Unveiling common hub genes in Dementia and Alzheimer's disease: Exploring Correlation & Functional role through Computational Analysis

Submitted by:

Nimra Ilyas

63-FOC/MSBI/F22

Supervised by:

Dr. Mehrosh Khalid

Department of Bioinformatics,
Faculty of Computing and Information Technology,
International Islamic University Islamabad.
2025


بسم الله الرحمن الرحيم

A thesis submitted to Department of Bioinformatics,
International Islamic University, Islamabad as a partial fulfillment of requirement for the award of the degree of M.S in Bioinformatics.

DEDICATION

This thesis is lovingly dedicated to my parents, whose unwavering support and heartfelt prayers have been the foundation of my journey and accomplishments. I also wish to express my sincere gratitude to my supervisor, Dr. Mehrosh Khalid, for their generous guidance, time, and encouragement throughout this project. Their insightful direction and constant moral support have been instrumental in bringing this work to fruition.

DECLARATION

We hereby declare that this thesis “Unveiling common hub genes in Dementia and Alzheimer’s disease: Exploring Correlation & Functional role through Computational Analysis” neither as a whole nor as a part has been copied out from any source. It is further declared that we have done this research with the accompanied report entirely based on our personal efforts, under the proficient guidance of our teachers, especially our supervisor Dr. Mehrosh Khalid. If any part of the system is proved to be copied out from any source or found to be reproduction of any project from any of the training institute or educational institutions, we shall stand by the consequences.

Ms. Nimra Ilyas
63-FOC/MSBI/F22

ACKNOWLEDGMENTS

I bow in deep reverence to "Almighty ALLAH", the most merciful, most beneficent, the most adorable, deity and paramount to whom I am indebted. First and foremost, I acknowledge my limitless thanks to Allah for His help and bless.

I would like to express my gratitude to my supervisor Dr. Mehrosh Khalid for her consistent support, encouragement, guidance and feedback throughout the project. Her patience and considerate nature made her accessible whenever I needed her assistance.

I would also like to thank my parents who always supported and encouraged me with their best wishes. Their love, patience and sacrifice are vested in every page of this study.

I extend my thanks to everyone who was concerned and has co-operated in completing this thesis.

THANK YOU ALL!

Nimra Ilyas

63-FOC/MSBI/F22

PROJECT IN BRIEF

Project Title:	Unveiling common hub genes in Dementia and Alzheimer's disease: Exploring Correlation & Functional role through Computational Analysis
Organization:	Department of Bioinformatics Faculty of computing and Information Technology International Islamic university H-10, Islamabad
Undertaken By:	Nimra Ilyas 63-FOC/MSB/F22
Supervised By:	Dr.Mehrosh Khalid
Start Date:	2023
Completion Date:	2025
Objective:	The aim of this study is to identify common hub genes in Dementia and Alzheimer Disease through computational analysis
Tools & Technologies:	RStudio, GEO2R, GENECARD, STRING, DISGENET.
Documentation Tool:	MSWord, MS PowerPoint, Mendeley
Operation System:	Windows 11
System Used:	Lenovo core i5

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

Acronym	Abbreviations
AD	Alzheimer Disease
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
DEGs	Differently Expressed Genes
MANCOVA	Multivariate Analysis of Covariance
LASSO	Least Absolute Shrinkage and Selection Operator
FDA	Food and Drug Administration
PET	Positron Emission Tomography
MoCA	Montreal Cognitive Assessment
MMSE	Mini-Mental State Examination
CT	Computed Tomography
CAS	Carotid Atherosclerosis
VaD	Vascular Dementia
DLB	Dementia with Lewy Bodies
WGCNA	Weighted Gene Co-Expression Network Analysis
EDA	Exploratory Data Analysis
NDDs	Neurodegenerative diseases
AUC	Area Under Curve
PPI	Protein-Protein Interaction
NCBI	The National Center for Biotechnology Information
GEO	Gene Expression Omnibus
CC	Cellular Component
BP	Biological Process
NGS	Next Generation Sequencing
MF	Molecular Function
WES	Whole Exome Sequencing

Abstract

Neurodegenerative disorders such as dementia and Alzheimer's disease (AD) share overlapping molecular mechanisms, yet the identification of common regulatory genes remains limited. In order to identify potential hub genes implicated in both conditions, we used a powerful computational pipeline in this study. Finding common pathogenic mechanisms and possible therapeutic targets is made possible by the combination of transcriptomic profiling and computational techniques. Gene expression datasets collected from public repositories were normalized and quality controlled using R's affyPLM package. Following preprocessing, differentially expressed genes (DEGs) that were significantly associated with both disorders were found using analysis of covariance (MANOVA). LASSO regression, which effectively penalizes irrelevant variables and enhances model interpretability, was performed using the glmnet package to decrease the number of candidate genes. Biological pathways and processes that were enriched among the DEGs were identified through functional enrichment analysis using the ClusterProfiler package.

This integrative approach led to the discovery of SOCS6 (Suppressor of Cytokine Signaling 6) as a common hub gene that is strongly associated with both dementia and AD. ROC curve analysis in RStudio also confirmed the diagnostic utility of SOCS6 by demonstrating the degree to which it differentiated between disease and control samples. SOCS6's role in the JAK/STAT signal transduction pathway, an important axis responsible for controlling inflammation, immune response, and cell survival, was also confirmed by pathway enrichment and literature data. The chronic neuroinflammation in AD and dementia could be induced by dysregulation of SOCS6, an inhibitor of this pathway.

The SOCS6 control of cytokine signaling is forecast to have significant impacts on the pathophysiology of neurodegenerative disorders by modulating synaptic plasticity, neuron survival, and microglial activation. This study presents SOCS6 as a novel and potential biomarker for neurodegeneration by integrative analysis with gene expression, machine learning, and pathway enrichment strategies. The role of SOCS6 in the JAK/STAT pathway provides a mechanistic explanation of how immunological dysregulation may be a root cause of dementia and Alzheimer's disease. These findings establish a foundation for subsequent experimental verification and therapeutic research to target SOCS6 in an effort to prevent or decelerate neurodegeneration.

Chapter 1
Introduction

In this section, we delve into the introduction of Alzheimer's and Dementia diseases, tracing their historical context, outlining the aims and objectives of our study, and proposing specific objectives to address the challenges posed by these diseases.

1.1 Dementia

Dementia refers to a broad range of symptoms affecting cognition, emotions, behavior, and physical health that are brought on by structural alterations in the brain brought on by degenerative illnesses or traumas for which there is currently no known treatment [1]. The primary cognitive symptoms include impaired verbal and visuospatial skills, diminished memory function, lower concentration, issues with planning and organization, and increased disorientation. Mood swings, impatience, and elevated anxiety are examples of emotional shifts [2]. Along with physical changes in food, sleeping patterns, and decreased mobility, dementia is known to cause behavioral changes such as increased hostility, roaming, and persistent questioning. One of the main causes of disability in older people is dementia, which can have a very severe social and emotional effect on those who have it, their care, and other family members [3].

There is currently no known treatment for dementia, although early detection and effective care can help both the individuals who are caring for the person with dementia and themselves live better lives. Medication to control symptoms, behavioral and cognitive therapy, and support services are all possible forms of treatment [4]. It's critical to remember that dementia is not a typical aspect of aging. Even while dementia becomes more likely as people age, many older folks can retain their cognitive abilities long into old age.

1.1.1 Symptoms of Dementia

The symptoms of dementia can vary widely, but common signs include:

1. Memory loss, especially short-term memory.
2. Difficulty in communicating or finding the right words.
3. Difficulty in reasoning or problem-solving.
4. Difficulty in handling complex tasks.
5. Confusion and disorientation.
6. Personality changes.
7. Depression and anxiety.

1.1.2 History of Dementia

Throughout ancient history, various cultures documented symptoms of dementia in their literature, though understanding of the condition as a distinct medical disorder emerged only in the late 19th and early 20th centuries [5]. As medical science advanced in the 20th century, researchers began to distinguish between different forms of dementia, including Alzheimer's disease, vascular dementia, and less common variants [6]. From the late 20th century to the present, dementia has undergone profound changes in diagnosis, treatment, and classification, driven by advancements in neuroscience, genetics, and imaging technology. This period has witnessed a shift towards a deeper understanding of the diverse types of dementia and their underlying mechanisms [7]. Moreover, there has been a surge in global awareness and advocacy efforts surrounding dementia, leading to increased research funding, activism, and the establishment of global initiatives aimed at improving dementia support services and care worldwide [8].

1.2 Alzheimer Disease

The most common cause of dementia in older adults is Alzheimer's disease (AD). It is a neurological condition that worsens over time. Dr. Alois Alzheimer originally described Alzheimer's disease in 1906. The disease gradually impairs memory and cognitive function, making it harder to carry out daily duties. [9].

Although the precise etiology of the condition is unknown, a combination of environmental, behavioral, and genetic factors that gradually damage the brain are thought to be responsible. The development of tangles of tau inside neurons and the buildup of beta-amyloid plaques outside of neurons are the two main characteristics of Alzheimer's disease [10]. The brain cells eventually die as a result of these alterations that impair communication between them.

Alzheimer's disease does not currently have a cure, but some therapies can help control symptoms and enhance quality of life [11]. In addition to non-pharmacological methods like cognitive stimulation and behavioral therapy, these therapies may involve drugs to temporarily enhance cognitive performance or control behavioral problems. The main goals of ongoing research are to comprehend the fundamental causes of Alzheimer's disease, provide new instruments for diagnosis, and investigate possible cures and preventative measures [12]. The best use of current medicines and support services depends on early diagnosis and action.

