P3_SingleCellCourse2023

2023-08-11

Prostate cancer single-cell data analysis: Reading the data, QC and Normalization

```
#Libraries
library(scRNAseq)
library(DropletUtils)
library(Matrix)
library(AnnotationHub)
library(scater)
library(BiocFileCache)
library(EnsDb.Hsapiens.v86)
library(dplyr)
library(scran)
```

#Setting the working directory

```
setwd("/Users/jrenewong/Desktop/P3_SC2023/")
getwd() #Confirmation
```

```
## [1] "/Users/jrenewong/Desktop/P3_SC2023"
```

#Reading input file and change it to a Single-Cell object

```
mat <- read.delim(
  "data/GSE157703_RAW/GSM4773521_PCa1_gene_counts_matrix.txt", sep = ' ')
mat <- as.matrix(mat)
sce <- SingleCellExperiment::SingleCellExperiment(
  assays = list(counts = mat))
sce
```

```
## class: SingleCellExperiment
## dim: 26069 2896
## metadata(0):
## assays(1): counts
## rownames(26069): RP11-34P13.7 FO538757.2 ... AC136352.4 AC007325.1
## rowData names(0):
## colnames(2896): AAACCTGAGCGTTCCG_1 AAACCTGCACACTGCG_1 ...
##   TTTGTCACAGCATACT_1 TTTGTCAGTTCGTTGA_1
## colData names(0):
## reducedDimNames(0):
## mainExpName: NULL
## altExpNames(0):
```

#Quality Control from mitochondrial genes

```
GeneNames <- rownames(sce@assays@data@listData$counts)
MitGenes <- GeneNames[which(grepl(pattern = "^MT-", x = GeneNames, perl = F)==TRUE)]
print(MitGenes)
```

```
##  [1] "MT-ND1"  "MT-ND2"  "MT-CO1"  "MT-CO2"  "MT-ATP8" "MT-ATP6" "MT-CO3"
##  [8] "MT-ND3"  "MT-ND4L" "MT-ND4"  "MT-ND5"  "MT-ND6"  "MT-CYB"
```

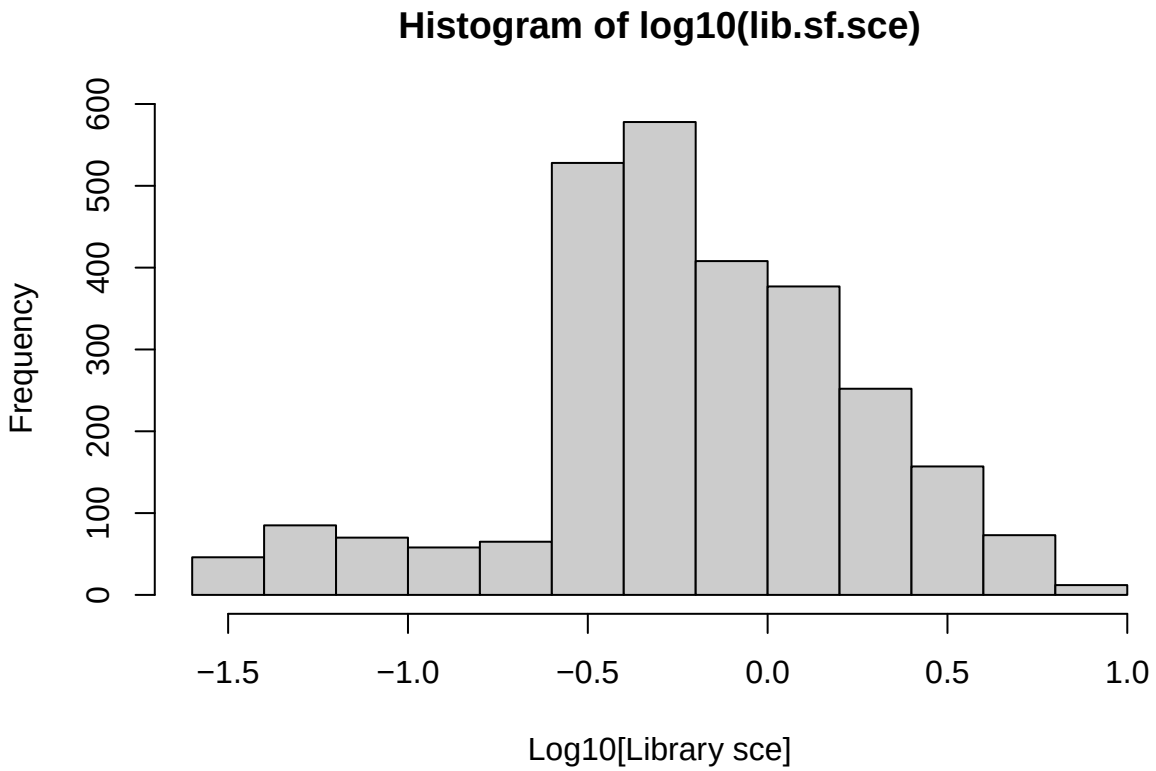
```
stats <- perCellQCMetrics(sce,
                          subsets = list(Mito = rownames(sce) %in% MitGenes)
)
high.mito <- isOutlier(stats$subsets_Mito_percent,
                       type = "higher"
)
sce <- sce[, !high.mito]
```

