

# Analysis of Mixtrue Experiment data set

Cynthia Liu and Matt Ritchie

18 December 2014

## Introduction

This data set was generated by Aliaksei Holik from Asselin-Labat Laboratory at the WEHI. FASTQ files from this experiment are available from GEO under accession number GSE64098. The matrix of counts analysed below was obtained by running Subread and featureCounts from the Rsubread package. The data set is from mixture experiment with total RNA libraries mapped to the reverse strand information. The R code below generated the results presented in Liu et al. (2014).

## Importing the data

```
library(edgeR)
library(limma)
load("mixtureExperiment.rda")
counts= tstrand$counts
Repl=as.factor(tstrand$samples[, "Replicate"])
group=as.factor(tstrand$samples[, "group"])

des=model.matrix(~0+group)
colnames(des) = c("mix0", "mix25", "mix50", "mix75", "mix100")

x=calcNormFactors(tstrand, method="TMM")
goodrep2 = 6:10
dodgyrep2 = 11:15

des2 = model.matrix(~tstrand$samples$group[-dodgyrep2]-1)
colnames(des2) = c("mix0", "mix25", "mix50", "mix75", "mix100")

#seperate data into different groups #
RSCE.good = calcNormFactors(tstrand[, -dodgyrep2])
RSCE.dodgy = calcNormFactors(tstrand[, -goodrep2])
RSCE.remove=calcNormFactors(tstrand[, -c(6:15)])

# Make an MDS plot: include all dodgy and good samples
plotMDS(x, labels=tstrand$sample$Sample, col=as.numeric(tstrand$sample$group))
```



```
summary(dtgoodonly)

##      100vs0 75vs0 50vs0 25vs0
## -1     6037  5112  3720  1053
##  0     10882 12371 14932 19714
##  1       6225  5661  4492  2377
```

