

Zadanie 3 - Geny MYB i CEN

Ksenia Kvitko

3.04.2020

1. Dane z etapu końca zadania 2

```
# Zadanie 1 -- start
raw_data <- read.csv("../source_files\\counts.txt", sep = "\t", skip = 1)
geneLengths <- raw_data[, c(1, 6:9)]
TPM_step1 <- geneLengths
TPM_step1$bam.flower.bam <- TPM_step1$bam.flower.bam / TPM_step1$Length
TPM_step1$bam.stem.bam <- TPM_step1$bam.stem.bam / TPM_step1$Length
TPM_step1$bam.leaf.bam <- TPM_step1$bam.leaf.bam / TPM_step1$Length
TPM_step2 <- TPM_step1
TPM_step2$bam.flower.bam <- TPM_step2$bam.flower.bam / (sum(TPM_step2$bam.flower.bam) / 1000000)
TPM_step2$bam.stem.bam <- TPM_step2$bam.stem.bam / (sum(TPM_step2$bam.stem.bam) / 1000000)
TPM_step2$bam.leaf.bam <- TPM_step2$bam.leaf.bam / (sum(TPM_step2$bam.leaf.bam) / 1000000)
dane_TPM <- TPM_step2[,c(1,3:5)]
colnames(dane_TPM)[2:4] <- c("liść_TPM", "pęd_TPM", "kwiat_TPM")
# Zadanie 1 -- koniec

# Zadanie 2 -- start
MYB <- read.csv("../source_files\\MYB.txt", sep = "\t")
library(dplyr)
dane_myb <- dane_TPM %>% filter(Geneid %in% MYB$Locus)
# Zadanie 2 -- koniec
```

2. Wczytanie danych o genach z podanych plików

(plik MYB.txt wczytano osobno od pliku z poprzedniego zadania - najprawdopodobniej zawiera te same dane, dla pewności utworzono jednak osobną zmienną)

```
MYB_Zad3 <- read.csv("../source_files\\MYB_Zad3.txt", sep = "\t")
CEN <- read.csv("../source_files\\CEN.txt", sep = "\t")
```

3. Rozwiązanie zadania

Filtrowanie opracowywanych danych wybierając tylko geny z grup MYB i CEN-like

```
dane_myb <- dane_TPM %>% filter(Geneid %in% MYB_Zad3$Locus)
dane_cen <- dane_TPM %>% filter(Geneid %in% CEN$Locus)
```

Scalenie tabel

1. Jako dane “z informacją o ekspresją genów MYB **oraz** CEN” (zgodnie z treścią polecenia)

```
dane_myb_AND_cen <- dane_myb %>% filter(Geneid %in% CEN$Locus)
dim(dane_myb_AND_cen)
```

```
## [1] 0 4
```

Z uwagi na brak genów spełniających warunek przynależności do obu grup (MYB i CEN-like) - polecenie zinterpretowano jako informację o ekspresji genów MYB **lub** CEN-like

2. Wybór danych

```
dane_myb_OR_cen <- rbind(dane_myb, dane_cen)
```

Kontrolna weryfikacja czy poprzez scalenie nie uzyskaliśmy duplikacji genów (choć jest to z zasady niemożliwe - patrz wynik podpunktu 3.1). Pozostawiono w sprawozdaniu celem zobrazowania toku myślenia.

```
sum(duplicated(dane_myb_OR_cen$Geneid))
```

```
## [1] 0
```

Liczba wymiarów uzyskanej tabeli, nazewnictwo zgodne z poleceniem do zadania 3

```
dane_myb <- dane_myb_OR_cen
dim(dane_myb)
```

```
## [1] 114 4
```