

Analyzing Shoal

Noor Pratap Singh

1/4/2022

```
suppressPackageStartupMessages(source("run_limma.R"))
dirSal <- "/fs/cbcb-lab/rob/students/noor/shoal_proj/swim/sal_best/vbprior=1e0"
load("/fs/cbcb-lab/rob/students/noor/Uncertainty/mikelove-swimdown-216aldd/simulate/data/simulate.rda")
library(pracma)
```

```
dirShoal <- "/fs/cbcb-lab/rob/students/noor/shoal_proj/swim/shoal_best_updated/vbprior=1e0/c=0.1"
suppressWarnings(dd <- runDE(dirSal, dirShoal, seq(0,1,0.05)))
df <- dd[["cob"]][["cob"]][[1]]@fdrtpr
inds <- seq(1,nrow(df),2)

print(trapz(df[inds, "FDR"], df[inds, "TPR"])) ##Area under roc curve for Salmon
```

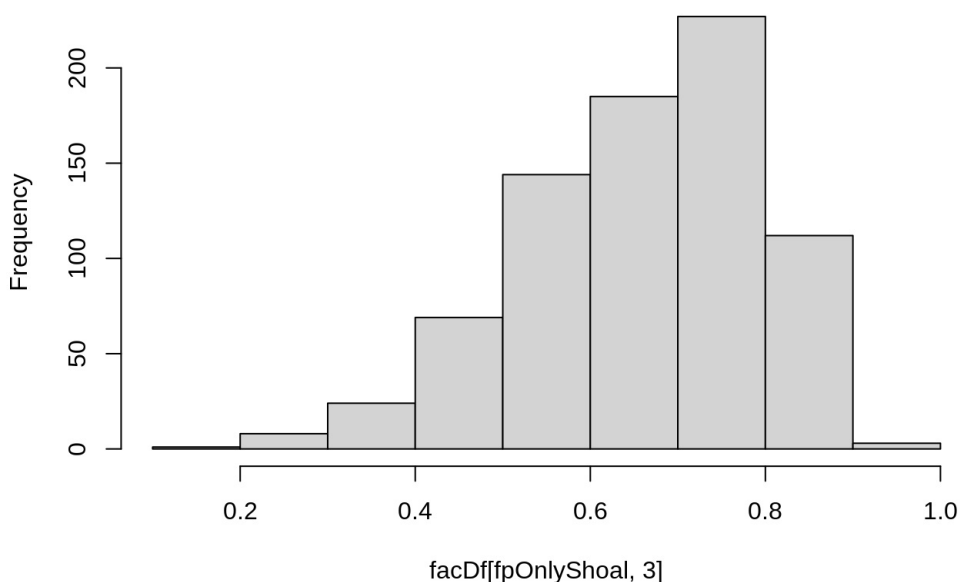
```
## [1] 0.6320981
```

```
print(trapz(df[inds+1, "FDR"], df[inds+1, "TPR"])) ##Area under roc curve for Shoal
```

```
## [1] 0.7062926
```

