

Interpreting LIGER results

Jean Fan

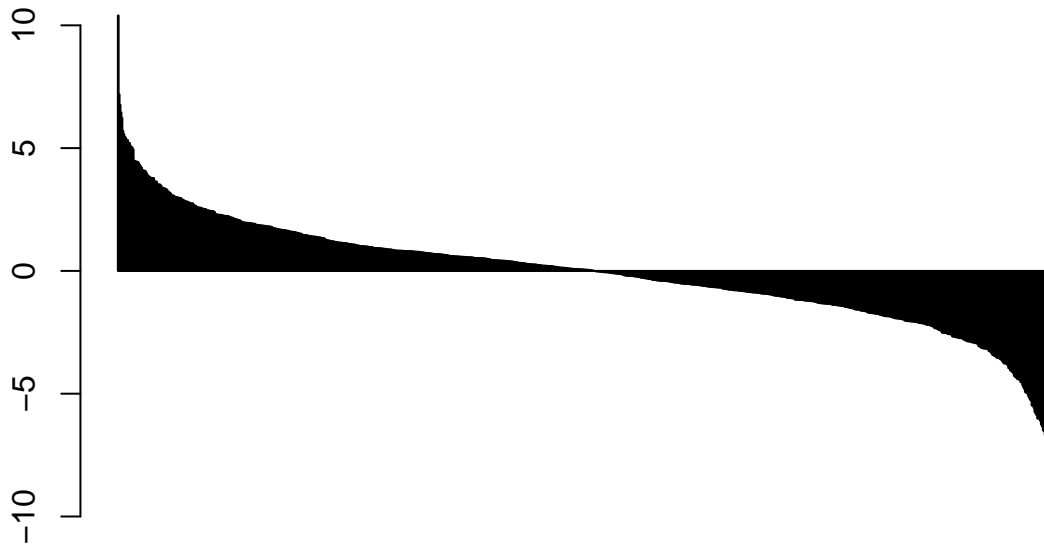
August 23, 2018

We will use a simulation to help us interpret the results of `liger`.

```
library(liger)
```

First, we will simulate some fake differential expression results. Perhaps we are comparing gene expression between two different conditions (X and Y) and we want to know which gene sets or pathways are being differentially up or down regulated. When we look at the sorted distribution of our gene ranking criteria (such as a log2 fold change), we will likely see a distribution similar to the following: a group of genes that are upregulated, and a group of genes that are downregulated. In this simulation, we will have 500 genes that are upregulated in condition X relative to condition Y and 500 genes that are upregulated in condition Y relative to condition X.

```
# Simulate differential expression
set.seed(0)
X <- abs(rnorm(1000, 1)) # gene expression in condition X
Y <- abs(rnorm(1000, 1)) # gene expression in condition Y
fc <- log2(X/Y) # log 2 fold change
fc <- sort(fc, decreasing=TRUE)
names(fc) <- paste0('gene', 1:length(fc))
barplot(fc, xaxt='n')
```



We will also simulate 4 gene sets or pathways.

```
genesets <- list(
  A = paste0('gene', seq(from=1, to=100, by=10)),
  B = paste0('gene', seq(from=1, to=750, by=10)),
  C = paste0('gene', seq(from=250, to=1000, by=10)),
  D = paste0('gene', seq(from=900, to=1000, by=10))
)
```

Now, we will use `liger` to run an iterative bulk gsea on all 4 gene sets. We find here that all 4 gene sets are significantly enriched (q -value < 0.05). But they are enriched in “different ways” and their interpretations are

slightly different. Let's take a closer look at the enrichment scores (`sscore`) and edge values (`edge`) to better understand what that means.

