flowType: Phenotyping Flow Cytometry Assays

Nima Aghaeepour

March 30, 2012

`naghaeep@bccrc.ca`

Contents

1 Licensing

Under the Artistic License, you are free to use and redistribute this software.

2 Introduction

This document demonstrates the functionality of the flowType package for phenotyping FCM assays. flowType uses a simple threshold, Kmeans, flowMeans or flowClust to partition every channel to a positive and a negative cell population. These partitions are then combined to generate a set of multi-dimensional phenotypes. For more details on how this package can be used in a complete analysis pipeline, please refer to the manuscript that describes analysis of 466 HIV[+] patients [1]. A Sweave example is available at `http://www.terryfoxlab.ca/flowsite/flowType`.

3   Loading the Library

We start by loading the library (for installation guidlines see the Bioconductor website).

```
> library(flowType)
> data(DLBCLExample)
```

4   Running flowType

We will use the *PropMarkers* and *MFIMarkers* arrays for measuring cell proportions and MFIs, respectively. Cell proportion of a given cell population is the number of cells in that population divided by the total number of cells:

```
> PropMarkers <- 3:5
> MFIMarkers <- PropMarkers
> MarkerNames <- c('FS', 'SS','CD3','CD5','CD19')
> Res <- flowType(DLBCLExample, PropMarkers, MFIMarkers, 'flowMeans', MarkerNames);
```
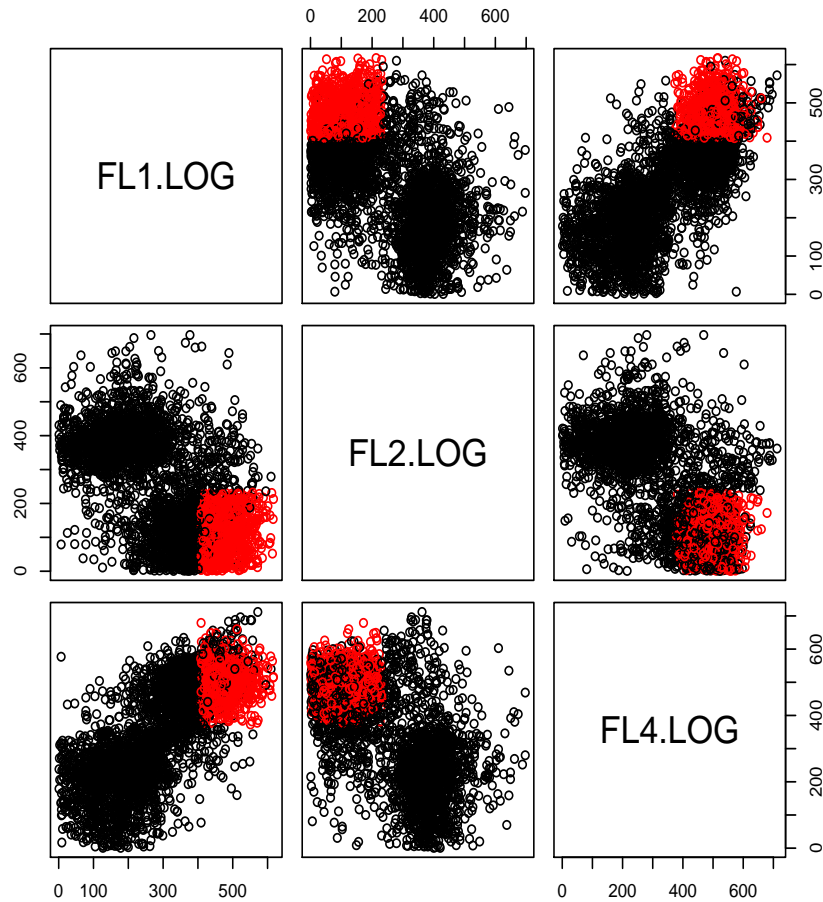
We can look at the single-dimensional partitions:

```
> plot(Res, DLBCLExample);
```