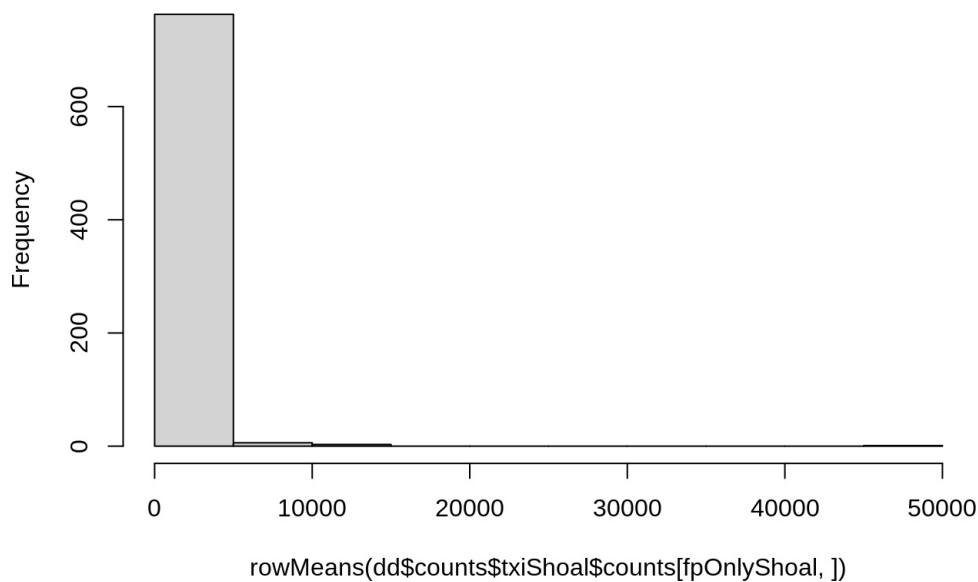
```
pShoal <- which(dd[["cob"]][["cd"]][@padj$Shoal <= 0.1)
pSal <- which(dd[["cob"]][["cd"]][@padj$Sal <= 0.1)
fpSal <- setdiff(pSal, which(dd[["cob"]][["cd"]][@truth$status==1))
fpShoal <- setdiff(pShoal, which(dd[["cob"]][["cd"]][@truth$status==1))
fpOnlyShoal <- rownames(dd[["cob"]][["cd"]][@truth][setdiff(fpShoal, fpSal)])
facDf <- read.delim("swim/shoal_best_updated/vbprior=1e0/c=0.1/prior.tsv", row.names = 1)
hist(facDf[fpOnlyShoal,3])
```

Histogram of facDf[fpOnlyShoal, 3]



```
hist(rowMeans(dd$counts$txiShoal$counts[fpOnlyShoal,]))
```

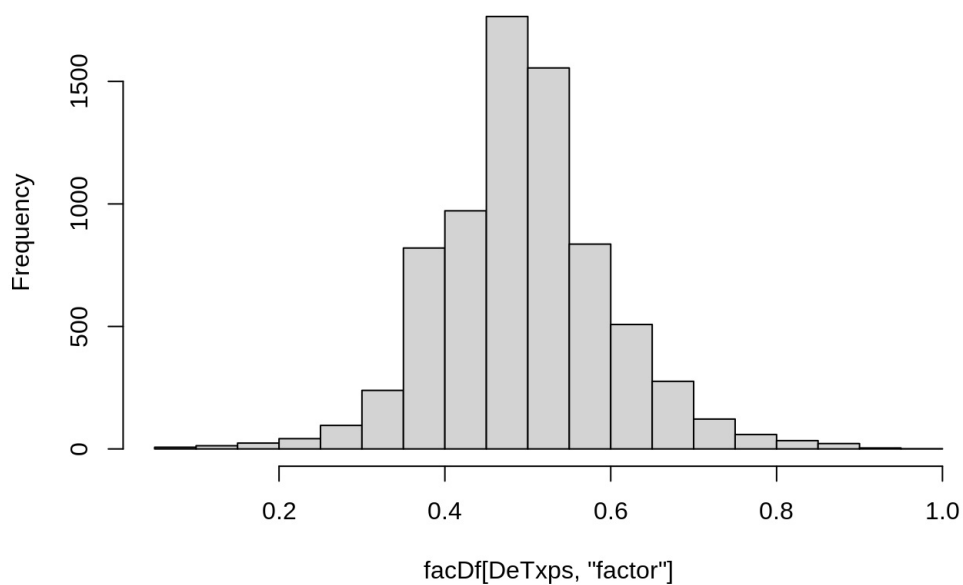
Histogram of rowMeans(dd\$counts\$txiShoal\$counts[fpOnlyShoal,])



