R Notebook

Loading Data

```
library(tximport)
library(DESeq2)
source("helper_func.R")
load("../../real_data/out_1/sim_counts_matrix.rda")
dir <- "../../real_data/out"

gsFiles <- file.path(dir, c("ERR188297_GS", "sample_01_GS", "sample_01_GS_150", "sample_01_GS_200"), "quant.
sf")
txiInfRepGS <- tximport(gsFiles, type = "salmon", txOut = TRUE)

bootFiles <- file.path(dir, c("ERR188297_B", "sample_01_B"), "quant.sf")
txiInfRepBoot <- tximport(bootFiles, type = "salmon", txOut = TRUE)

txiInfRepGS <- computeConfInt(txiInfRepGS)
txiInfRepBoot <- computeConfInt(txiInfRepBoot)

sum(rownames(txiInfRepBoot$conf[[2]]) != rownames(txiInfRepGS$conf[[2]])) == 0 ## Checking transcripts names
match across BS and GS
```

```
## [1] TRUE
```

```
sum(rownames(txiInfRepGS$conf[[2]]) != rownames(txiInfRepGS$counts[,2])) == 0 ## Checking transcripts names
match across the counts
```

```
## [1] TRUE
```

```
mInds <- match(rownames(counts_matrix), rownames(txiInfRepBoot$conf[[2]])) ##Indexes of transcripts of simul
ated data within txi

reqMat <- cbind(txiInfRepGS$counts[mInds,2], counts_matrix[,1])
colnames(reqMat) <- c("Estimated", "True")
print(head(reqMat))
```

```
##                    Estimated True
## ENST00000371588.9    930.430 1288
## ENST00000466152.5    108.936  130
## ENST00000371582.8    131.695  179
## ENST00000371584.8     29.938   56
## ENST00000367771.10    93.047  112
## ENST00000367770.5    181.972  171
```

```
sizeFac <- estimateSizeFactorsForMatrix(reqMat)
print(sizeFac)
```

```
## Estimated      True
## 0.8902486 1.1232818
```

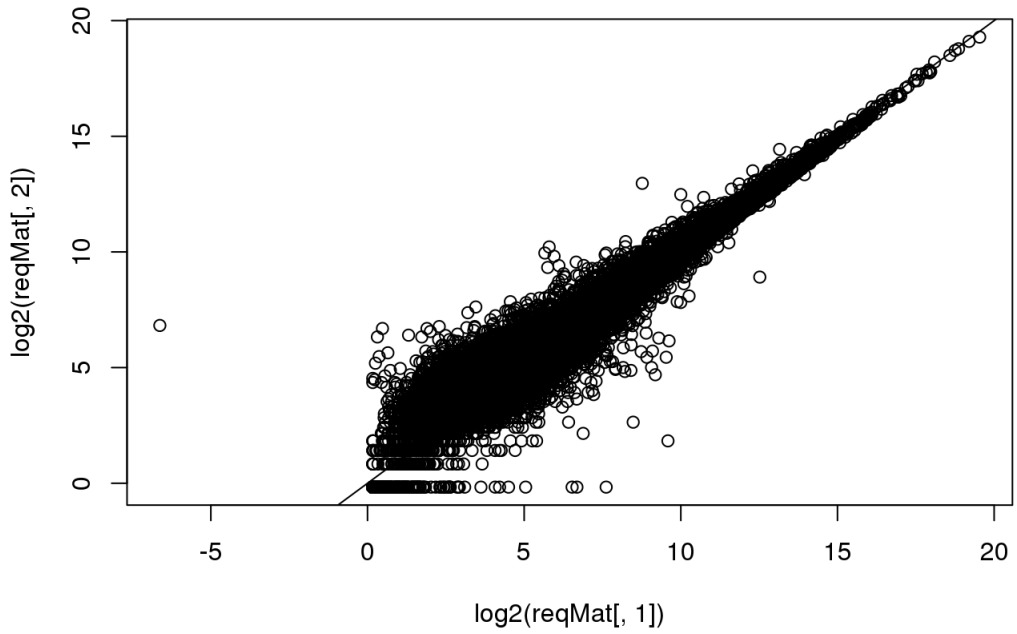
```
reqMat[,1] <- reqMat[,1]/sizeFac[1]
reqMat[,2] <- reqMat[,2]/sizeFac[2]
print(head(reqMat))
```

```
##                     Estimated       True
## ENST00000371588.9  1045.13504 1146.64018
## ENST00000466152.5   122.36582  115.73232
## ENST00000371582.8   147.93059  159.35450
## ENST00000371584.8    33.62881   49.85392
## ENST00000367771.10  104.51800   99.70784
## ENST00000367770.5   204.40583  152.23251
```

