`testerccdashboard` Package Example

Sarah A. Munro

November 6, 2013

This vignette describes the use of the erccdashboard R package to analyze External RNA Control Consortium (ERCC) spike-in controls in gene expression experiments. Two types of data from the SEQC/MAQC III project were analyzed.

1. Rat toxicogenomics treatment and control samples for different drug treatments
2. Human reference RNA samples from the MAQC I project, Universal Human Reference RNA (UHRR) and Human Brain Reference RNA (HBRR)

1 Rat Toxicogenomics Example: MET treatment

1.1 initDat function to create expDat

Load the testerccdashboard package.

```
> library( "testerccdashboard" )
```

Load Rat Toxicogenomics Data set.

```
> load(file = system.file("data/SEQC.RatTox.Example.RData",
+                         package = "testerccdashboard"))
```

The R workspace should now contain 5 count tables and for each count table a corresponding total reads vector. Take a look at the data for the MET experiment.

```
> head(COH.RatTox.ILM.MET.CTL.countTable)
```

```
         Feature MET_1 MET_2 MET_3 CTL_1 CTL_2 CTL_3
16499 ERCC-00002 16629 18798 26568 36600 45436 25163
16500 ERCC-00003  1347  1565  1983  3048  3447  2195
16501 ERCC-00004  4569  5570  6755  1240  1484   902
16502 ERCC-00009   811   869  1123   909  1073   537
16504 ERCC-00013     3     1     2     1     5     1
16508 ERCC-00019    24    32    43     5    13     4
```

```
> COH.RatTox.ILM.MET.CTL.totalReads

[1] 41423502 46016148 44320280 38400362 47511484 33910098
```

The first column of the count table, "Feature", contains unique names for all the transcripts that were quantified in this experiment. The remaining columns represent sample replicates and each column heading is unique

The total reads vectors will be used for library size normalization of the count tables. Total reads can either represent the total number of reads in FASTQ files or total mapped reads. In the examples provided with this package FASTQ file total reads are used.

Control rat samples are labeled CTL and the labels 3ME, MET, NAP, THI, and NIT represent the Count tables for 5 different rat toxicogenomics experiments and vectors with total reads for each da Define the list, expDat, to contain experimental data and metadata.