Factors for differential and non differential transcripts

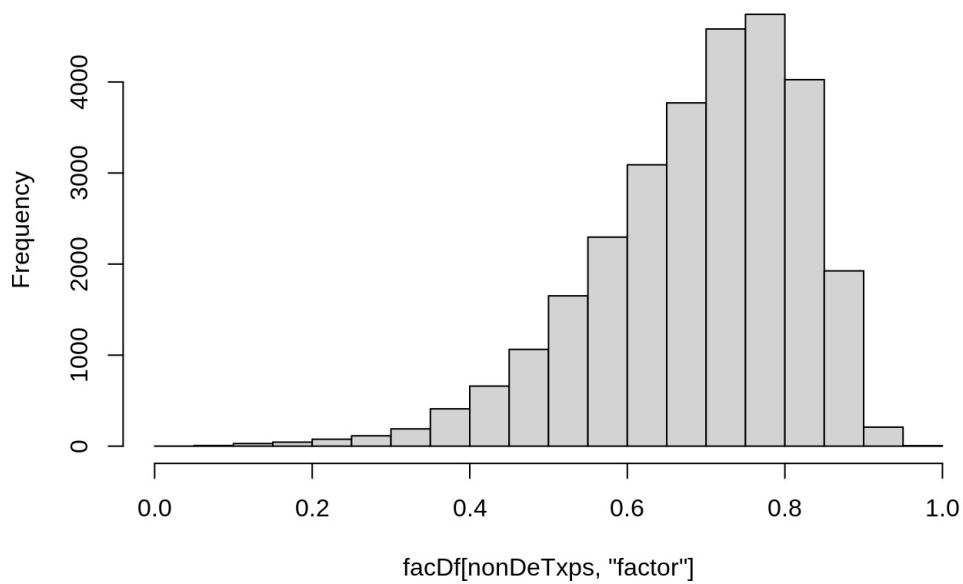
```
DeTxps <- rownames(dd[["cob"]][["cd"]>@truth][dd[["cob"]][["cd"]>@truth==1]  
nonDeTxps <- rownames(dd[["cob"]][["cd"]>@truth][dd[["cob"]][["cd"]>@truth==0]  
  
hist(facDf[DeTxps, "factor"])
```

Histogram of facDf[DeTxps, "factor"]



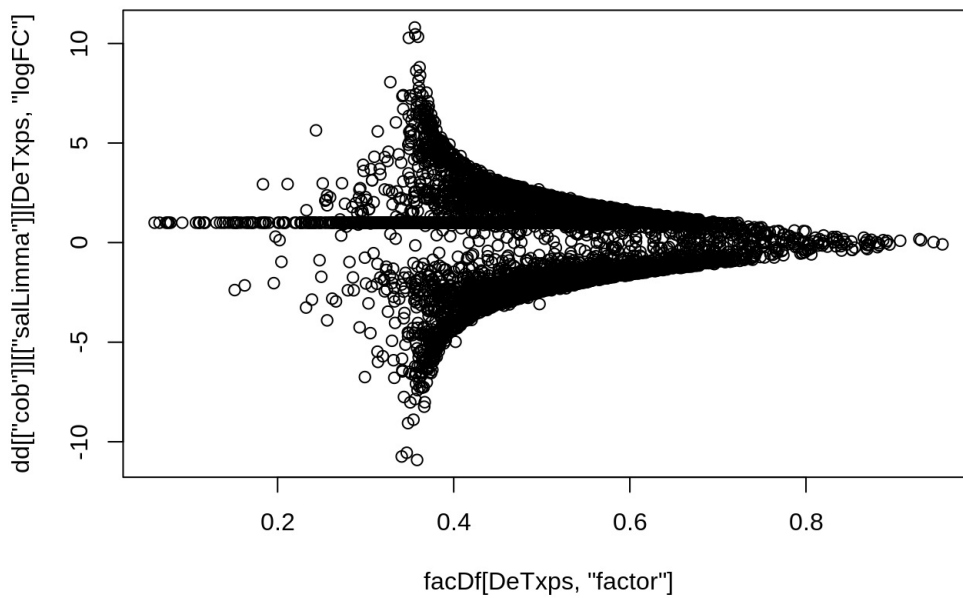
```
hist(facDf[nonDeTxps, "factor"])
```

Histogram of facDf[nonDeTxps, "factor"]