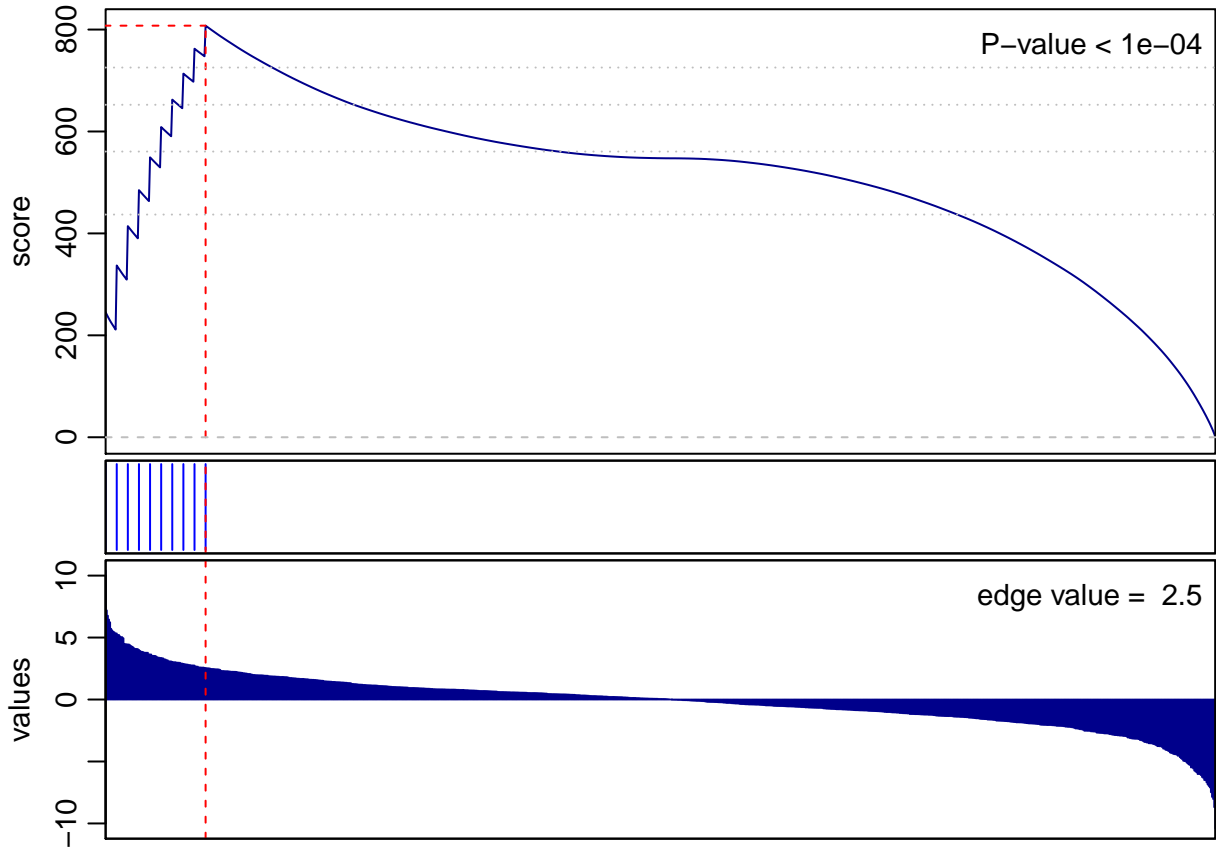
```
iterative.bulk.gsea(values = fc, set.list = genesets)
```

```
## initial: [1e+02 - 4] [1e+03 - 4] [1e+04 - 4] done
```

```
##      p.val    q.val    sscore    edge
## A 9.999e-05 9.999e-05  1.649078  2.543420
## B 9.999e-05 9.999e-05  1.824068 -1.218316
## C 9.999e-05 9.999e-05 -1.566468  1.093849
## D 9.999e-05 9.999e-05 -1.594859 -2.714962
```

Gene set A has a positive enrichment score and a positive edge value. This can be interpreted as: genes upregulated in condition X are enriched in gene set A.

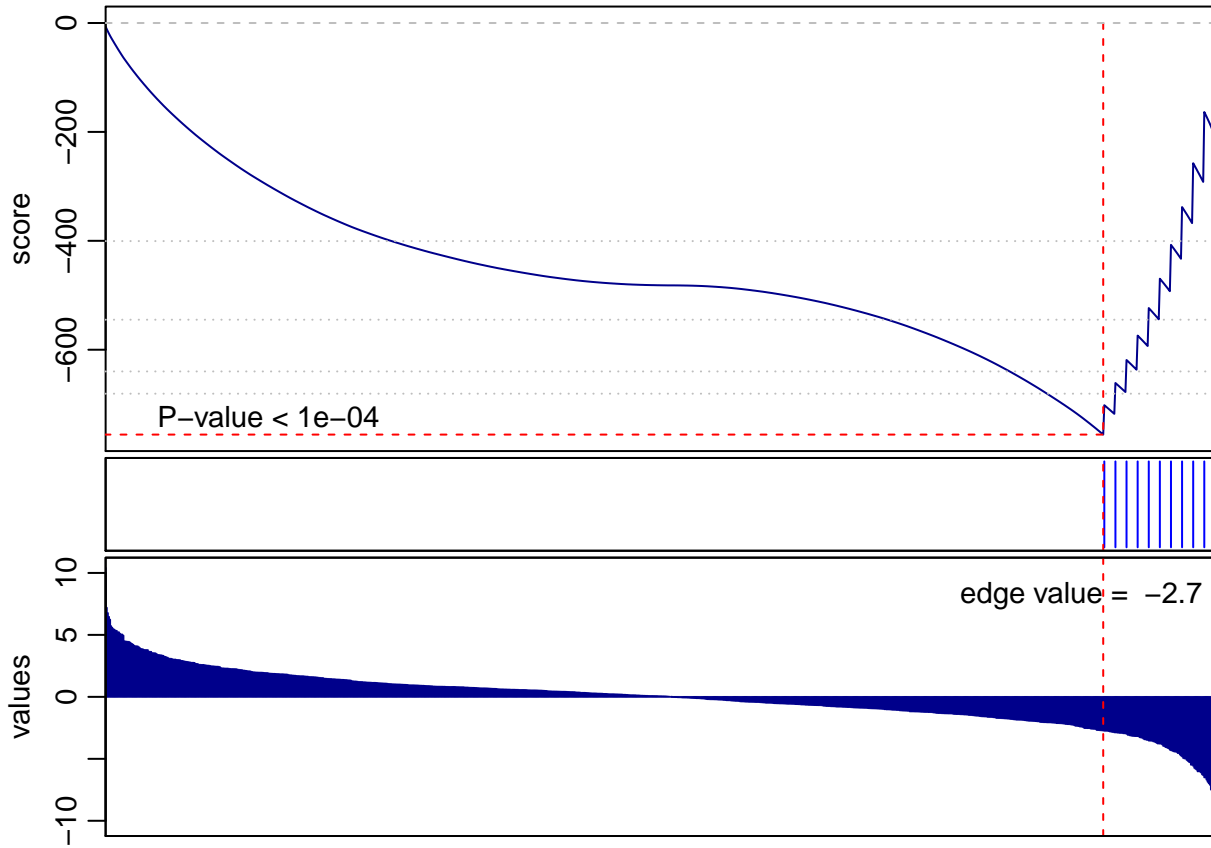
```
gsea(values = fc, geneset = genesets$A)
```



```
## [1] 1e-04
```

