R Notebook

True Counts

```
load("/fs/cbcb-lab/rob/students/noor/Uncertainity/mikelove-swimdown-216a1dd/simulate/data/simulate.rda")
source("run_limma.R")
trueDir <- "/fs/cbcb-lab/rob/students/noor/Uncertainity/swim_data"
dirSal <- "/fs/cbcb-lab/rob/students/noor/shoal_proj/swim/sal_best/vbprior=1e0"
samp <- paste(outer(paste("out", c(1:6), sep = "_"), paste(paste("out", c(1:2), sep = ""), "counts.tsv", sep = "_
"), function(x,y) file.path(x,y)))
files <- file.path(trueDir, samp)

trueDf <- lapply(files, function(f) read.delim(f, header = T))
trueDf <- Reduce(function(...) merge(..., all=T, by="Transcript"), trueDf)
colnames(trueDf) <- c("Transcript", outer(c(1:6), c(1,2), function(x,y) paste(x,y,sep="_")))
rownames(trueDf) <- trueDf[,1]
trueDf <- trueDf[,-1]
trueDf[is.na(trueDf)] <- 0
```

Shoal Requant Oracle

```
dirShoal <- "/fs/cbcb-lab/rob/students/noor/shoal_proj/swim/shoal_best_true_de_fac/vbprior=1e0/c=0.1"
shoalDf <- getQuantFiles(dirShoal, type='shoal')
cShoalRequant <- tximport(shoalDf[["files"]], type="salmon", txOut=T)[["counts"]]
df <- computeStat(trueDf, cShoalRequant, method = "Shoal Oracle")
```

```
suppressWarnings(dd <- runDE(dirSal, dirShoal))
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)]
dfFPs <- computeStat(trueDf, cShoalRequant, method = "Shoal Oracle", gName = fpOnlyShoal)
```

Salmon Counts

```
salDf <- getQuantFiles(dirSal, type='salmon')
cSal <- tximport(salDf[["files"]], type="salmon", txOut=T)[["counts"]]
df <- rbind(df, computeStat(trueDf, cSal, method = "Salmon"))
dfFPs <- rbind(dfFPs, computeStat(trueDf, cSal, method = "Salmon", gName = fpOnlyShoal))
```

Shoal Counts

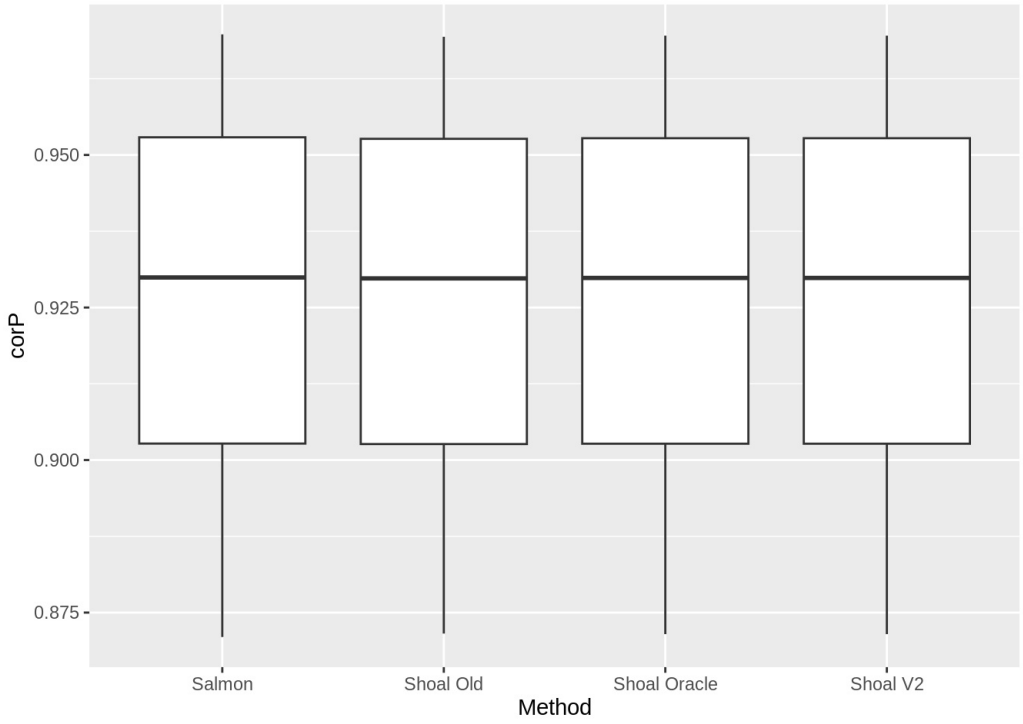
```
dirShoal <- "/fs/cbcb-lab/rob/students/noor/shoal_proj/swim/shoal_best/vbprior=1e0/c=100"
shoalDf <- getQuantFiles(dirShoal, type='shoal')
cShoal <- tximport(shoalDf[["files"]], type="salmon", txOut=T)[["counts"]]
df <- rbind(df, computeStat(trueDf, cShoal, method = "Shoal Old"))
dfFPs <- rbind(dfFPs, computeStat(trueDf, cShoal, method = "Shoal Old", gName = fpOnlyShoal))
```

