FLAMEPreprocess Documentation

Module Name: FLAMEPreprocess
Description: Performs preprocessing of .fcs or .txt flow cytometric data files
Author: Xinli Hu (Broad Institute), gp-help@broad.mit.edu

Summary:

FLAME (FLow analysis with Automated Multivariate Estimation) uses finite mixture model clustering techniques with novel algorithms and models to identify and characterize discrete populations in flow cytometric data [1]. A pipeline of GenePattern modules implements the method: FLAMEPreviewTransformation, FLAMEPreprocess, FLAMEMixtureModel, FLAMEChooseOptimalClusterNumber, FLAMEMetacluster, FLAMEContourDataGenerator, and FLAMEViewer.

The FLAMEPreprocess module performs a series of preprocessing operations on flow cytometric data files, including column/channel selection, bi-exponential transformation, optional live-cell gating, and optional data transformation. The module outputs a zip file of preprocessed data files, which is used as the input file for the FLAMEMixtureModel module.

For column/channel selection, the user specifies a list of the columns (corresponding to antibodies/channels) to keep and a list of the corresponding antibody/channel names. For transforming the raw data using bi-exponential transformation, the user specifies whether the original data were 4-decade data or 18-bit data.

Optionally, the user chooses to perform automatic live-cell gating by setting the *remove dead* parameter to yes. In this case, the module clusters the Forward- and Side-Scatter intensities for each sample and eliminates the cluster with the lowest mean scatter intensities as "dead cells." Currently, this function works well only if the sample consists of cell lines or clones (one single cell type), as opposed to whole blood or PBMC samples, so use caution when selecting this function. Alternatively, perform gating outside GenePattern and use the FLAME modules to analyze only live-cell events.

> **Note**: Gating (outside GenePattern) is usually performed in FlowJo (http://www.flowjo.com/). Gated data can be exported as either .fcs or .txt files. However, currently there exist some unresolved problems in exporting certain FCS3.0 files in .fcs format. These modified/exported FCS files may fail to be read successfully in FLAME. Therefore we suggest exporting gated data from FlowJo as .txt files until the problem is resolved.

For data transformation, the user selects a transformation (logicle, arcsinh, both or none) and a parameter of the transformation called the cofactor that determines the spread of the points in the sample. The effects of the transformations and their cofactors are

described in the FLAMEPreviewTransformation module documentation. Typically, one uses FLAMEPreviewTransformation to choose which transformation (if any) to use.

References:

1. Saumyadipta Pyne, Xinli Hu, Kui Wang, Elizabeth Rossin, Tsung-I Lin, Lisa M. Maier, Clare Baecher-Allan, Geoffrey J. McLachlan, Pablo Tamayo, David A. Hafler, Philip L. De Jager, and Jill P. Mesirov. (2009). Automated High-dimensional Flow Cytometric Data Analysis. *PNAS* 106:8519-8524.

Parameters:

Name	Description
dataset	A .zip file containing flow files in .txt or .fcs format. (The files should be all .txt or all .fcs, but not both.) The .fcs files can be 2.0 or 3.0 format (file formats available under “standards” at http://www.isac-net.org/). The .txt files should contain a matrix of fluorescent intensities in all colors, where each row is data for one cell and each column is one color/antibody. There should also be a header row containing the color/antibody names. The fields should be either space or tab delimited.
file type	File type of input sample data. fcs (default) or txt.
data scale	The original data scale. 18-bit (default) or 4-decade.
remove dead	Whether to attempt removing dead cells from the samples. No (default) or yes.
channels	A comma-separated list of channel numbers indicating which columns (antibodies) to keep (e.g., 1, 2, 3, 7).
channel names	A comma-separated list of channel/antibody names that corresponds to the channel numbers (e.g., FSC, SSC, CD4, CD45RA).
scatter channels	A comma-separated list of channel/antibody names that corresponds to the channel numbers (e.g., FSC, SSC, CD4, CD45RA). Default: 1,2.
transformation	The transformation to apply: logicle (default), arcsinh, all or none.
logicle cofactor	A cofactor that tunes the logicle transformation. The cofactor is usually greater than 0 and less than 10. Default: 3.

arcsinh cofactor	A cofactor that tunes the arcsinh transformation. The cofactor is 0 usually between 0 exclusively and a few thousand. Default: 250.
output.prefix	A prefix for output files.

Output File:

Preprocessed (transformed and optionally filtered) samples in a zip file.

Example Data:

The example data is a subset of the data described in Pyne et al. (2009) [1]:

Input parameter	**Value**
dataset	ftp://ftp.broad.mit.edu/pub/genepattern/example_files/FLAME/SMALL_phospho.lymphgated.fcs.zip
file type	fcs
data scale	4-decade
remove dead	no
channels	3,4,5,7
channel names	SLP76,ZAP70,CD4,CD45RA
scatter channels	1,2
transformation	logicle
logicle cofactor	5
output prefix	SMALLphospho

Platform dependencies:

Task type: FLAME
CPU type: Any
OS: Any
Java JVM level:
Language: R (2.7.0 or later)