#Deconvolution normaliztion

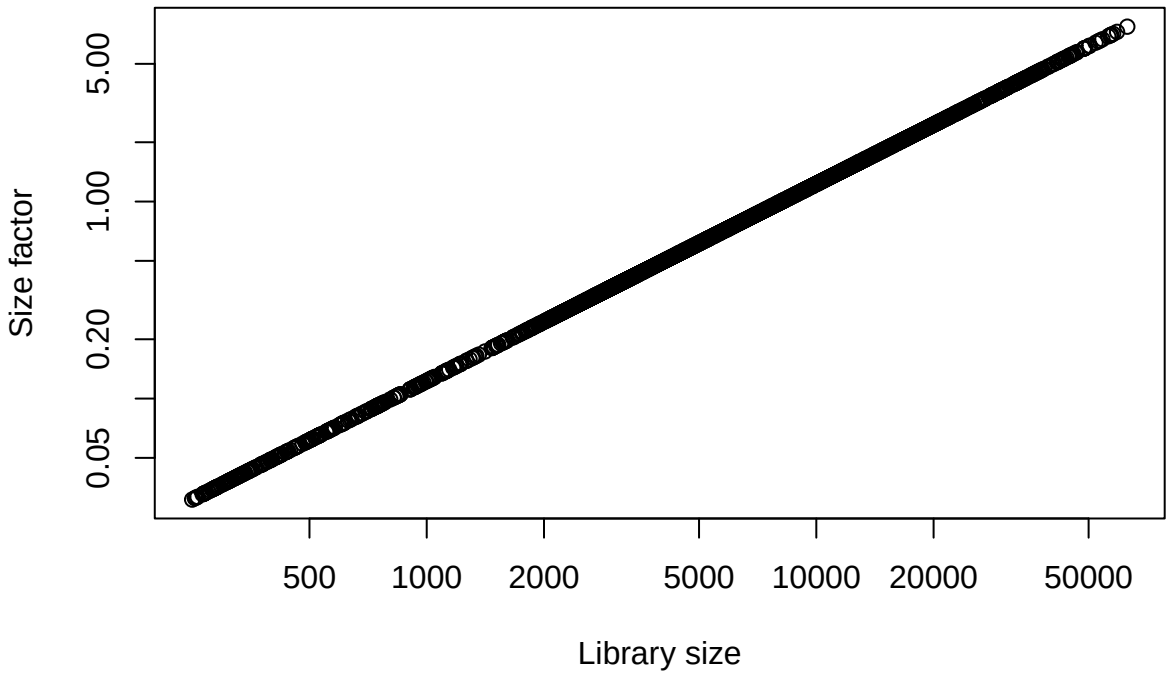
```
# Estimation of normalization factors
lib.sf.sce <- librarySizeFactors(sce)
# Examination of library sizes that we estimated
summary(lib.sf.sce)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.03071 0.36411 0.57700 1.00000 1.24318 7.71926
```

```
hist(log10(lib.sf.sce), xlab = "Log10[Library sce]", col = "grey80")
```




```
## Calculando el tamaño de las librerias
ls.sce <- colSums(counts(sce))
plot(ls.sce, lib.sf.sce, log="xy",
     xlab="Library size", ylab="Size factor")
```