Shoal Requant Counts

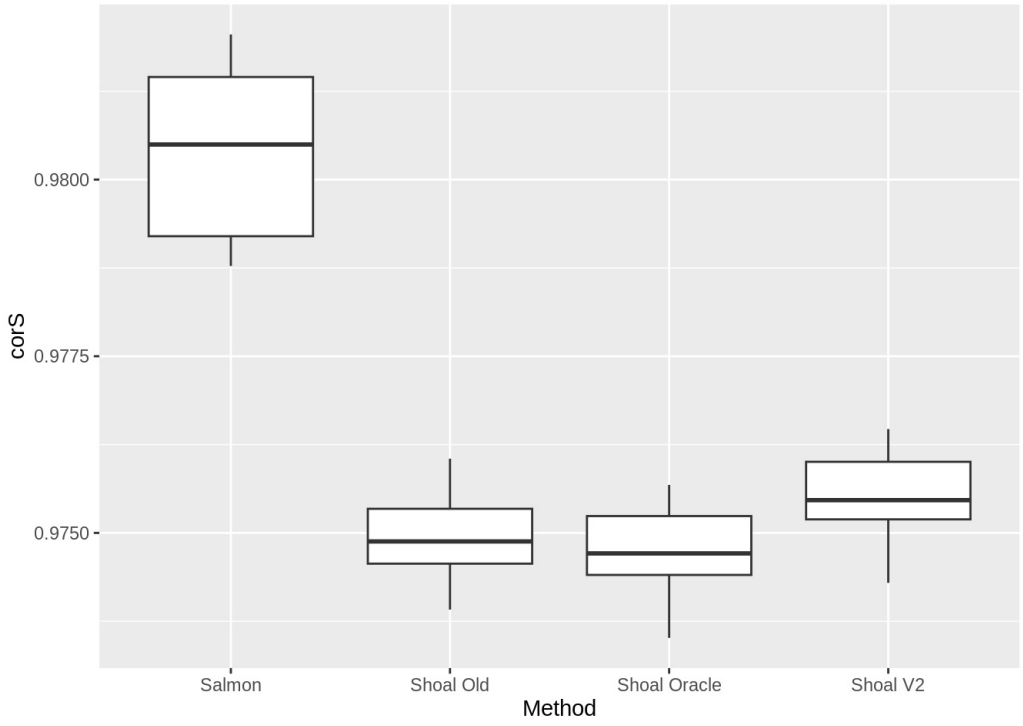
```
dirShoal <- "/fs/cbcb-lab/rob/students/noor/shoal_proj/swim/shoal_best_updated/vbprior=1e0/c=0.1"
shoalDf <- getQuantFiles(dirShoal, type='shoal')
cShoalRequant <- tximport(shoalDf[["files"]], type="salmon", txOut=T)[["counts"]]
df <- rbind(df, computeStat(trueDf, cShoalRequant, method = "Shoal V2"))
dfFPs <- rbind(dfFPs, computeStat(trueDf, cShoalRequant, method = "Shoal V2", gName = fpOnlyShoal))
```

