Hints

- Caution must be exercised with the K-S algorithm as it will erroneously report that two halves of the same population are distinct (every other cell makes up one of the halves while the cells in between make up the other half).
- The minimum value of T(X) that has biological significance depends on the nature of the data being analyzed and therefore needs to be determined empirically. Only populations which have T(X) values larger than this empirical minimum can be considered to be different.
- Several populations can be compared in order to determine the minimum T(X) value. Machine stability during the collection, as well as inherent variability in the FACS data are just two reasons why the comparison of a population to itself can give a T(X) > 0.
- You can compare a population to itself by opening the Population Comparison platform on a sample and dragging the same sample to the control box. FlowJo compares the two halves of this population (one half made up of every other cell while the other half is made up of the cells in between).
- You can also compare the same sample collected twice using the Population Comparison platform (at the beginning and end of the sample collection best determines the machine stability).
- As the number of parameters being compared increases, more events may need to be collected in order to distinguish subtle variations in the populations. However, inclusion of parameters in the comparison which do not vary between populations does not degrade the ability to distinguish the populations.
- Note that the computations in the Multivariate Population Comparison platform can be time and memory intensive. You may need to allocate more memory to FlowJo.
- Determining the samples to be concatenated in a Multi-sample Comparison is best approached by an iterative process. One can concatenate all the control samples and compute the distances of each control sample to the average of all of them. Thus, those control samples which are outliers can be removed from the control set. Caution is warranted since reduction of the number of samples entered as controls can lead to sampling bias.

Links and References


Abstract

Structural based approaches to calculate the differences between populations are a powerful tool for analyzing flow cytometric data. Algorithms can be used as an objective approach for the comparison of distributions. Manual analysis is more prone to error, or in some cases, cannot be accurately done because there is no distinguishable difference in the distribution of the populations. FlowJo’s Population Comparison platform provides four of the most commonly used published algorithms to calculate differences in distributions of a single parameter. Using an additional sophisticated algorithm, FlowJo’s Population Comparison platform also provides for an objective statistical based approach for analyzing the difference in distributions of multiple parameters in order to identify and indicate very specific populations. FlowJo has the richest set of population comparison features of any flow cytometry analysis program.

Introduction

In flow cytometry analyses, it is common to compare distributions of events between two or more samples. Whether for quality control purposes or biological applications, statistical based approaches that provide an objective analysis are useful. For example, when monitoring immune activation, DNA distributions, protein up-regulation or cell tracking, it is often difficult to determine the percent positive or difference between an experimental ‘unknown’ sample and the control. These user cases and numerous others provide the impetus to developing algorithms that can objectively calculate the difference between distributions in one or more parameters. The compared populations can either be subsets of the same sample, or more commonly, equivalent populations in different samples.

There are currently four algorithms within FlowJo’s Population Comparison platform for comparing distributions between two samples in a single parameter. These are the Super Enhanced Danux (SED), Overton Subtraction I, Probability Binning Comparison and Kolmogorov-Smirnov (K-S). All of the algorithms calculate differences in the distributions slightly differently, so it is important to consider them all (more information on page 4). FlowJo has also implemented an additional, multi-parameter or Multivariate platform to further assist users in objectively quantifying differences in distributions. The Multivariate display uses the Probability Binning Comparison algorithm to compare multiple parameters at the same time. A gate can be created under this display once the number of bins is established and the X2 threshold is determined. This feature set is highlighted in this technical document and each of the algorithms is briefly explained.
Select the parameters that you want to compare (they will become shaded).

Select the Multivariate tab. As described for the Univariate comparison, define the control population by dragging it from the Workspace window to the box in the top left corner of the platform’s window.

Set the number of bins according to the number of events collected.

Create a gate. In order to create gates based on the differences between the control and the test sample(s), the Gating Cut-off value must be selected. The distribution of X2 scores for each bin are graphed versus the percentage of cells that would fall into the gate if that X2 cut-off value is chosen. The Probability Binning Comparison algorithm estimates the probability that your test population(s) are different from a control population.

The default threshold is the ChiSq value corresponding to those bins that are at least two standard deviations above random variation; i.e., T(X) = 2. The bins that fall above this X2 cut-off value are shown by red dots.

The threshold/cut-off can be altered by drawing a new gate (click and drag) or by entering the upper and lower X2 values in the boxes.

Within the Univariate Population Comparison platform, the Overton Subtraction and SED algorithms are used to calculate the percentage of positive cells found in the sample and not the control. The Overton cumulative histogram subtraction algorithm subtracts histograms on a channel-by-channel basis to provide the percent of positive cells. The SED algorithm is a modification of Enhanced Normalized Subtraction that provides a more precise measurement of the percent of positive cells. However, these methods do not provide an indication of the probability with which two distributions are different; nor do they provide confidence intervals. The K-S and Probability Binning Comparison algorithms are used to determine the statistical difference between samples.