Samreen Mehta
Sherwood High School
Sandy Spring, Maryland

Using Bioinformatics Tools to Identify Possible Target Genes for Alzheimer's Disease Treatments

ABSTRACT

Background
Alzheimer's disease is a progressive, advanced brain disorder that mainly targets older adults. Alzheimer's disease causes the accumulation of amyloid plaques which significantly affects memory, thoughts, and behavior. With no cure, this study aims to find target genes for the treatment of Alzheimer's disease using bioinformatics to analyze gene enrichment and expression.

Methods
I utilized bioinformatics tools including NCBI GEO2R and SR Plot, for analysis of KEGG and GO, to study dataset **GSE52022**. Within the dataset, I defined the samples into groups entitled littermate and mutant. GEO2R was used to identify differentially expressed genes, these differentially expressed genes were then further analyzed using SR Plot for GO and KEGG pathway enrichment.

Results
Genes with a P-value greater than 0.5 were excluded to find DEGS, and fold change values were utilized to identify upregulated and downregulated genes. There were 30 top Differentially Expressed Genes, with 15 upregulated genes and 15 downregulated genes. The GO analysis indicated enrichment in the biological process ontology, and importation into the cell was the most enriched function. The GO analysis also revealed key genes including TREM2 and TYROBP. KEGG analysis emphasized the Tuberculosis pathway and genes including Dectin-1 (CLEC7A), CALM, and CathepsinS.

Conclusion
Genes like TREM2, TYROBP, and CLEC7A are vital in microglial activation and neuroinflammation in Alzheimer's disease. Targeting these genes may enhance the clearance of

amyloid-beta plaques and reduce neurodegeneration, providing insights that could inform future research and therapeutic strategies.

INTRODUCTION

Despite significant advancements in our knowledge of Alzheimer's Disease (AD), one persistent issue is the absence of a cure or disease-modifying treatments, which is what this research aims to address (1). The scientific question that is being investigated is determining what genes are involved in the risk of developing Alzheimer's disease.

Alzheimer's disease is a progressive, advanced brain disorder that mainly targets older adults (2). Further, AD causes the accumulation of amyloid plaques which has significant effects on memory, thoughts, and behavior. It is the leading cause of dementia, caused by the buildup of abnormal proteins that disrupt brain function. Over time, it can lead to severe cognitive decline, preventing individuals from living independently. Alzheimer's disease has impacted over 15 million people nationwide and poses a dangerous public health crisis (3), as cases of this disease expand, it will continue to put financial and emotional strain on family members and healthcare organizations. Therefore, addressing and understanding Alzheimer's is a staple in decreasing its harmful impact and improving brain health across the board.

So far, we know that Alzheimer's disease causes progressive cognitive decline, which gravely impacts the amount of care and treatment that a diagnosed individual would need.

Bioinformatics studies on Alzheimer's disease aim to better comprehend the biological mechanisms driving the conversion of normal/MCI into AD pathology, which may ultimately open the door to disease-modifying treatments or cures (4). Research has identified multiple genes that are linked to Alzheimer's risk and examined how genetic variations impact the progression of the disease. Bioinformatics tools, including NCBI and GEO2R (5), analyze data including brain imaging (6), protein interaction, and even modified mice to find potential indicators of the disease. Moreover, these bioinformatics tools aid with predicting early signs of Alzheimer's which supports the process of treatment strategies.

A major setback with Alzheimer's is that there is currently no cure for this disease (1). While there are medications that people can take to reduce the symptoms, they do not offer a permanent cure or stop the progression of Alzheimer's (1). This impacts bioinformatics studies because researchers have devoted innumerable hours to identifying genes and molecular mechanisms that can lead to a treatment or a cure for this disease.

The goal of my research is to help identify significant genes with altered expressions that can help us identify treatment for this disease.

I hypothesize that there are significant differences in gene expression that will reveal new molecular pathways and targets involved in Alzheimer's disease and will help find a cure.

Researching a cure for Alzheimer's is crucial because it can prevent suffering, improve quality of life, and reduce the immense emotional and financial burden on families and healthcare systems. Advancing our understanding of the disease can lead to effective treatments and a cure that will benefit millions globally.