CD3

CD5
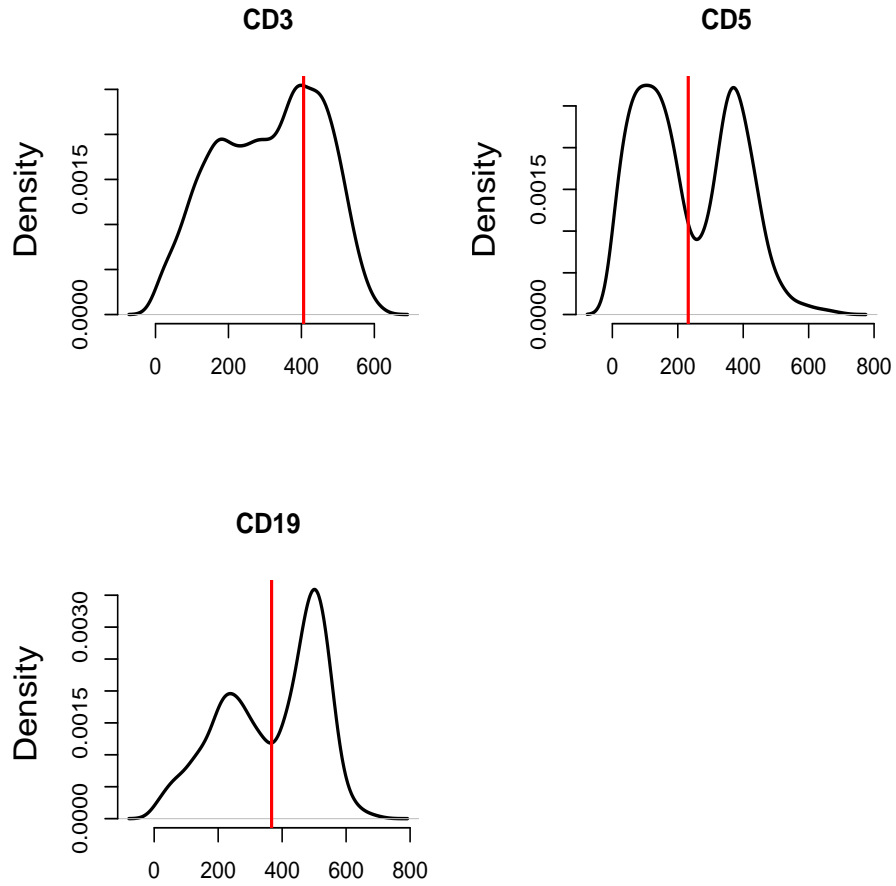
CD19

And we can plot a specific cell population:

```
>    plot(Res, "CD3+CD5-CD19+", Frame=DLBCLExample);
```

Next we will plot the 20 largest phenotypes. The first phenotype includes all of the cells.

```
> MFIs=Res@MFIs;
> Proportions=Res@CellFreqs;
> Proportions <- Proportions / max(Proportions)
> index=order(Proportions,decreasing=TRUE)[1:20]
> bp=barplot(Proportions[index], axes=FALSE, names.arg=FALSE)
> text(bp+0.2, par("usr")[3]+0.02, srt = 90, adj = 0, labels = names(Proportions[index]), xp
> axis(2);
> axis(1, at=bp, labels=FALSE);
> title(xlab='Phenotype Names', ylab='Cell Proportion')
```
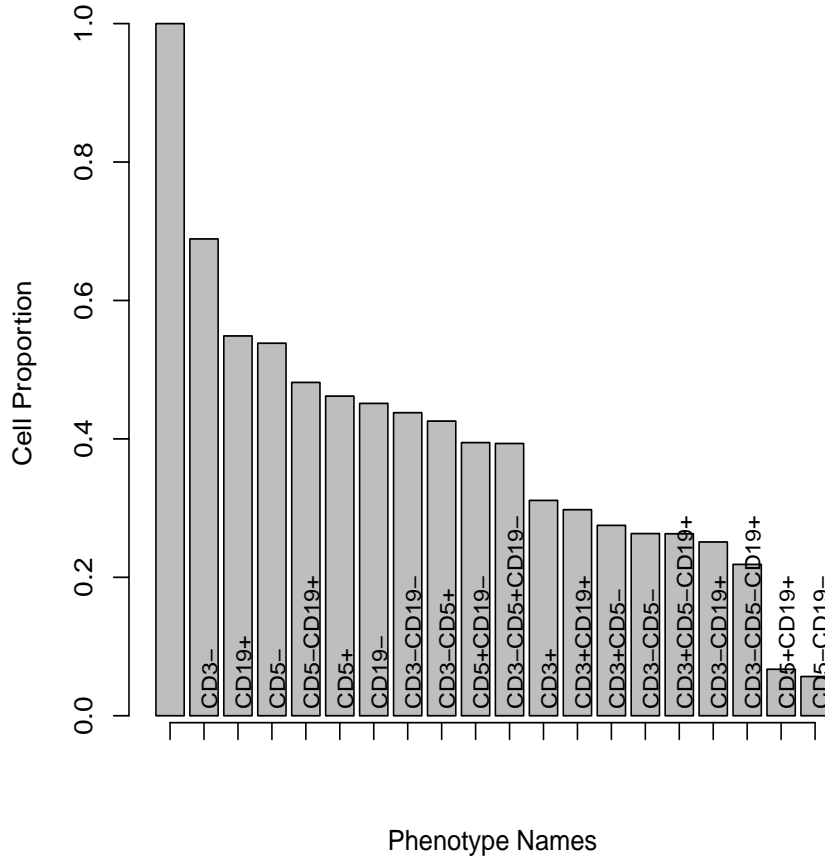
These phenotypes can now be analyzed using a predictive model (*E.g.,* classification or regression).