Alzheimer's patients and their families, as well as those who care for them, can experience significant effects from the illness [13]. Navigating the difficulties posed by this degenerative condition requires knowledge, assistance, and access to resources.

1.2.1 Symptoms of Alzheimer's Disease

Some common symptoms associated with Alzheimer's disease:

1. Misplacing Items
2. Mood and Personality Changes
3. Loss of Initiative
4. Difficulty with Spatial Relationships
5. Hallucinations and Delusions

1.2.2 History of Alzheimer's Disease

Since its initial observation in 1906 by German psychiatrist and neurologist Dr. Alois Alzheimer, the understanding of Alzheimer's disease has undergone significant evolution. Dr. Alzheimer's study of Auguste Deter marked the first documented case of the illness, characterized by symptoms such as disorientation, memory loss, and language difficulties [14]. Throughout the early to mid-20th century, Alzheimer's remained relatively obscure until the latter half of the century when scientific advancements, particularly in neuropathology and medical imaging, began shedding light on its distinct nature as a form of dementia [15]. By the 1970s and 1980s, significant strides were made in identifying the degenerative changes within the brain associated with the disease, notably the discovery of tau protein tangles and amyloid-beta plaques [16]. From the 1990s to the present day, research efforts intensified, focusing on unraveling the underlying mechanisms of Alzheimer's, developing diagnostic tools, and exploring potential treatments [17]. While a cure remains elusive, medications have been developed to enhance quality of life and temporarily alleviate symptoms. Despite ongoing challenges, the collective efforts of researchers continue to deepen our understanding and offer hope for future breakthroughs in the fight against Alzheimer's disease [18].

1.3 Dementia vs Alzheimer's Disease

Dementia is a syndrome defined by a deterioration in cognitive function that impairs thinking, memory, and reasoning. Memory loss, language problems, poor judgment, and behavioral abnormalities are some of the symptoms of dementia [19]. It includes a range of underlying causes, such as Lewy body dementia, vascular dementia, Alzheimer's disease, and others. While the symptoms and course of each type of dementia vary, they all have cognitive impairment that interferes with day-to-day functioning [20]. On the other hand, Alzheimer's disease pertains particularly to a neurodegenerative condition characterized by the gradual loss of brain cells as a result of aberrant protein accumulations such as tau tangles and amyloid-beta plaques [21]. Memory loss, linguistic challenges, poor judgment, and personality changes are typical signs. Alzheimer's disease is a neurological ailment with unique pathological symptoms, diagnosis techniques, and treatment approaches, whereas dementia is a syndrome [22].

1.4 Prevalence of Dementia and AD

Due to their increasing incidence, especially in older persons, dementia and Alzheimer's disease (AD) have become major global public health concerns. The number of individuals worldwide believed to have dementia in 2020 was 50 million, and it was predicted that this estimate would nearly double every 20 years to reach 82 million by 2030 and 152 million by 2050 [23]. The prevalence of people with AD from 2020 to 2050 is shown in Figure 1.1. People with dementia are affected by dementia at varied rates in different countries and locations. About 60–80% of dementia cases are caused by Alzheimer's disease, making it the most prevalent dementia cause. Worldwide, 5-7% of those 60 and older were predicted to have Alzheimer's disease in 2020 [24]. The bulk of occurrences occur in those 65 years of age and older, and the prevalence rises with age. These rates of prevalence highlighted the critical need for more education, funding, and research to address the growing effects of dementia and Alzheimer's disease on people, families, healthcare systems, and societies across the globe [25]. The most accurate and recent information on the global incidence of dementia and Alzheimer's disease can be found by consulting more recent data from reliable sources like the World Health Organization (WHO) or Alzheimer's associations [26]. It's important to keep in mind that these numbers may have changed since my last update. The Global Prevalence of AD and Dementia by age is shown in Figure 1.2.

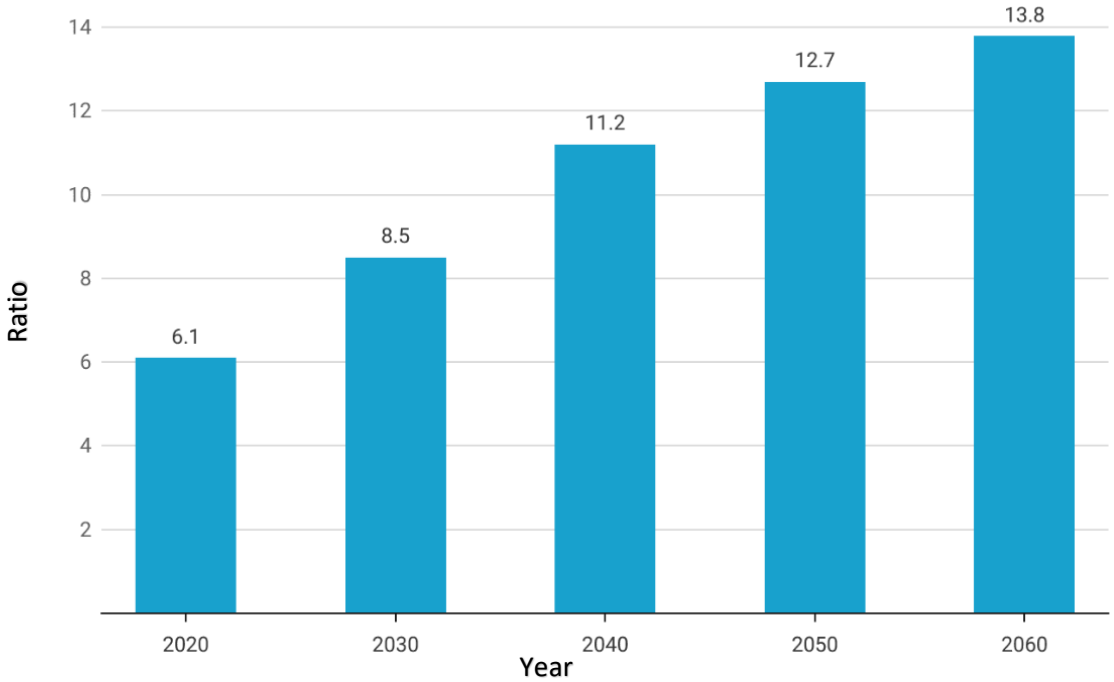

Figure 1.1: The Prevalence of people (2020 – 2050) in millions with AD [27]

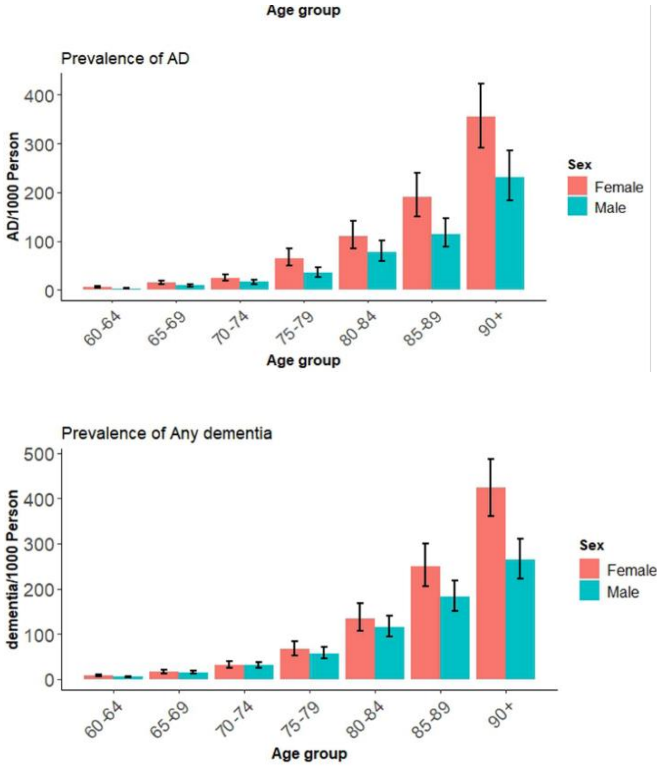

Figure 1.2: Global Prevalence of AD and Dementia by Age [28].

1.5 Diagnosis

A dementia diagnosis can be determined through a thorough evaluation that takes into account the patient's medical history, cognitive tests, neurological examinations, and occasionally imaging scans [29]. The process includes the following steps.

1.5.1 Cognitive Assessment

A variety of tests and assessments are available to measure memory, attention, language, and other cognitive abilities.

1.5.2 Imaging Studies

Brain imaging methods, such as computed tomography (CT) scans or magnetic resonance imaging (MRI), can be used to identify anatomical alterations in the brain, such as shrinkage or the presence of tumors, and strokes.

1.5.3 Neuropsychological Testing

Comprehensive neuropsychological testing can be carried out to evaluate particular cognitive abilities and offer further details regarding the type and degree of cognitive impairment.

1.5.4 Analysis of Cerebrospinal Fluid:

A lumbar puncture, often known as a spinal tap, may be necessary in some circumstances to examine brain fluid for biomarkers of Alzheimer's disease, including tau and amyloid-beta proteins.

1.6 Drug Treatment

The main objectives of Alzheimer's and dementia disease treatment are to reduce the rate of the illness's progression and enhance quality of life. Alzheimer's disease and the majority of other dementias have no known cure, although there are strategies that can assist manage particular symptoms and promote general well-being [30].