Plotting Estimated vs True

Plotting Estimated vs True

```
plot(log2(reqMat[,1]), log2(reqMat[, 2])) + abline(coef = c(0,1))
```




```
## integer(0)
```

Computing the Coverage for the 1st sample of simulated data

```
confMat <- list("Boot100" = txiInfRepBoot$conf[[2]][mInds,]/sizeFac[1], "GS100" = txiInfRepGS$conf[[2]][mInds,]/sizeFac[1])

covOverall <- sapply(confMat, function(mat) computeCoverage(reqMat[,"True"], mat, list(seq(nrow(counts_matrix)))))

cInds100 <- extractBinInds(reqMat[,"True"], breaks = 100)
cov100 <- createCovDf(confList = confMat, counts = reqMat[,"True"], cInds100)

cIndsAll <- extractBinInds(reqMat[,"True"], breaks = NULL)
covAll <- createCovDf(confList = confMat, counts = reqMat[,"True"], cIndsAll)
```

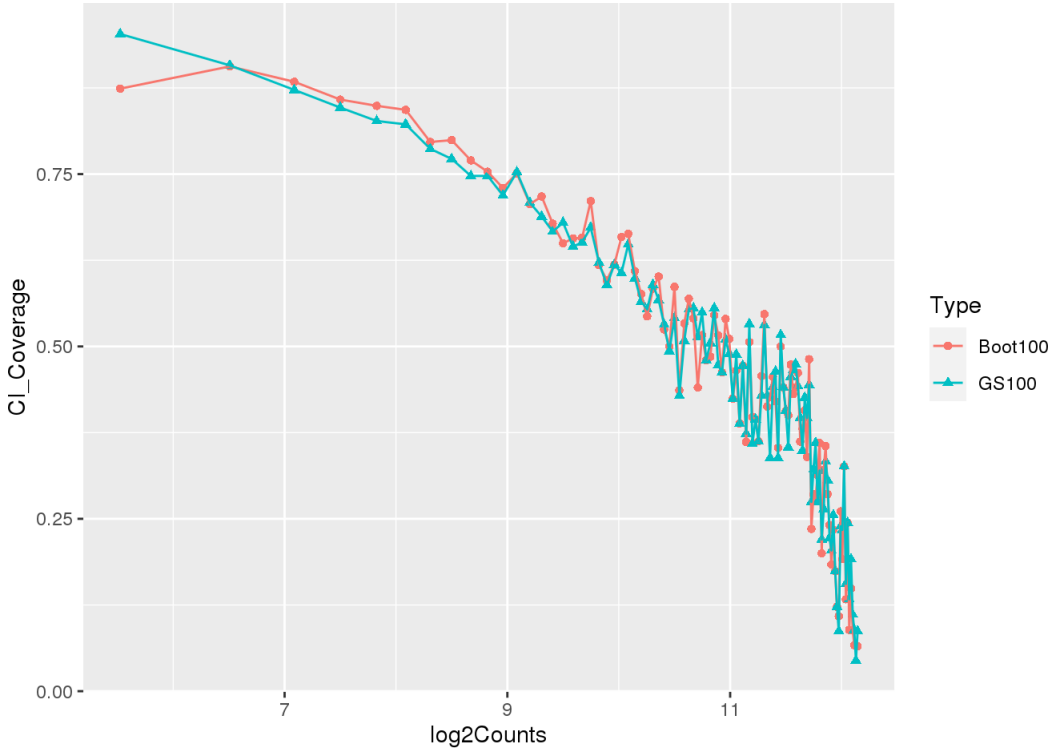
Overall Coverage

```
print(covOverall)
```

