Disclaimer: I collaborated with Aron Diamond to search for this paper. However, all code, analyses, and report content were created individually.

Introduction

The paper I researched, “Multi-omics analysis reveals the key factors involved in the severity of the Alzheimer’s disease”, looked into Alzheimer’s disease (AD), a progressive neurodegenerative disorder with a pathogenesis that still remains largely misunderstood. Building on evidence that age, metabolic abnormalities, and neurotoxic substances are potential risk factors, and that molecular and microbial differences exist between AD patients and healthy controls, this comprehensive analysis of 87 AD patients was motivated by growing evidence that psychological factors, particularly motivation, may influence cognitive decline and dementia.

The data used in this study consisted of a lot of metrics for each patient. These metrics for each patient include demographic data, such as age, gender, education level, race, measures of cognition, motivational variables, neuropsychological tests, clinical parameters (ex. triglycerides), and genetic information (such as ApoE genotyping). The analysis of this comprehensive dataset aims to elucidate the role of these factors in the cognitive decline in the studied population in hope to potentially identify new therapeutic targets to intervene with the progression of AD and cognitive decline.

Data Analysis and Code Structure

<u>UMAP plot:</u>

The first plot I aimed to create was the UMAP plot created based on significant clinical parameters to visualize inter-group differences. To create this plot, plot parameters were first defined. To determine which parameters to use for the UMAP clustering, Mann-Whitney U tests were run to compare the distributions of each parameter between ADAS groups. To do this, the raw sheet of the clinic spreadsheet was loaded in as a data frame. Next, for each parameter missing values were filtered out to avoid errors when computing statistics. Then, the log2 fold change (log2FC) was calculated between group averages, and the Mann-Whitney U test was performed to calculate the test statistic and the p-values. These tests were performed three times, for low vs medium, low vs high, and medium vs high ADAS groups. The results for each were filtered with a p value threshold of 0.05 to determine significance. 6 parameters were significant in total: 'Triglycerides', 'HbA1c', 'Insulin', 'ALT', 'Platelet', and 'Albumin'. These made lots of sense as well, as they also have vast clinical relevance to many contributors to cognitive decline.

Next, the imputed sheet from the clinical data was loaded in, and these parameters along with ADAS group columns were selected. The imputed sheet was used for the creation of the UMAP due to the algorithm requiring a complete dataset without any missing values to ensure proper dimensionality reduction. Colors were then set for each ADAS group by creating a color map and iterating through each item in the df and assigning a color based on the group. When creating the UMAP model, the parameters described in the paper were used (25 neighbors, 2 components, and a random seed of 0). The model was then fitted to generate 2d coordinates (embedding). Finally, the UMAP was plotted using the x and y

coordinates stored in the embedding from the previous step with the colors from the color set defined earlier. Finally, tick marks were hidden and UMAP axis labels were added.

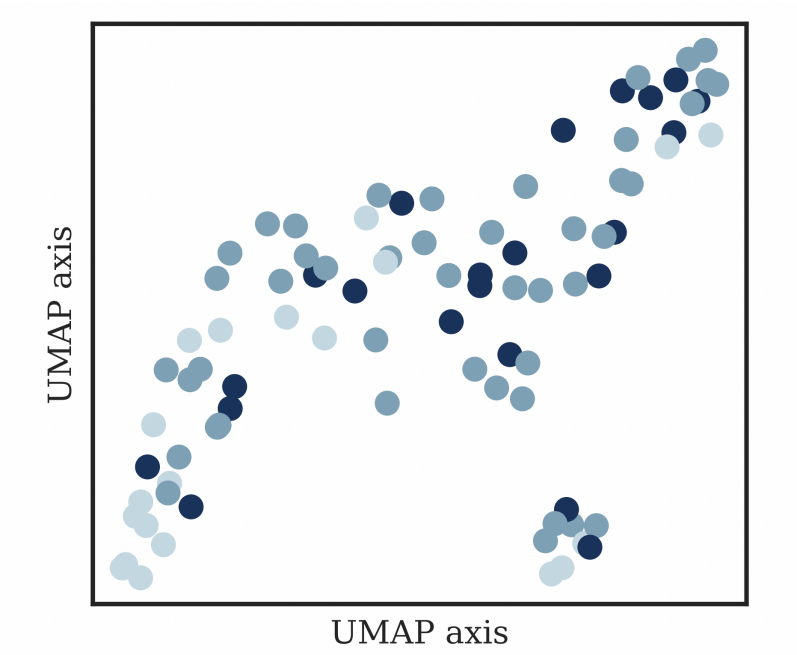
Volcano plot:

