Results and Discussion & Conlcusion

2022-10-02

Results

Following the conclusion of exploratory data analysis, multiple important observations can be made about the data set. These will be examined by order of importance to answering the research question.

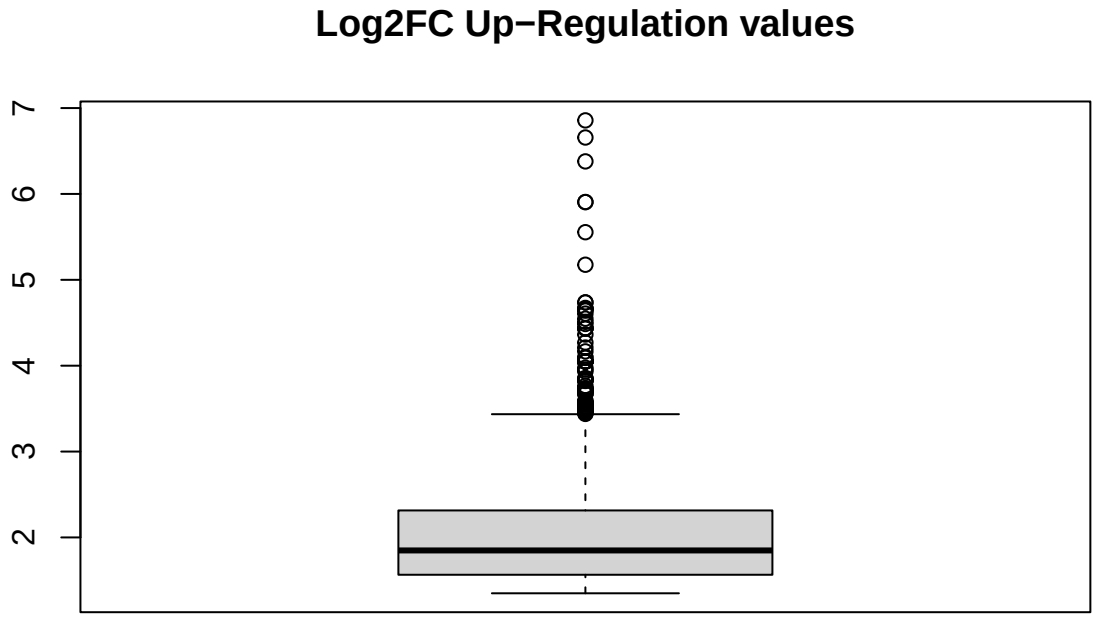
Starting with the significant fold change values.

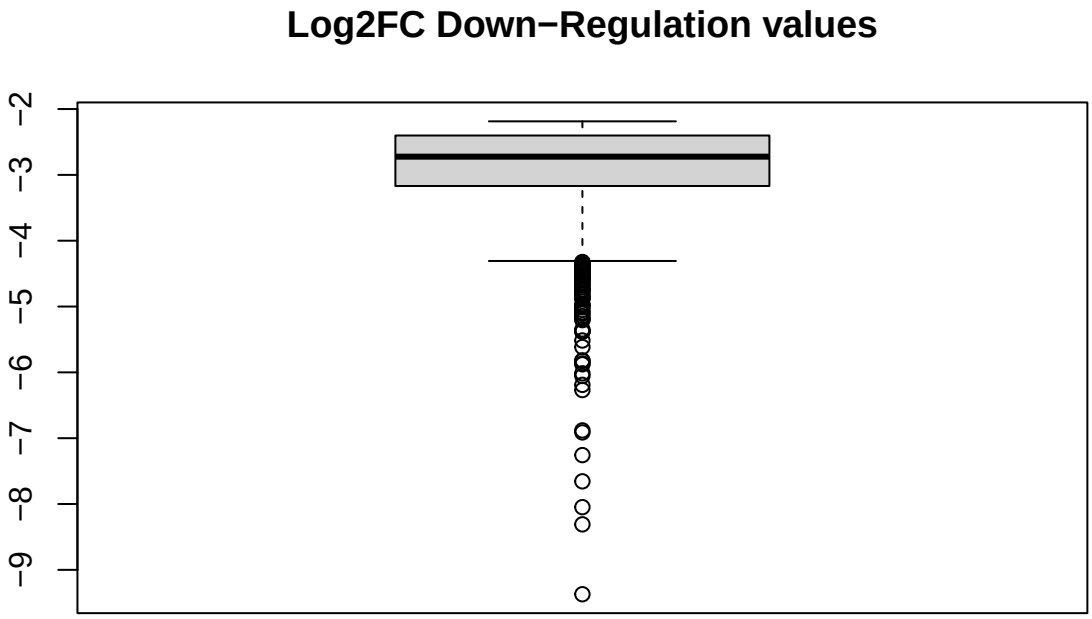
```
##       log.2.fold.change    SYMBOL
## 15347          6.856410      BMP5
## 21720          6.657270     CASP1
## 14353          6.378964      MMP1
## 37774          5.905406     ITGB6
## 37775          5.905406 LINC02478
## 20067          5.554548      CD69
## 21719          5.175524     CASP1
## 332            4.736748    CARD16
## 333            4.736748     CASP1
## 21331          4.672425       HGF
```

```
##       log.2.fold.change SYMBOL
## 11737         -9.370551 RPS4Y1
## 14897         -8.308893  DDX3Y
## 14288         -8.047485 EIF1AY
## 21020         -7.655585 ZNF257
## 14287         -7.256954 EIF1AY
## 40891         -6.912372    HRK
## 44064         -6.881689 TXLNGY
## 51508         -6.268827   <NA>
## 14898         -6.189748  DDX3Y
## 39797         -6.054623  USP9Y
```

