Navigating Alzheimer's Complexity: Genetic and Neurological Perspectives

Buddy Pair 18: Anna Serbina (2835638) and Kristine McLaughlin (2837098)

Introduction and Motivation

Alzheimer's disease (AD) is a widely known progressive brain disorder characterized by a gradual decline in memory, cognitive abilities, behavior, and social skills, ultimately significantly impairing an individual's daily functioning capabilities [1]. Given its debilitating impact, understanding Alzheimer's disease on a biological and genetic level is crucial in the quest for potential cures and effective treatments. Therefore, throughout the past 8 weeks, we were motivated to use a variety of openly available AD datasets to perform analyses on genes and brain regions implicated in AD. In this exploratory data report, we summarize our processes and findings as we examine different types of studies, all in effort to extract key genetic and neurological variation between AD patients and control groups. This report is targeted toward researchers, clinicians, and stakeholders in the field of Alzheimer's disease research, with the aim of highlighting the significance of the identified genes and proteins in AD pathology. Additionally, it serves as a call to action for further exploration and implementation of machine learning models for early detection and personalized treatment strategies, emphasizing the potential for advancing clinical practice and improving patient outcomes in the fight against AD.

Initial Exploration using GWAS Results

Genome-Wide Association Studies (GWAS) are commonly used to pinpoint genetic variants associated with increased risk for certain traits or diseases, providing us with a favorable entry point for the first stage of our process. We leveraged data from one of the largest GWAS studies on AD via the platform, FUMA [2]. We utilized the tabular GWAS data to filter for genes that surpassed a significance threshold of 10^{-8}, arriving at a list of 44 significant genes associated with AD. Table 1 presents a view of the 10 most statistically significant of the list, including their corresponding chromosome, z-statistic, and p-value. These significant genes acted as the foundation of the rest of our analysis, particularly the APOE gene which is firmly recognized in the scientific community to contribute to a higher risk of AD.

GENE	CHR	ZSTAT	P	SYMBOL
{'ENSG00000130204'}	19	21.921	8.1347e−107	{'TOMM40' }
{'ENSG00000130202'}	19	21.742	4.1268e−105	{'PVRL2' }
{'ENSG00000130208'}	19	20.284	8.8846e−92	{'APOC1' }
{'ENSG00000130203'}	19	17.106	6.639e−66	{'APOE' }
{'ENSG00000007047'}	19	12.952	1.1511e−38	{'MARK4' }
{'ENSG00000069399'}	19	12.302	4.4453e−35	{'BCL3' }
{'ENSG00000104856'}	19	11.391	2.3249e−30	{'RELB' }
{'ENSG00000266958'}	19	11.214	1.7413e−29	{'AC006126.3'}
{'ENSG00000104853'}	19	10.957	3.0595e−28	{'CLPTM1' }
{'ENSG00000073008'}	19	8.5859	4.5065e−18	{'PVR' }

Table 1. GWAS Genes Associated with Alzheimer's Disease: Z-Statistics and P-Values

Expanding Findings with Gene Expression Data

After identifying 44 statistically significant genes in AD from the GWAS data, we further studied these genes by examining their expression levels between individuals with and without AD. We evaluated gene expressions for all combinations of each significant gene within each brain region, performing a two-sample t-test with each pair to determine the statistical significance of the differences between the control and patient groups. To account for multiple comparisons made simultaneously, we applied the Bonferroni correction by adjusting the p-value threshold of 0.05 based on the total number of significant gene and brain region combinations. With the adjusted p-value, we isolated 12 genes exhibiting the most

significant effects and found that they are predominantly situated in three key regions: the Piriform Cortex (PC), Superior Frontal Gyrus (SFG), and Hippocampus (HIP). The PC is mainly involved in olfactory responsibilities, which we found fascinating due to the relationship between smell and memory. Studies indicate AD patients often experience olfactory dysfunction early on, including reduced odor discrimination and memory loss. These findings reinforce our conclusion regarding the significant role of PC regulation in AD [3].

Utilizing Proteomics and Protein Interactions

Wanting to take a step further in our exploration of significant genes, we decided to introduce proteomics, the study of proteins in a biological context, to our analysis. Given that proteins define the functions of cells, we believed examining Protein-Protein-Interactions (PPIs) of our significant GWAS genes would lead to further insights on the function and effects of these genes in the context of AD. Working with the HENA dataset, we analyzed PPIs of our 44 significant genes and identified 128 PPIs of interest. Continuing on our integration of datasets, we then performed a two sample t-test to see if these proteins showed statistically significant differences in gene expression levels in the Middle Temporal Gyrus (MTG), a brain region thought to be involved in early AD pathology. Table 2 displays a preview of these gene hits based on our PPIs of interest.