METHODS

Data Collection and Analysis of GEO2R Data

To begin the research process, I selected the dataset **GSE52022** from the National Center for Biotechnology Information (NCBI). This dataset was titled, "Genome-wide analysis of transcriptome and microRNAs in early stage of Alzheimer's disease (mRNA)," it was created to explore mRNA and miRNA expression changes in early-stage Alzheimer's disease using Tg6799 AD Model mice.

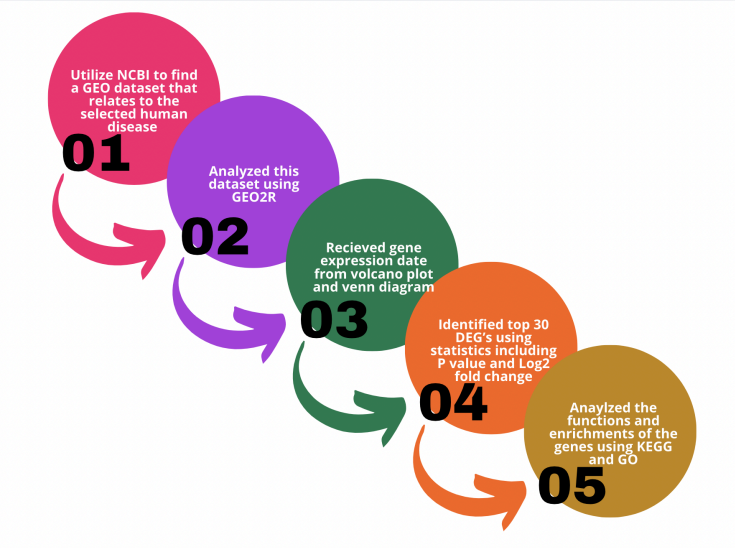


Figure 1: Research Methodology: Provides a summary of the procedures used in this study.

Identification of the Top Differentially Expressed Genes

To identify the most significant differentially expressed genes (the top 30 or 40), statistical analysis was applied. This process used p value and log fold change to prioritize the most important genes based on their differential expression across samples.

Data Analysis Using SRPlot, KEGG, and GO Bioinformatics Tools

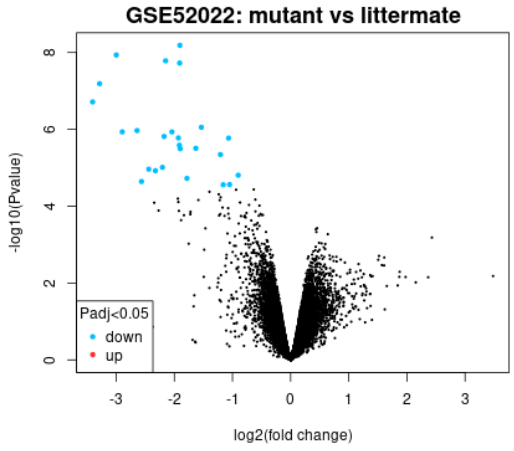
Then SRPlot (7), KEGG (8), and GO (9) bioinformatics tools and databases were utilized to analyze the functions of these top genes. SRPlot is a bioinformatics tool that provides a wide variety of plots and different graph types for scientists to analyze (7). The KEGG bioinformatics tool that organizes information regarding genes and their behaviors (8). The GO tool is another large resource meant for cataloging genes and their functions (9). Altogether, these tools helped uncover the potential roles of the genes in Alzheimer's Disease.

RESULTS

Identification of Differentially Expressed Genes

The GEO2R bioinformatics tool was utilized to determine differentially expressed genes (DEGs). The visualization of the analysis results was in the form of a Venn diagram and a volcano plot. Based on the results from the volcano plot, genes that were expressed differently between the two groups, mutant and littermate, can be identified.

In this plot, the blue dots represent genes that were downregulated; and less active. The black dots represent genes that did not display significant differences in expression. In this case, there are no red dots, but if there were, the red dots would be a representation of genes that are upregulated; and more active. Originally, the Venn diagram produced a sum of 45076 genes. Furthermore, 25 genes overlapped each other within the sample groups.