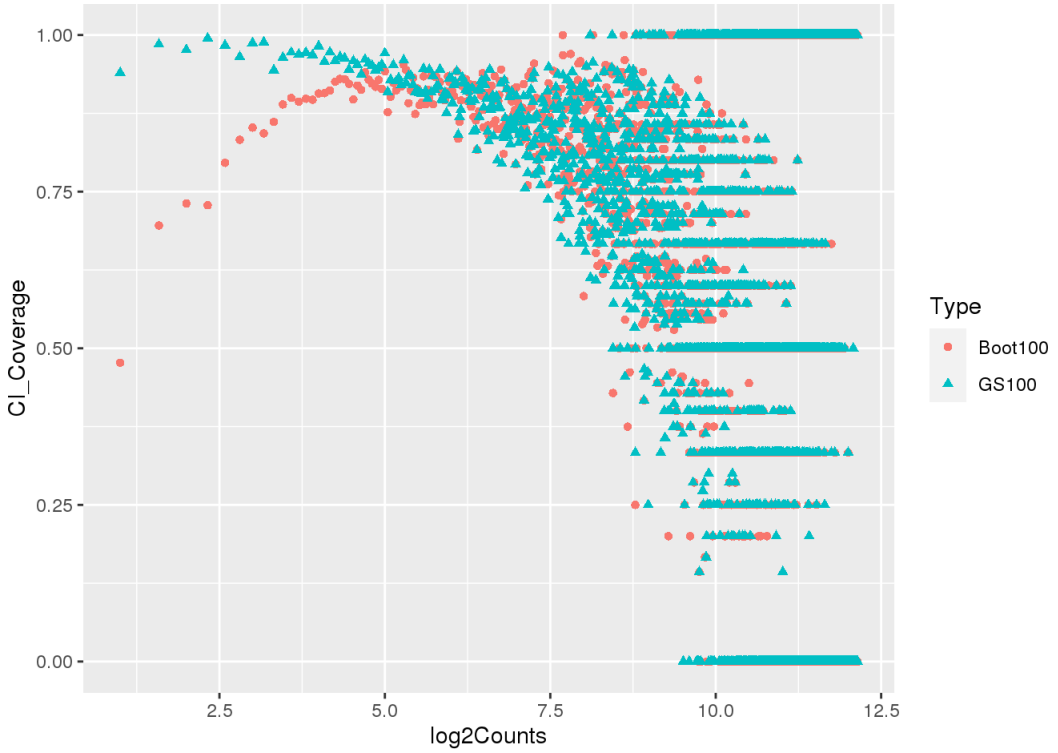
```
##   Boot100     GS100
## 0.7953375 0.8195825
```

Plotting Coverage

```
plotCovDf(cov100, line = T)
```




```
plotCovDf(covAll)
```




```
inds <- cInds100[[100]]

df <- cbind(data.frame(estCounts = reqMat[inds,"Estimated"], trueCount = reqMat[inds, "True"],
                       Cov_GS100 = computeCoverage(reqMat[, "True"], confMat[["GS100"]], list(inds), F)),co
nfMat[["GS100"]][inds,1:2], cov_Boot100 = computeCoverage(reqMat[, "True"], confMat[["Boot100"]], list(inds)
, F),
            confMat[["Boot100"]][inds,1:2])
colnames(df)[4:5] <-  paste("GS100", colnames(df)[4:5], sep = "_")
colnames(df)[7:8] <-  paste("BS100", colnames(df)[7:8], sep = "_")
df <- df[order(df$estCounts, decreasing = T),]
write.table(df, file = "High_exp_int_deseq.csv", sep = " ")

sum(df[,"estCounts"] > df[,"trueCount"])
```

```
## [1] 39
```

```
print(nrow(df))
```

```
## [1] 46
```