## Voom with sample weights

```
# Analysis of Degraded samples using combined voom and sample quality weights
#figure 8A
vdodgya = voomWithQualityWeights(RSCE.dodgy, des2, plot=F)

contr = makeContrasts("100vs0"=mix100-mix0,"75vs0"=mix75-mix0,
                      "50vs0"=mix50-mix0,"25vs0"=mix25-mix0,levels=des2)

fitdodgya = lmFit(vdodgya)
cfitdodgya = contrasts.fit(fitdodgya, contr)
cfitdodgya = eBayes(cfitdodgya)

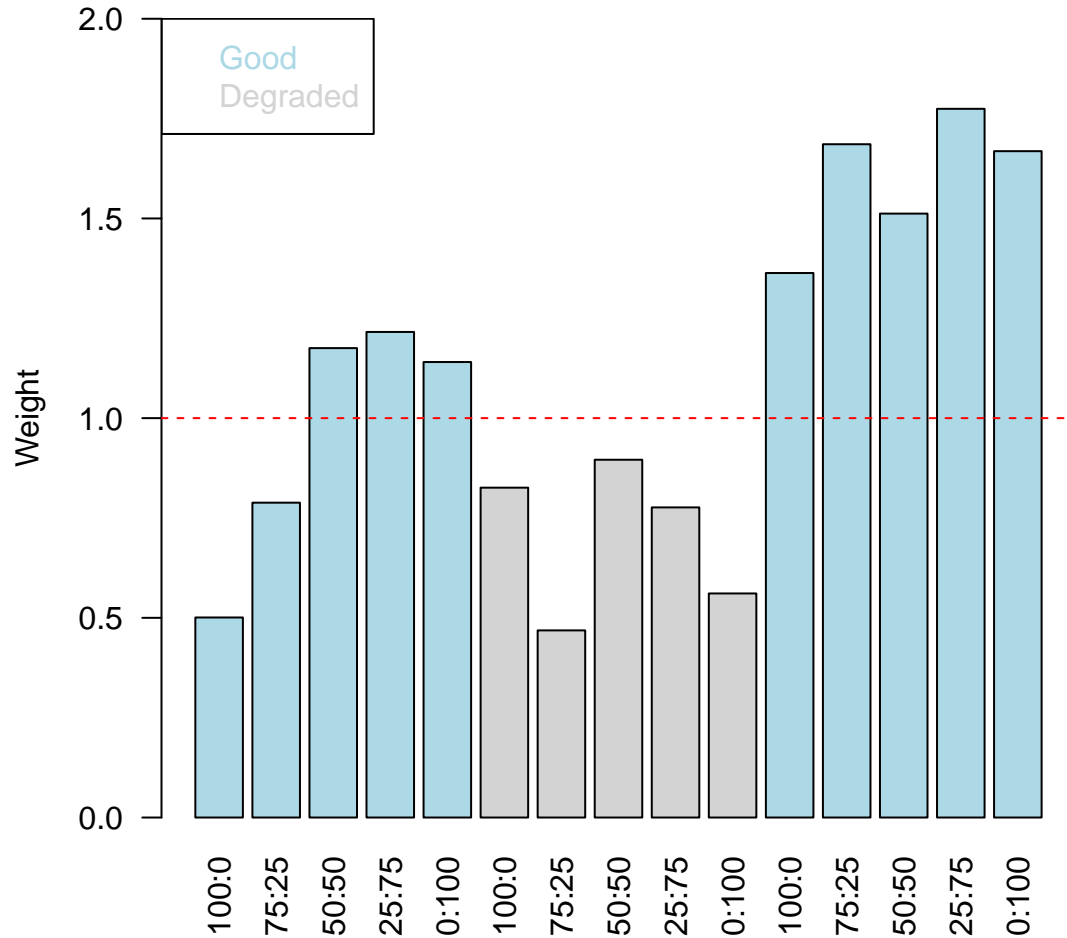
dtdodgya = decideTests(cfitdodgya)

summary(dtdodgya)

##      100vs0 75vs0 50vs0 25vs0
## -1     5895  4849  3250   589
##  0     11272 12894 15703 20620
##  1       5977  5401  4191  1935

barplot(vdodgya$sample, main="Figure 8(A)", col=rep(c("lightblue", "lightgrey", "lightblue"),
           times=c(5,5,5)), ylab="Weight",
        names=rep(c("100:0", "75:25", "50:50", "25:75", "0:100"), times=3), las=2, ylim=c(0,2))
legend("topleft", legend=c("Good", "Degraded"), text.col=c("lightblue", "lightgrey"))
abline(h=1, col=2, lty=2)
```

Figure 8(A)



## Voom after removing outlier samples

```
# Analysis where outlier samples have been removed
des3=model.matrix(~tstrand$samples$group[-c(6:15)]-1)
colnames(des3) = c("mix0", "mix25", "mix50", "mix75", "mix100")

vremove = voom(RSCE.remove, des3, plot=F)

contrRemove = makeContrasts("100vs0"=mix100-mix0,"75vs0"=mix75-mix0,
                             "50vs0"=mix50-mix0,"25vs0"=mix25-mix0,levels=des3)
fitremove = lmFit(vremove)
cfitremove = contrasts.fit(fitremove, contrRemove)
```

```
cfitremove = eBayes(cfitremove)

dtremove = decideTests(cfitremove)

summary(dtremove)

##      100vs0 75vs0 50vs0 25vs0
## -1    5231  4042  2413   393
##  0    12610 14552 17494 21480
##  1     5303  4550  3237  1271
```