gene_name	GENexpdstats
{'STK11' }	-4.1942
{'ACSF2' }	-3.5501
{'SMIM3' }	-3.177
{'PLP1' }	-3.0755
{'DLGAP4' }	-2.7272
:	:
{'TEX264' }	4.2097
{'REPS2' }	4.3737
{'STX8' }	5.416
{'COMMD1' }	7.6965
{'ADCYAP1'}	9.837

Table 2. Genes and t-scores from Protein-Protein-Interactions in GWAS Significant Genes

One interesting gene we found with one of the most notable differences in expression is ADCYAP1, a gene that codes for proteins essential in regulating neuroendocrine stress responses [4]. Stress is known to have damaging effects in the development and progression of diseases, and researchers examining the relationship between stress and dementia/AD in particular have concluded its impact with earlier onset and worse progression of this disorder [5]. Upon further investigation, several independent studies in the past few years have brought up ADCYAP1 as a new potential biomarker of AD, proving to be an intriguing path for further research [6].

Beyond individual PPIs, we can also create PPI networks to model the relationships between proteins and did so with a list of proteins involved in AD by experts. From this network, we were able to determine that the protein with the most interactions was ENSG00000142192 with a degree of 28. This protein is an amyloid beta precursor protein [7], whose relation to AD is confirmed by multiple studies. ENSG00000142192 is crucial in Alzheimer's disease (AD) due to its role in producing amyloid beta (Aβ)

peptides. APP is cleaved by β-secretase and γ-secretase, generating Aβ peptides that form amyloid plaques, a key feature of AD. These plaques disrupt neural communication and cause neuronal damage, contributing to cognitive decline. Mutations in the APP gene are linked to early-onset familial Alzheimer's disease, increasing Aβ production and plaque formation. As a result, APP and its processing enzymes are major targets for therapeutic interventions aimed at reducing Aβ production and clearing plaques [8].

Exploring Correlations with Brain Regions

After investigating relationships between genes and proteins, we were curious about potential correspondence between the brain regions affected by significant genes and their functions in AD. Using data from the Allan Human Brain institute (AHB), we created a visual plot illustrating the transcriptomic profile of APOE across the cortex in the left hemisphere, as the right hemisphere was incompletely covered (Figure 1). Notably, we observed the highest expression of APOE in the entorhinal cortex (EC) with an expression level of 2.193. Given that the entorhinal cortex (EC) is widely recognized as a key player in memory processing [9], it follows that dysregulation of APOE in this region, associated with AD, may manifest as memory issues - the main symptom of AD. Expanding this analysis to the rest of our significant genes, we found the brain region with the highest expression for each gene and displayed a segment of this list in Table 3. In this segment, the pericalcarine cortex appears several times, which is not as intuitive of a finding due to its main relation to blindness, but certainly is an interesting correlation to further inspect.

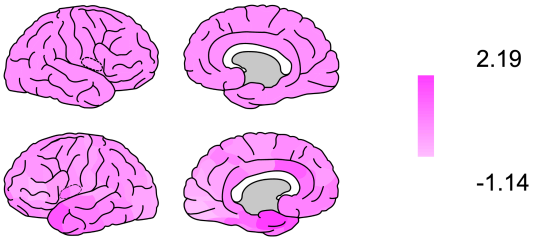

Figure 1. APOE Expression Across the Cortex

GENESYMBOL	TOPREGION_AHBAEXPRESSION
{'SORL1' }	{'ctx-lh-pericalcarine_1' }
{'SLC24A4' }	{'ctx-lh-pericalcarine_1' }
{'SLTM' }	{'ctx-lh-precentral_2' }
{'ZNF232' }	{'ctx-lh-pericalcarine_1' }
{'ABCA7' }	{'ctx-lh-pericalcarine_1' }
{'PVR' }	{'ctx-lh-precentral_1' }
{'CEACAM19'}	{'ctx-lh-lingual_1' }
{'CEACAM16'}	{'ctx-lh-rostralmiddlefrontal_3' }
{'BCL3' }	{'ctx-lh-pericalcarine_1' }
{'CBLC' }	{'ctx-lh-supramarginal_1' }
{'BCAM' }	{'ctx-lh-lateralorbitofrontal_1' }
{'TOMM40' }	{'ctx-lh-precentral_4' }
{'APOE' }	{'ctx-lh-entorhinal_1' }

