THvsFascin

Sarthak Sharma

May 7, 2018

Load the counts data

```
x <- read.delim("~/bulkRNAseq/HY7YJBGX2.dedup.matrix.full.txt", sep = ",")
samples = read.delim("~/bulkRNAseq/HY7YJBGX2.samples.txt")
genes = read.delim("~/bulkRNAseq/HY7YJBGX2.genes.txt", sep = ",")
```

Group the data

We had counts for 8 conditions (twice for replicates). For the present study, we used Fascin (group 1) and TH (group 7) counts data.

```
# adding grouping factors
group <- factor(c(1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8))
```

The following steps follow the procedure specified in edgeR documentation -

Building the design matrix

Test whether a set of genes is highly ranked relative to other genes in terms of differential expression. http://bioconductor.org/packages/release/bioc/html/edgeR.html

```
y <- DGEList(counts = x, group = group, samples = samples, genes = genes)
y <- calcNormFactors(y)
design <- model.matrix(~group)
y <- estimateDisp(y,design)
```

Exact Test

Exact Tests for Differences between Two Groups of Negative-Binomial Counts

```
# Again, Fascin was group 1 and TH was group 7
th_vs_fasc <- exactTest(y,pair = c(1,7))
```

Write to an output file

Write the output to csv file for downstream analysis.

```
write.csv( topTags(th_vs_fasc,n=nrow(th_vs_fasc),sort.by = "none"),
           file = "THxFasc.csv", fileEncoding = "UTF-8")
```