Results and Discussion & Conlcusion

2022-10-03

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.

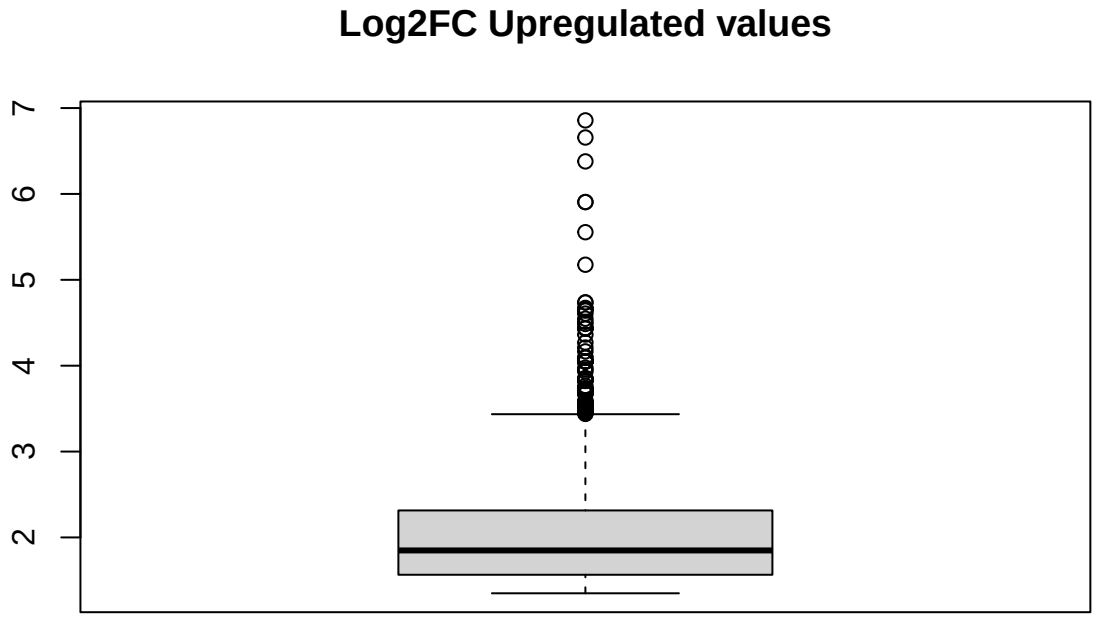
Starting with the significant fold change values.

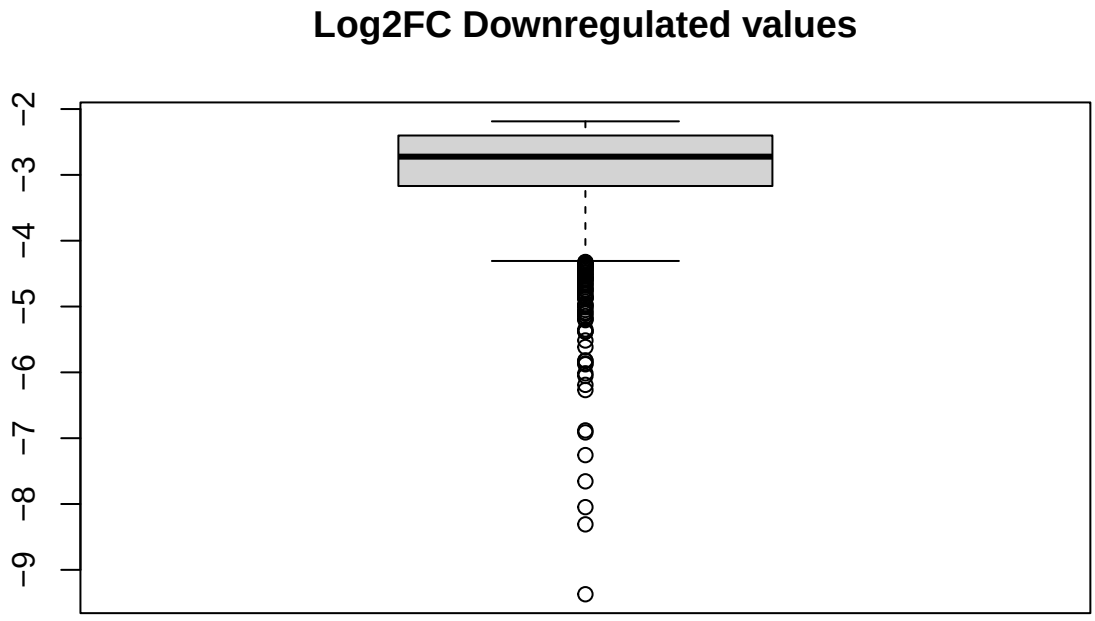
```
##       log.2.fold.change    SYMBOL
## 15347          6.856410      BMP5
## 21720          6.657270     CASP1
## 14353          6.378964      MMP1
## 37774          5.905406     ITGB6
## 37775          5.905406 LINC02478
## 20067          5.554548      CD69
## 21719          5.175524     CASP1
## 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
```