Gene set D has a negative enrichment score and a negative edge value. This can be interpreted as: genes upregulated in condition Y are enriched in gene set D.

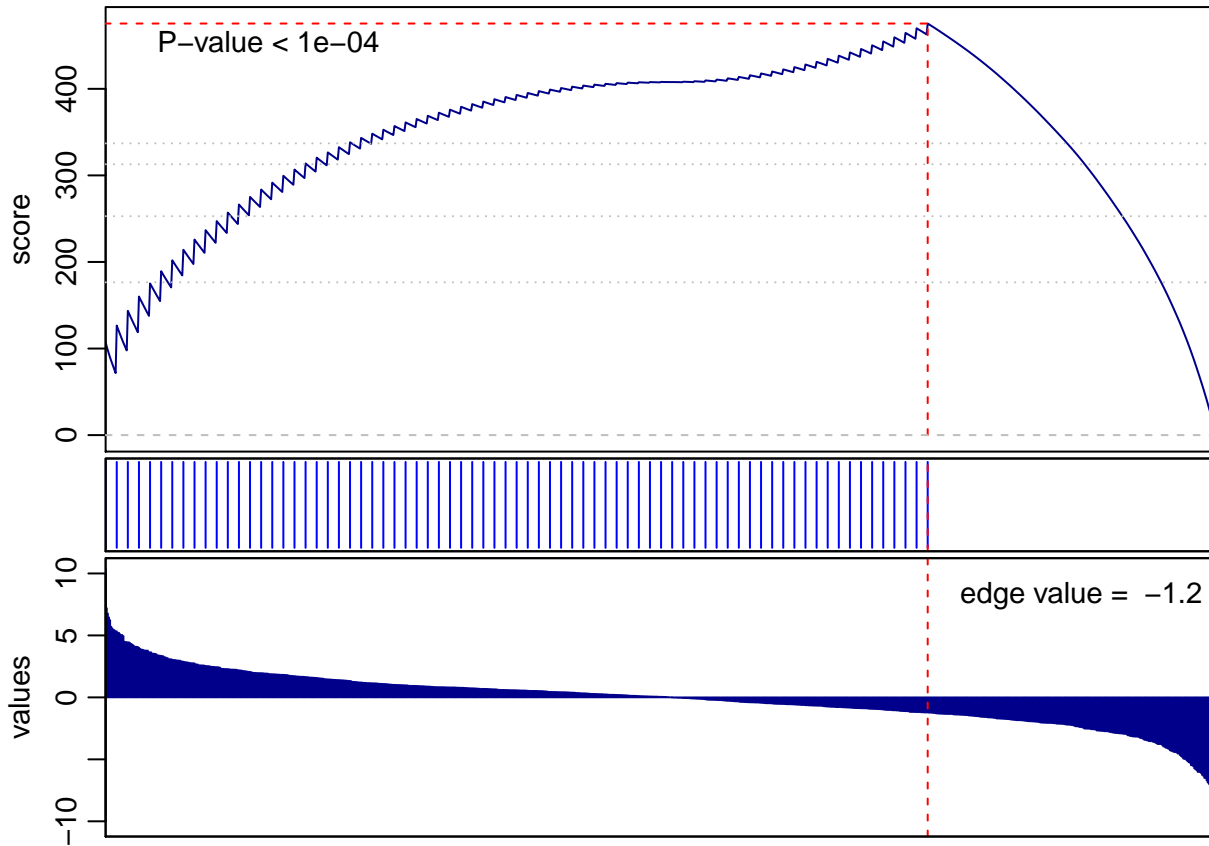
```
gsea(values = fc, geneset = genesets$D)
```



```
## [1] 1e-04
```