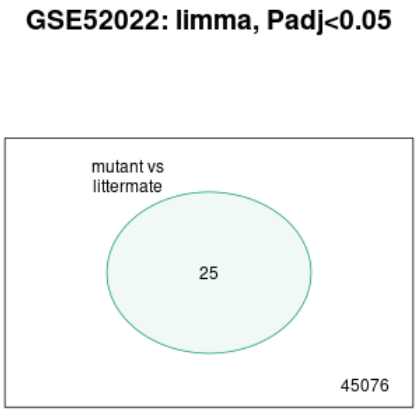


Figure 2: Differentially Expressed Genes: The volcano plot shows us how genes are expressed differently between the two groups. while the Venn diagram shows if any genes overlap or if they are different between the two groups.

Identification of (30 or 40 or 50) Statistically Significant Differentially Expressed Genes (DEGs)

To narrow down the genes, genes with a P-value greater than 0.5 were excluded, and fold change values were utilized to identify upregulated and downregulated genes. Based on the statistical methods used, P-value and fold change, there were 30 top Differentially Expressed Genes (DEGs) from the total of 45102. Out of the 30 DEGs, there were 15 upregulated genes and 15 downregulated genes. Gene Dataset- Alzheimers

Potential Functions and Enrichment of the Identified Genes and/or pathways

To determine the potential functions of the genes, the SR Plot enrichment tool was utilized. From the KEGG results, I received an overall pathway analysis and I identified a tuberculosis pathway. The pathway analysis KEGG result indicated that Tuberculosis (TB) was the most prominent In Alzheimer's disease (AD). The bacterium that causes TB can result in inflammation in areas of the nervous system as well as prolonged immune responses, which contribute to the progression of AD. Through the Tuberculosis path, significant genes including Dectin-1, CALM, and CathepsinS were identified.

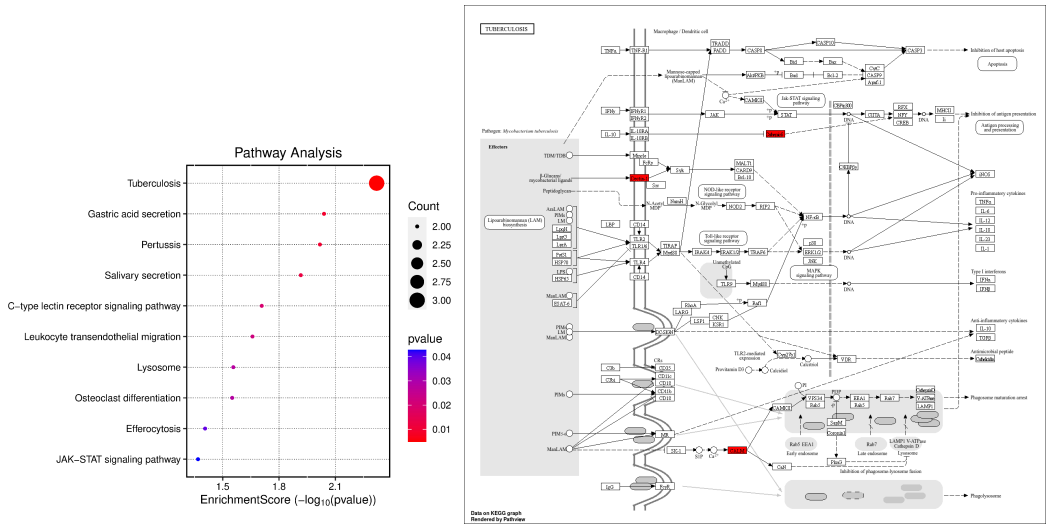


Figure 3: Functional Analysis from KEGG: The KEGG Pathway analysis results indicate the biological pathways that are significantly enriched within a set of genes. In this case, Tuberculosis is extremely prominent. In the Tuberculosis Pathway, the

significant genes within the pathway are highlighted in red. These genes include Dectin-1, CALM, and CathepsinS.

From the GO results of the Three Ontologies, the bar graph shows three ontologies and how enriched they are. The blue represents Molecular Functions (MF), the green represents Cellular Components (CC), and the orange represents Biological processes (BP). In this display, BP is slightly more enriched than the other two ontologies. The most significantly enriched BP was 'import into cell.' The cnet plot provided is used to visualize the relationship between genes and their associated pathways.

This cnet plot displays the most significant genes in red, which represents a high-fold change. From this graph two main genes are highlighted: TREM2 and TYROBP.