## Voom with block weights

```
# Analysis of Degraded samples using combined voom and blocked sample quality weights
#figure 8B
Z <- cbind(c(1,1,1,1,1,0,0,0,0,0,1,1,1,1,1),c(0,0,0,0,0,1,1,1,1,1,0,0,0,0,0))

vdodgyb = voomWithQualityWeights(RSCE.dodgy, design=des2,
  normalization="none", var.design=Z, plot=F)
contr = makeContrasts("100vs0"=mix100-mix0,"75vs0"=mix75-mix0,
  "50vs0"=mix50-mix0,"25vs0"=mix25-mix0,levels=des2)

fitdodgyb = lmFit(vdodgyb)
cfitdodgyb = contrasts.fit(fitdodgyb, contr)
cfitdodgyb = eBayes(cfitdodgyb)

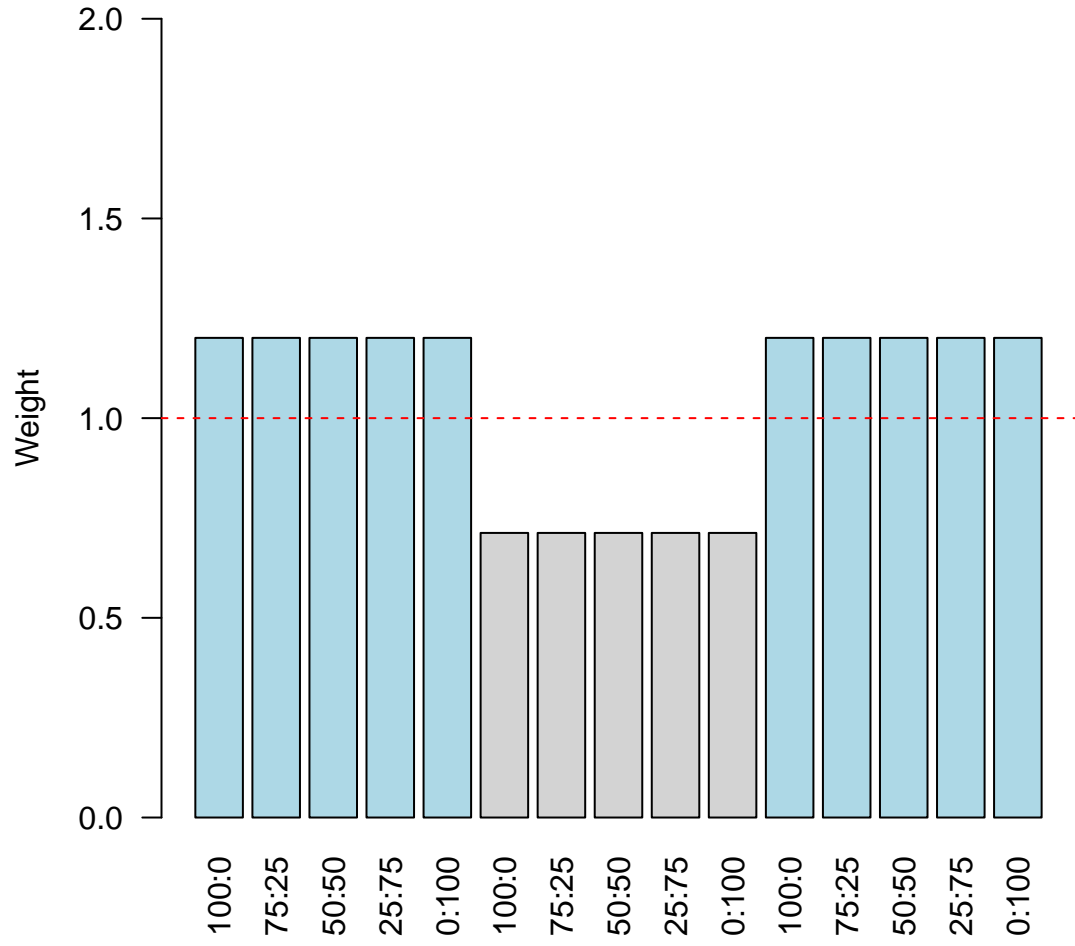
dtdodgyb = decideTests(cfitdodgyb)

summary(dtdodgyb)

##      100vs0 75vs0 50vs0 25vs0
## -1      5760  4673  3090   622
##  0     11482 13208 15992 20540
##  1       5902  5263  4062  1982

barplot(vdodgyb$sample, main="Figure 8(B)", col=rep(c("lightblue", "lightgrey", "lightblue"),
  times=c(5,5,5)), ylab="Weight",
  names=rep(c("100:0", "75:25", "50:50", "25:75", "0:100"), times=3), las=2, ylim=c(0,2))
abline(h=1, col=2, lty=2)
```