countTable, totalReads, designFactors, sample1Name, sample2Name, erccMixes, ERCCdilution, spikeVol, totalRNAmass, printPDF, filenameRoot, DEtest, choseFDR, totalSeqReads, libeSizeNorm, myYLimMA, myXLim, myYLim

```
> expDat <- initDat(countTable=COH.RatTox.ILM.MET.CTL.countTable,
+                   totalReads=COH.RatTox.ILM.MET.CTL.totalReads,
+                   designFactors = c("Sample","Library"), sample1Name = "MET",
+                   sample2Name = "CTL", erccMixes = "Ambion4plexPair",
+                   ERCCdilution = 1/100, spikeVol = 1, totalRNAmass = 0.500,
+                   printPDF = F, filenameRoot = "COH.ILM", DEtest = T,
+                   choseFDR = 0.1, totalSeqReads = T, libeSizeNorm = T,
+                   myYLimMA = c(-3.5,3.5), myXLim = c(-10,15),myYLim = NULL)

[1] "COH.ILM.MET.CTL"
[1] "Feature" "MET_1"   "MET_2"   "MET_3"   "CTL_1"   "CTL_2"   "CTL_3"
[1] "Library sizes:"
[1] 41.42350 46.01615 44.32028 38.40036 47.51148 33.91010
[1] "Using total sequencing reads mean library size = "
[1] 41.93031

> print(str(expDat))

List of 14
 $ sampleInfo     :List of 19
  ..$ designFactors: chr [1:2] "Sample" "Library"
  ..$ sample1Name  : chr "MET"
  ..$ sample2Name  : chr "CTL"
  ..$ choseFDR     : num 0.1
  ..$ ERCCdilution : num 0.01
  ..$ spikeVol     : num 1
  ..$ totalRNAmass : num 0.5
  ..$ printPDF     : logi FALSE
```

```
 ..$ DEtest       : logi TRUE
 ..$ totalSeqReads: logi TRUE
 ..$ libeSizeNorm : logi TRUE
 ..$ myYLimMA     : num [1:2] -3.5 3.5
 ..$ myXLim       : num [1:2] -10 15
 ..$ myYLim       : NULL
 ..$ filenameRoot : chr "COH.ILM.MET.CTL"
 ..$ idColsSRM    :'data.frame':        96 obs. of  6 variables:
 .. ..$ Feature: Factor w/ 96 levels "ERCC-00002","ERCC-00003",..: 1 2 3 4 5 6 7 8 9 10 ...
 .. ..$ Length : int [1:96] 1061 1023 523 1135 984 994 808 1957 844 1136 ...
 .. ..$ GC     : num [1:96] 0.51 0.33 0.34 0.46 0.47 0.51 0.43 0.44 0.48 0.51 ...
 .. ..$ Ratio  : Factor w/ 4 levels "a","b","c","d": 4 4 1 NA 2 3 4 4 3 1 ...
 .. ..$ Conc1  : num [1:96] 15000 938 7500 0 938 ...
 .. ..$ Conc2  : num [1:96] 30000 1875 1875 0 938 ...
 ..$ MixDef       :'data.frame':        96 obs. of  4 variables:
 .. ..$ Feature              : Factor w/ 96 levels "ERCC-00002","ERCC-00003",..: 1 2 3 4 5
 .. ..$ Ratio                : Factor w/ 4 levels "a","b","c","d": 4 4 1 NA 2 3 4 4 3 1 ...
 .. ..$ Mix1Conc.Attomoles_ul: num [1:96] 15000 938 7500 0 938 ...
 .. ..$ Mix2Conc.Attomoles_ul: num [1:96] 30000 1875 1875 0 938 ...
 ..$ FCcode       :'data.frame':        4 obs. of  2 variables:
 .. ..$ Ratio: Factor w/ 4 levels "a","b","c","d": 1 2 3 4
 .. ..$ FC   : num [1:4] 4 1 0.667 0.5
 ..$ legendLabels : chr [1:4] "4:1" "1:1" "1:1.5" "1:2"
$ totalReads     : int [1:6] 41423502 46016148 44320280 38400362 47511484 33910098
$ Transcripts    :'data.frame':        11583 obs. of  7 variables:
 ..$ Feature: chr [1:11583] "ERCC-00002" "ERCC-00003" "ERCC-00004" "ERCC-00009" ...
 ..$ MET_1  : int [1:11583] 16629 1347 4569 811 3 24 162 42 3 1 ...
 ..$ MET_2  : int [1:11583] 18798 1565 5570 869 1 32 184 54 7 1 ...
 ..$ MET_3  : int [1:11583] 26568 1983 6755 1123 2 43 227 74 6 8 ...
 ..$ CTL_1  : int [1:11583] 36600 3048 1240 909 1 5 323 50 1 4 ...
 ..$ CTL_2  : int [1:11583] 45436 3447 1484 1073 5 13 446 62 0 4 ...
 ..$ CTL_3  : int [1:11583] 25163 2195 902 537 1 4 218 29 0 3 ...
$ designMat      :'data.frame':        6 obs. of  3 variables:
 ..$ countSet: Factor w/ 6 levels "CTL_1","CTL_2",..: 4 5 6 1 2 3
 ..$ Sample  : Factor w/ 2 levels "CTL","MET": 2 2 2 1 1 1
 ..$ Library : Factor w/ 3 levels "1","2","3": 1 2 3 1 2 3
$ sampleNames    : chr [1:2] "MET" "CTL"
$ idCols         :'data.frame':        63 obs. of  6 variables:
 ..$ Feature: Factor w/ 96 levels "ERCC-00002","ERCC-00003",..: 1 2 3 5 7 12 13 16 17 18 ..
 ..$ Length : int [1:63] 1061 1023 523 984 808 644 751 1994 1130 1138 ...
 ..$ GC     : num [1:63] 0.51 0.33 0.34 0.47 0.43 0.49 0.47 0.5 0.51 0.48 ...
 ..$ Ratio  : Factor w/ 4 levels "a","b","c","d": 4 4 1 2 4 1 4 2 1 2 ...
 ..$ Conc1  : num [1:63] 318.3 19.1812 78.45 18.45 0.0148 ...
 ..$ Conc2  : num [1:63] 636.6 38.3625 19.6125 18.45 0.0296 ...
$ totalReads     : int [1:6] 41423502 46016148 44320280 38400362 47511484 33910098
$ expressDat     :'data.frame':        63 obs. of  7 variables:
```

```
  ..$ Feature: chr [1:63] "ERCC-00002" "ERCC-00003" "ERCC-00004" "ERCC-00009" ...
  ..$ MET_1  : num [1:63] 401.4388 32.5178 110.2997 19.5783 0.0724 ...
  ..$ MET_2  : num [1:63] 408.5088 34.0098 121.0445 18.8847 0.0217 ...
  ..$ MET_3  : num [1:63] 599.4547 44.7425 152.4133 25.3383 0.0451 ...
  ..$ CTL_1  : num [1:63] 953.116 79.374 32.291 23.672 0.026 ...
  ..$ CTL_2  : num [1:63] 956.316 72.551 31.235 22.584 0.105 ...
  ..$ CTL_3  : num [1:63] 742.0503 64.73 26.5997 15.836 0.0295 ...
 $ libeSize       : num [1:6] 41.4 46 44.3 38.4 47.5 ...
 $ ERCCxlabelIndiv:  expression(paste("Log2 ERCC Spike Amount(attomol nt/", mu, "g total RNA
 $ ERCCxlabelAve  :  expression(paste("Log2 Average ERCC Spike Amount(attomol nt/", mu, "g t
 $ spikeFraction  : num 0.02
 $ mnLibeFactor   : num 41.9
 $ sampleLibeSums : int [1:6] 41423502 46016148 44320280 38400362 47511484 33910098
NULL
```