

How does `gene.locator` work in the `corplot` and `corplot2` functions

Sarah Song and Mik Black

June 29, 2006

1 Overview

This document demonstrates the use of the `gene.locator` argument in both the `corplot` and `corplot2` functions. The AML/ALL data of Golub et al. (1999) is again used to demonstrate the use of these functions.

The `gene.locator` argument helps users identify genes from the correlation image plots for a specific gene set. This allows genes that are highly correlated and differentially expressed between two biological conditions to be found from the correlation image plots. By default the `gene.locator` argument is set to `false` in the `corplot` and `corplot2` functions. It can be activated by setting `gene.locator=T`.

```
> library(multtest)
Loading required package: survival
Loading required package: splines
> library(hu6800)
> data(golub)
> rownames(golub) <- golub.gnames[,3]
> colnames(golub) <- golub.cl
> KEGG.list <- as.list(hu6800PATH)
> imat <- getImat(golub, KEGG.list, ms=10)
> colnames(imat) <- paste("KEGG", colnames(imat), sep="")
> ## transform multiple probesets data into average probesets data ##
> pathlist <- as.list(hu6800PATH)
> pathlist <- pathlist[match(rownames(golub), names(pathlist))]
> ids <- unlist(mget(names(pathlist), env=hu6800SYMBOL))
> output <- aveProbe(golub, imat, ids)
> newdata <- output$newx
> newimat <- output$newimat
> newimat <- newimat[,apply(newimat, 2, sum)>=10]
> selgene <- rownames(newimat)[newimat[,match("KEGG04620",colnames(newimat))]==1]
```

Figure 1 shows the process of using gene locator to find genes that are highly correlated and differentially expressed. The correlation image plot (top

right) shows the correlation and expression of genes in the KEGG04620 Toll-like receptor signaling pathway (one of the significant pathways detected via the PCOT2 methodology). Two image plots (the top-left and bottom-right ones) can be generated by calling `gene.locator=T` in the `corplot2` function. They are the correlation plots of genes for a specific gene group for the two distinct classes. On each plot the red block on the diagonal shows genes that are highly correlated. Gene expression values are displayed on the color bar in the middle of the top right plot, with green and red denoting down- and up-regulation in class 1 respectively. To select the genes users click the beginning and end points (the white circles in Figure 1) on the image plots by using the left key of the mouse. The names of these genes can be saved as a variable and then displayed.

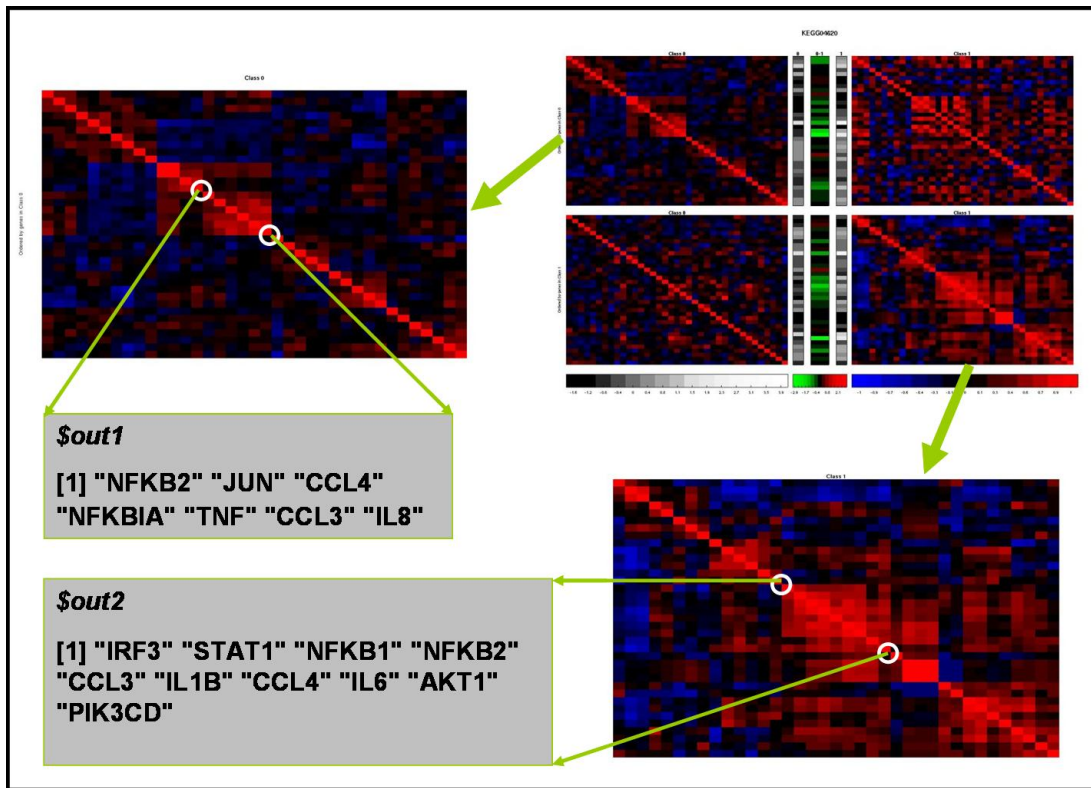


Figure 1: Using `gene.locator` to find genes from the correlation plot

```
> out<-corplot2(newdata,selgene,golub.cl,main="KEGG04620",add.name=F,gene.locator=T)
You have chosen 7 genes in Class0
You have chosen 10 genes in Class1

> out
$out1
[1] "NFKB2" "JUN" "CCL4"
     "NFKBIA" "TNF" "CCL3" "IL8"

$out2
[1] "IRF3" "STAT1" "NFKB1" "NFKB2"
     "CCL3" "IL1B" "CCL4" "IL6" "AKT1"
     "PIK3CD"
```

```
[1] "NFKB2" "JUN" "CCL4" "NFKBIA" "TNF" "CCL3" "IL8"
```

```
$out2
```

```
[1] "IRF3" "STAT1" "NFKB1" "NFKB2" "CCL3" "IL1B" "CCL4" "IL6"  
[9] "AKT1" "PIK3CD"
```

References

- [1] Golub T.R., Slonim D.K., et al.(1999) Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring, *Science*, **286**, 531-537.