

# R Notebook

## Loading Data

```
library(tximport)
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
```

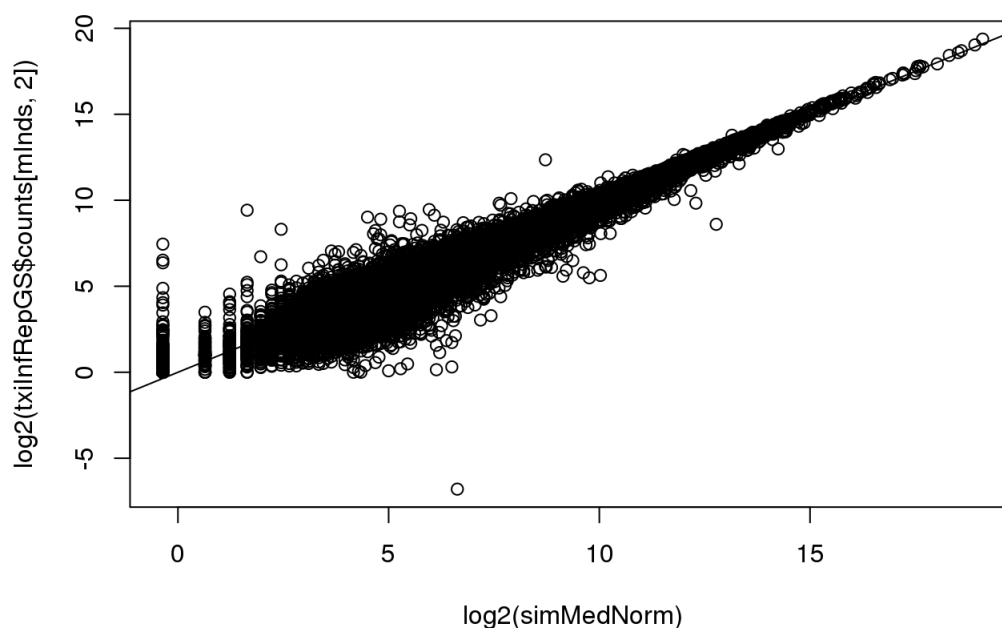
```
## [1] TRUE
```

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

sizeFac <- exp(median(log(counts_matrix[,1]) - log(txiInfRepGS$counts[mInds,2])))
simMedNorm <- counts_matrix[,1]/sizeFac
```

## Plotting Simulated (True) vs Generated

```
plot(log2(simMedNorm), log2(txiInfRepGS$counts[mInds, 2])) + abline(coef = c(0,1))
```



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

## Computing the Coverage for the 1st sample of simulated data

```
confBootReq <- txiInfRepBoot$conf[[2]][mInds,]
confGSReq <- vector(mode = "list", 3)
names(confGSReq) <- c("GS100", "GS150", "GS200")
for(i in c(2:4))
  confGSReq[[i-1]] <- txiInfRepGS$conf[[i]][mInds,]

confMat <- list(confGSReq[[1]], confGSReq[[2]], confGSReq[[3]], confBootReq)
names(confMat) <- c("GS100", "GS150", "GS200", "BS100")

covOverall <- sapply(confMat, function(mat) computeCoverage(simMedNorm, mat, list(seq(nrow(counts_matrix))))
)

cInds100 <- extractBinInds(simMedNorm, breaks = 100)
cov100 <- createCovDf(confList = confMat, counts = simMedNorm, cInds100)

cIndsAll <- extractBinInds(simMedNorm, breaks = NULL)
covAll <- createCovDf(confList = confMat, counts = simMedNorm, cIndsAll)
```

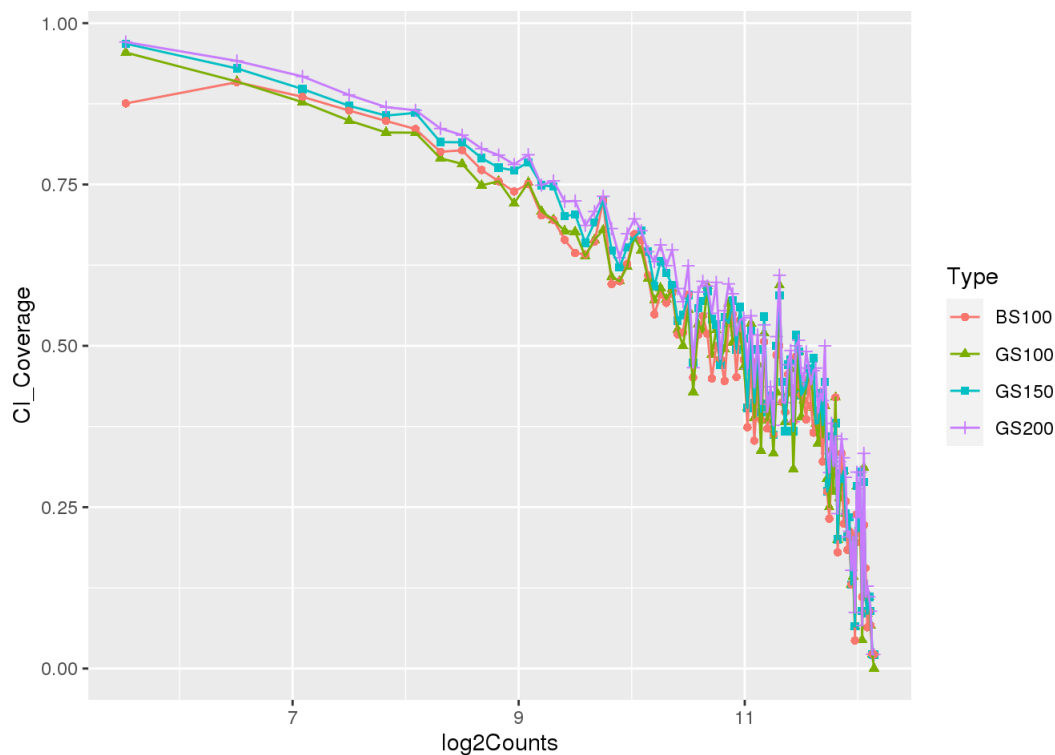
## Overall Coverage