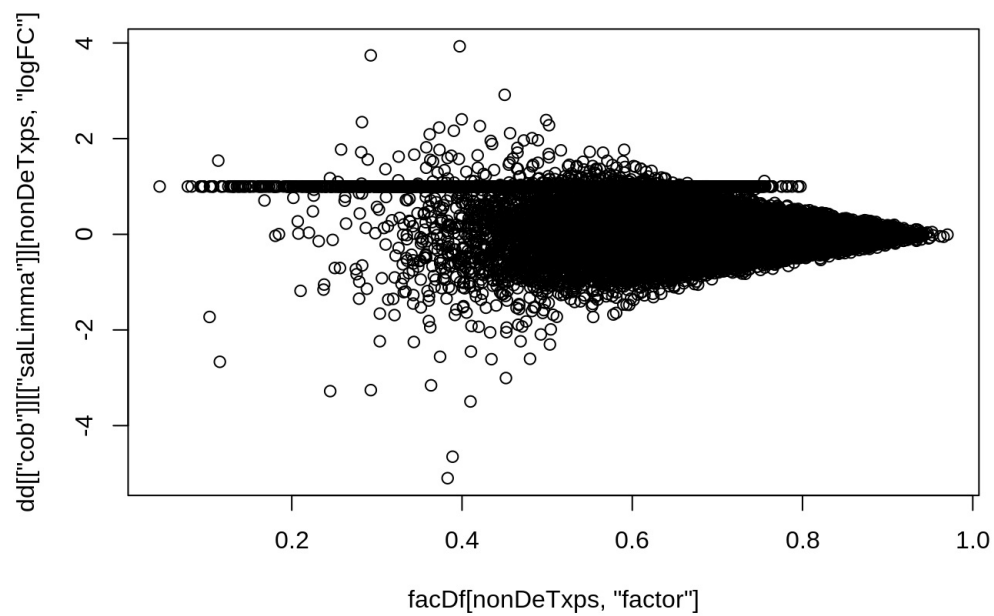


logFC changes with factors Shoal

```
plot(facDf[DeTxps, "factor"], dd[["cob"]][["salLimma"]][DeTxps, "logFC"])
```

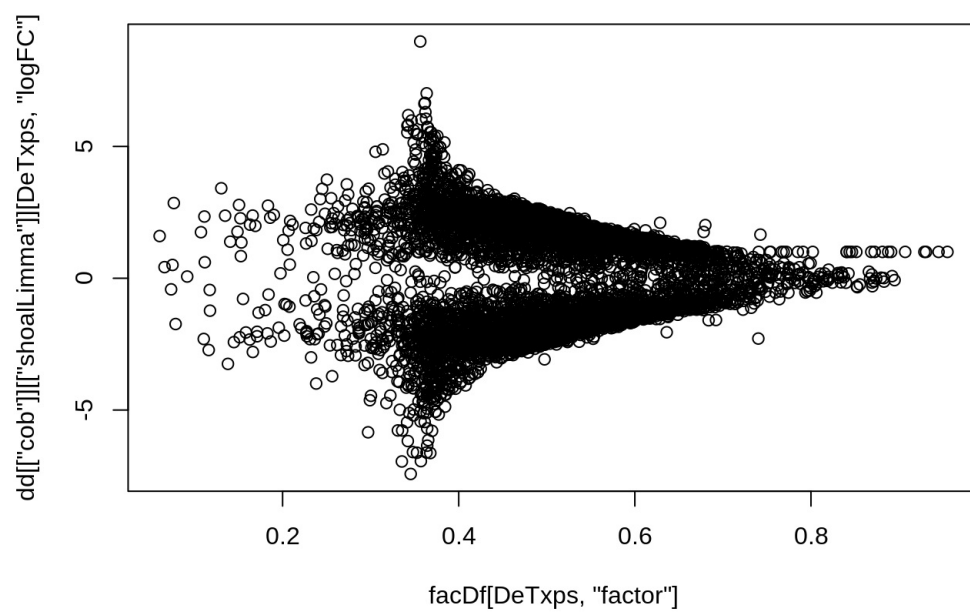


```
plot(facDf[nonDeTxps, "factor"], dd[["cob"]][["salLimma"]][nonDeTxps, "logFC"])
```

