Introduction to FlowSOM analyse

Anja Bråthen Kristoffersen

23 March 2022

This document describes the different lines of the file “runFlowSOM\_analyse.R”. To perform your own flowSOM analysis save the files “runFlowSOM\_analyse.R” and “FlowSOM\_analyse.R” from github <https://github.com/folkehelseinstituttet/cytof> give it a new name and change the necessary paths and values throughout the document to use it on your own data. The file “FlowSOM\_analyse.R” is sourced from the file “runFlowSOM\_analyse.R” see also the explanatory file called “Introduction to FlowSOM\_analyse.R”. These documents only make a clustering of your data and some additional files that later will be used when analysing the flowSOM data together with metadata, examples of analysing of the clustering data together with metadata can be found in “runResults\_Regression\_Clusters.R” and “resultater\_Regression\_clusters.Rmd”.

This document starts with removing everything that you may already have in R to make sure you start with no previous data. The removing is done by using the functions rm() and ls(). The function ls() list all objects in R while rm() remove everything in the list included.

rm(list = ls())

Then some librarys are loaded. If you do not have these libraries installed you have to install them by using the function install.packages(“LIBRARY NAME”)

library(ComplexHeatmap)  
library(gridExtra)  
library(ggplot2)  
library(grid)

This chunk is optional. If you want the clustering of the median signal to be clustered on markers the order will not be used. But if you do not want to cluster the markers this will give the order you want the markers to be shown in.

orderPanel1 <- c("X89Y\_CD45", "X116Cd\_CD3", "X145Nd\_CD4", "X113Cd\_CD8", "X111Cd\_CD19", "X152Sm\_TCRgd", "X166Er\_TCRVa7.2", "X149Sm\_CD25",   
 "X158Gd\_CD27", "X160Gd\_CD28", "X143Nd\_CD127", "X172Yb\_CD38", "X167Er\_CCR7", "X155Gd\_CD45RA", "X114Cd\_HLADR", "X146Nd\_IgD", "X159Tb\_IgG",  
 "X156Gd\_CXCR3", "X153Eu\_CCR4", "X141Pr\_CCR6", "X171Yb\_CXCR5", "X168Er\_ICOS", "X142Nd\_KLRG1", "X150Nd\_CD134\_OX40", "X154Sm\_TIGIT",  
 "X161Dy\_CD160", "X164Dy\_CD161", "X162Dy\_CD95", "X163Dy\_CRTH2", "X165Ho\_CD85j", "X169Tm\_NKG2A", "X173Yb\_CD141", "X174Yb\_CD279\_PD.1",  
 "X175Lu\_CD14", "X148Nd\_CD16", "X176Yb\_CD56", "X106Cd\_CD57", "X209Bi\_CD11b", "X147Sm\_CD11c", "X112Cd\_CD5", "X144Nd\_CD15", "X170Er\_CD169", "X151Eu\_CD123")  
  
orderPanel2 <- c("X89Y\_CD45", "X116Cd\_CD3", "X113Cd\_CD8", "X145Nd\_CD4", "X111Cd\_CD19", "X153Eu\_CXCR5.CD185",   
 "X114Cd\_HLADR", "X175Lu\_CD14", "X209Bi\_CD16", "X176Yb\_CD56", "X106Cd\_CD57", "X158Gd\_CD27",  
 "X160Gd\_CD28", "X169Tm\_CD25", "X172Yb\_CD38", "X112Cd\_CD44", "X143Nd\_CD127.IL7Ra", "X167Er\_CCR7.CD197",  
 "X155Gd\_CD45RA", "X162Dy\_FoxP3", "X170Er\_CTLA.4.CD152", "X163Dy\_CD33", "X154Sm\_CD272.BTLA", "X110Cd\_CD107a",  
 "X161Dy\_IL.17A", "X166Er\_IL.10", "X164Dy\_Perforin", "X171Yb\_GranzymeB", "X148Nd\_CD274.PD.L1",   
 "X173Yb\_CD273.PD.L2", "X159Tb\_GM.CSF", "X168Er\_CD154.CD40L", "X174Yb\_CD279.PD1", "X141Pr\_CD223.LAG3",   
 "X147Sm\_TIM.3", "X151Eu\_CD137.4.1BB", "X165Ho\_IFNg", "X142Nd\_IL.1b", "X156Gd\_IL.6", "X166Er\_IL.10",   
 "X152Sm\_TCRgd", "X151Eu\_CD137.4.1BB", "X149Sm\_IL.12p70", "X150Nd\_MIP.1b")

then the scriptPath has to be updated to where you have saved your script. Read the data you want to analyse. and decide which files you want to use in this analyses.