Although there is still no known treatment for AD, some medications have been shown to reduce symptoms and decrease the disease's progression. Physicians classify symptoms into "cognitive" and "behavioral and psychiatric" categories when starting treatment for AD patients [31]. This makes it possible to provide therapy that is tailored to the individual's symptoms. Cognitive

symptoms impact thought processes, judgment, language, and memory. A patient's behavior and emotions are changed by behavioral symptoms [32].

1.6.1 Treatment for Cognitive Impairment

In treating cognitive disorders, the impact of chemical messengers in the brain is modified. For this purpose, the Food and Drug Administration (FDA) has approved two types of drugs.[33]. The first kind is known as a cholinesterase inhibitor, and it works by blocking the enzyme that breaks down acetylcholine in the brain. An essential neurotransmitter for memory and learning is acetylcholine. Periodic amnesia is brought on by a modest drop in acetylcholine levels that occurs with normal aging [34]. On the other hand, in AD, the concentration might drop as much as 90%, which causes a notable impairment in memory and behavior. These medications work by facilitating neural cell-to-neuron transmission, which raises acetylcholine levels. Donepezil, galantamine, and rivastigmine are the three cholinesterase inhibitors that are currently often prescribed [35].

Apart from cholinesterase inhibitors, memantine is an additional medicine licensed for the treatment of AD. Glutamate activity in the brain is regulated by memantine. An excitatory neurotransmitter involved in memory and learning is glutamate [36]. The neuron degeneration observed in AD may be brought on by glutamate overstimulation of nerves, which is known as excitotoxicity. N-methyl-D-aspartate (NMDA) receptors on the surface of brain cells are bound by glutamate. Memantine works by inhibiting NMDA receptors, which shields the nerves from too much glutamate excitation [37]. Memantine can temporarily halt the progression of cognitive symptoms and is recommended for the treatment of moderate to severe AD.

1.6.2 Treatment for Behavioral and Psychiatric Issues

AD and Dementia can result in significant behavioral and mental symptoms in addition to cognitive and functional impairment. Anxiety, insomnia, agitation, hallucinations, and delusions are some of these symptoms [42]. Medication to address the symptoms could be used in addition to non-drug therapies as potential therapy strategies. Changing the surroundings to remove barriers and boost security is a successful non-drug strategy [43]. Examining any possible drug interactions that can harm the patient's behavior or mental health is an additional option. Medication can be necessary if these therapies are ineffective in treating the symptoms. Depending on the symptoms,

several drugs may be selected. For instance, an antidepressant like Prozac or Zoloft may be administered if the patient is depressed. It is possible to take antipsychotics and anxiolytics to lessen hallucinations and anxiety [44].

1.7 Research Gap

Although progress has been made in understanding neurodegenerative diseases such as dementia and Alzheimer's disease (AD), substantial research gaps remain. There is limited application of multivariate analysis of covariance (MANCOVA) to explore complex interactions among clinical and molecular factors. Additionally, the identification of common hub genes using advanced computational techniques remains underexplored, and machine learning methods for predicting the common gene are largely unexplored. Addressing these gaps could enhance the identification of molecular mechanisms of common gene.

1.8 Problem Statement

Dementia and Alzheimer’s disease (AD) are prevalent neurodegenerative disorders with complex molecular mechanisms that remain largely unexplored, hindering effective therapeutic development. Computational approach analyzing gene expression data in these conditions often rely on univariate methods like Limma, lacking comprehensive integration of multivariate approaches such as multivariate analysis of covariance (MANCOVA) and advanced machine learning techniques. This limited scope restricts the identification of common hub genes and impedes a deeper understanding of shared gene expression patterns.

1.9 Research Questions

The research questions, which will address in this study are

1. What are the common hub genes involved in both dementia and AD, and how do they contribute to the molecular mechanisms underlying these neurodegenerative disorders?
2. How can multivariate analysis of covariance (MANCOVA) and machine learning algorithms, such as LASSO regression, be applied to transcriptomic data to identify key molecular pathways and potential therapeutic targets?
3. How do common hub genes identified in dementia and Alzheimer’s disease influence the biological pathways responsible for neurodegeneration?

1.10 Aim and Objectives

This research aim to identify common hub genes involved in both dementia and Alzheimer's disease and to explore their molecular significance using advanced computational techniques.

The objectives of this study are:

- Integrate bulk tissue and single-cell transcriptomic data to identify DEGs and common hub genes shared between dementia and Alzheimer's disease.
- Employ MANCOVA and LASSO regression to analyze the relationship between hub gene expression and clinical/demographic factors in dementia and Alzheimer's disease.
- Explore the functional role of common hub gene in molecular mechanism of dementia and AD.

1.11 Proposed Solution

This study uses MANCOVA and machine learning (LASSO regression) to identify common hub genes between dementia and Alzheimer's disease, focusing on neuroinflammation and synaptic dysfunction. By analyzing transcriptomic data, it aims to uncover key pathways and predict drug candidates for improved diagnostics and therapies.

1.12 Proposed Methodology

In the research study, machine learning (ML) techniques play a crucial part in deciphering intricate biological information linked to these neurodegenerative conditions. In high-dimensional datasets, ML approaches enable the selection of features and dimensionality reduction, which in turn allows the identification of pertinent genetic and clinical variables. A statistical method MANCOVA is used to find the connection between many dependent and several independent variables while accounting for one or more covariates. MANCOVA may be used as a component of the analysis pipeline in the context of identifying Differentially Expressed Genes (DEGs) in the research to enhance the accuracy of DEG identification. Supervised learning models like LASSO help with predictive modeling and classification within Alzheimer's and dementia disease populations. To find common genes linked to Alzheimer's and dementia disease while minimizing overfitting and lowering the possibility of false positives, LASSO can be used as a component of the analysis pipeline. While removing genes with weak or insignificant effects, LASSO can assist in identifying a group of genes that are most closely associated with the illness phenotype. The working flow of

the proposed study is mentioned in Figure 1.3.

Functional annotation and pathway analysis illuminate the molecular roles of genes, providing direction for future research and potential therapeutic approaches. By utilizing machine learning technology, the research advances our knowledge of Alzheimer's disease and dementia, which may lead to better methods for diagnosing, treating, and predicting these difficult neurological illnesses.

1.13 Scope & Limitations

The study's objectives include identifying common hub genes that are dysregulated in both illnesses, investigating the relationships between these genes and pathological and clinical aspects, and using computational analysis to find the functional roles of these genes. Through the examination of extensive datasets that comprise genetic, clinical, and pathological data, the study seeks to identify possible biomarkers and targets for the treatment of these intricate neurodegenerative conditions. Finding shared hub genes could help us better understand how diseases work and open the door to the creation of innovative diagnostic techniques and focused treatments.

However, while evaluating the study's conclusions, several limitations need to be taken into account. Firstly, data availability and quality are critical since computational analysis greatly depends on the accuracy and comprehensiveness of the datasets utilized. Furthermore, the heterogeneity of individuals with dementia and Alzheimer's disease is a problem because variations in patient characteristics and disease subtypes may affect how broadly applicable the findings are. Confirming the functional importance of identified hub genes and their relevance to disease pathophysiology requires biological confirmation of computational findings. Furthermore, whereas correlation analysis might show links between hub genes and symptoms of an illness, further experimental research is needed to prove causation. Thorough validation in human populations via clinical trials and longitudinal investigations is also necessary for the therapeutic translation of computational findings.

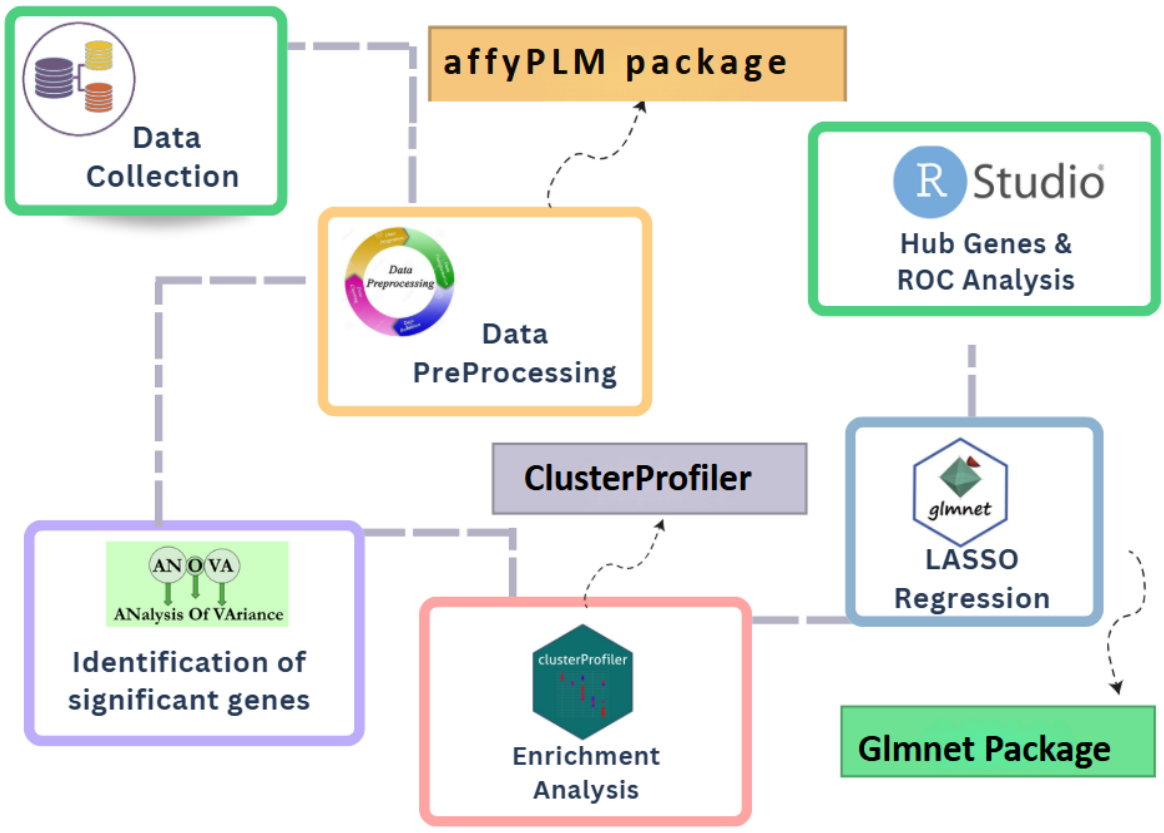

Figure 1.3: The workflow of this study. It includes data collection from the GEO database, gene enrichment, and significance analysis, LASSO regression for gene screening, and hub gene identification.