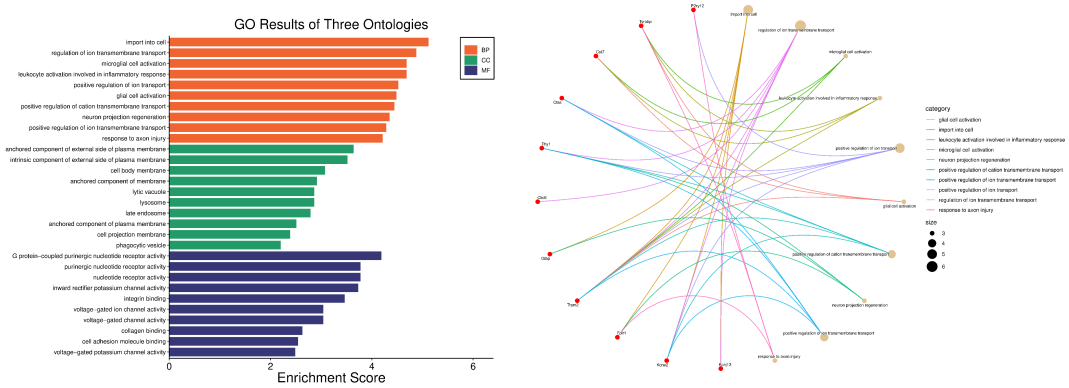


Figure 4: BP, CC, and MF are the three ontologies that are included in this graph. The graph reveals that the top 30 DEGS are most enriched in BP and MF. Overall, the enrichment scores were greater than or around 3, (middle of the x-axis). The Cnet plot is a display of key Biological Process genes and their relation to specific functions and activities. The red dots represent significant, upregulated genes.

DISCUSSION

Summary of Findings

The main goal of this research was to identify significant genes that could be used as potential targets for a cure or effective treatment. Results from the analyses conducted in GEO2R and SR plots helped to define key genes that showed high enrichment and were involved in prevalent pathways.

Interpretation of Results

The key genes that were identified in the graphs could potentially be used as targets for AD treatments. TREM2, TYROBP, and Dectin-1, also known as CLEC7A (Figures 5 and 7) play key roles in AD.

Comparison with Previous Studies

TREM2 was a significant gene that appeared in the BP cnet plot (Figure 7). This gene plays a crucial role in a variety of things, including positive regulation of ion transmembrane transport, positive regulation of cation transmembrane transport, glial cell activation, positive regulation of ion transport, and much more. Enhancing TREM2 signaling could result in a cure for AD. One distinguishing characteristic of AD is a lack of TREM2 signaling or a decreased efficiency because it impairs microglial function that is meant to clear amyloid-beta plaques (10). Moreover, if therapies that enhanced TREM2 signaling were engineered, researchers could potentially improve microglial phagocytosis of amyloid-beta which would reverse the progression of this disease.

TYROBP was another key gene that occurred in the BP cnet plot (Figure 7). It interacts with TREM2 and plays a crucial role in glial cell activation. Targeting its role in immune response regulation could offer a potential cure for AD. TYROBP is heavily involved in enhancing the phagocytic activity of microglia which helps to clear amyloid-beta peptides and apoptotic neurons, which contribute to the progression of AD. Moreover, when there is neuroinflammation, TYROBP plays a role by restraining it by reducing the production of microglia-mediated cytokine (11). Because these actions can protect against neuroinflammation and plaque building, factors that contribute to AD, enhancing TYROBP's role could help reverse disease progression.

Dectin-1 (CLEC7A) is a receptor that is activated around amyloid plaques. Normally, microalgia like CLEC7A is helpful to clear waste and protect the brain. In this case, it is too activated, which worsens AD progression. Researchers found that CLEC7A microalgia in model mice contributed to neuroinflammation and blood vessel damage. When these microalgia were blocked, blood vessels were shielded and inflammation decreased significantly (12). This evidence suggests that blocking or controlling CLEC7A when it is too activated could help reduce brain inflammation, which would help treat AD. For instance, treatments that limit CLEC7A activity could be a potential solution to AD.

