Background

According to the World Health Organization, one person dies due to suicide every 40 seconds globally. As suicide rate climbs with over 16 per 100,000 individuals dying by suicide annually, Major depressive disorder (MDD) is projected to become the leading disease burden globally by 2030. The pathogenesis of depression and suicide is not clear and there is a high non-response rate for current solutions involving antidepressants. Thus, under the mentorship of Dr. Maura Boldrini and Dr. Hanga Galfalvy, our team leveraged brain proteomics data to understand at a molecular level how proteins are differentially expressed in individuals with MDD, MDD with treatment, and controls, as well as depressed individuals who died by suicide versus did not. Our data consisted of postmortem brain samples measuring various protein levels and collected from the Dentate Gyrus (DG) of 36 individuals, as well as demographic data for each of the 36 individuals (Figure 1).

Methods

First, a Random Forest Classifier which predicted ‘suicide’ vs. ‘non-suicide’ using protein expression levels was used to narrow down the number of proteins to be tested during the feature selection phase (Figure 2). Next, the Benjamini-Hochberg procedure with a maximum false positive rate of 5% for both suicide and group differences revealed 10 significant proteins which may signal greater risks for depression and suicide (Table 1). Lastly, the non-parametric Kruskal-Wallis test yielded VGF_HUMAN protein levels as statistically significant amongst all three groups with p< 0.05 (Table 1). The 11 key proteins found using these non-parametric tests are listed in Table 1. Natural Language Processing and Text Mining techniques were also used to analyze protein function text data scraped from uniprot.org and uncover relationships between proteins.

Results

The three network graphs shown in Figure 4a, 4b, and 4c visualize clustering patterns of samples based on protein level correlations.

(i) Selected key protein features effectively discriminate non-suicide vs. suicide samples and control vs. depressed (MDD, MDD + treated) groups (Figure 4a, 4b).

(ii) The entangled network in Figure 4c demonstrates the difficulty in distinguishing MDD from MDD + treated samples.

The NLP/text mining results revealed interesting findings related to the behavior of proteins and their involvement with depression and suicide.

(i) Through word distributions, HS71L and ATD3A were both found to be highly involved with mitochondrial protein synthesis.

- Lower ATD3A levels for suicide samples may indicate a mitochondrial protein synthesis gap which adversely affects ATP production downstream and further escalates MDD symptoms (i.e. lethargy).

(ii) Through word frequency distributions, RB66I2 was found to be heavily involved with neurotransmitter regulation.

- Excess RB66I2 levels observed in depressed samples may signal neurotransmitter system dysfunction and reflect as depression-like symptoms due to neurological molecular imbalances.

(iii) PRS8_HUMAN and HS71L_HUMAN (row 2, row 6 in heatmap) have 0.8 pairwise textual similarity score and have similar roles in misfolded protein formation and/or damage and ATP-dependent processes (Figure 5).

Conclusions

This proteomics and data mining approach helps us to uncover molecules involved in the pathogenesis of MDD and suicide and identify new molecular treatment targets to develop novel drugs to treat these severe conditions and underlying dysfunctions in ATP-dependent processes and neurotransmitter regulation.

Acknowledgements

We would like to thank our capstone advisors, Dr. Boldrini, and Dr. Galfalvy of the Department of Psychiatry and the faculty at the Data Science Institute at Columbia University for all of their support and guidance.

References