Gene set B has a positive enrichment score but a negative edge value. This can be interpreted as: genes upregulated in condition Y are depleted in gene set B. Note that just because genes upregulated in condition Y (ie. downregulated in condition X) are depleted in gene set B, doesn't mean that genes upregulated in condition X are necessarily enriched for gene set B. Visually, we can see that genes in this gene set are equally dispersed throughout the ranking of genes upregulated in condition X but completely missing from the top set of genes upregulated in condition Y.

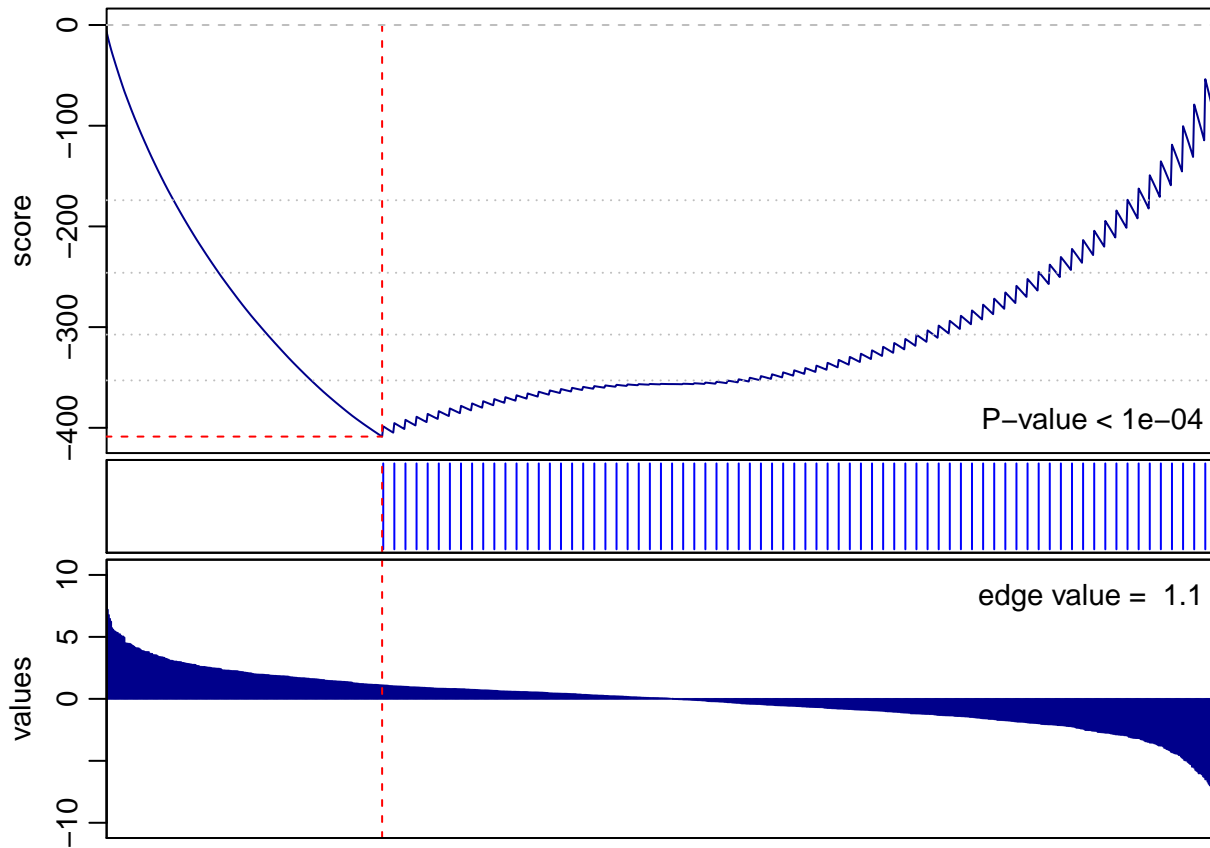
```
gsea(values = fc, geneset = genesets$B)
```



```
## [1] 1e-04
```

Similarly, gene set C has a negative enrichment score but a positive edge value. This can be interpreted as: genes upregulated in condition X are depleted in gene set C.

```
gsea(values = fc, geneset = genesets$C)
```



```
## [1] 1e-04
```