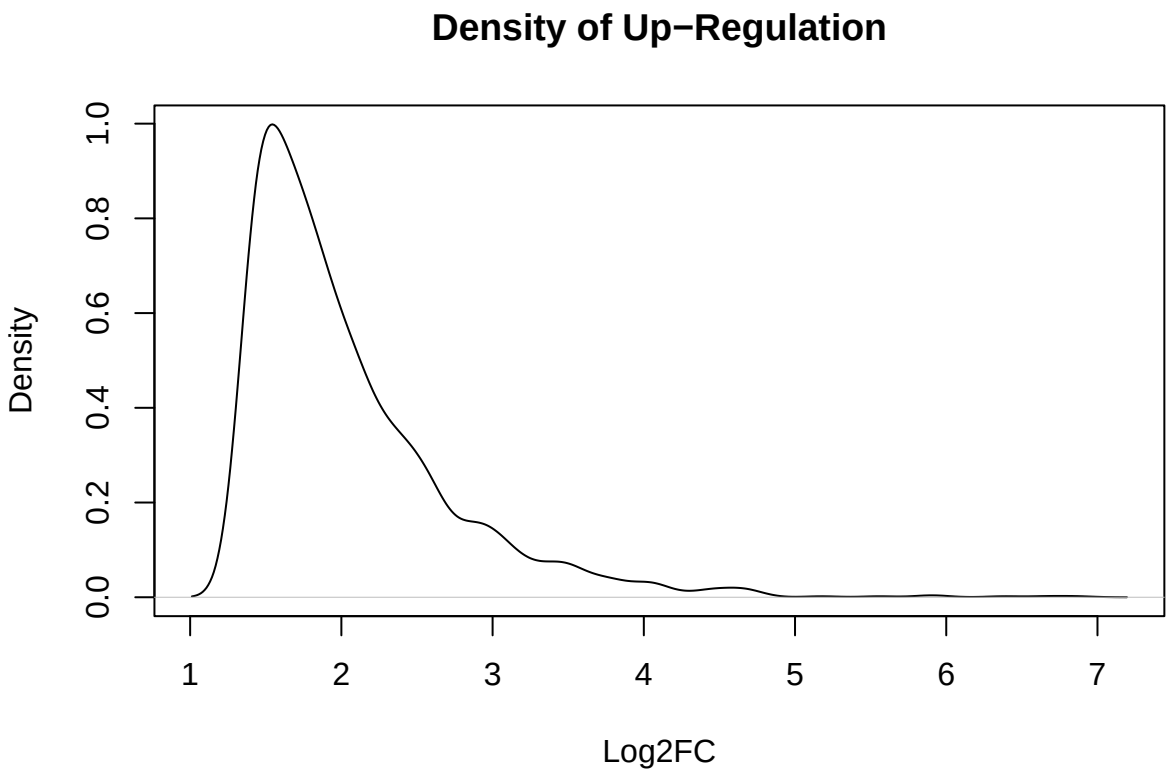
These are the 10 most significant genes, for both the higher and lower ends. There is a total of 57182 probe reads in the initial dataset. Using the pnorm function within R, the genes with Log2FC values on both ends were extracted. The higher end consisting of the upper 2.5% of the distribution, which totals 1761 of the 57182 genes. The lower 2.5% consists of 1945 genes. These two numbers are close, reinforcing the conclusion made in the EDA that the data is normally distributed. The higher the value, the more up-regulated the gene is in the mutant samples. The lower values are therefore indicators of down-regulation in the mutant samples.

