

coverage.Rmd

Loading the data

```
suppressPackageStartupMessages(library(tximeta))
suppressPackageStartupMessages(library(SummarizedExperiment))
dir <- "../real_data/out"
poleeDir <- "../real_data/polee"
load("../real_data/out_1/sim_counts_matrix.rda")
source("helper_func.R")

samp <- "sample_01"
#nInf <- c(10, 100, 1000, 2000)
nInf <- 100
infTypes <- c("B", "GS", "polee")
seListOld <- vector(mode = "list", length(nInf))
names(seListOld) <- as.character(nInf)

for(n in nInf)
{
  files <- file.path(dir, paste(samp, infTypes[1:2], n, sep = "_"), "quant.sf")
  coldata <- data.frame(files = files, names = paste(samp, infTypes[1:2], n, sep = "_"), infType = infTypes[1:2])
  se <- tximeta(coldata)
  mInds <- match(rownames(counts_matrix), rownames(se))
  seListOld[[as.character(n)]] <- se[mInds,]

  ### Integrating polee
  poleeFiles <- file.path(poleeDir, samp, paste("boot", n, sep = "_"), "abundance.h5")
  coldata <- data.frame(files = poleeFiles, names = paste(samp, "polee", n, sep = "_"), infType = "polee")
  se <- tximeta(coldata, type = "kallisto", txOut = T)
  mInds <- match(rownames(counts_matrix), rownames(se))
  se <- se[mInds,]
  mInds <- match(rownames(se), rownames(seListOld[[as.character(n)]]))
  se <- se[mInds,]
  rowData(se) <- rowData(seListOld[[as.character(n)]]))
  rowRanges(se) <- rowRanges(seListOld[[as.character(n)]]))
  seListOld[[as.character(n)]] <- cbind(seListOld[[as.character(n)]], se)
}

r <- length(nInf)*length(infTypes)
scaleTypes <- c("DESeq2", "medScale", "depthScale")

dfCoverageOld <- data.frame(NoScale = rep(0, r), DeScale = rep(0, r), MedScale = rep(0, r), depthScale = rep(0, r))
rownames(dfCoverageOld) <- as.vector(t(outer(nInf, infTypes, paste, sep = "_")))

dfLogCoverageOld <- data.frame(NoScale = rep(0, r), DeScale = rep(0, r), MedScale = rep(0, r), depthScale = rep(0, r))
rownames(dfLogCoverageOld) <- as.vector(t(outer(nInf, infTypes, paste, sep = "_")))
```

Below we compute coverage using different conditions ### Without scaling factors

```
start = 1
for(i in seq_along(seListOld))
{
  end = (start + length(infTypes)-1)
  seListOld[[i]] <- computeSizeFactors(seListOld[[i]], counts_matrix[,1])
  seListOld[[i]] <- computeConfInt(seListOld[[i]], sf = F)
  dfCoverageOld[start:end,1] <- t(computeCoverage(counts_matrix[,1], seListOld[[i]], list(seq(nrow(counts_matrix), 1))))
  for(j in seq_along(scaleTypes))
  {
    seListOld[[i]] <- computeConfInt(seListOld[[i]], sf = T, type = scaleTypes[j])
    dfCoverageOld[start:end,j+1] <- t(computeCoverage(counts_matrix[,1], seListOld[[i]], list(seq(nrow(counts_matrix), 1))))
  }
  print(c(start, end))
  start = end + 1
}
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 1 3
```

```
save(seListOld, file = "environment/seListOld.RData")
save(dfCoverageOld, file = "environment/dfCoverageOld.RData")
load("environment/seListOld.RData")
load("environment/dfCoverageOld.RData")
print(dfCoverageOld)
```

```
##           NoScale  DeScale  MedScale  depthScale
## 100_B      0.5367596 0.7598813 0.7557698 0.7601356
## 100_GS     0.5605807 0.7830667 0.7829607 0.7821342
## 100_polee  0.5235350 0.7189573 0.7186818 0.7187454
```

Using logs

```
start = 1
for(i in seq_along(seListOld))
{
  end = (start + length(infTypes)-1)
  seListOld[[i]] <- computeConfInt(seListOld[[i]], sf = F, log = T)
  dfLogCoverageOld[start:end,1] <- t(computeCoverage(counts_matrix[,1], seListOld[[i]], list(seq(nrow(counts_matrix), 1))))
  for(j in seq_along(scaleTypes))
  {
    seListOld[[i]] <- computeConfInt(seListOld[[i]], sf = T, log = T, type = scaleTypes[j])
    dfLogCoverageOld[start:end,j+1] <- t(computeCoverage(counts_matrix[,1], seListOld[[i]], list(seq(nrow(counts_matrix), 1))))
  }
  print(c(start, end))
  start = end + 1
}
```

```
## [1] 1 3
```

```
save(dfLogCoverageOld, file = "environment/dfLogCoverageOld.RData")
load("environment/dfLogCoverageOld.RData")
print(dfLogCoverageOld)
```

```
##           NoScale   DeScale   MedScale   depthScale
## 100_B      0.5362297 0.7593727 0.7553248  0.7595634
## 100_GS     0.5604959 0.7831726 0.7830455  0.7822189
## 100_polee 0.5233867 0.7188937 0.7187030  0.7188089

# seListOld[["100"]] <- computeConfInt(seListOld[["100"]], sf = T)
# cInds100 <- extractBinInds(counts_matrix[,1], breaks = 100)
# covDf <- createCovDf(seListOld[["100"]], counts_matrix[,1], cInds100)
# p <- plotCovDf(covDf, line=T)
# print(p)
# ggsave(filename = "../presentations/Oct7/coverage.png", p)
```