scriptPath <- fs::path("H:", "git", "cytof")  
#scriptPath <- fs::path("C:", "CyToF data", "fra github")  
  
source(fs::path(scriptPath, "readDataToAnalysePanel2.R"))  
  
tamed <- as.character(dInfo$filenames)  
tamed <- tamed[!grepl("FHI005", tamed)] #denne oppfører seg helt rart og tas ut av analysen.  
tamed <- tamed[!grepl("Ref1", tamed)] #denne oppfører seg helt rart og tas ut av analysen.  
tamed

ks is then assigned to be a vector with the number of clusters you want to analyse. Here you could use one or more numbers. FlowSOM requires the dimensions of the SOM to be specified, which is a default of 100 total clusters. Then, metaclusters are formed by combining these small clusters. Prior to the metaclustering, the 100 clusters are all the same. After different types of metaclustering (different ks), some of these final clusters may consist of the exact same initial clusters and some will be different. xdim and ydim tells how many nodes that flowSOM will use that later will be clustered to k clusters defined in ks. By setting xdim and ydim to 14 there will be 14\*14 = 196 nodes, so the ks has to be less than 196 and probably as a rule at most half of the number of nodes preferably less. The more nodes and cells you choose the longer time will the analysis take. But you can start the script and leave it running alone. The variable n\_per\_file tells how many cells to use per file.

ks <- c(40, 50, 60)  
n\_per\_file <- 25000  
xdim <- 14  
ydim <- 14

Here the path to where you want to save the data is assigned, this has to be changed according to your data. Together with a seed that ensure that you get the same result if you run the script one more time. The seed has to be changed to randomly draw another set of cells to analyse. If selectedEvents are TRUE channel has to be given as well. selectedEvents tells if you will draw from all cells or only from a selected number of cells here by using channel equal to the colnames in posNeg (the list for each file of 0 (negative), 1 (positive/low), 2 (high))

utSti <- fs::path("F:", "Forskningsprosjekter", "PDB 2794 - Immune responses aga\_", "Forskningsfiler", "JOBO", "CyTOF", "Analyse i R OUS", "Resultat\_Panel\_2\_ALLE")  
seed <- 12345 #nb må endres vil man vil gjøre et annet uttrekk  
selectedEvents <- FALSE  
channel <- NULL  
o <- orderPanel2  
column\_cluster <- FALSE  
  
source(fs::path(scriptPath, "FlowSOM\_analyse.R"))

By choosing another seed and source again you will have another run of the same script, just with another random generation of cells.

seed <- 54321 #nb må endres vil man vil gjøre et annet uttrekk  
source(fs::path(scriptPath, "FlowSOM\_analyse.R"))

The abow code made a run for all cells and sourced the file “FlowSOM\_analyse.R” which produced the clusters see “introduction to FlowSOM\_analyse.docx”. Below only B cells, selected by gating for CD3 positive, CD45 positive and CD4. positive. Not all files have that many B cells so the number of cells to be used for clustering per file has to be reduced. When number of cells for clustering is reduced also the dimension of the SOM should be reduced, here xdim = ydim = 10 is choosen, resulting in 100 nodes. Also, the number of different clusters expected when only clustering on B cells are reduced so the chosen ks are her 15, 20, 25, 30, 35 and 40.

for(j in 1:length(posNeg)){  
 posNeg[[j]]$CD3CD45CD4 <- posNeg[[j]]$CD3 & posNeg[[j]]$CD45 & posNeg[[j]]$CD4  
}  
  
utSti <- fs::path("F:", "Forskningsprosjekter", "PDB 2794 - Immune responses aga\_", "Forskningsfiler", "JOBO", "CyTOF", "Analyse i R OUS", "Resulta\_Panel\_2\_ALLE\_CD4\_test")  
seed <- 45903#nb må endres vil man vil gjøre et annet uttrekk  
selectedEvents <- TRUE  
channel <- "CD3CD45CD4"  
n\_per\_file <- 4250  
ks <- c(15, 20, 25, 30, 35, 40)  
xdim <- 10  
ydim <- 10  
o <- orderPanel2  
column\_cluster <- FALSE  
  
source(fs::path(scriptPath, "FlowSOM\_analyse.R"))

And then an example for T cells where CD8 are used for gating instead of CD4

for(j in 1:length(posNeg)){  
 posNeg[[j]]$CD3CD45CD8 <- posNeg[[j]]$CD3 & posNeg[[j]]$CD45 & posNeg[[j]]$CD8  
}  
  
utSti <- fs::path("F:", "Forskningsprosjekter", "PDB 2794 - Immune responses aga\_", "Forskningsfiler", "JOBO", "CyTOF", "Analyse i R OUS", "Resulta\_Panel\_2\_ALLE\_CD8")  
seed <- 45902#nb må endres vil man vil gjøre et annet uttrekk  
selectedEvents <- TRUE  
channel <- "CD3CD45CD8"  
n\_per\_file <- 4000  
ks <- c(15, 20, 25, 30, 35, 40)  
xdim <- 10  
ydim <- 10  
o <- orderPanel2  
column\_cluster <- FALSE  
  
  
source(fs::path(scriptPath, "FlowSOM\_analyse.R"))