sets but if you expect there is more heterogeneity in your data, you can increase the grid size. Check the box **Calculate EMD & MAD** to calculate Earth Mover’s Distance (EMD) and Median absolute deviation (MAD). EMD is used to compare the distribution of each marker within SOM modes across batches, before and after correction. The resulting EMD reduction value is a relative measure, but generally speaking, a small value here is good. MAD is used to detect potential information loss by quantifying the variability of each marker within batches before and after correction. Because the MAD score quantifies the information loss, ideally it has a small value. These calculations can be memory intensive and may fail for large data sets over 300k cells.

After the plugin completes, you will get a new normalized .fcs file loaded back into the FlowJo workspace. Before drawing gates or running additional plugins on this file, the transforms for each parameter should be checked carefully and adjusted as needed using the Transform button in the FlowJo graph window. In addition to the new .fcs file, there will be image(s) loaded into the layout editor to show expression patterns for each marker pre and post correction, for each batch. If you chose to perform EMD and MAD calculations, there will be additional images returned to the layout editor.

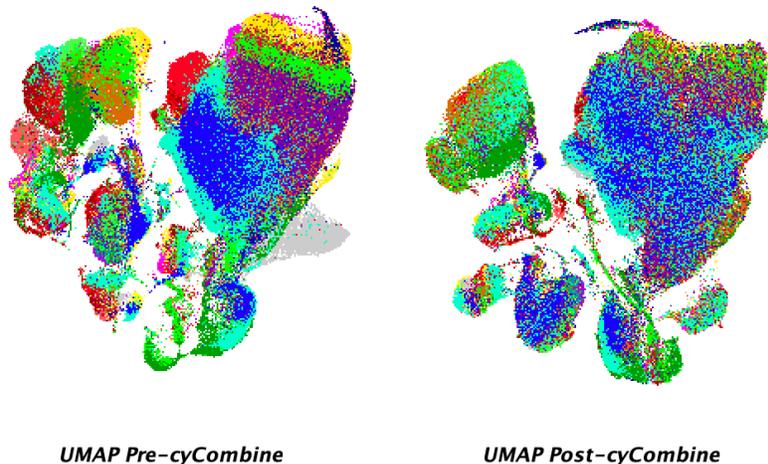


Figure 1: Comparison of data overlaid by batches showing overall better integration of data in cyCombine normalized UMAP plot.

Leave us your feedback

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

References

1. Pedersen, C.B., Dam, S.H., Barnkob, M.B. et al. “cyCombine allows for robust integration of single-cell cytometry datasets within and across technologies.” Nat Commun 13, 1698 (2022) <https://doi.org/10.1038/s41467-022-29383-5>