Chapter 2

Literature Review

In this section, previous research on the identification of hub genes associated with AD and dementia are discussed:

2.1. Bioinformatics identification of potential biomarkers and therapeutic targets in carotid atherosclerosis and vascular dementia [45]**.** 60 upregulated and 159 downregulated DEGs were found in this study's bioinformatics analysis of differentially expressed genes (DEGs) in the carotid atherosclerosis (CAS) and vascular dementia (VD) datasets (GSE43292 and GSE122063). Notably, seven hub genes were shown to have substantial connectivity: TYROBP, CSF1R, C1QB, ITGB2, LY86, and FCGR3A. All of these genes displayed downregulation in both CAS and VD. Immune and inflammatory responses, as well as phagosome and osteoclast differentiation pathways, were implicated by functional enrichment analysis. These results provide molecular insights into the development of CAS and VD and point to possible biomarkers and treatment targets for additional experimental confirmation.

2.2. Identification of key genes and signaling pathways associated with dementia with Lewy bodies and Parkinson's disease dementia using bioinformatics [46]**.** The aim of this study was to identify the molecular mechanisms that distinguish Parkinson's disease dementia (PDD) from dementia with Lewy bodies (DLB), also known as Lewy body dementia (LBD). By examining the GSE150696 dataset, the researchers discovered 1,864 differentially expressed genes (DEGs) in PDD and DLB. These DEGs identified pathways associated with vesicle localization, neurodegeneration, glycerolipid metabolism, and immunological signaling. Weighted gene co-expression network analysis (WGCNA) was used to identify modules associated with specific LBD subtypes; the blue module showed a substantial positive correlation with PDD, while the yellow module did the same with DLB. The integration of DEGs and WGCNA revealed seven hub genes: SNAP25, GRIN2A, GABRA1, GRIA1, SLC17A6, and SYN1, which are associated with synaptic function and neurotransmitter regulation.

2.3. Identification of Molecular Signatures and Candidate Drugs in Vascular Dementia by Bioinformatics Analyses [47]**.** It aims to use bioinformatics techniques to find similar transcriptome patterns in the frontal and temporal cortex of vascular dementia (VaD). The study found 159 differentially expressed genes (DEGs) enriched in angiogenesis, synapse pruning, inflammation, innate immunity, and more by examining a microarray dataset (GSE122063). In

addition to possible medications (maraviroc, cenicriviroc, PF-04634817, efalizumab) for the treatment of VD, 10 hub genes were found, including GNG13, CD163, C1QA, SST, C1QB, CCR5, CRH, TLR2, ITGB2, and TAC1. The results shed light on the molecular mechanisms behind VD and suggest possible treatment targets and medications. Nevertheless, additional experimental verification is required to validate the functions of discovered genes and medications in VD.

2.4. Identification of the Hub Genes in Alzheimer's Disease [48]**.** The present difficulties in comprehending the pathophysiology of AD and the dearth of efficient therapies are discussed in this work. The work uses thorough bioinformatics analysis to improve knowledge of AD molecular processes. Targeted therapies for AD may be developed using the discovered hub genes, including GAPDH, RHOA, RPS29, and RPS27A. RT-qPCR validation and additional investigation of the identified genes in animal models are among the constraints. Future investigations to confirm and explore the processes of these important genes in AD are encouraged by the study.

2.5. A Comprehensive Analysis Identified Hub Genes and Associated Drugs in Alzheimer's Disease [49]**.** In order to find possible genes linked to Alzheimer's disease (AD), this study uses sophisticated bioinformatics to examine expression profiles from samples of the temporal lobe cortex and entorhinal cortex. Protein-protein interaction and functional enrichment studies were performed on the 158 differentially expressed genes (DEGs) that were found to be shared by both areas. FPR3, CXCL3, APLNR, GABRB2, GABRG2, and GABRA1 were the six important hub genes that were found. According to drug-gene interaction study, 26 currently available medications that specifically target GABA-A receptor subunits may be used to treat AD. Furthermore, a prediction model that was based on machine learning algorithms showed promise in identifying AD patients. These results highlight the significance of GABAergic dysfunction in the pathophysiology of AD and offer insights into possible treatment targets and medication candidates. In an effort to slow the progression of the disease and enhance patient outcomes, the study suggests a risk prediction model for early AD screening and treatments. Clinical trials and molecular tests are recommended for additional validation.

2.6. Identification of hub genes associated with cognition in the hippocampus of Alzheimer's Disease [50]**.** The study uses the profile of gene expression GSE48350 from the Gene Expression Omnibus (GEO) database to conduct a bioinformatics analysis in order to uncover hub genes linked

to cognitive dysfunction in the hippocampus of Alzheimer's disease (AD). 96 differentially expressed genes (DEGs) with a protein cluster associated with cognition and enriched in synapse-related alterations were found by the study. Additional verification in APP/PS1 mice verified cognitive impairment and downregulation of Calbindin 1 (CALB1), Tachykinin precursor 1 (TAC1), Cholecystokinin (CCK), and Cannabinoid receptor type 1 (CNR1). Expressed in GABAergic neurons, these hub genes are linked to neurotransmitter activity and synaptic function, indicating their potential as therapeutic targets and their role in the underlying mechanism of cognitive loss in AD. The study offers fresh perspectives on how to comprehend and manage cognitive deterioration brought on by AD.

2.7. Identification of Key Pathways and Genes in Dementia via Integrated Bioinformatics Analysis [51]**.** This study used high throughput sequencing data (GSE153960) from the Gene Expression Omnibus to do a thorough bioinformatics analysis in order to decipher the molecular complexities of dementia. Functional enrichment analyses were made possible by the discovery of 948 differentially expressed genes (DEGs) in dementia patients, of which 475 were upregulated and 473 were downregulated. Neutrophil degranulation, ion transport, defensive response, and the neural system were identified to be the main functions of the DEGs. Key players in the form of hub genes, miRNAs, and transcription factors linked to dementia were identified through the development of protein-protein interaction networks, as well as miRNA-hub gene and TF-hub gene regulatory networks. Notably, hub genes with promising diagnostic efficacy according to ROC analysis, including CDK1, TOP2A, MAD2L1, and NOTCH3, surfaced as possible biomarkers. By illuminating possible therapeutic targets and biomarkers for more research and validation, this study offers significant insights into the molecular foundations of dementia.

2.8. Identification of Hub Genes Related to Alzheimer's Disease and Major Depressive Disorder [52]**.** The authors of this study sought to identify the pathophysiological link between major depressive disorder (MDD) and Alzheimer's disease (AD). They discovered 171 differentially expressed genes (DEGs) associated with AD and 79 DEGs shared by AD and MDD using gene expression data from several datasets. For common genes, functional analysis showed enrichment in pathways linked to long-term depression signaling and circadian entrainment. They discovered five hub genes—DYNC1H1, TTBK2, ITGB1, MAPRE3, and WASL—through

network analysis that could be used as potential targets for AD and MDD diagnosis and treatment. Subsequent verification with a separate dataset revealed notable alterations in these hub genes' expression in MDD patients receiving venlafaxine, indicating their applicability in the context of antidepressant treatment.

The study highlights the possible significance of these hub genes in the intricate interactions between AD and MDD and offers insights into the molecular pathways that these two neurological illnesses share.

2.9. Role of Genetics and Epigenetics in the Pathogenesis of Alzheimer's Disease and Frontotemporal Dementia [53]**.** The intricate relationships between genetics and epigenetics in the etiology of Alzheimer's disease (AD) and frontotemporal dementia (FTD) are examined in this study. Genetic factors, which include mutations in genes like APP, PSEN1, and PSEN2 in AD and MAPT, GRN, and C9orf72 in FTD, are primarily responsible for familial types of AD and FTD. Sporadic AD is linked to genes like APOE. Examples of epigenetic processes that control gene expression and influence the significance of neurodegenerative pathways include histone modifications, DNA methylation, and non-coding RNAs. Environmental variables also influence these epigenetic modifications, potentially hastening the progression of the illness.

The study underlines the complexity of the pathophysiology of AD and FTD and the necessity of all-encompassing strategies that use genetic and epigenetic knowledge for improved diagnosis and therapeutic measures. This is because genetics and epigenetics interact delicately.