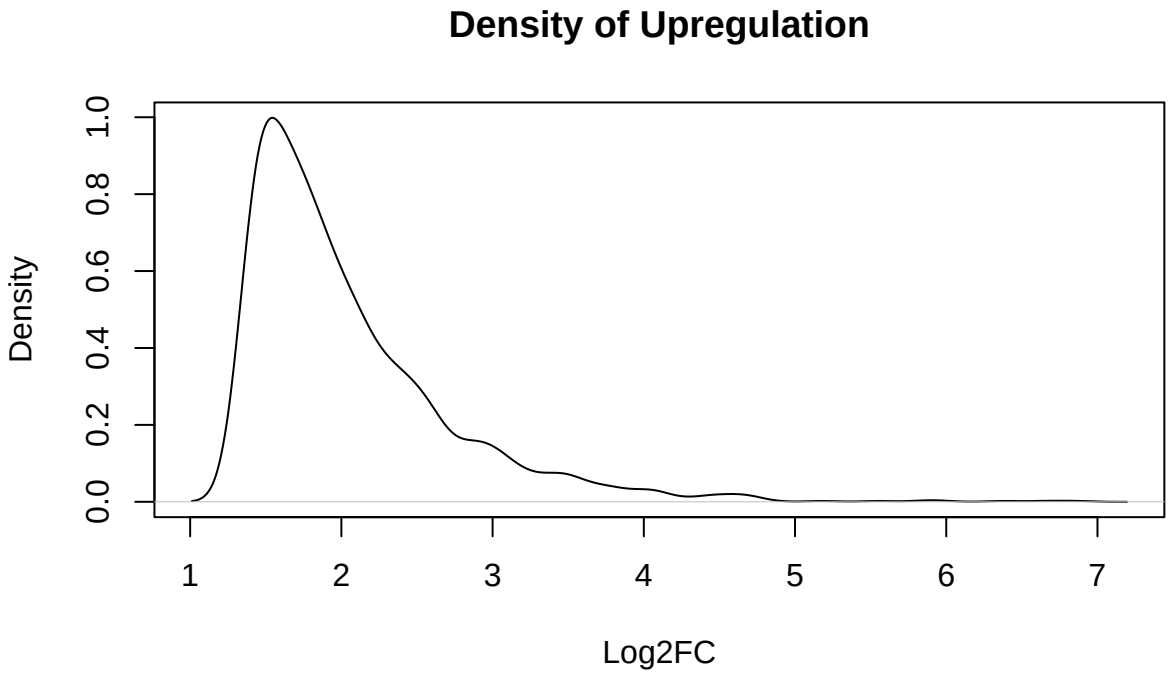
These are the 10 most significant genes, for both the higher and lower ends. There is a total of 57182 probe reads in the initial dataset. Using the pnorm function within R, the genes with Log2FC values on both ends were extracted. The higher end consisting of the upper 2.5% of the distribution, which 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 upregulated the gene is in the mutant samples. The lower values are therefore indicators of downregulation in the mutant samples.

