

# Analiza Transkryptomu - Zadanie 2 (Geny MYB)

Ksenia Kvitko

3.04.2020

## 1. Dane z etapu końca zadania 1

```
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")
```

## 2. Wczytanie z pliku informacji o genach grupy MYB

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

## 3. Rozwiązanie zadania

Filtrowanie opracowywanych danych wybierając tylko geny z grupy MYB

```
library(dplyr)
dane_myb <- dane_TPM %>% filter(Geneid %in% MYB$Locus)
```

Liczba wymiarów przefiltrowanej tabeli - finalna tabela zawiera 4 kolumny i 37834 wierszy

(kolumny zgodne z oryginalnymi, mniejsza liczba wierszy adekwatnie do wyników filtrowania)

```
dim(dane_myb)
```

```
## [1] 111 4
```

```
ncol(dane_myb)
```

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

```
nrow(dane_myb)
```

```
## [1] 111
```