Figure 8(B)



## Overlap DE genes between good samples and four methods

```

#Checking the agreement between DE genes from analysis of good sample and DE genes from anal-
ysis with dodgy samples#
n=n1=matrix(NA,4,5)
for(i in 1:4){
  coegoodonly=cfitgoodonly$genes$GeneID[dtgoodonly[,i]!=0]

  coedodgyonly=cfitdodgyonly$genes$GeneID[dtdodgyonly[,i]!=0]

  coedodgya=cfitdodgya$genes$GeneID[dtdodgya[,i]!=0]
  coedodgyb=cfitdodgyb$genes$GeneID[dtdodgyb[,i]!=0]

  coeremove=cfitremove$genes$GeneID[dtremove[,i]!=0]

  n[i,]=c(length(intersect(coegoodonly,coedodgyonly)),
          length(intersect(coegoodonly,coedodgya)),length(intersect(coegoodonly,coedodgyb)),
          length(intersect(coegoodonly,coeremove)),length(coegoodonly))
  n1[i,]=c(length(coedodgyonly),length(coedodgya),length(coedodgyb),length(coeremove),length(coegoodonly)
          )
}
n=apply(n,2,function(x){paste(round(x,1),"%",sep="")})

colnames(n)=c("Degraded-Voom", "Dodgy-Voom+Sample",
              "Degraded-voom+block", "Remove Outliers", "Good analysis")
colnames(n1)=c("Degraded-Voom", "Dodgy-Voom+Sample",
               "Degraded-voom+block", "Remove Outliers", "Good analysis")
rownames(n)=rownames(n1)=c("100vs0", "75vs0", "50vs0", "25vs0")
n

##          Degraded-Voom Dodgy-Voom+Sample Degraded-voom+block
## 100vs0  "88.7%"          "90.4%"          "90.4%"
## 75vs0   "84.8%"          "87.2%"          "86.8%"
## 50vs0   "78.1%"          "81.5%"          "80.8%"
## 25vs0   "64%"           "64.6%"          "67.3%"
##          Remove Outliers Good analysis
## 100vs0  "82.7%"          "100%"
## 75vs0   "76.6%"          "100%"
## 50vs0   "65.6%"          "100%"
## 25vs0   "44.8%"          "100%"

n1

##          Degraded-Voom Dodgy-Voom+Sample Degraded-voom+block
## 100vs0          11403          11872          11662
## 75vs0           9692          10250           9936
## 50vs0           6924           7441           7152
## 25vs0           2479           2524           2604
##          Remove Outliers Good analysis

```



```
## 100vs0      10534      12262
## 75vs0       8592      10773
## 50vs0       5650       8212
## 25vs0       1664       3430
```

```
sessionInfo()
```

```
## R version 3.1.1 (2014-07-10)
## Platform: x86_64-unknown-linux-gnu (64-bit)
##
## locale:
## [1] LC_CTYPE=en_AU.UTF-8      LC_NUMERIC=C
## [3] LC_TIME=en_AU.UTF-8      LC_COLLATE=en_AU.UTF-8
## [5] LC_MONETARY=en_AU.UTF-8  LC_MESSAGES=en_AU.UTF-8
## [7] LC_PAPER=en_AU.UTF-8     LC_NAME=C
## [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_AU.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods
## [7] base
##
## other attached packages:
## [1] edgeR_3.8.2  knitr_1.8   limma_3.23.1
##
## loaded via a namespace (and not attached):
## [1] evaluate_0.5.5 formatR_1.0  highr_0.4   stringr_0.6.2
## [5] tools_3.1.1
```