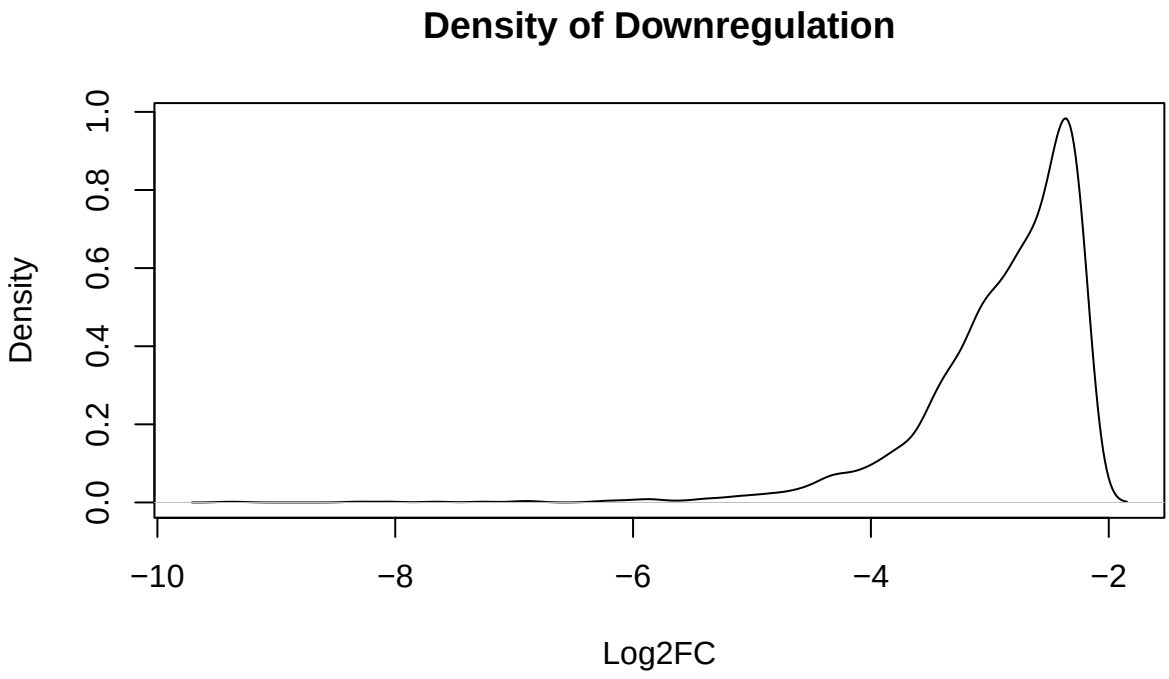
A comparison between the two groups may be made. The following box plots display summaries of both ends.





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





Density plots also indicate that the downregulation is on average of bigger effect than upregulation.

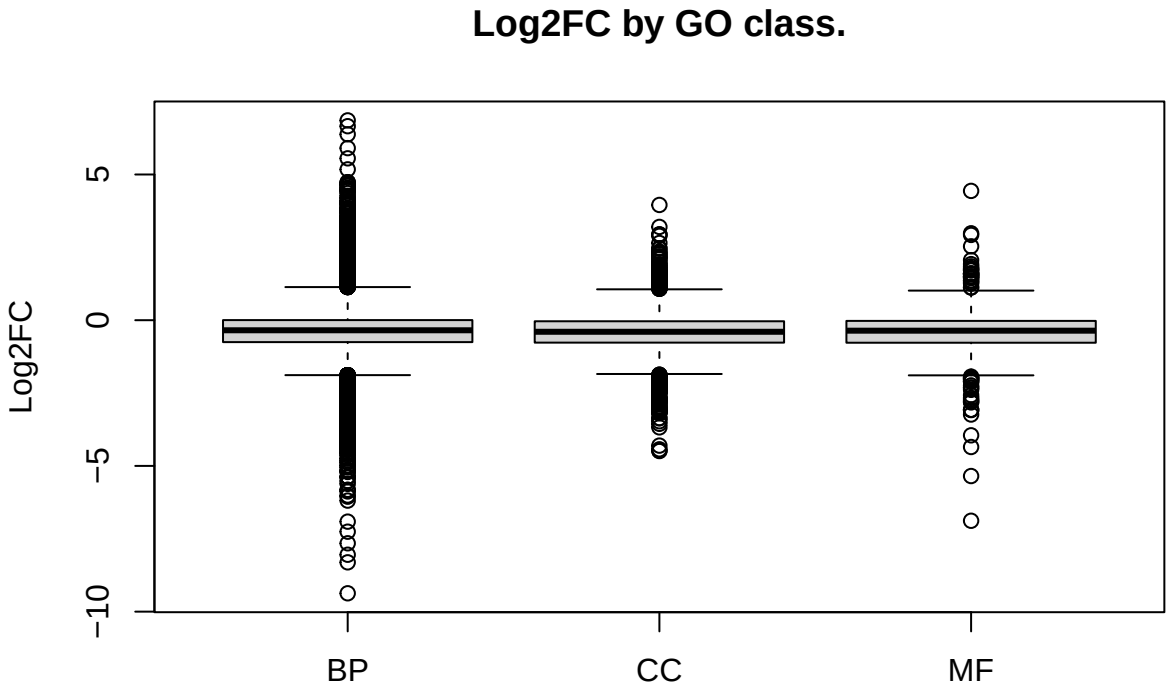
The results show above can also be demonstrated with a numerical summary