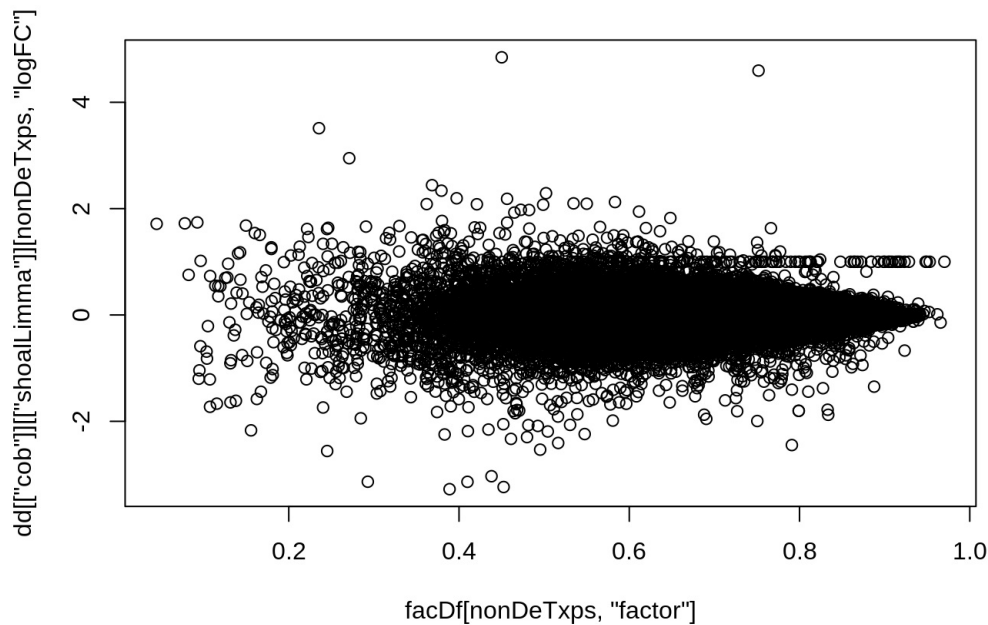


logFC changes with factors Shoal

```
plot(facDf[DeTxps, \"factor\"], dd[[\"cob\"]][[\"shoalLimma\"]][DeTxps, \"logFC\"])
```



```
plot(facDf[nonDeTxps, \"factor\"], dd[[\"cob\"]][[\"shoalLimma\"]][nonDeTxps, \"logFC\"])
```