The next plot I aimed to recreate was the volcano plot illustrating the plasma proteins that are significantly altered between high and low ADAS (Alzheimer's Disease Assessment Scale) cognitive subscale groups. As such, I began my analysis by loading in the clinical data, specifically the "imputed" sheet, along with the proteomics sheet as dataframes using pandas (both were excel spreadsheets). The imputed sheet was chosen because it contains a complete dataset where missing values have been filled, ensuring compatibility with downstream analyses, and the proteomics sheet was selected because it provides plasma protein measurements, which are central to identifying significant alterations between the ADAS groups. Next, I merged the two by adding the ADAS group for each patient ID in the clinical df to the proteomics df. This was done to enable the comparison of proteomics data between each ADAS group (low and high in this case). After, I extracted the gene names from each column of the merged df. I removed the section of the name following the "-" to increase readability when creating the plot. I then split the df into two, one containing the rows with patients with a high ADAS label, and another containing those with a low label.

Next came the statistical analysis of the data. To create the volcano plot, the log2FC was calculated for each gene and added to a list. Next, Kruskal-Wallis tests were performed to determine if expression values differed between low and high ADAS groups. A data frame of the results was created for the creation of the volcano plot. In that dataframe, significance levels were computed and -log10 transformations were applied to p values for ease of viewing.

Finally, the volcano plot was created. Significant points were annotated by selecting points that had a -log10 p value transformation greater than 2.4 as the threshold, and a negative log2FC (which indicates that gene is downregulated). The other conditions were if the -log10 exceeded 2 and the log2FC was positive (which indicated that the gene is upregulated). These thresholds were chosen as they are commonly accepted values of significance in biological research.

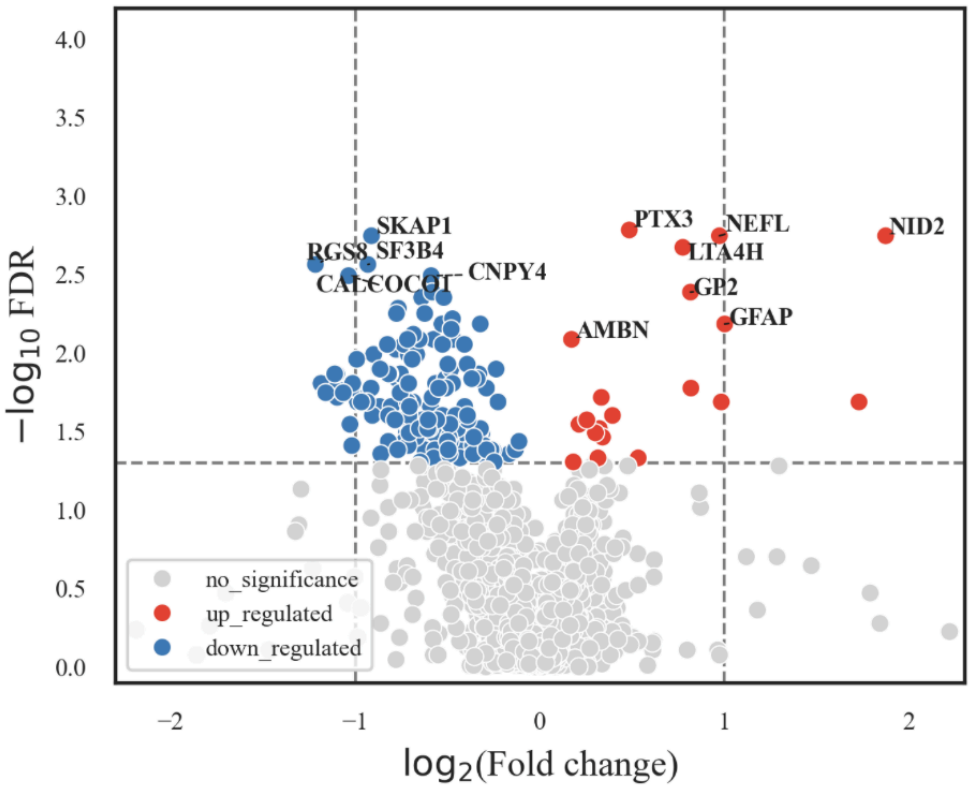
Results



The UMAP plot visualized inter-group differences based on the significant clinical parameters outlined in the methods section. Based on the output of this plot, it was observed that patients from each ADAS group were well mixed. The projection shows data points that are distributed without clear distinct clusters. However, some localized groups were present, such as the one in the bottom-right corner of the projection. The overlap of different color gradients (ADAS groups), however, indicates overlap between the groups as opposed to strict separation. The overall takeaway from this projection is that the clinical parameters selected do in fact contribute to a mixed distribution, implying that while there could be some group-specific trends, the boundaries between the different ADAS groups of patients are not sharply defined.