2.10. Alzheimer's polygenic risk scores, APOE, Alzheimer's disease risk, and dementia-related blood biomarker levels in a population-based cohort study followed over 17 years [54]**.** The association between AD risk factors, including the APOE gene and polygenic risk scores (PRS), and dementia-related blood biomarkers is investigated in this 17-year population-based cohort study. Through the use of longitudinal data, the study aims to elucidate the relationship between genetic susceptibility as measured by PRS and APOE genotype and the risk of developing AD, as well as its association with specific blood biomarkers suggestive of dementia. By analyzing a diverse cohort over an extended period of time, the study aims to discover potential biomarker signatures associated with AD pathology and their relationship to genetic risk factors.

The investigation may reveal associations between the risk of Alzheimer's disease or dementia and blood biomarker levels, the APOE genotype, and Alzheimer's PRS during the course of the 17-year follow-up period. For example, blood biomarker levels associated with AD pathology may be higher in those with the APOE ε4 allele or those with higher PRS. The findings of this study could contribute to a better understanding of the genetic and molecular mechanisms underlying the onset of AD, facilitating the identification of individuals at risk and directing the creation of novel early detection and intervention strategies.

2.11. Comprehensive genetic screening of early-onset dementia patients in an Austrian cohort suggesting new disease-contributing genes [55]**.** This study details a comprehensive genetic screening study conducted on a group of early-onset dementia patients in Austria. Using state-of-the-art genetic sequencing methods, the research aims to find additional genes associated with the ailment that contribute to the illness. Through thorough research, the study finds unexpected genomic changes, highlighting the genetic heterogeneity widespread among patients with early-onset dementia. These findings advance our knowledge of the disease's genetic landscape, improve diagnosis accuracy, and make it possible to develop customized therapy regimens based on each patient's unique genetic profile. By shedding light on previously unknown genetic factors, this study significantly advances ongoing efforts to improve clinical care and pharmaceutical treatments for early-onset dementia.

2.12. Study of Alzheimer's disease- and frontotemporal dementia-associated genes in the Cretan Aging Cohort [56]**.** Using exome sequencing, this study looked at 196 participants from the Cretan Aging Cohort who had Alzheimer's disease (AD), mild cognitive impairment (MCI), and cognitively normal controls. The study discovered that the APOE ε4 allele was significantly more prevalent in AD patients than in controls, and that PSEN2 variants were present in both AD and MCI patients. Notably, the TARDBP gene variant was found in AD patients and those with the ALS/FTD phenotype. Although this discovery lacks statistical significance in replication cohorts, the lower frequency of the GLUD2 variant in AD patients compared to controls also suggested a possible protective impact against AD. Understanding the underlying genetics of AD and related illnesses in the elderly requires genetic testing, as the study highlights. A detailed summary of the literature are mentioned in Table 2.

Table 2.1: Summary of Literature Review.

Title	Aim	Datasets	Methodology	Findings
(Li et *al*., 2023)	Identify molecular markers and therapeutic targets associated with carotid atherosclerosis and vascular dementia.	GSE43292 GSE122063	Identify DEGs Functional Enrichment Analysis PPI network Therapeutic targets interaction Visualization	TYROBP, CSF1R, C1QB, ITGB2, LY86,
(Xu et *al*., 2023)	Understand the molecular interaction shared between Dementia with Lewy Bodies (DLB) and Parkinson's Disease Dementia (PDD.	GSE150696	Identify DEGs Functional Enrichment Analysis PPI network (WGCNA) Visualization	SNAP25, GRIN2A, GABRA1, GRIA1,
(Fan et *al*., 2022)	Utilize bioinformatics approaches to characterize the molecular signatures associated with vascular dementia.	GSE122063	Identify DEGs Functional Enrichment Analysis PPI network TFs and miRNA interaction analysis Visualization	GNG13, CD163, SST, C1QB, ITGB2, and TAC1
(Gui et *al*., 2021)	Understanding the underlying molecular mechanisms of the disease	GSE63061	Identify DEGs Functional Enrichment Analysis PPI network Visualization	GAPDH, RPS29, and

(Zhou et *al*., 2021)	Central genes (hub genes) that play a crucial role in AD and explore therapeutic intervention	GSE118553	Identify DEGs Functional Enrichment Analysis PPI network (MCODE) Drug gene interaction Visualization	FPR3, CXCL3, GABRB2, GABRG2
(Yuan et *al*., 2021)	Identify key genes (hub genes) that play a central role in the cognitive impairment observed in AD within the hippocampus	GSE48350	Identify DEGs Functional Enrichment Analysis PPI network Animal experiment Visualization	CNR1, GAD1, CALB1, GNG2, CCK, TAC1, SST, SCG2
(Fang et *al*., 2021)	Understand the molecular mechanisms, pathways, and genes that are implicated in dementia.	GSE153960	Identify DEGs Functional Enrichment Analysis PPI network TFs and miRNA interaction analysis Visualization	CDK1, TOP2A, RSL24D1, CDKN1A, MYB, PWP2, WNT7B
(Cheng et *al*., 2021)	Identify common genetic factors or molecular pathways that may be shared between AD and Major Depressive Disorder (MDD)	GSE48350, GSE5281, GSE18309, GSE98793, GSE32280	Identify DEGs Functional Enrichment Analysis PPI network Visualization	DYNC1H1, MAPRE3, ITGB1

2.14 Genetics of Dementia and AD

Understanding the genetic variables implicated in dementia and Alzheimer's disease is vital for identifying individuals at elevated risk, creating preventive interventions, and promoting personalized medicine techniques. For those who have a family history of dementia or Alzheimer's disease, genetic testing and counseling may be advised to determine their risk and make well-informed decisions on their future care [57]. There is hope for the development of novel medicines and interventions that target particular genetic pathways implicated in the etiology of these disorders through ongoing research into the genetics of these diseases. Current Research is based on Next Generation Sequencing (NGS) technologies employed in Whole Exome Sequencing (WES), and Whole Genome Sequencing (WGS). Causative and risk factor genes involved in both diseases are mentioned in Table 2.2.

2.15 Critical Analysis

Current approaches for hub gene identification in neurodegenerative diseases, particularly those employing Cyto-Hubba and Limma, have limitations that may impact their effectiveness in capturing complex disease mechanisms. Cyto-Hubba, focused on network centrality, effectively identifies prominent genes within interaction networks but lacks sensitivity to temporal dynamics and subtle molecular interactions, which are critical in conditions like Dementia and Alzheimer’s disease [58]. Limma, while widely used for identifying differentially expressed genes, relies on univariate analysis for two-group comparisons and does not consider interactions between genes or provide a multivariate analysis of variance and covariance, which can be essential for multifactorial diseases. Furthermore, existing gene expression studies often draw from limited datasets focusing on brain regions such as the temporal and frontal lobes and the hippocampus. This restriction in data scope may limit the generalizability and robustness of findings, highlighting the need for more comprehensive approaches that integrate temporal and spatial aspects of gene expression across broader populations. The critical analysis of identified hub genes is mentioned in Table 2.3.

Table 2.2: Causative genes and risk factor genes involved in Dementia and AD.

AD Causative genes	Risk Factors (AD)	Dementia Causative genes	Risk Factors (Dementia)	Common Risk Factor Genes
APP	APOE	MAPT	APOE	TREM2 (2024)
PSEN1	TREM2	GRN	TMEM106B	APOE (2023)
PSEN2	BIN1	C9ORF72	HLA	TARDBP (2023)
	ABCA7		TREM2	
	CLU		CHMP2B	
	TARDBP		BACE1	
	CR1		SQSTMI	
	PICALM		CHCHD10	
	MS4A6A		TARDBP	
	CD33		FUS	
	UNC5C		UBQLN2	
	PLD3		TUBA4A	
			LRRK2	

Table 2.3: Critical Analysis of Identified Hub Genes.

Literature Studies	Identifying DEGs	Enrichment analysis	PPI Network	Hub genes identified	Visualization	Hub Genes
(Li *et al*., 2023)	Limma in GEO2R	DAVID	STRING	Cyto-Huba	Cytoscape	TYROBP, CSF1R, C1QB, ITGB2, LY86,
(Xu *et al*., 2023)	Limma in GEO2R	Metascape	STRING	WGCNA	Cytoscape	SNAP25, GRIN2A, GABRA1, GRIA1,
(Fan *et al*., 2022)	Limma in GEO2R	Metascape	STRING	Cyto-Huba	Cytoscape	GNG13, CD163, SST, C1QB, ITGB2, and TAC1
(Gui *et al*., 2021)	Limma	DAVID	STRING	Cyto-Huba	Cytoscape	GAPDH, RPS29, and
(Zhou *et al*., 2021)	Limma	Metascape	STRING	MCODE in Cytoscape	Cytoscape	FPR3, CXCL3, GABRB2, GABRG2
(Yuan *et al*., 2021)	Limma in GEO2R	Cluster Profiler	Metascape	Cyto-Huba	Cytoscape	CNR1, GAD1, CALB1, GNG2, CCK, TAC1, SST, SCG2
(Fang *et al*., 2021)	Limma	Cluster Profiler	STRING	Cyto-Huba	Cytoscape	CDK1, TOP2A, RSL24D1, CDKN1A, MYB, PWP2, WNT7B
(Cheng *et al*., 2021)	Limma	ToppGene	STRING	Cyto-Huba	Cytoscape	DYNC1H1, MAPRE3, ITGB1

Chapter 3
Materials and Methods

This section provides all the information about the packages and tools used in this study and outlines the methodology followed. In this chapter, we will be discussing our material, data, and processing of data.