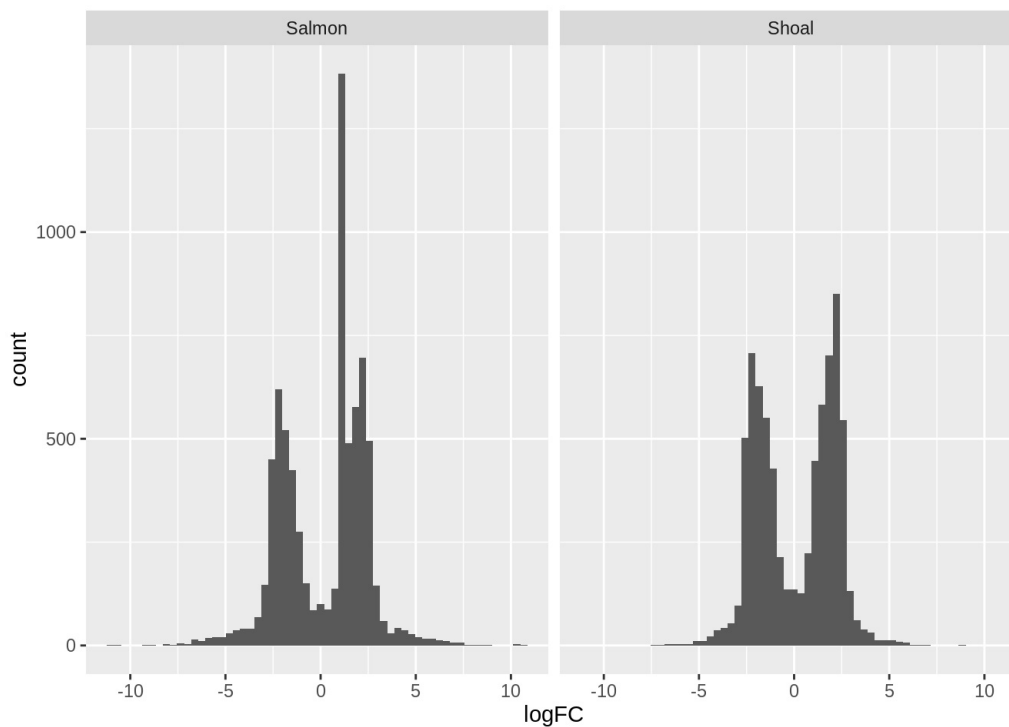


Creating data frame

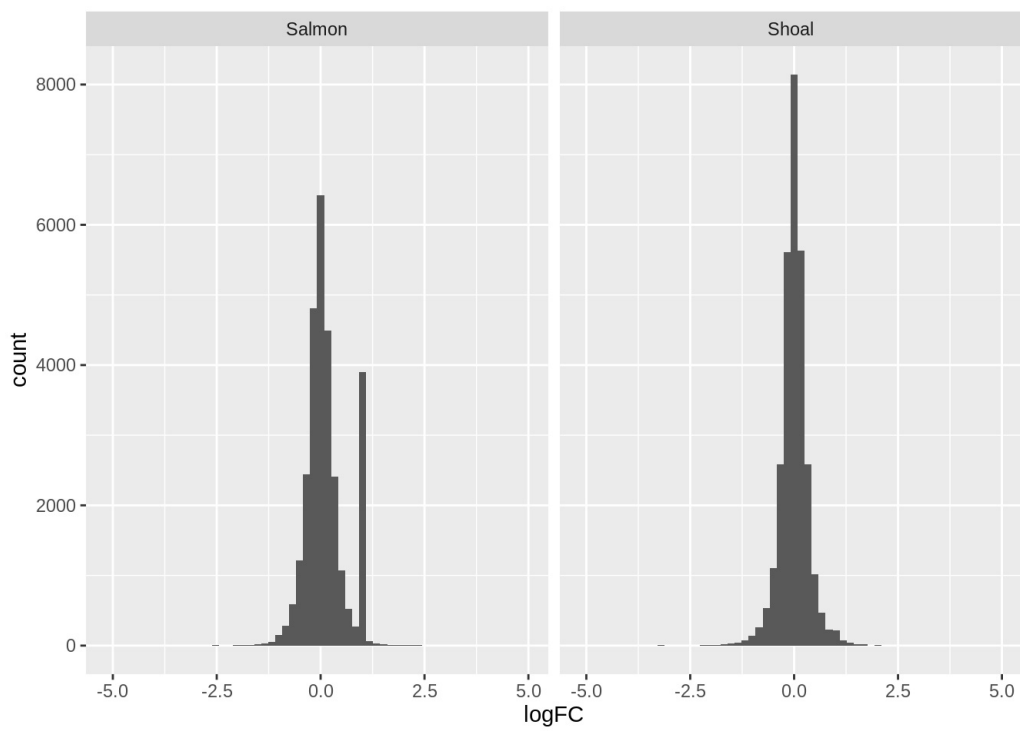
```
dfQ <- data.frame(logFC = c(dd[["cob"]][["salLimma"]][DeTxps, "logFC"],
                             dd[["cob"]][["salLimma"]][nonDeTxps, "logFC"],
                             dd[["cob"]][["shoalLimma"]][DeTxps, "logFC"],
                             dd[["cob"]][["shoalLimma"]][nonDeTxps, "logFC"]),
                  DEStatus = rep(rep(c("DE", "NonDE"), c(length(DeTxps), length(nonDeTxps))), 2),
                  quantMethod = rep(c("Salmon", "Shoal"), c(nrow(dd[["cob"]][["salLimma"]]), nrow(dd[["cob"]][["shoalLimma"]])),
                  txps = rep(c(DeTxps, nonDeTxps), 2))
```

Histogram of logFCs

```
library(ggplot2)
library(dplyr)
dfQ %>% filter(DEStatus=="DE") %>%
  ggplot(aes(x=logFC)) + geom_histogram(bins=60) +
  facet_grid(~quantMethod)
```