Table 3. Top Expression Region for GWAS Significant Genes

Structural Brain Differences

Inspired by the results of our APOE cortex map, we set out to explore potential structural disparities between the brains of healthy individuals and those diagnosed with Alzheimer's disease (AD), and used MRI data of the two groups in this analysis. We implemented a specific focus on the middle temporal gyrus (MTG) due to its speculated involvement in early AD, and our findings revealed a significant reduction in gray matter volume among AD patients, as exhibited in the violin plot in Figure 2. The width of the violin plots are comparable, but there is a noticeable difference in the median and quartile ranges. This finding could be cause for investigation if a decrease in MTG volume could serve as an early detector for AD. Once again applying this investigation to other regions, below in Table 4, we summarize the top 10 regions observing the most statistically significant differences in mean gray matter volume based on the t-test scores comparing healthy individuals and AD patients. The largest difference in gray matter volume is observed in the left hemisphere in the parahippocampal region, which plays an

important role in both spatial memory and navigation [10]. One study indicates gray matter volume decline in this region could be an early biomarker of AD. We also noted significant differences in entorhinal regions in both the left and right hemispheres, which correlates with our findings on APOE expression. Additionally, it makes sense that the volume effects are the only region to span both hemispheres in AD patients due to its central role of memory processing.

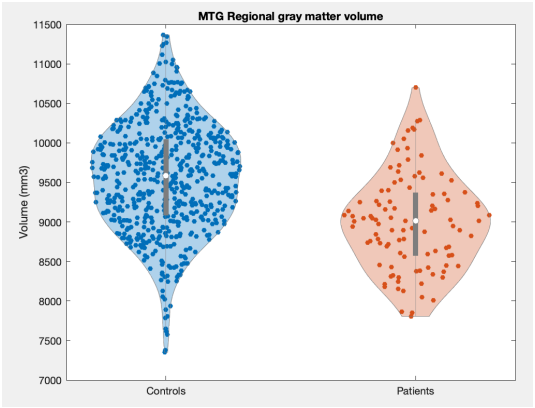

Figure 2. MTG Regional Gray Matter Volume

Region	TTestScore
{'ctx-lh-parahippocampal'}	8.4432
{'ctx-rh-paracentral' }	6.7375
{'ctx-rh-fusiform' }	6.6232
{'ctx-rh-bankssts' }	6.5486
{'ctx-rh-entorhinal' }	6.3367
{'ctx-lh-entorhinal' }	6.1715
{'ctx-rh-frontalpole' }	6.1142
{'ctx-lh-parsorbitalis' }	6.029
{'ctx-rh-lingual' }	6.0142
{'ctx-lh-supramarginal' }	5.9667

Table 4. Differences in Gray Matter Volume by Brain Region

Circling back to our investigation on APOE, our next curiosity included if the expression pattern of APOE across cortical regions was associated with the pattern of structural brain changes in AD patients we had just observed. We created the graph in Figure 3 to answer this question, mapping our regions to the regions we had AHB expression data for and averaging the atrophy pattern across both hemispheres to create the AD pathology score. The red lines indicate our 90% and 95% confidence interval.

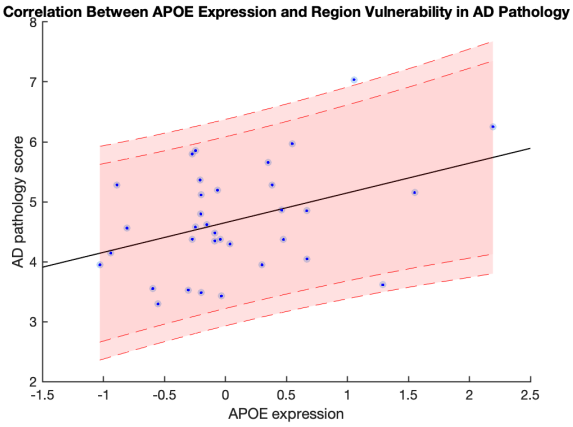

Figure 3. Correlation Between APOE Expression and Region Vulnerability in AD Pathology

The p-value for coefficient significance in this regression model, which evaluates the presence of a significant linear relationship between APOE expression and AD pathology score while considering other variables, was found to be 0.0318. This value indicates a significant positive correlation between APOE expression and AD pathology score, as it is below the conventional significance level of 0.05. This statistically significant result allows us to conclude that the APOE gene expression across the brain is related to the pattern of brain atrophy in Alzheimer's disease, confirming APOE's strong involvement in AD and encouraging further research into this gene in hopes to better understand AD.

Applying Machine Learning to Classify AD

Finally, we explored the possibility of using machine learning to try to differentiate patients with AD and control subjects based on their MRI data, more specifically on the cortical volume of various brain regions. We employed supervised learning as we trained a simple neural network on a small set of data (60 training patients, 60 training controls, 40 test patients and 40 test controls), and tested different versions of this model by modifying parameters of learning rates, regularization methods, and momentum. Our best model settings included a learning rate of 0.1, the L2 regularization method to prevent overfitting, and a momentum of 0.95. With this model, our training accuracy landed around 83.43% and our testing accuracy at 84.62%. We generated a confusion matrix to further analyze our model, which informed us of 18 false negatives and 9 false positives compared to 144 correctly classified data points. There is certainly room for improvement in our model, but these results are generally positive given the simple architecture and small dataset.