3.1 Dataset Selection (Microarray Data)

The gene expression profiling datasets related to AD and Dementia disease were gained from Gene Expression Omnibus (GEO) of The National Center for Biotechnology Information (NCBI) in the public domain with the series accession number (GSE13162, GSE117586) [59].

- The AD (GSE117586) dataset was based on the platform GPL25371 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 [60].
- The Dementia (GSE13162) dataset was based on the GPL571 platform [HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array [61].

To study AD and Dementia disease, the following publicly available microarray datasets were included in the analysis. The dataset details are mentioned in Table 3.1.

3.1.1 GSE117586

AD is characterized by a prolonged prodromal phase lasting decades, during which pathological changes accumulate in the brain before dementia onset. The specific nature, timing, and molecular mechanisms underlying these alterations remain unknown.

- The dataset comprised 10 samples, including 5 from patients with sporadic AD and 5 from age-matched controls [62].

This study investigates neurons and neural progenitor cells (NPCs) derived from induced pluripotent stem cells (iPSCs) of AD patients and age-matched controls. Transcriptome analysis did not distinguish between age-matched controls and iPSCs from sporadic AD (SAD) patients, but revealed significant variations in gene networks related to:

- Synaptic transmission
- Neuronal development
- NPC characteristics from iPSCs

Neurons generated from SAD NPCs exhibited increased electrical excitability and enhanced synapse formation due to accelerated neuronal development. Network analysis of the

transcriptome implicated the transcriptional repressor REST/NRSF and two components of the polycomb repressive complex 2 (PRC2), SUZ12 and EZH2.

3.1.2 GSE13162

Expression data from postmortem human brain samples were analyzed, focusing on those with and without frontotemporal lobar degeneration with ubiquitin-positive inclusions (FTLD-U), which is the most prevalent pathological correlate of frontotemporal dementia, a type of neurodegenerative dementia [63].

- This data from the frontal cortex, hippocampus, and cerebellum included 39 FTD samples and 17 non-demented controls (Control).

We analyzed the worldwide expression of FTLD-U brain tissues using microarrays. Postmortem brain samples were extracted from three groups of patients: progranulin-mutant FTLD-U patients, FTLD-U patients without progranulin gene mutations, and normal controls. Regional dissections from the cerebellum, hippocampus, and frontal cortex were performed.

Table 3.1: Dataset Attributes

Dataset	Disease	Samples	Array	Platform	Probe ID	File
GSE117586 [60]	Alzheimer	10 Samples (5 Control Samples 5 AD Samples)	[HG-U133A_2] Affymetrix Human Genome Array	GPL25371	54675 IDs	.CEL
GSE13162 [61]	Dementia	56 Samples (17 Control Samples & 39 Dementia Samples)	[HG-U133A_2] Affymetrix Human Genome Array	GPL571	22277 IDs	.CEL

3.2 Identification of Differentially Expressed

We used affycoretools, affy, gcrma, affyPLM, genefilter, and GEOquery R libraries to process our AD and Dementia datasets. To use these libraries, we first install the above-mentioned packages through BiocManager in R studio.

3.2.1 Input the files

The CEL files of both datasets (AD & Dementia) were fetched from NCBI GEO through the GEOquery library in R version (4.3.2).

To obtain supplementary files for the datasets, we first call the getGEOSuppFiles() function with the argument "GSE13162" and store the result in the variable. We repeat this process for the dataset "GSE117586", storing the result again. Next, we extract the contents of the "GSE13162_RAW.tar" file into a directory using the untar() function with appropriate arguments. We then list all CEL files within the working directory by calling the list.files() function and "CEL" as the pattern. Following this, we decompress each CEL file by iterating over the cels list, concatenating with each file name to get the full path, and passing this full path to the gunzip() function.

3.2.2 Read the Raw data

To read the Affymetrix CEL files, we first call the ReadAffy() function with the argument verbose set to FALSE and the filenames parameter set to the list of CEL files stored in the variable. This function call stores the resulting data in the variable raw data. This function is used for analyzing the microarray gene expression data. After loading the data, we simply reference the raw. Data to view it. To retrieve the column names of the data, we call the colnames() function with raw data as the argument. This process allows us to access and inspect the structure of the raw Affymetrix data, including its column names. The output summary of raw data on AD and dementia is shown in Figure 3.1.

(A)

```
AffyBatch object
size of arrays=1164x1164 features (21 kb)
cdf=HG-U133_Plus_2 (54675 affyids)
number of samples=10
number of genes=54675
annotation=hgu133plus2
```

(B)

```
AffyBatch object
size of arrays=732x732 features (35 kb)
cdf=HG-U133A_2 (22277 affyids)
number of samples=56
number of genes=22277
annotation=hgu133a2
```

Figure 3.1: (A) Summary of array used in AD Dataset **(B)** Summary of Array used in dementia Dataset

3.2.3 Boxplot of Raw Data

Boxplot demonstrates how expression values are distributed throughout all probes or genes. This is necessary to evaluate the quality of the data and identify deviations. We used the using the RColorBrewer library in R for accessing the different colors in the graph.

To create and save a boxplot of the raw data, we start by calling the tiff () function with the filename parameter set to boxplots_raw_data.tif to initiate saving the plot as a TIFF file. We then generate

the boxplot by calling the boxplot () function with the expression set exprSet, setting the color of the boxes with col set to color, adjusting the axis labels to be perpendicular to the axis with las set to 3, scaling the axis text with cex. Axis set to 0.5, and providing a title for the plot with the main set to Boxplot of raw data. To add a legend, we call the legend () function with coordinates -2, 15 for the placement, horiz set to TRUE for a horizontal legend, and specifying the legend labels as control and DM with corresponding colors green and yellow. Finally, we complete and save the plot by calling the dev. off () function, which closes the TIFF device and writes the plot to the file.

3.2.4 Data Preprocessing

Before performing downstream analyses, data preprocessing is an essential step in microarray analysis that aims to improve the quality and reliability of the data. Some techniques are used in this stage to address the different sources of noise, biases, and technical errors present in microarray experiments. Preprocessing consisted of three steps: Background correction, normalization, and summarization. The affyPLM package was used to achieve the quality and reliability of data. To get the normalized data, the GCRMA library is also required under the affyPLM package.

To process the raw data using normalization method, we first load the gcrma library. Then perform RMA normalization by calling the justRMA() function and storing the result. Next, we perform GCRMA normalization by calling the justGCRMA() function and storing the result. Finally, we use the expresso() function to apply the dChip normalization method, specifying the normalization method as an invariant set, and disabling background correction with bg. Correct set to FALSE, setting the perfect match correction method to pmonly, and using the liwong method for summarization. The result is stored in the variable. Figure 3.2 shows the processing of normalization of AD and Dementia.

(A)

```
normalization: invariantset
PM/MM correction : pmonly
expression values: liwong
normalizing...done.
54675 ids to be processed
```

(B)

```
normalization: invariantset
PM/MM correction : pmonly
expression values: liwong
normalizing...done.
22277 ids to be processed
```

Figure 3.2: (A) Processing of normalization in AD**. (B)** Processing of normalization in Dementia.

3.2.5 Boxplot of Normalized Data

A boxplot of normalized data was plotted. Which shows the normalized expression values across the samples. The boxplot () Function is used to create a boxplot of normalized data by setting the color and title.

3.2.6 Filtering Absent Genes

In microarray data analysis, filtering out genes that are absent or have low expression levels is a typical preprocessing step. Absent genes are often defined as having low expression levels that are comparable to background noise and may not be reliably detected. Filtering alters the results of statistical testing. If the filtering is unspecific, then no bias has been introduced to statistical testing and the result will be valid. Genefilter package was used for filtering the absent genes. kOverA() R function was used under the genefilter package to create the filters. This function uses the absolute number of samples during filtering. Using the MAS5.0 calls, gene expression data were further filtered to eliminate all missing genes. Log filtering was used to further filter the dataset to eliminate the skewed character of the gene expression data. Since it equalizes the distance between all the samples, all those genes with expression values above log2 (10) were filtered out.

Then extracts the expression values from the filtered dataset and checks its dimensions. The expression values of the data passing the log filter are extracted, and its dimensions are checked. A variance filter is applied to retain rows with variance above the 1000th highest variance threshold, resulting in the variable. The full expression data is stored, and dimensions are checked.

3.2.7 Statistical analysis (MANCOVA)

MANCOVA was used for the identification of differential Expressed Genes. MANCOVA calculates an average estimate of variability between groups taking into account all the groups. The multivariate covariance function was used for the gene expression matrix to calculate the distribution of all the genes among all the samples. Initially, a design matrix containing factors for fitting linear models was established through as.matrix() function in R. To change the vector group to factor group as.factor() function was used [62].

Initializing the empty result variable. It then sets up a variable with a sequence of values representing different categories (Control & Diseased). The expression data is read from a file. The gene names are extracted from the file and stored in genes. Next, a try block is used to handle

any errors that might occur during the execution of a loop. Inside the loop, which iterates over each row of expression gene, the current row is extracted into row_i and converted into a matrix m. The expression values from columns of the matrix are selected and stored in the m2 variable. A linear model m is then fitted with m2 as the response variable and results as the predictor. The summary of the linear model is generated, and the t-value and p-value are extracted. These values, along with the corresponding gene name, are combined into a matrix cp. This matrix is appended to the result variable, which accumulates the results for all genes.