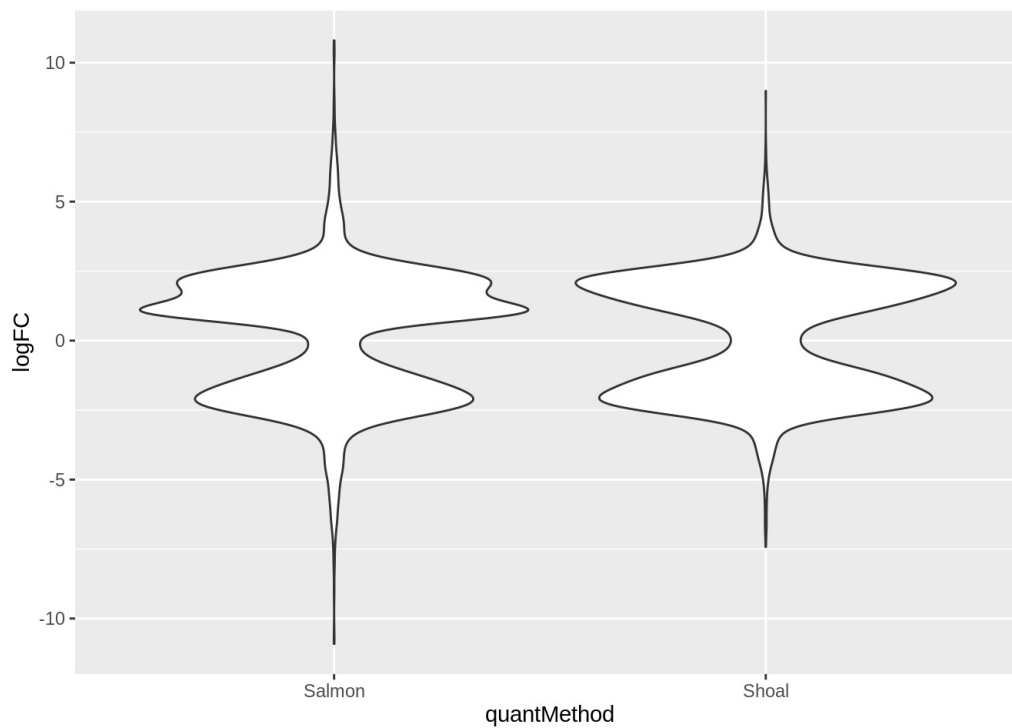


```
dfQ %>% filter(DEStatus=="NonDE") %>%
  ggplot(aes(x=logFC)) + geom_histogram(bins=60) +
  facet_grid(~quantMethod)
```

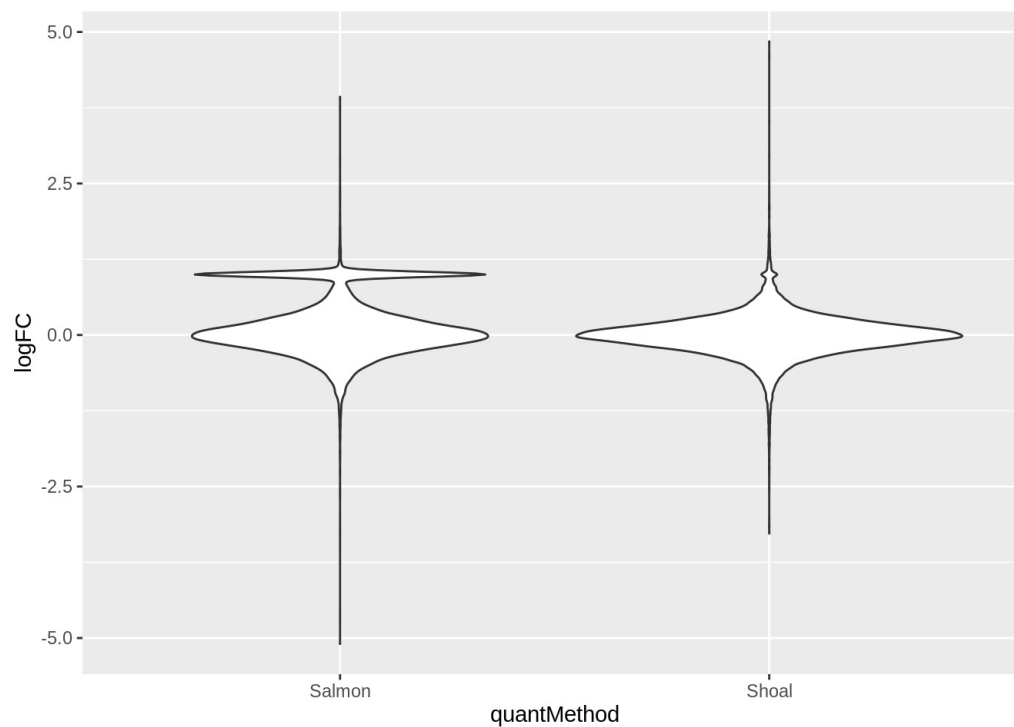


Violin plot of logFCs

```
dfQ %>% filter(DEstatus=="DE") %>%  
  ggplot(aes(y=logFC, x=quantMethod)) + geom_violin()
```