Global Statistics

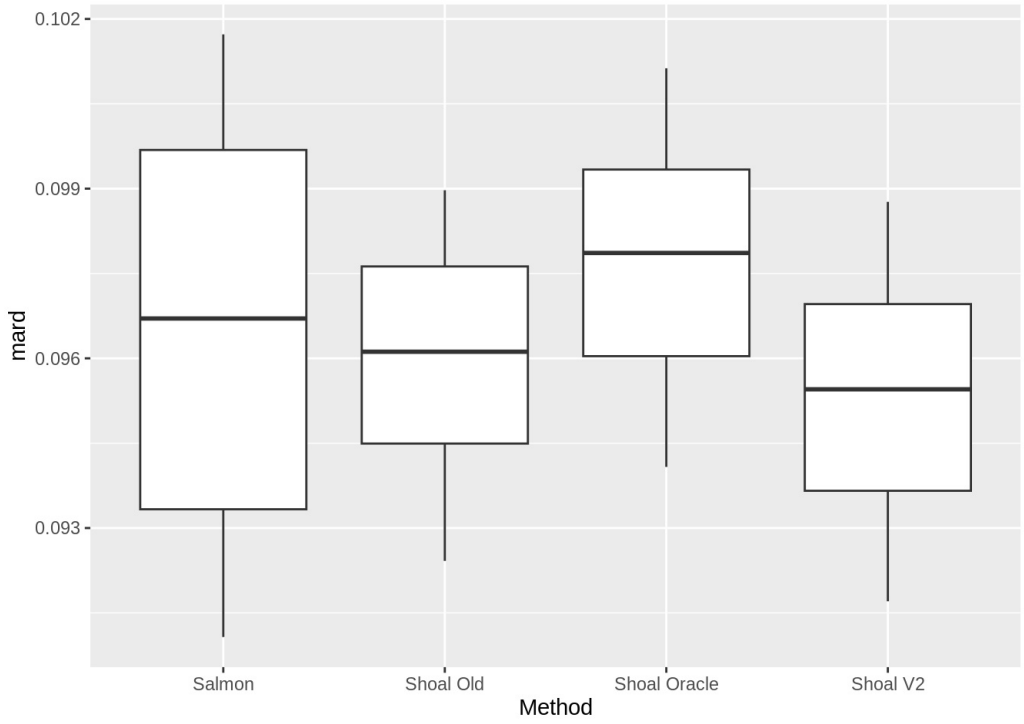
```
library(ggplot2)
#### Pearson Correlation
ggplot(df, aes(x=Method, y=corP)) + geom_boxplot()
```




```
#### Spearman Correlation
ggplot(df, aes(x=Method, y=corS)) + geom_boxplot()
```

