SigClust

Maziyar Baran Pouyan, Mehrdad Nourani

March 3, 2016

Licensing

Install

At first, you need to install the following packages:

```
install.packages(c("Rtsne","feature","apcluster","cluster","pbapply","ggplot2"))
```

The SigClust R package is free available to download from:http://www.utdallas.edu/~nourani/Bioinformatics/SigClust. When you download the package (SigClust_1.0.tar.gz), put it in a folder (e.g. your R working directory). Then, you can install it with the following instruction:

```
install.packages("path_to_SigClust_Folder/SigClust_1.0.tar.gz", repos = NULL, type="source")
```

Overview

We apply a non-parametric and fast clustering technique to identify homogeneous cell populations from single-cell flow cytometry data. This is done based on a novel technique that estimates the initial number of clusters in high dimension and identifies the final clusters by merging clusters using their phenotypic signatures in low dimension. The proposed method is called SigClust. The technical details of SigClust can be found in: http://www.utdallas.edu/~nourani/Bioinformatics/SigClust. SigClust includes 4 main functions which described below:

Functions

1. SigClust

The *SigClust* function performs the main clustering process:

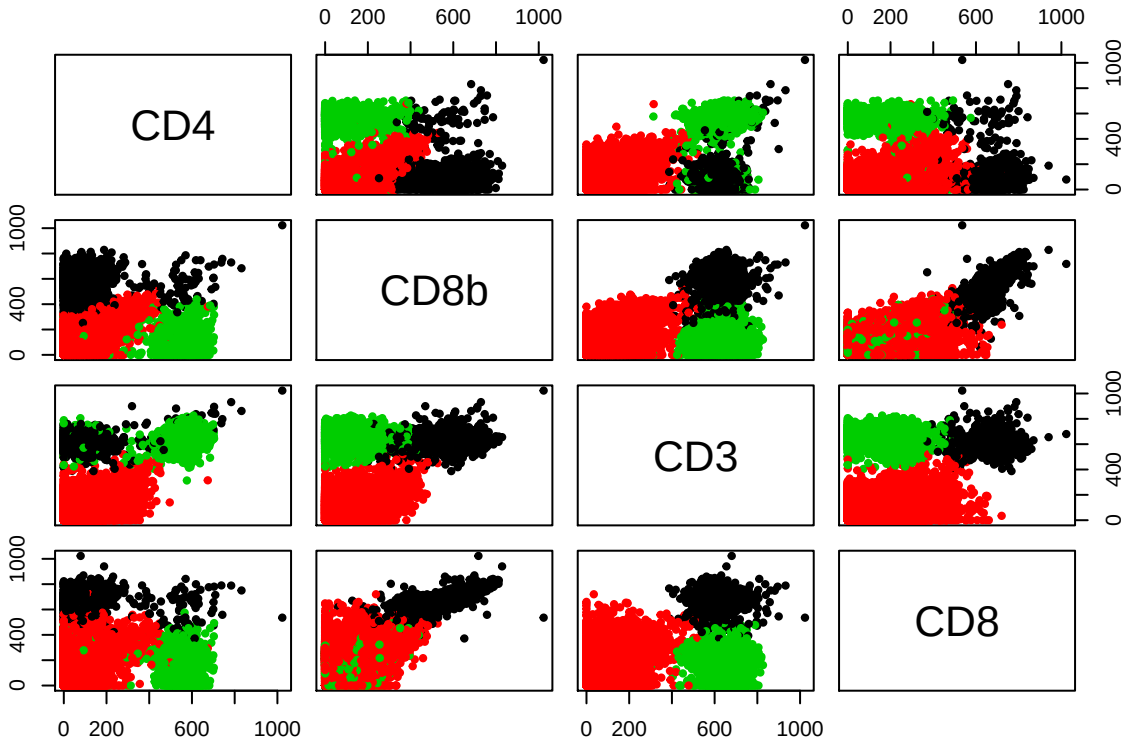
```
library(SigClust)
data(x)
Label<-SigClust(x)
```

The *Label* value includes the list of cluster membership labels for the single cells. The clustered cell data is automatically saved as $Clustered - Data.csv$ in your home directory.

2. BiaxPlots

This function provides all the possible biaxial plots made by cell markers. The value Label is used to visualize these populations with different colors.

```
library(SigClust)
data(x)
Label<-SigClust(x)
BiaxPlots(x,Label)
```



3. tsne2dplot

This function visualizes all identified cell populations on a 2D map using t-stochastic neighbor embedding (t-SNE) approach.

```
library(SigClust)
data(x)
Label<-SigClust(x)
tsne2dplot(x,Label)
```

The identified populations are visualized with different colors.

4. PhynoSigPattern

This function calculates phenotypic signatures as bar plots for each identified population.

```
library(SigClust)
data(x)
Label<-SigClust(x)
PhynoSigPattern(x,Label)
```

The computed phenotypic signatures will be generated automatically stored in *PhenPattern* folder in your working directory. In these barplot patterns, the blue line shows the median of that markers all over the cells and the green bar heights indicates the median expression values of cellular markers. Median deviation (error) is also shown on top of green bars using black line. The computed bars present the cellular phenotypic properties of each cellular populations. Graphically speaking, a phenotypic pattern of an extracted population shows a natural distinctive pattern of that population.