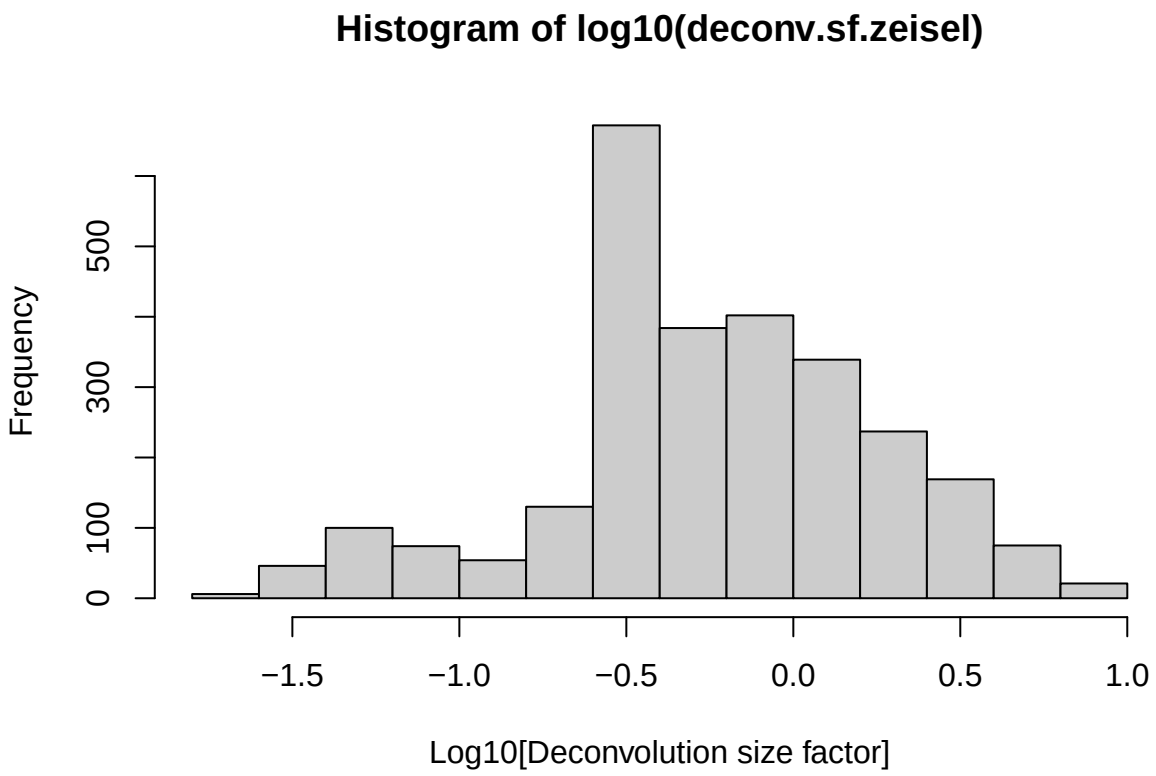

```
# Deconvolution
clust.sce <- quickCluster(sce)

# Calculation of size factors for deconvolution
deconv.sf.zeisel <-
  calculateSumFactors(sce, clusters = clust.sce, min.mean = 0.1)

# Distribution
summary(deconv.sf.zeisel)
```

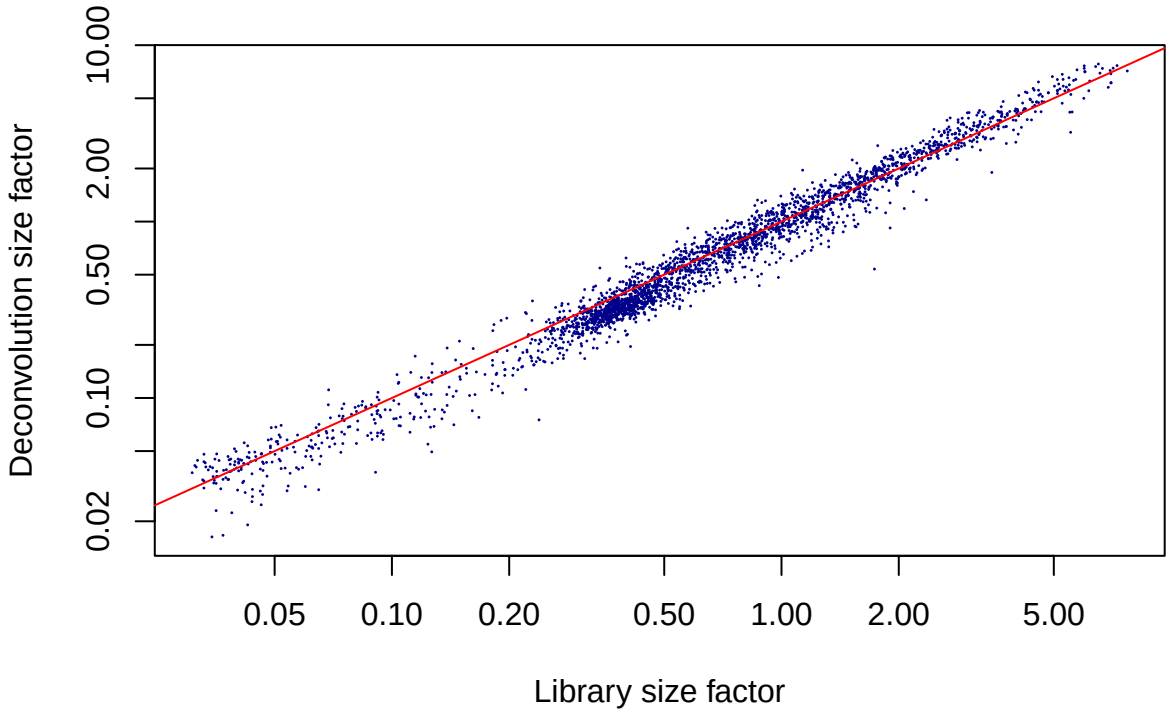
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01631 0.31095 0.54546 1.00000 1.21051 7.82132
```

```
hist(log10(deconv.sf.zeisel),
     xlab = "Log10[Deconvolution size factor]",
     col = "grey80"
)
```




```
plot(lib.sf.sce,
     deconv.sf.zeisel,
     xlab = "Library size factor",
     ylab = "Deconvolution size factor",
     log = "xy",
     pch = 16,
     cex = 0.2,
     col="darkblue")

abline(a = 0, b = 1, col = "red")
```


Deconvolution size factor
0.02
0.10
0.50
2.00
10.00
0.05
0.10
0.20
0.50
1.00
2.00
5.00
Library size factor