```
print(covOverall)
```

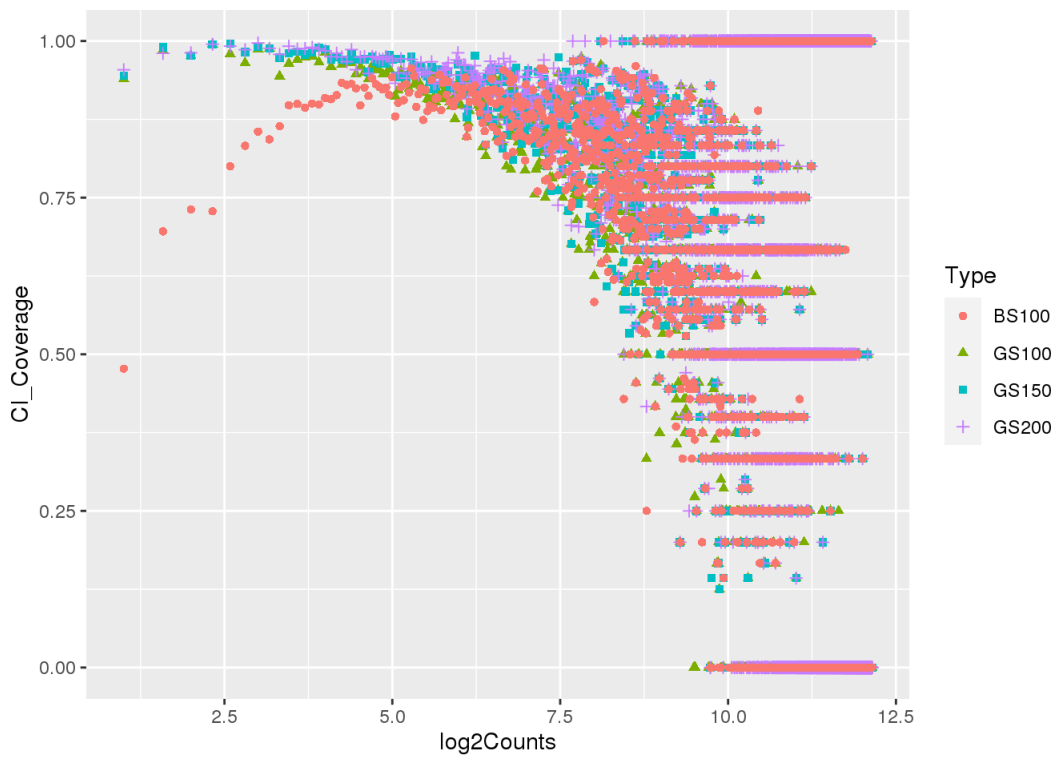
```
##      GS100      GS150      GS200      BS100
## 0.8218290 0.8426831 0.8534704 0.7952103
```

## Plotting Coverage

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



```
plotCovDf(covAll)
```



```
ind <- 100
df <- cbind(data.frame(estCounts = txiInfRepGS$counts[mInds,2][cInds100[[ind]]], trueCount = simMedNorm[cInds
100[[ind]]],
             Cov_GS100 = computeCoverage(simMedNorm, confMat[[1]], list(cInds100[[ind]]), F), con
fMat[[1]][cInds100[[100]],1:2], cov_Boot100 = computeCoverage(simMedNorm, confMat[[4]], list(cInds100[[ind]]
), F),
             confMat[[4]][cInds100[[100]],1:2])
colnames(df)[4:5] <- paste("GS100", colnames(df)[4:5], sep = "_")
colnames(df)[7:8] <- paste("BS100", colnames(df)[7:8], sep = "_")
write.table(df, file = "High_exp_int_med.csv", sep = " ")
```