

Here is a step by step guide for preparing the K-Means algorithm for submission to flowCAP using the R scripting language:

```
#Parse the CSV file
cells=read.csv('GvHD/CSV/001.csv')

#Use kmeans to find 4 populations
km=kmeans(cells,4)

#Make the output folder
system('mkdir GvHD/Output')

#Write the output file
write.csv(km$clusters, col.names=FALSE, row.names=FALSE,
paste("GvHD/Output/", sprintf("%.3d", 1), ".csv", sep=""));
```

Now, the R script needs to accept input parameters from the shell script:

```
args=(commandArgs(TRUE))
print(args)
if(length(args)==0){
  stop("No arguments supplied.")
}else{
  if(length(args)==2){
    stop("The kmeans algorithm requires the number
of populations to be pre-identified")
  }
  InputFolder=args[[1]];
  OutputFolder=args[[2]];
  n=as.numeric(args[[3]]);
}
system(paste('mkdir ', OutputFolder, sep=""))
ListOfFiles=dir(paste(InputFolder, '/CSV', sep=""))
for (i in ListOfFiles){
  cells=read.csv(paste(InputFolder, '/CSV/', i, sep=""))
  km=kmeans(cells,n)
  write.csv(km$cluster, col.names=FALSE, row.names=FALSE,
paste(OutputFolder, '/', i, sep=""));
}
```

Assuming that the script above is saved in *flowcap-kmeans.R*, the following shell script (*flowcap-kmeans.sh*) can be used to run it:

```
Rscript flowcap-kmeans.r $1 $2 $3
```

Finally, flowcap organizers can run the code in a unix shell:

```
./flowcap-kmeans.sh GvHD GvHD/Output 10
```