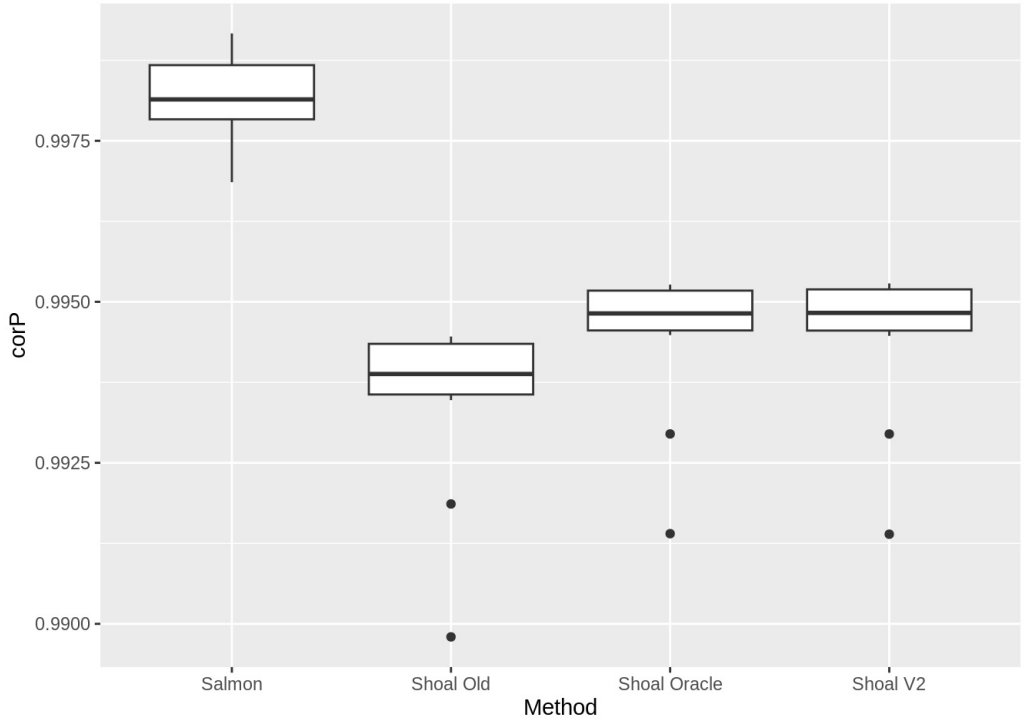



```
#### MARD
ggplot(df, aes(x=Method, y=mard)) + geom_boxplot()
```