Table 1. Summary of Key Genes Identified and their Connection to Alzheimer's Disease

Gene	Function	Gene Name	Connection
TREM2	-Contributes to microagial activation -Regulates inflammation -Neuroprotection	Triggering Receptor Expressed on Myeloid Cells 2	Clears amyloid-beta plaques that play a part in AD
TYROBP	-Helps with glial cell activation -Enhances the phagocytic activity of microalgia -Exhibits inflammatory response regulation	TYRO3, AXL, and MER tyrosine kinase binding protein	Restrains neuroinflammation by reducing the production of microglia-mediated cytokine
CLEC7A	-Helps microaglia become overactive and harmful -Leads to Inflammation -Can cause Blood Vessel Damage	C-type lectin domain family 7 member A	If regulated, the microglial ability to clear amyloid beta plaques and reduce neuroinflammation will be enhanced.

Implications

Regulating TREM2 and TYROBP could enhance microglial activity to clear amyloid-beta plaques, while controlling CLEC7A may reduce harmful inflammation and amyloid beta plaques. These findings could lead to targeted drugs, gene therapies, or early interventions to slow or prevent Alzheimer's disease.

Limitations

In our case, since we use **bioinformatics datasets** from microarray experiments conducted by other researchers, one limitation is that the identified genes will need to be further studied in the laboratory or clinical environment before developing a cure.

Future Directions

The identified genes can be tested in the laboratory by scientists who can conduct clinical trials or other studies to determine if any treatments or a cure could be determined using the pathways we have identified.

REFERENCES

1. Lane, C.A., Hardy, J. and Schott, J.M. (2018), Alzheimer's disease. Eur J Neurol, 25: 59-70. https://doi.org/10.1111/ene.13439
2. Selkoe DJ. Defining Molecular Targets to Prevent Alzheimer Disease. *Arch Neurol.* 2005;62(2):192–195. doi:10.1001/archneur.62.2.192
3. Blennow K, de Leon MJ, Zetterberg H. Alzheimer's disease. *Lancet*. 2006;368(9533):387-403. doi:10.1016/S0140-6736(06)69113-7.
4. Khanna S, Domingo-Fernández D, Iyappan A, Emon MA, Hofmann-Apitius M, Fröhlich H. Using multi-scale genetic, neuroimaging and clinical data for predicting Alzheimer's disease and reconstruction of relevant biological mechanisms. *Sci Rep*. 2018;8:11173. doi:10.1038/s41598-018-29433-3.
5. Tanya Barrett, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, Maxim Tomashevsky, Kimberly A. Marshall, Katherine H. Phillippy, Patti M. Sherman, Michelle Holko, Andrey Yefanov, Hyeseung Lee, Naigong Zhang, Cynthia L. Robertson, Nadezhda Serova, Sean Davis, Alexandra Soboleva, NCBI GEO: archive for functional genomics data sets—update, *Nucleic Acids Research*, Volume 41, Issue D1, 1 January 2013, Pages D991–D995, https://doi.org/10.1093/nar/gks1193
6. Reiman EM, Jagust WJ. Brain imaging in the study of Alzheimer's disease. *Neuroimage*. 2012;61(2):505-516. doi:10.1016/j.neuroimage.2011.11.075.
7. Tang, D., Chen, M., Huang, X., Zhang, G., Zeng, L., Zhang, G., Wu, S., & Wang, Y. (2023). SRplot: A free online platform for data visualization and graphing. *PLOS ONE*, *18*(11), e0294236. https://doi.org/10.1371/journal.pone.0294236
8. Kanehisa, M. The KEGG Database. 91-103. https://doi.org/10.1002/0470857897.ch8
9. Louis du Plessis, Nives Škunca, Christophe Dessimoz, The what, where, how and why of gene ontology—a primer for bioinformaticians, *Briefings in Bioinformatics*, Volume 12, Issue 6, November 2011, Pages 723–735, https://doi.org/10.1093/bib/bbr002
10. Ulrich JD, et al. Elucidating the role of TREM2 in Alzheimer's disease. Neuron. 2017;94(2):237-248.
11. Ma, J., Jiang, T., Tan, L. *et al.* TYROBP in Alzheimer's Disease. *Mol Neurobiol* 51, 820–826 (2015). https://doi.org/10.1007/s12035-014-8811-9
12. Shi, H., Yin, Z., Koronyo, Y. *et al.* Regulating microglial miR-155 transcriptional phenotype alleviates Alzheimer's-induced retinal vasculopathy by limiting Clec7a/Galectin-3^{+} neurodegenerative microglia. *acta neuropathol commun* 10, 136 (2022). https://doi.org/10.1186/s40478-022-01439-z