

How to MNN

Introduction

The Mutual Nearest Neighbor algorithm (MNN) is a tool designed to correct variation due to batch processing and occurring within shared populations across different samples for high dimension data. The algorithm was originally designed for use on Single Cell RNA sequencing data but we have implemented it here for use in Flow Cytometry data. The tool was originally developed by Haghverdi et al.

More information about the tool can be read here:

[Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol. 2018 Jun;36\(5\):421-427. doi: 10.1038/nbt.4091. Epub 2018 Apr 2. PMID: 29608177; PMCID: PMC615](#)

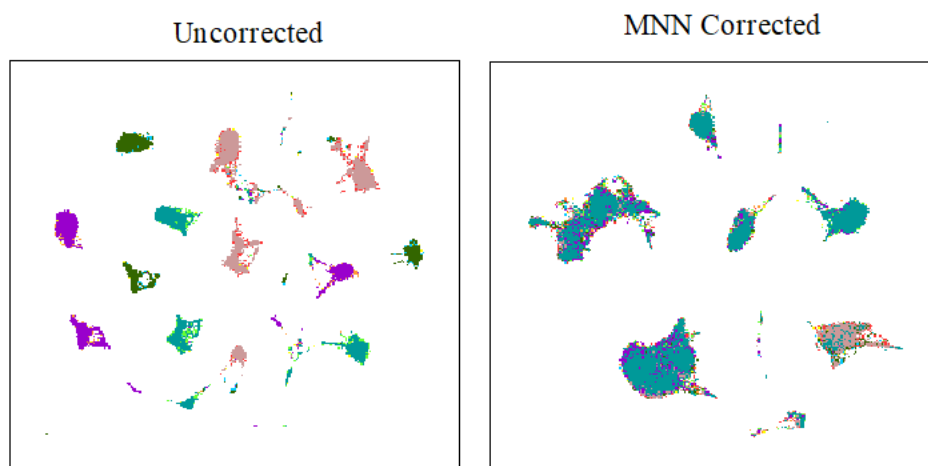
This version is an implementation of the [C++ mnncorrect library](#) which is based on the [batchlr algorithm](#) developed by Aaron Lun and Laleh Haghverdi.

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

Installing MNN

- 1) To use the MNN plugin, you will need FlowJo 10.9.0 or newer.
- 2) Download MNN from our website.
- 3) Unzip the download folder and place the “MNN.jar” file into the Plugins folder.
- 4) Restart FlowJo and MNN should be available in your “Plugins” dropdown menu in the workspace tab.

*Note: MNN is an executable. Some institutions security may prevent MNN from running. If that is the case, the user’s IT department can give MNN permission to run by going to the “cpp_mnn.exe” file created in the output folder next to the workspace and giving that file read/write/execute permissions.

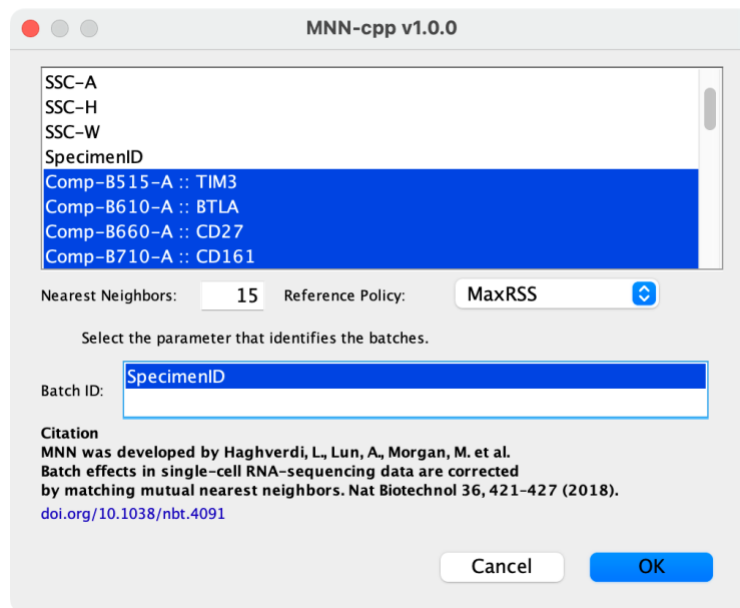


Using MNN

Before running MNN, 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. The keyword is an integer value (ie 1, 2, 3) representing which batch each sample belongs to. Keywords with alphabetic characters (such as A, B, C) will not work. Please review our video on FlowJo University on how to include keywords during concatenation [HERE](#).