Based on the results from this specific dataset, the significant difference observed in gray matter volume between the control and patient groups indicates the effectiveness of utilizing a linear classifier, such as SVM, for identifying potential new AD patients. The SVM is likely a better choice here compared to a neural network due to the small dataset size, which may lead to overfitting. However, it is worth noting that a neural network of the chosen architecture yielded poorer results, likely due to potential overfitting. Nevertheless, machine learning remains a promising approach for identifying new patients. In future research, integrating all findings - including those related to genes, brain regions, and structural differences - into a neural network for predicting potential AD candidates would be intriguing. This approach is feasible given the identification of multiple genes associated with AD through overexpression or mutation, along with specific brain regions that exhibit differential susceptibility to the effects of these genes.

Future directions

Looking ahead, we envision integrating diverse datasets and exploring advanced machine learning techniques for early AD detection and personalized medicine approaches. Longitudinal studies and biomarker development will be pivotal in tracking disease progression and developing prognostic tools. While our findings have uncovered significant insights into the genetic and neurological aspects of AD, synthesizing all this data is infeasible for an individual researcher. This is where big data and machine learning can come in handy, allowing us to process and analyze vast amounts of information to reach groundbreaking results. Through these efforts, we aim to deepen our understanding of AD and pave the way for improved diagnosis and treatment strategies tailored to individual patients.

Data Management Plan

Genome-Wide Association Studies (GWAS) data was obtained from FUMA platform, accessible via txt files for raw data and csv files for metadata. Gene expression data was sourced from the National Center for Biotechnology Information, available in txt files. Protein-Protein Interaction (PPI) dataset was retrieved from a publication on Nature, downloadable as a txt file. STRING database was utilized for visualizing PPIs, accessible at STRING database. Transcriptomic levels of tissue samples were obtained from the Human Brain Map. Induced Pluripotent Stem Cells (iPSCs) data was sourced from the NCBI GEO database. Magnetic Resonance Imaging (MRI) data was acquired from the OASIS and ADNI

databases, available in csv files. Connectivity Data from the CLA_CCN_APOE_DTI dataset was accessed from the UMCD Human Connectome Project, provided in csv files.

In efforts to prioritize the FAIR principles, all data sources are openly accessible through the provided URLs, facilitating replication and analysis by any interested parties. To enhance reusability, we offer comprehensive metadata detailing data formats, processing methods, and analysis workflows. Processing of all data was conducted using our provided MatLab scripts, and can be employed by other researchers for similar analyses, thus promoting reproducibility and enabling further exploration of our findings.

References

1. Alzheimer's disease - Symptoms and causes - Mayo Clinic. (2024, February 13). Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447
2. K. Watanabe, E. Taskesen, A. van Bochoven and D. Posthuma. Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. 8:1826. (2017).
3. Zhang, Jingjing et al. "Olfactory Evaluation in Alzheimer's Disease Model Mice." Brain sciences vol. 12,5 607. 6 May. 2022, doi:10.3390/brainsci12050607
4. "ADCYAP1 Adenylate Cyclase Activating Polypeptide 1 [Homo Sapiens (Human)] - Gene - NCBI." National Center for Biotechnology Information, U.S. National Library of Medicine, www.ncbi.nlm.nih.gov/gene/116. Accessed 28 May 2024.
5. Justice, Nicholas J. "The relationship between stress and Alzheimer's disease." Neurobiology of stress vol. 8 127-133. 21 Apr. 2018, doi:10.1016/j.ynstr.2018.04.002
6. Tian, Yu et al. "Identification of diagnostic signatures associated with immune infiltration in Alzheimer's disease by integrating bioinformatic analysis and machine-learning strategies." Frontiers in aging neuroscience vol. 14 919614. 29 Jul. 2022, doi:10.3389/fnagi.2022.919614
7. Summary - Homo_sapiens - Ensembl Genome Browser 112, www.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000142192%3Br. Accessed 28 May 2024.
8. O'Brien, Richard J, and Philip C Wong. "Amyloid precursor protein processing and Alzheimer's disease." Annual review of neuroscience vol. 34 (2011): 185-204. doi:10.1146/annurev-neuro-061010-113613
9. Garcia, A. D., & Buffalo, E. A. (2020). Anatomy and Function of the Primate Entorhinal Cortex. Annual review of vision science, 6, 411–432. https://doi.org/10.1146/annurev-vision-030320-041115
10. Echávarri, C et al. "Atrophy in the parahippocampal gyrus as an early biomarker of Alzheimer's disease." Brain structure & function vol. 215,3-4 (2011): 265-71. doi:10.1007/s00429-010-0283-8