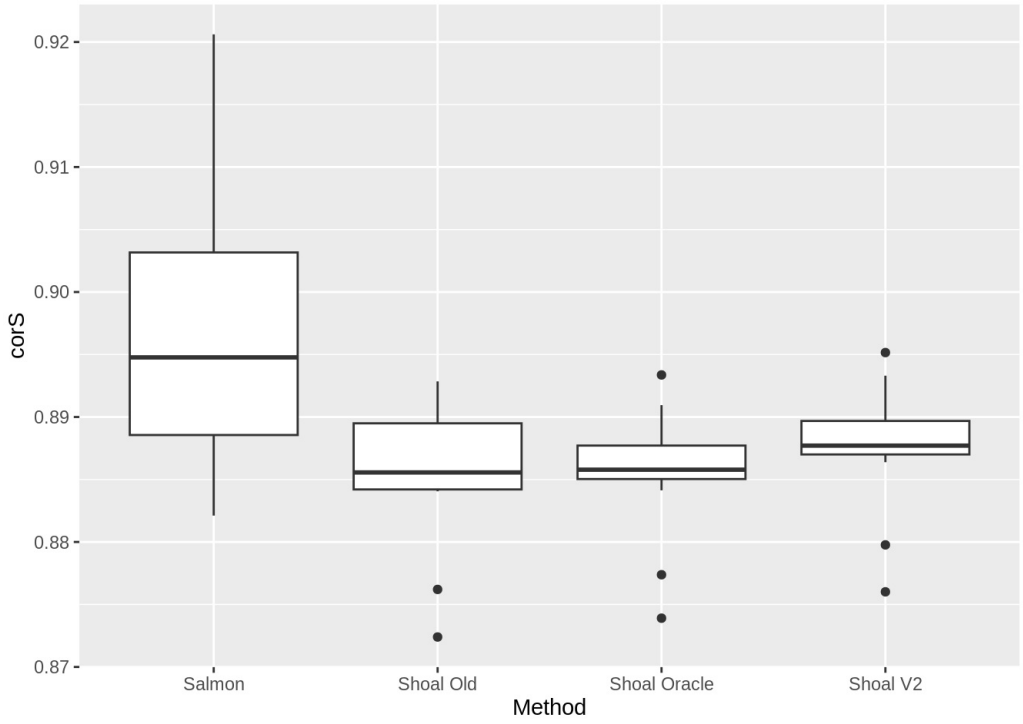


Only FP

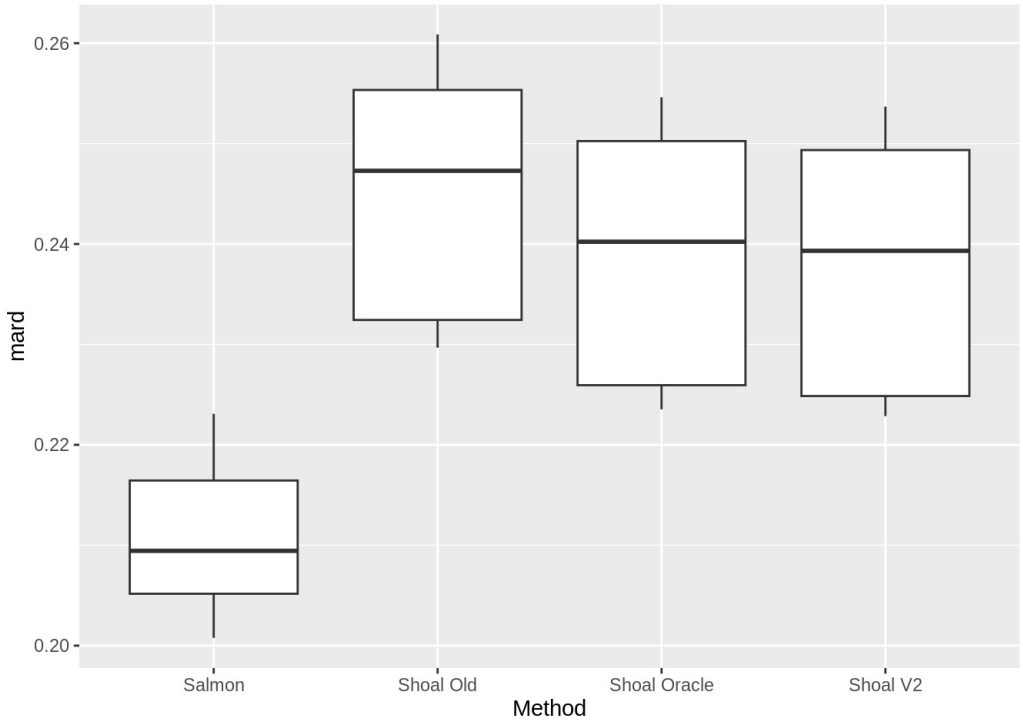
```
library(ggplot2)
#### Pearson Correlation
ggplot(dfFPs, aes(x=Method, y=corP)) + geom_boxplot()
```



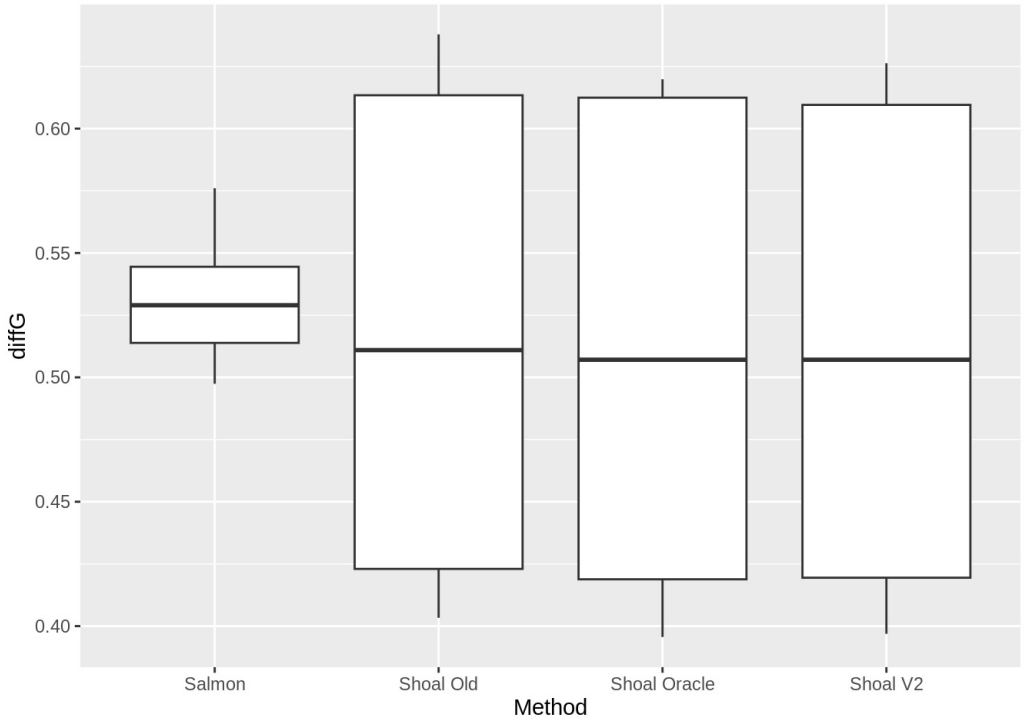

```
#### Spearman Correlation
ggplot(dfFPs, aes(x=Method, y=corS)) + geom_boxplot()
```



