**Heatmap how-to for a labmate**

**20150903 by Niki**

#Clustering for genomic data needs to happen on the server. It will crash your personal.

#Your clustered input file can look like this. Make sure you remove all spaces and weird characters from the headings. Save as a .csv.

#green highlighted text is what I actually entered in the terminal

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Fibroblasts | MGT1week | MGT2weeks | MGT3weeks | cluster | order |
| ENSMUSG00000005774 | 1.93 | 1.13 | 0.4 | 0.54 | 100 | 6 |
| ENSMUSG00000003873 | 23.39333333 | 6.94 | 7.12 | 9.51 | 100 | 9 |
| ENSMUSG00000021243 | 24.29333333 | 22 | 4.54 | 15.78 | 100 | 10 |
| ENSMUSG00000028195 | 297.74 | 249.44 | 268.6 | 220.88 | 100 | 13 |
| ENSMUSG00000027346 | 6.83 | 4.63 | 2.98 | 3.57 | 100 | 14 |
| ENSMUSG00000085663 | 6.493333333 | 3.02 | 0.39 | 3.31 | 100 | 16 |
| ENSMUSG00000089901 | 22.86333333 | 17.07 | 12.42 | 10.17 | 100 | 19 |
| ENSMUSG00000078784 | 17.22666667 | 12.69 | 7.32 | 16.57 | 100 | 20 |
| ENSMUSG00000047675 | 81.45333333 | 26.57 | 43.92 | 26.24 | 100 | 21 |

#this is all done in the R console. First, figure out where you are

getwd()

#go to the directory where your data files are

setwd("/Users/stone/Desktop")

#make sure you actually went there

getwd()

#load gplots

library(gplots)

#set your color scheme; this is Christina’s favorite. It’s color blind friendly.

hmcolsby <- colorRampPalette(c("cornsilk","lightblue","cornflowerblue","darkblue","darkblue"))(256)

#make a variable name for your new data matrix and read in your data

ind\_matrix <- read.csv("20150903\_tamer\_File2\_fib\_MGT1\_2\_3Weeks\_2.csv")

#check that it read in correctly

head(ind\_matrix)

#make a new matrix out of just the fpkm data columns

ind\_matrix\_forHM <- as.matrix(ind\_matrix[,**2:5**])

#check that you have just the numbers in this, not the ensemble IDs or clusters

head(ind\_matrix\_forHM)

#now make your heatmap, using all the data columns from ind\_matrix. This will take a short while. Setting RowSideColors to the “cluster” column in the original data file makes a key for your clusters, so you can see where they are (it isn’t always that clear). I don’t know how to not have the junk on the right pop up. I just remove it in .ppt or illustrator. Similarly, the cluster key is a rainbow – I just use it as a guide and make my own brackets later.

heatmap(ind\_matrix\_forHM[,**1:4**], col = hmcolsby, Colv=NA, Rowv=NA, scale = "row", RowSideColors=as.character(ind\_matrix[,"cluster"]))

#export your figure as a pdf so it will be a vectorized image you can work with easily in illustrator or anywhere

### these settings are going to give you a heatmap where there is no commonly expressed gene cluster, across all conditions. This is very suspicious. It probably only makes sense to use these settings when you’ve preselected only the genes that are differentially expressed in a certain condition. Just keep it in mind when you discuss a figure made like this.

What I needed to enter:

setwd("~/Desktop")

library(gplots)

hmcolsby <- colorRampPalette(c("cornsilk","lightblue","cornflowerblue","darkblue","darkblue"))(256)

ind\_matrix <- read.csv("20150903\_tamer\_File2\_fib\_MGT1\_2\_3Weeks\_2.csv")

ind\_matrix\_forHM <- as.matrix(ind\_matrix[,2:5])

heatmap(ind\_matrix\_forHM[,1:4], col = hmcolsby, Colv=NA, Rowv=NA, scale = "row", RowSideColors=as.character(ind\_matrix[,"cluster"]))

What I actually entered:

20150903\_forTamer\_heatmapping.Rhistory

Moving output files from the server back to your desktop

scp nstone@10.1.101.251:~/RNA/HOPACH/20150829\_File3\_lineend.txt\_hopach\_clusters\_gene\_avg.csv /Users/stone/Desktop/Niki\_RNAseq\_Final\_analysis/20150829\_forHOPACH\_forTamer/