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

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

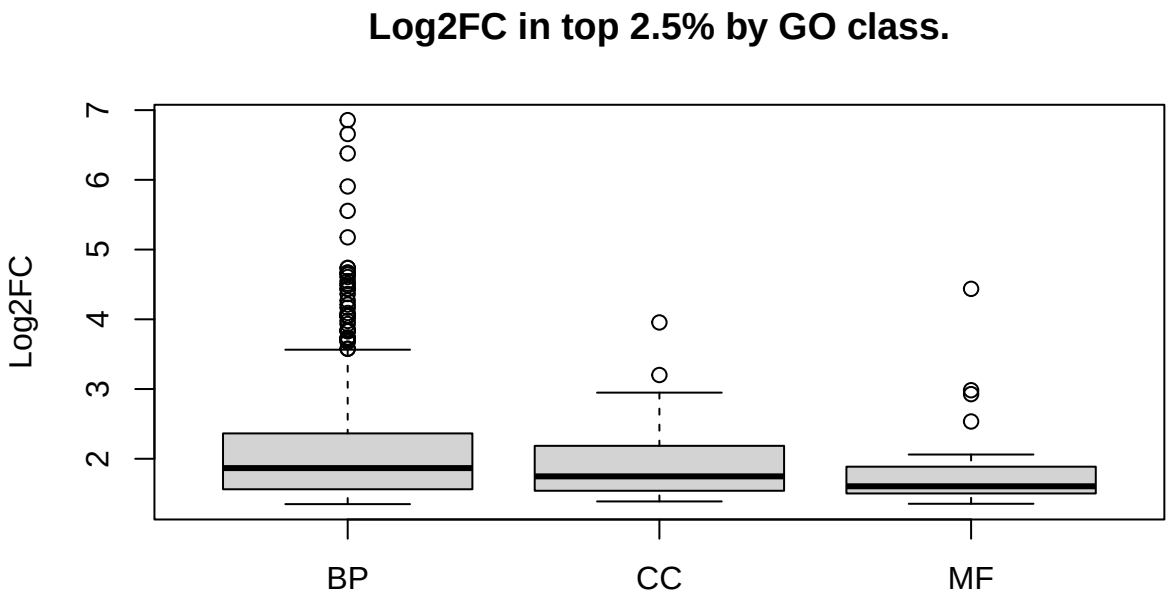
As can be seen, the downregulation has a more extreme mean than that of the upregulation.

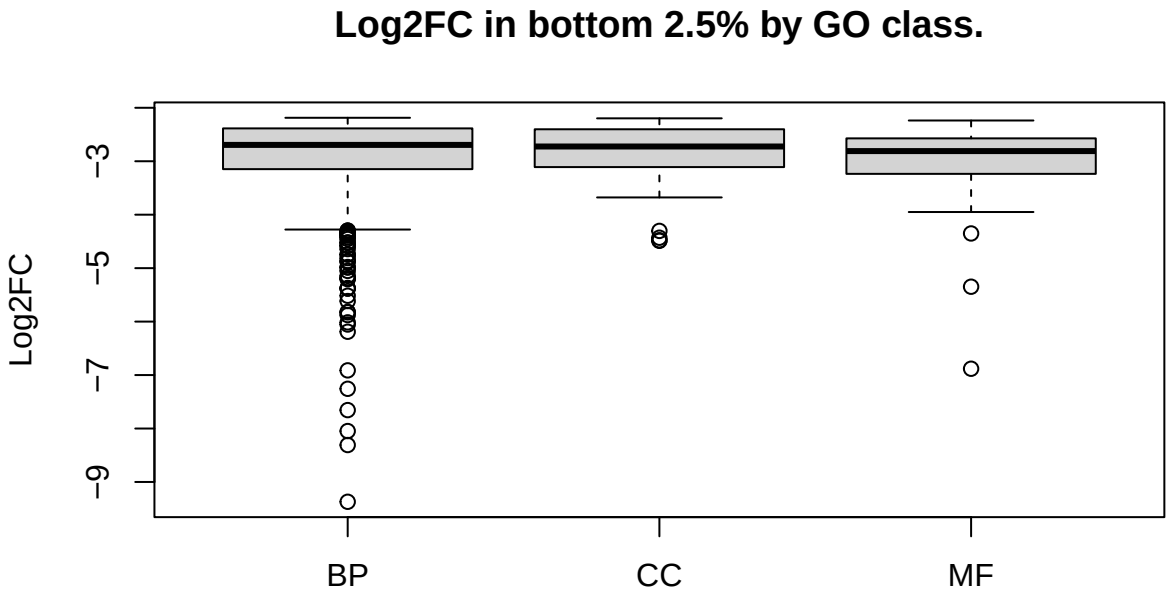
In addition to the individual genes, ontological classes were analysed. The following spread in ontological classes was observed.