3.2.8 Significant Genes

Genes with p-values less than 0.05 are regarded as the most significant genes. P-values are used to quantify the relevance of the acquired data [63]. All of the genes in this study that had a p-value below a significant threshold which is 0.002 were chosen as the differentially expressed genes.

Start with initializing an empty variable sig to store significant results. It then iterates over each row of the result variable. For each row, it checks if the p-value in the third column is less than 0.05. If the condition is met, the entire row is appended to the variable. This process filters out only the significant results based on the p-value threshold and stores them. Significant gene expression values are visualized through a heatmap.

3.2.9 GeneList Annotation

Gene annotation was done after statistical testing. Hgu133plus2 annotation package was used according to the array chip type. The vector of gene names was given as input for annotation which was obtained from MANCOVA results. XML, Annotate, and R2HTML packages were used to fetch the gene list annotation. This annotation file holds the information of gene ID, gene SYMBOL, gene NAME, and Gene ENSEMBL_ID.

Loading the necessary libraries: hgu133plus2.db, annotate, and R2HTML. It retrieves the feature names from the significant genes dataset and stores them in the ID variable. The gene symbols corresponding to these IDs are obtained using the getSYMBOL() function and stored in the variable. The gene names are fetched using lookup() and stored in the Name variable. Similarly, the Ensembl IDs are retrieved using lookup() and stored in the Ensembl variable.

If an Ensembl ID is "NA", it is replaced with NA; otherwise, it is converted into a hyperlink to the Ensembl website for the corresponding gene. A data frame tmp is created containing the ID,

Symbol, Name, and Ensembl columns, with stringsAsFactors set to FALSE. Any "NA" values in the data frame are replaced with NA. At the end, the data frame tmp is written to an HTML file named Sig_Genes.html without appending to any existing file.

3.2.10 Correlation of Genes

Examining the associations between differentially expressed genes (DEGs) can reveal potential regulatory or functional linkages among them. Positive correlation indicates the activated interactions and negative correlation indicates the inhibitory interactions. The interaction matrix was created. The symmetric and diagonal matrix that stores the dissimilarity values of all the gene expression values with one another is produced using the Euclidean distance between all the gene expression vectors in R. A covariance matrix including all the gene pair multivariate covariance values is calculated from the gene expression matrix using the cov () function in R. The contingency matrix was created in R which holds the activation and inhibition interactions of genes.

3.3 Enrichment Analysis of DEGs (AD & Dementia)

GO enrichment analysis was carried out using a cluster profiler (V3.16.0) package to examine more thorough biological information for the DEGs. This analysis included biological processes (BP), cellular components (CC), and molecular function (MF) terms. Using the cluster profiler (V3.16.0), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment was also carried out. Enrichment analysis was performed to investigate the biological meaning of genes.

3.3.1 Gene Ontology (GO) Enrichment Analysis

3.3.1.1 Biological Process

Biological processes offer important insights into the roles and activities of genes and gene products in living things. We used clusterProfiler R Package to get the biological processes of DEGs of both AD & Dementia. Org. Hs.eg.db R package was used to fetch data about human genomes. It is the Entrez Gene database for the human genome. We used an adjusted 0.05 p-value cutoff to filter genes and get statistically significant results.

Loading the necessary libraries: clusterProfiler, org.Hs.eg.db, and ggplot2. It then reads the significant gene data from a file into a variable called data. Using the enrichGO() function, then it

performs Gene Ontology (GO) enrichment analysis for the Biological Process (BP) category on the gene symbols in the data. The enrichment analysis uses the org.Hs.eg.db database, with SYMBOL as the key type, and applies the Benjamini-Hochberg (BH) method to adjust p-values, setting a p-value cutoff of 0.05. Finally, a dot plot of the top 20 enriched biological processes is created using the dot plot () function.

3.3.1.2 Molecular Function

The molecular activities of gene products, such as their structural characteristics, catalytic activities, and binding capacities, are described by their molecular functions. Performing Gene Ontology (GO) enrichment analysis for the Molecular Function (MF) category. It uses the same set of gene ID symbols stored in the data variable. The enrich () function is called again, this time with the ont parameter set to "MF" for Molecular Function and results are stored. We used pvalueCutoff 0.05 for molecular function.

3.3.1.3 Cellular Component

Within the Gene Ontology (GO), the subcellular structures or places where gene products are located or active within a cell are described by the cellular component ontology.
It uses the same set of gene ID symbols. The enrichGO() function is called again, this time with the ont parameter set to "CC" for Cellular Component and results are stored. We used pvalueCutoff 0.05 for the Cellular component.

3.3.1.4 Networks of GO Enrichment Analysis

Understanding the biological pathways and processes connected to a group of genes or proteins may be gained by visualizing the outcomes of KEGG pathway enrichment and Gene Ontology (GO) analysis. Network visualization, in which enriched phrases or routes are represented as nodes and their relationships as edges is a popular method of visualizing the findings of enrichment analyses. It uses a list of gene IDs stored in the list$ID variable. The enrichGO() function is called with the following parameters: gene is set to list$ID, the universe is set to the names of list$ID, OrgDb is set to org.Hs.eg.db, and ont is set to "CC" for Cellular Component. The p-value adjustment method is specified as Benjamini-Hochberg (BH), with a p-value cutoff of 0.01 and a q-value cutoff of 0.05. The readable parameter is set to TRUE to convert gene IDs to gene symbols

if possible. The results of the enrichment analysis are stored in the variable. In the end, a GO plot of the enrichment results is created using the goplot () function.

3.3.2 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment

KEGG Pathway is an assembly of meticulously selected pathway that depict interactions between molecules and reaction networks seen in cells. A particular biological activity or pathway, such as metabolic processes, signaling, or genetically encoded information processing, is shown on each pathway map. Using p. adjust <.05, the map for the top 10 important KEGG pathways was created using the ggplot2 R Package.

First uses the bitr () function to convert gene symbols to Entrez IDs. It specifies fromType as "SYMBOL" and toType as "ENTREZID", with the OrgDb parameter set to org.Hs.eg.db. Then, the enrichKEGG () function is called to perform KEGG pathway enrichment analysis. It takes the Entrez IDs from Gene_ids$ENTREZID and specifies "hsa" as the organism (for Homo sapiens). Finally, a bar plot of the top 20 enriched KEGG pathways is created using the barplot () function with the KEGG results.

3.4 Identification of Hub Genes

Hub genes are essential components within a biological network that are crucial for preserving the network's structure and function. Discovering these hub genes can offer significant insights into the fundamental mechanisms of diseases and aid in pinpointing potential therapeutic targets. We used the LASSO algorithm for the identification of hub genes.

3.4.1 Least Absolute Shrinkage and Selection Operator (LASSO)

A feature gene screen was conducted using the least absolute shrinkage and selection operator (LASSO) model. In R, LASSO was carried out using the glmnet package. When variables have less of an impact on outcomes, LASSO uses a penalty function to reduce the regression coefficient until it is 0. The process can converge toward a prediction model that is more accurate in this way. First, load the necessary libraries glmnet, and plotmo. Then, read the gene expression data from the file with headers into a variable where rows represent samples and columns represent gene expression levels. Display the dimensions of the data frame. Rename the column with index 194

to 'y', which will be the target variable. Convert the data into a matrix. Extract the target variable y from the y column in dat. Fit a LASSO regression model to x and y with alpha set to 0.1 in AD while in dementia alpha was set to 1 and family set to binomial for binary classification by using glmnet () Function. Plot the coefficients of the LASSO model using plot_glmnet(). Also, plot the LASSO model coefficients to the lambda parameter, showing the top 20 labels. Next, perform cross-validated LASSO regression with alpha set to 0.04 in AD and Dementia alpha value set to 0.7 by using cv.glmnet () Function. We can find the ideal value of λ by accessing lasso_model$lambda.min after fitting a LASSO model with the cv. glmnet() function. This value is utilized to build the final model and is selected during the cross-validation procedure. Extract the coefficients from the cross-validated LASSO model and identify the selected features by extracting the row names from the coefficients where they are non-zero. These non-zero features will be selected as hub genes.

3.4.2 Common Hub gene

After applying the LASSO algorithm, we get the common hub genes by intersecting the genes. First, load the VennDiagram library. Create two lists, containing the selected features from LASSO regression for Alzheimer's Disease (AD) and dementia (DM), respectively. Identify the common genes through the reduce() Function by finding the intersection of list1 and list2. Generate a Venn. Diagram plot () Function using the Venn diagram library, where AD is represented by list1 and DM is represented by list2. Set the fill colors to orange and blue, and the transparency (alpha) to 0.5 for both. Adjust the category label size (cat. cex) and the main label size (cex) to 1.5 and set the main title.

Create a new page in the grid and draw the initial Venn diagram plot. Examine the default plot by looking at the names in the plot object v, specifically focusing on the labels. Update the intersection label in the plot object v with the common genes identified earlier, excluding the first element, and join them into a single string with newline characters. Create a new page in the grid and redraw the updated Venn diagram plot with the intersection label showing the common genes.

