

Task 5: Exploratory Analysis

Dongyuan Wu

2020-06-10

First of all, load data and packages what I need.

```
library(DESeq2)
library(ggplot2)
library(tidyverse)

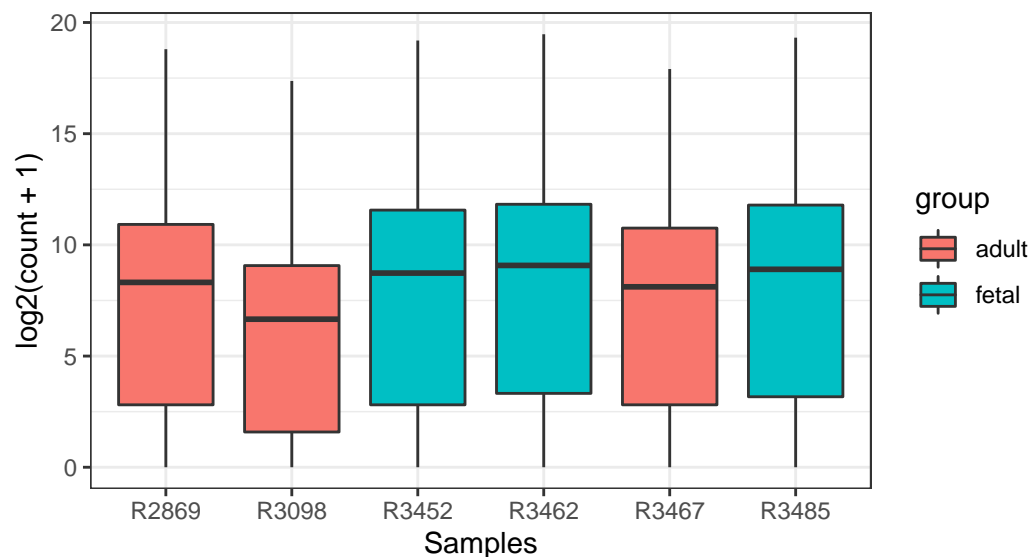
rawcount <- read.table("featurecounts.txt", header=TRUE)
rownames(rawcount) <- rawcount$GENEID
rawcount <- rawcount[, -1]
phenotype <- read.table("phenotype.txt", header=TRUE)
```

I used DESeq2 to do the exploratory analysis. When drawing boxplots for each sample, $\log_2 + 1$ transformation was used because there are some zeros in raw expression matrix.

```
dds <- DESeqDataSetFromMatrix(countData=rawcount, colData=phenotype, design=~group)
log2count <- as.data.frame(log2(counts(dds) + 1))

datforplot <- log2count %>% gather(sample, value, colnames(log2count)[1]:colnames(log2count)[6])
datforplot <- left_join(datforplot, phenotype[, c("sample", "group")], by="sample")

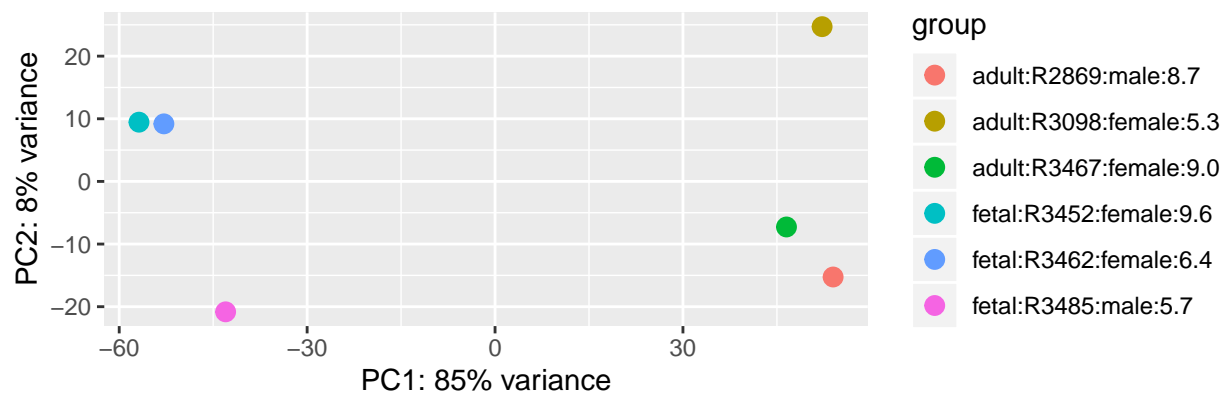
ggplot(data=datforplot) + aes(x=sample, y=value, fill=group) +
  geom_boxplot() + theme_bw() + labs(x="Samples", y="log2(count + 1)")
```



As we can see, the expression levels of fetal group may be a little larger than adult group.

The scatter plot of PCA is displayed below. In this plot, samples of fetal group gather on the left while adult group on the right, which means the value of PC1 in fetal group is lower than in adult group. In addition, male may have lower value of PC2 than female.

```
plotPCA(rlog(dds, blind=FALSE), intgroup=c("group", "sample", "sex", "RIN"))
```



```
plotPCA(rlog(dds, blind=FALSE), intgroup=c("RIN", "sex"))
```

