

SP1 Analysis

Orfeas Gkourlias

3/10/2022

3. Exploratory Data Analysis

The data will be need to be extracted from the count file first. Since it's in .tsv format, the seperator is going to be tab based. The inclusion of headers adds an X to the sequence ID's, because R is unable to make headers out of just integers.

3.1 Loading the data

```
file <- c("../data\\GSE152262_RNAseq_Raw_Counts.tsv")
raw_data <- read.table(file, sep = '\t', header = TRUE)
raw_data[1,]
```

```
##                X X4275 X4277 X4279 X4280 X4280a X4281
## 1 ENSG00000000003    23    30    11    43    31     8
```

The data is now loaded in as a data frame. Every row shows the raw counts of a specific gene being expressed. 4275, 4277 and 4281 are the variant types.