This shows that the Log2FC spread differs for the classes. They do retain roughtly the same averages, but have different lower and higher ends.

This may also be observed for the significant DEGs.





Discussion & Conclusion

The examination of the data and plots shows that the mutation indeed affects several genes. Both in being up and down regulated, with a slight bias towards downregulation. In answering the research question, there needs to be a correlation between genes and the development of FAD. This correlation seems to be present, considering the mutation levels across the three mutation samples are similar. This may further be researched with addition of other control groups, and more FAD patient samples.

Which genes are correlated to FAD can also be deduced from the data. There's hundreds of genes which are expressed significantly different in the mutant samples. While all of these are important, there's still extremes with an alpha value of 0.05. These extreme downregulated genes are, in order of significance: RPS4Y1, DDX3Y, EIF1AY, ZNF257, EIF1AY, HRK and TXLNGY. For upregulation these are: BMP5, CASP1, MMP1, ITGB6, LINC02478, CD69.

Upregulation

Looking at the individual genes and their transcripts may point to proteins which would make sense for FAD patients to be functioning differently. Looking at RPS4Y1: a gene encoding for ribosomal protein S4. This specific gene, and more related to ribosome components, have been observed to be differentially expressed in multiple forms of Alzheimer's disease, not merely familial Alzheimer's disease. Translational gene mapping of cognitive decline

There's more relevant DEGs which affect the transcriptional capabilities of patients. These have also been discovered in prior papers. DDX3Y is related to translation. EIF1AY and ZNF257 are both genes encoding for transcription components, just like RPS4Y1.Identification of potential blood biomarkers for early diagnosis of Alzheimer's disease through RNA sequencing analysis

The two remaining upregulated genes are neither related to the central dogma of cells. HRK encodes for a gene related to activation and inhibition of apoptosis. This has also been observed in a prior paper on AD genetics as a possibly important gene for AD. The hippocampus being mainly affected. Genomic Variants, Genes, and Pathways of Alzheimer's Disease: An Overview

TXLNGY is a gene detrimental to syntaxin binding activity. This gene appears to be differentially expressed in AD patients who experience choroid plexus epithelium failure. Gene expression and functional annotation of human choroid plexus epithelium failure in Alzheimer's disease

Downregulation

In the case of downregulation, the BMP5 gene has the lowest log2FC value. This gene encodes for a protein which is part of developing cartilage and bones. The literature on this specific gene is quite limited. Cartilage and bone development intuitively does not allude to any involvement with development of Alzheimer's disease.

CASP1 is a gene that encodes for an enzyme which mediates the cleavage of the inactive precursor of IL-1beta. The literature on the link between this and familial Alzheimer's disease is a topic of debate. This paper:
Caspase-1 genetic variation is not associated with Alzheimer's disease risk seems to acknowledge the differential expression correlation of this gene, but does not accept the causative association with Alzheimer's disease.

MMP1 encodes for a metalloproteinase enzyme, which can cause tissue breakdown and secondary damage by several diseases. According to this paper: Brain matrix metalloproteinase 1 levels are elevated in Alzheimer's disease, there might be a link with increased MMP1 levels and AD. The keyword being elevated. This DEG data set indicates that the gene is down regulated in case of FAD. This is a surprising result, considering the referenced paper shows the opposite.

The literature on the link between [ITGB6] and FAD is limited, but the differential expression of this gene is a recurring observation. The protein for which the gene encodes is an integrin.

LINC02478 is a bit of a different gene to the others. This is a long intergenic non-coding rna gene. This might be cause for further analysis, because this gene lacks sufficient research to make any yet conclusions.

The last gene, CD69, encodes for a membrane-bound, type 2 C-lectin receptor. The paper: Elevation of CD69+ monocyte/macrophages in patients with Alzheimer's disease Indicates that this is indeed a relevant DEG. But just like MMP1, there's a difference in up/down regultaion between the data in the referenced paper, and the data set used in this one. While the reference shows that there is a link between upregulation and AD, the data from this project shows downregulation.

It is not clear why some of the genes in this project seem to have opposing regulation trends with the data seen in other papers. This might be because both up and down regulation of these genes are pathogenic, and may lead to the loss of function in proteins. Which would explain the phenomenon.