Unfortunately, this UMAP projection suggests that the parameters selected may not be strong indicators of cognitive decline on their own. There is overlap present between the different ADAS groups, indicating the parameters do not provide clear separation between groups, and thus that they alone likely lack the discriminatory power to distinguish between levels of cognitive function. This also suggests that the other clinical parameters used throughout the study likely do not have the power to distinguish between ADAS groups as well, as they came up as being insignificant between groups when performing Mann Whitney U tests.

*It should be noted that my figure does not exactly match the UMAP in the paper due to UMAP being a stochastic algorithm, which means that variations in library versions, random initialization, and computational precisions may lead to differences in plots even when using the same dataset. While the plot looks different, the groupings and general trends are conserved.



On the other hand, the volcano plot illustrated the plasma proteins that are significantly altered between high and low ADAS-Cog groups. From this figure it was identified that the top 5 downregulated proteins were SKAP1, SF3B4, CALCOCO1, RGS8, and CNPY4, while the top upregulated proteins were PTX3, NEFL, NID2, LTA4H, GFAP, AMBN and GP2 between high and low ADAS groups. Diving deeper into these results and what they mean, the proteins such as PTX3 and GFAP are markers of inflammation and glial activation, both of which are usually found to be elevated in neurodegenerative conditions such as AD. On the other hand, proteins such as SKAP1 and CNPY4 downregulated in the high ADAS group can be hypothesized to have some neuroprotective features. If this were to be true, it would make sense that these plasma proteins would be downregulated in patients in the high ADAS group. More downstream in the study, further analyses indicated that SKAP1 could play an important role in the development of AD. In previous studies, it was found that SKAP1 deficient platelets were shown to exhibit impaired activation and aggregation in response to various different stimuli (Kasirer-Friede et al., 2006). SKAP1 was stated to play a crucial role in the regulation of dendritic spine formation, dendritic spine actin dynamics and maintenance. This suggests that these processes could be tightly correlated with the appearance and progression of AD.

On the other hand, further analysis showed that NEFL was one of the proteins correlated most strongly with ADAS-Cog scores. It was found that NEFL was found to also be negatively associated with triglycerides, insulin, and ALT, but positively correlated with ADAS-Cog scores. According to genecards.org, the NEFL protein is a light chain neurofilament protein, and mutations in the gene that encodes have been known to cause Charcot-Marie-Tooth disease (CMT), a hereditary group of nerve disorders that affects the peripheral nervous system. This is very fascinating, as the brain is part of the

central nervous system as opposed to the peripheral nervous system where this protein is known to be linked to disease pathologies.

References

Meng, L., Jin, H., Burak Yulug, Ozlem Altay, Li, X., Lutfu Hanoglu, Seyda Cankaya, Coskun, E., Ezgi Idil, Rahim Nogaylar, Ahmet Ozsimsek, Saeed Shoaie, Hasan Turkez, Nielsen, J., Zhang, C., Borén, J., Mathias Uhlén, & Adil Mardinoglu. (2024). Multi-omics analysis reveals the key factors involved in the severity of the Alzheimer's disease. *Alzheimer S Research & Therapy*, *16*(1). https://doi.org/10.1186/s13195-024-01578-6

Kasirer-Friede A, Moran B, Nagrampa-Orje J, Swanson K, Ruggeri ZM, Schraven B, Neel BG, Koretzky G, Shattil SJ. ADAP is required for normal αIIbβ3 activation by VWF/GP Ib-IX-V and other agonists. Blood. 2006;109(3):1018–25. https://doi.org/10.1182/blood-2006-05-022301.

Database, Gene. "NEFL Gene - GeneCards | NFL Protein | NFL Antibody." *Genecards.org*, 2024, www.genecards.org/cgi-bin/carddisp.pl?gene=NEFL. Accessed 10 Dec. 2024.