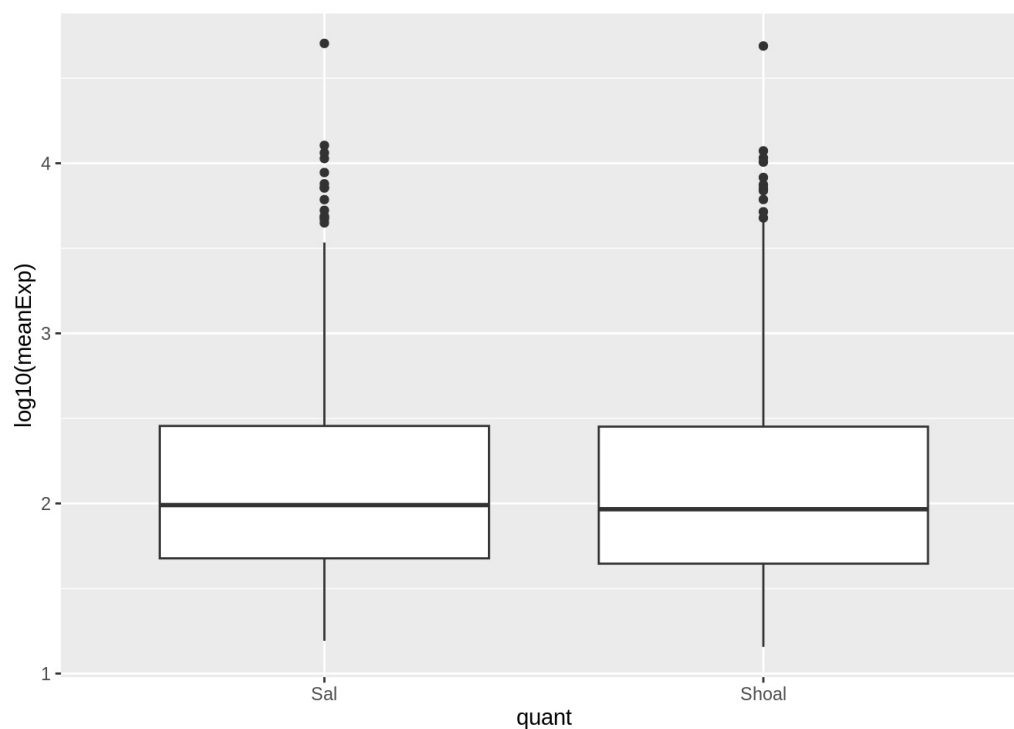
```
dfQ %>% filter(DEstatus=="NonDE") %>%  
  ggplot(aes(y=logFC, x=quantMethod)) + geom_violin()
```



Plotting boxplot shoal false positives

```
rMSal <- rowMeans(dd$counts$txiSal$counts[fpOnlyShoal,])
rMShoal <- rowMeans(dd$counts$txiShoal$counts[fpOnlyShoal,])

dfMean <- data.frame(meanExp=c(rMSal, rMShoal), quant = rep(c("Sal", "Shoal"), c(length(rMSal), length(rMSal))))
ggplot(dfMean, aes(x=quant, y=log10(meanExp))) + geom_boxplot()
```



Plotting boxplot per condition shoal false positives

```
rMSalC1 <- rowMeans(dd$counts$txiSal$counts[fpOnlyShoal,1:6])
rMSalC2 <- rowMeans(dd$counts$txiSal$counts[fpOnlyShoal,6+1:6])
rMShoalC1 <- rowMeans(dd$counts$txiShoal$counts[fpOnlyShoal,1:6])
rMShoalC2 <- rowMeans(dd$counts$txiShoal$counts[fpOnlyShoal,6+1:6])

dfMean <- data.frame(meanExp=c(rMSalC1, rMSalC2, rMShoalC1, rMShoalC2),
  quant = rep(c("Sal", "Shoal"), c(length(rMSal)*2, length(rMSal)*2)),
  cond = rep(rep(c("1", "2"), c(length(rMSal), length(rMSal))), times=2))
ggplot(dfMean, aes(x=cond, y=log10(meanExp))) + geom_boxplot() + facet_grid(~quant)
```

