

Zadanie 9 - Analiza Differential Expression (DE)

Ksenia Kvitko

9 05 2020

1. Dane z etapu końca zadania 8

Importowanie danych wygenerowanych przez program featureCounts na serwerze oraz normalizacja danych

```
#BiocManager::install(c("DESeq2"))
library(DESeq2)

data22 <- read.csv('../source_files/counts_22.txt', sep = '\t', skip = 1)
dataERCC92 <- read.csv('../source_files/counts_ERCC92.txt', sep = '\t', skip = 1)

normalizeDDS <- function(data){
  countData <- data[,7:12]
  rownames(countData) = data$Geneid
  samples <- names(countData)
  cond_1 <- rep("cond1", 3)
  cond_2 <- rep("cond2", 3)
  condition <- factor(c(cond_1, cond_2))
  colData <- data.frame(samples = samples, condition = condition)
  dds <- DESeqDataSetFromMatrix(countData = countData, colData = colData, design = ~condition)

  return(dds)
}

dds_22 <- normalizeDDS(data22)
dds_ERCC92 <- normalizeDDS(dataERCC92)
```

2. Rozwiązanie zadania

Przygotowanie funkcji generującej analizę i jej filtracja wg podanych kryteriów

```
summarizeDDS <- function(ddsData){
  dds <- ddsData

  dds <- DESeq(dds)
  res <- results(dds)

  r <- res[res$baseMean != 0,] # pominięcie genów bez ekspresji (zerowej)
```

```

r <- r[r$log2FoldChange > 1 | r$log2FoldChange < -1,] # założenie, że istotne są geny z co najmniej d
x <- !is.na(r$padj) # usunięcie danych brakujących (NA)
r <- r[x,]
r <- r[r$padj < 0.05,] # wybór genów o istotności statystycznej < 5%
print(paste("Geny o istotnie statystycznie zmienionym poziomie ekspresji pomiędzy badanymi warunkami:
print(r)
}

```

Przeprowadzenie analiz

```
summarizeDDS(dds_22)
```

```

## [1] "Geny o istotnie statystycznie zmienionym poziomie ekspresji pomiędzy badanymi warunkami: 245"
## log2 fold change (MLE): condition cond2 vs cond1
## Wald test p-value: condition cond2 vs cond1
## DataFrame with 245 rows and 6 columns
##
##          baseMean    log2FoldChange      lfcSE
##          <numeric>      <numeric>      <numeric>
## POTEH          7.51544814382448  4.19641624255765  1.18815109645868
## DUXAP8          190.85715523607  4.37986153892418  0.245867240229179
## LL22NC03-N64E9.1 12.5087538584591  4.99825333967629  1.12282268650184
## BMS1P22         19.7568886702301  3.77055055841666  0.658753435974667
## LL22NC03-N14H11.1 49.9298664333367  4.46940491095316  0.762421578001015
## ...
## SBF1           1203.85266360978 -1.26001910429072  0.0631380543534769
## ADM2           23.2618362099514  4.27388517372684  0.683224945031972
## MAPK8IP2       563.990778684874 -5.66638807876333  0.178682045776597
## SHANK3         480.297042650272 -3.99616356493982  0.12603121839484
## RPL23AP82      82.9462987004674 -1.26266341637979  0.2143837813449
##
##          stat          pvalue          padj
##          <numeric>      <numeric>      <numeric>
## POTEH          3.53188769935507  0.000412604538243108  0.00113629979976475
## DUXAP8          17.8139289107471  5.51025648891956e-71  1.09260514380291e-69
## LL22NC03-N64E9.1  4.45150725912777  8.5269649853446e-06  2.7652867756211e-05
## BMS1P22          5.72376605950888  1.04188194468107e-08  4.25332982122742e-08
## LL22NC03-N14H11.1 5.86211754744854  4.57001309933617e-09  1.93389578715811e-08
## ...
## SBF1          -19.9565716301065  1.31425162875461e-88  3.37811344576184e-87
## ADM2           6.25545833009192  3.96350162716321e-10  1.74093046155144e-09
## MAPK8IP2       -31.712128961451  1.05725543134108e-220 1.20534707534355e-218
## SHANK3         -31.7077277823368  1.21576794343009e-220 1.20534707534355e-218
## RPL23AP82      -5.88973386166942  3.86818054003525e-09  1.64694312563464e-08

```

```
summarizeDDS(dds_ERCC99)
```

```

## [1] "Geny o istotnie statystycznie zmienionym poziomie ekspresji pomiędzy badanymi warunkami: 12"
## log2 fold change (MLE): condition cond2 vs cond1
## Wald test p-value: condition cond2 vs cond1
## DataFrame with 12 rows and 6 columns
##
##          baseMean    log2FoldChange      lfcSE

```

```

##          <numeric>          <numeric>          <numeric>
## ERCC-00004 3918.98719921685 2.55760523577421 0.11640927167085
## ERCC-00019 14.2667163428672 1.46746567372637 0.621978976077996
## ERCC-00028 6.73992866973277 2.79204728530912 1.18471735892956
## ERCC-00062 77.4886630526754 2.55866701060085 0.364308288705593
## ERCC-00092 310.791194556933 2.43243862633193 0.174497749551288
## ...          ...          ...
## ERCC-00116 952.57953992746 2.2206659192427 0.148500492628805
## ERCC-00130 29681.8244237545 2.22586041484259 0.0895717309143729
## ERCC-00131 134.742367710255 1.92479201914061 0.235761424674267
## ERCC-00136 1898.3382995277 2.37165375964103 0.127301092296586
## ERCC-00170 15.8240228951514 2.59173551510514 0.653047654581001
##          stat          pvalue          padj
##          <numeric>          <numeric>          <numeric>
## ERCC-00004 21.9708035198941 5.47906515298811e-107 1.56153356860161e-105
## ERCC-00019 2.35934931913575 0.0183070138740007 0.0375327388745645
## ERCC-00028 2.35672016136563 0.018437134885751 0.0375327388745645
## ERCC-00062 7.02335656345324 2.166005022573e-12 1.23462286286661e-11
## ERCC-00092 13.9396561421956 3.63688155885512e-44 3.45503748091236e-43
## ...          ...          ...
## ERCC-00116 14.9539296465064 1.46823396793068e-50 1.67378672344097e-49
## ERCC-00130 24.850032394377 2.58378239978189e-136 1.47275596787568e-134
## ERCC-00131 8.16415162828246 3.23702090844188e-16 2.05011324201319e-15
## ERCC-00136 18.6302703052661 1.82599789882994e-77 3.46939600777688e-76
## ERCC-00170 3.96867747234773 7.22726220162091e-05 0.000257471215932745

```