5 Cross-sample Analysis

This document demonstrates the functionality of the flowType package for performing cross-sample analysis,

The dataset used here is provided by the Scott laboratory of the Simon Fraser University and Spina laboratory of University of California, San Diego. This analysis is performed as a proof of principle and is not a complete analysis of this dataset. The data is transformed, compensated, and the lymphocytes are manually gated. The flowFrames have been downsampled to 1000 events.

The dataset consists of 19 HIV$^+$ and 13 normal subjects. Raw FCS files are

available. The meta-data is stored in a matrix (which consists of FCS filename, tube number, and patient label). In this example, we are interested in the second tube only.

```
> data(HIVMetaData)
> HIVMetaData <- HIVMetaData[which(HIVMetaData[,'Tube']==2),];
```

We convert the subject labels so that HIV$^+$ and normal subjects are labeled 2 and 1, respectively.

```
> Labels=(HIVMetaData[,2]=='+')+1;
```

Load the data and run flowType:

```
> library(sfsmisc);
> library(flowCore);
> data(HIVData)
> PropMarkers <- 5:10
> MFIMarkers <- PropMarkers
> MarkerNames <- c('Time', 'FSC-A','FSC-H','SSC-A','IgG','CD38','CD19','CD3','CD27','CD20',
> ResList <- fsApply(HIVData, 'flowType', PropMarkers, MFIMarkers, 'kmeans', MarkerNames);
```

Extract all cell proportions from the list of flowType results and normalize them by the total number of cells:

```
> All.Proportions <- matrix(0,3^length(PropMarkers)-1,length(HIVMetaData[,1]))
> for (i in 1:length(ResList)){
+   All.Proportions[,i] = ResList[[i]]@CellFreqs
+   All.Proportions[,i] = All.Proportions[,i] / max(All.Proportions[which(names(ResList[[i]]
+ }
```

We use a t-test to select the phenotypes that have a significantly different mean across the two groups of patients (FDR=0.05). Remember that in real world use-cases the assumptions of a t-test must be checked or a resampling-based alternative (e.g., a permutation test) should be used. P-value correction for multiple testing (e.g., bonferonni's method) and sensitivity analysis (e.g., bootstrapping) are also necessary.

```
> Pvals <- vector();
> EffectSize <- vector();
> for (i in 1:dim(All.Proportions)[1]){
+   if (length(which(All.Proportions[i,]!=1))==0){
+     Pvals[i]=1;
+     EffectSize[i]=0;
+     next;
+   }
+   temp=t.test(All.Proportions[i, Labels==1], All.Proportions[i, Labels==2])
+   Pvals[i] <- temp$p.value
```

```
+    EffectSize[i] <- abs(temp$statistic)
+ }
> Selected <- which(Pvals<0.05);
> print(length(Selected))

[1] 179
```

179 phenotypes have been selected. After P-value adjustment, only 5 of them remain in the list:

```
> Selected <- which(p.adjust(Pvals)<0.05);
> library(xtable)
> MyTable=cbind(names(ResList[[1]]@CellFreqs)[Selected], format(Pvals[Selected],digits=2), 1
> colnames(MyTable)=c('Phenotype Name', 'p-value', 'adjusted p-value', 'cell frequency')
> print(xtable(MyTable, caption='The selected phenotypes, their p-values, adjusted p-values,
```

Table 1: The selected phenotypes, their p-values, adjusted p-values, and cell frequencies

Phenotype Name	p-value	adjusted p-value	cell frequency
IgG-CD38-CD19-CD27+CD20-	7.3e-05	0.0492	0.243
IgG-CD38-CD19-CD27+	7.4e-05	0.0499	0.243
IgG-CD38-CD27+CD20-	7.3e-05	0.0495	0.243
IgG-CD19-CD27+CD20-	7.3e-05	0.0492	0.252
IgG-CD19-CD27+	7.4e-05	0.0499	0.252

References

[1] Nima Aghaeepour, Pratip K. Chattopadhyay, Anuradha Ganesan, Kieran O'Neill, Habil Zare, Adrin Jalali, Holger H. Hoos, Mario Roederer, and Ryan R. Brinkman. Early Immunologic Correlates of HIV Protection can be Identified from Computational Analysis of Complex Multivariate T-cell Flow Cytometry Assays. *submitted to Bioinformatics*, 2011.