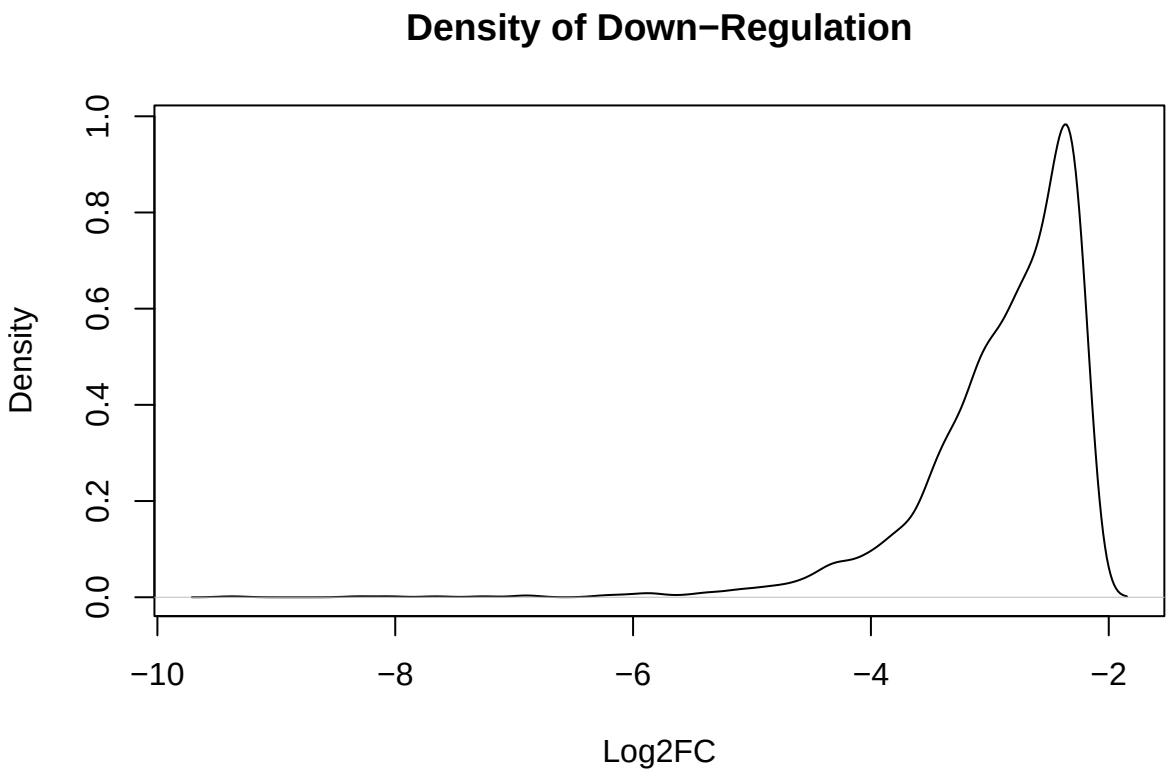
A comparison between the two groups may be made. The following box plots display summaries of both ends.





The boxplot shows the usual summary values. It is noteworthy that there are extreme values in both instances. 6 Of the up-regulated values being higher than the others in the plot. The same can be seen in the down-regulation plot, but with 7 values.


Density of Up−Regulation
Density
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
7
Log2FC



Density plots also indicate that the down-regulation is on average of bigger effect than up-regulation. In addition to plotting, this can also be demonstrated with a numerical summary

```
## [1] "test"
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.350   1.565   1.848   2.046   2.314   6.856
```

```
## [1] "test"
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -9.371  -3.169  -2.723  -2.899  -2.402  -2.187
```

Discussion & Conclusion