When performing batch correction with MNN, 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 MNN 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 batch correct. We recommend selecting all of your compensated parameters for batch effect correction. The BatchID selector at the bottom is where you will select the keyword that identifies your batches in the concatenated file. Choose only one.

The “Nearest Neighbors” option defines the minimum number of mutual nearest neighbors that two events are required to have to be an MNN pair. Larger values for the Nearest Neighbors increase the stability of the correction but reduce the resolution of matching subpopulations across batches or reduce sensitivity to local variations when calculating correction vectors. For most Flow Cytometry data, an initial Nearest Neighbor value of 15 is appropriate.

The Reference Policy, shown to the right on the Nearest-Neighbor value, is the way the program chooses the first reference batch when automatically merging. The first reference should be a sample that contains all of the populations found in the other samples.

- Input - use the first supplied batch. This is useful in cases where one batch is known to contain most subpopulations and should be used as the reference, but there is no obvious ordering for the other batches.
- MaxRSS - use the batch with the greatest residual sum of squares (RSS). This is similar to MaxVariance but it puts more weight on batches with more cells, so as to avoid picking small batches with few cells and unstable population structure.

- MaxSize - use the largest batch. Ensures that the initial reference has enough cells for stable correction.
- MaxVariance - use the batch with the largest variance. This improves the likelihood of obtaining a reference that contains a diversity of subpopulations and thus is more likely to form sensible MNN pair with subsequent batches.

After the plugin completes, it will add the new batch corrected parameters with “MNN_” at the beginning of their names to the sample. An additional RunID string will be appended to the end of the parameter names to identify separate runs of the MNN algorithm. These parameters can then be used for further analysis and machine learning. 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.

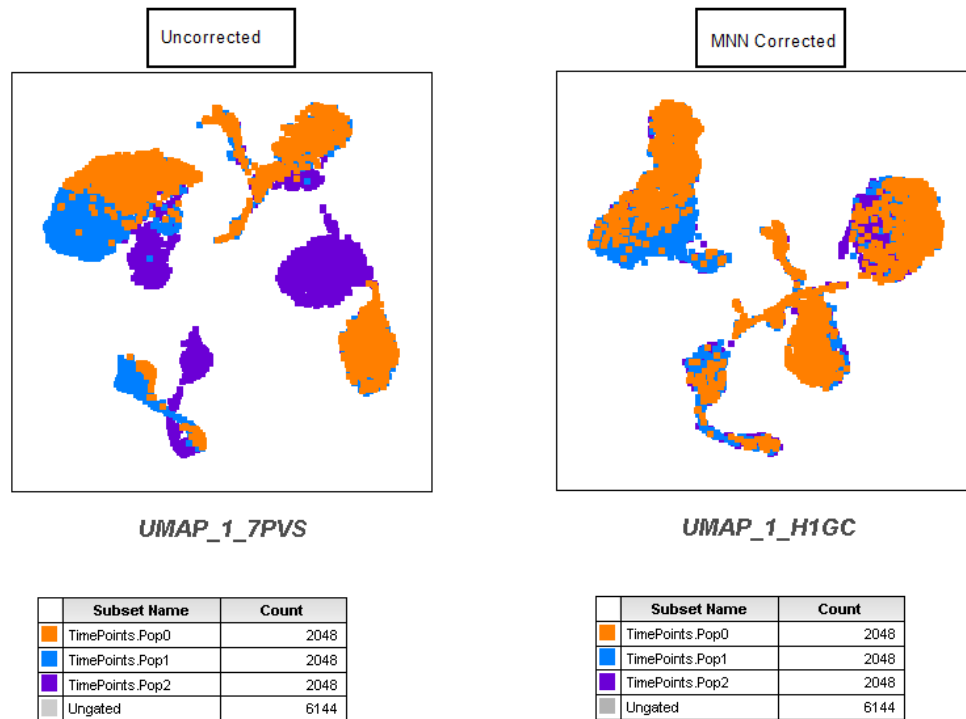


Figure 2. A comparison of a set of samples before and after correction.

Leave us your feedback

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

References

- The MNN plugin uses the [MnnCorrect](#) package
- [Batch effects in single-cell RNA sequencing data are corrected by matching mutual nearest neighbours](#)
Haghverdi L, Lun ATL, Morgan MD, Marioni JC (2018). *Nat. Biotechnol.* 36(5):421-427