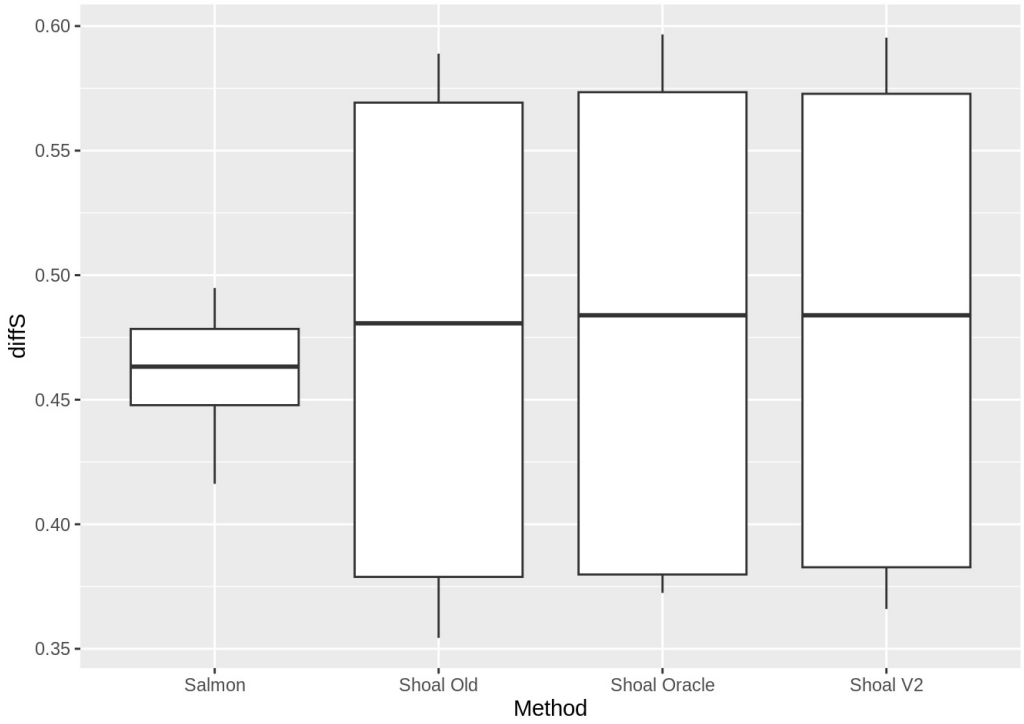

```
#### MARD
ggplot(dfFPs, aes(x=Method, y=mard)) + geom_boxplot()
```




```
#### Overestimated compared to true counts
ggplot(dfFPs, aes(x=Method, y=diffG)) + geom_boxplot()
```




```
#### Underestimated compared to true counts
ggplot(dfFPs, aes(x=Method, y=diffS)) + geom_boxplot()
```



FPs over and down regulated

```
print(sum(dd[["cob"]][["shoalLimma"]][fpOnlyShoal, "logFC"] > 0))
```

```
## [1] 281
```

```
print(sum(dd[["cob"]][["shoalLimma"]][fpOnlyShoal, "logFC"] < 0))
```

```
## [1] 495
```