3.4.3 Protein-Protein Interaction of Hub Genes

In this study, we constructed and analyzed the PPI network of common hub gene in the STRING database. The STRING database collects both predicted protein interactions and those that are

experimentally validated. Hub genes of AD and Dementia identified through the LASSO Algorithm were put into the STRING database, from which the interaction partners of their high-confidence interactions were retrieved. The generation of the interaction network was based on the combined score, taking into account all the sources of evidence due to co-expression, experimental data, and computational predictions. This promised not only deeper insight into molecular mechanisms and potentially underlying functions but also their tractability in a more reduced fashion.

3.4.4 AUC Curve

First, the outcome variables in datasets were converted into factors to prepare for modeling. Subsequently, each dataset was subsetted to include only common genes identified earlier, and the outcome variable was retained. Logistic regression models were then fitted to predict the outcome variable using all remaining variables in each subset, employing the binomial family due to the binary nature of the response. Predictions of probabilities were generated from these models to assess their predictive performance. ROC curves were plotted for both datasets (AD in blue and DM in red) by using the pROC library, illustrating the trade-offs between sensitivity and specificity. The area under the ROC curve (AUC) was calculated for each model to quantify its discriminatory power, with values rounded to two decimal places. These AUC values were subsequently added to the plot legend, providing a clear comparison of model performance between Alzheimer's disease (AD) and dementia (DM) datasets based on the common genes analyzed.

3.5 Disease- Gene Association

To validate the gene role and disease association of the identified common gene, we utilized databases such as DisGeNET [64], GeneCards [65], and the Comparative Toxicogenomics Database (CTD) [66], along with several literature sources [67]. This comprehensive approach ensured the reliability of our gene identification.

Chapter 4

Results and Discussion

This chapter provides all the results of the methodology which we get through tools and software used in this methodology. In this chapter, we will be discussing our results of data.

4.1 Dataset of AD

The GSE117586 dataset of Alzheimer's disease that we used in our study is shown in Table 4.1.

Table 4.1: Dataset of Alzheimer's Disease.

GSM Accession	Sample Name	Description	Age/ Gender	Disease Status	Coriell #
GSM3304289	NP NL1.1	neural progenitor cells	60F	normal	AG04455
GSM3304290	NP NL2.1	neural progenitor cells	64M	normal	AG08125
GSM3304291	NP NL3.1	neural progenitor cells	72M	normal	AG08379
GSM3304292	NP NL4.1	neural progenitor cells	73M	normal	AG08509
GSM3304293	NP NL6.1	neural progenitor cells	92F	normal	AG09173
GSM3304294	NP_SAD1.1	neural progenitor cells	60F	sporadic AD	AG06869
GSM3304295	NP SAD2.1	neural progenitor cells	60M	sporadic AD	AG07376
GSM3304296	NP_SAD3.1	neural progenitor cells	69F	sporadic AD	AG21158
GSM3304297	NP SAD4.1	neural progenitor cells	72M	sporadic AD	AG08243
GSM3304298	NP_SAD5.1	neural progenitor cells	87F	sporadic AD	AG08245

4.2 Dataset of Dementia

The GSE13162 dataset of Dementia disease that we used in our study is shown in Table 4.2.

Table 4.2: Dataset of Dementia Disease.

GSM ID	Sample ID	Tissue Type
GSM329660	N1A	Normal-frontal
GSM329661	N1C	Normal-hippocampus
GSM329662	N1B	Normal-cerebellum
GSM329663	N2A	Normal-frontal
GSM329664	N5A	Normal-frontal
GSM329665	N5B	Normal-cerebellum
GSM329666	N6A	Normal-frontal
GSM329667	N7A	Normal-frontal
GSM329668	N7B	Normal-cerebellum
GSM329669	N8C	Normal-hippocampus
GSM329670	N10A	Normal-frontal
GSM329671	N10B	Normal-cerebellum
GSM329672	N11A	Normal-frontal
GSM329673	N11B	Normal-cerebellum
GSM329674	N12A	Normal-frontal
GSM329675	N13B	Normal-cerebellum
GSM329676	N14B	Normal-cerebellum
GSM329677	P1A	Progranulin-frontal
GSM329678	P1C	Progranulin-hippocampus
GSM329679	P2A	Progranulin-frontal
GSM329680	P2C	Progranulin-hippocampus
GSM329681	P3A	Progranulin-frontal
GSM329682	P3B	Progranulin-cerebellum
GSM329683	P4A	Progranulin-frontal
GSM329684	P4B	Progranulin-cerebellum
GSM329685	P4C	Progranulin-hippocampus
GSM329686	P5A	Progranulin-frontal
GSM329687	P5B	Progranulin-cerebellum
GSM329688	P5C	Progranulin-hippocampus
GSM329689	P7A	Progranulin-frontal
GSM329690	P7B	Progranulin-cerebellum
GSM329691	P7C	Progranulin-hippocampus
GSM329692	S1A	Sporadic-frontal
GSM329693	S1C	Sporadic-hippocampus
GSM329694	S2A	Sporadic-frontal
GSM329695	S2B	Sporadic-cerebellum
GSM329696	S2C	Sporadic-hippocampus
GSM329697	S3A	Sporadic-frontal
GSM329698	S3B	Sporadic-cerebellum
GSM329699	S3C	Sporadic-hippocampus
GSM329700	S4A	Sporadic-frontal

GSM ID	Sample ID	Tissue Type
GSM329701	S4B	Sporadic-cerebellum
GSM329702	S4C	Sporadic-hippocampus
GSM329703	S5A	Sporadic-frontal
GSM329704	S6A	Sporadic-frontal
GSM329705	S6B	Sporadic-cerebellum
GSM329706	S6C	Sporadic-hippocampus
GSM329707	S7A	Sporadic-frontal
GSM329708	S7C	Sporadic-hippocampus
GSM329709	S8A	Sporadic-frontal
GSM329710	S8C	Sporadic-hippocampus
GSM329711	S9A	Sporadic-frontal
GSM329712	S9B	Sporadic-cerebellum
GSM329713	S9C	Sporadic-hippocampus
GSM329714	S10A	Sporadic-frontal
GSM329715	S10B	Sporadic-cerebellum

4.3 Identification of Differential Expressed Genes

In section 3.2 we identified the DEGs in both AD and dementia. The identification of DEGs in AD and dementia has yielded valuable insights into the molecular mechanisms that underlie these illnesses. Certain genes may have an impact on the pathogenesis of these neurodegenerative illnesses. Comprehending these DEGs can facilitate the creation of focused treatments and diagnostic instruments for AD and dementia.

4.3.1 Data Pre-Processing in AD.

Preprocessing is considered a necessary step in gene expression analysis because it prepares the dataset to be free of inaccuracies for further, reliable analyses. Data preprocessing requires assessing the quality of raw data, normalization of the raw data, and the elimination of technical variability. Results of data preprocessing are provided in Figures 4.1 and 4.2 that include the raw data and normalized data, respectively.

Figure 4.1 represents boxplot showing the distribution of raw data across samples with different colors representing different samples. Each boxplot, indicated by the x-axis labels, represents a single sample and is constructed based on the median, interquartile range, and outliers. The y-axis is the value obtained for each experimental condition. The designation "AD" represented in yellow may refer to one of the groups or conditions listed in the legend. For instance, the distributions of data for the different samples usually show a high degree of likeness, with little chance of considerable differences arising between the groups.

Figure 4.2 represents a boxplot containing normalized data from different samples where the samples are represented using different colors as mentioned in the x-axis labels. It ranges normalized values that were gotten after utilizing the RMA expression measure in the y-axis. There exist data normalization procedures on the probe IDs for 54675. Each boxplot will represent how the median, interquartile, with the eventual outliers will look like. Most values of data are spread within a relatively similar range across the conditions. It can be seen that although compared to raw data, normalized data looks much more uniform, most of the data points are close to the median and only a few outliers are found.

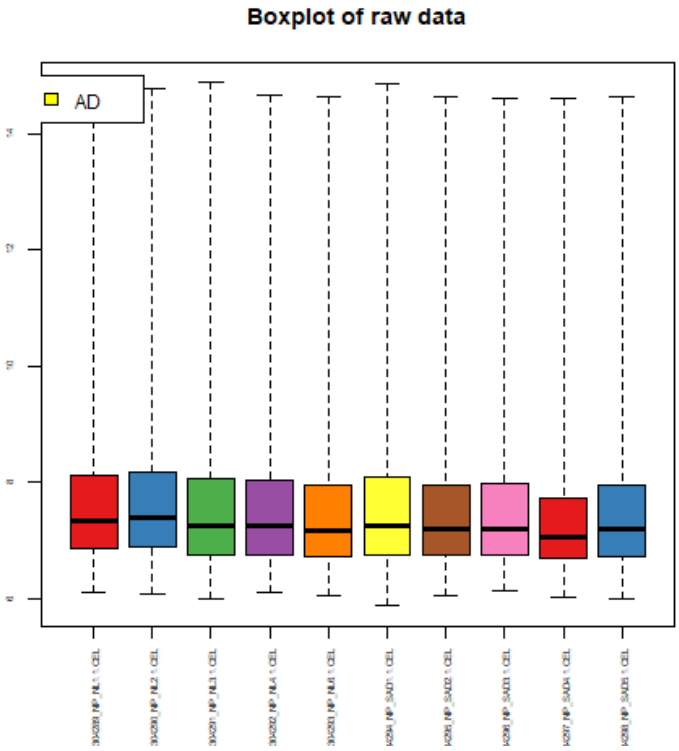

Figure 4.1: Boxplot of AD raw data.

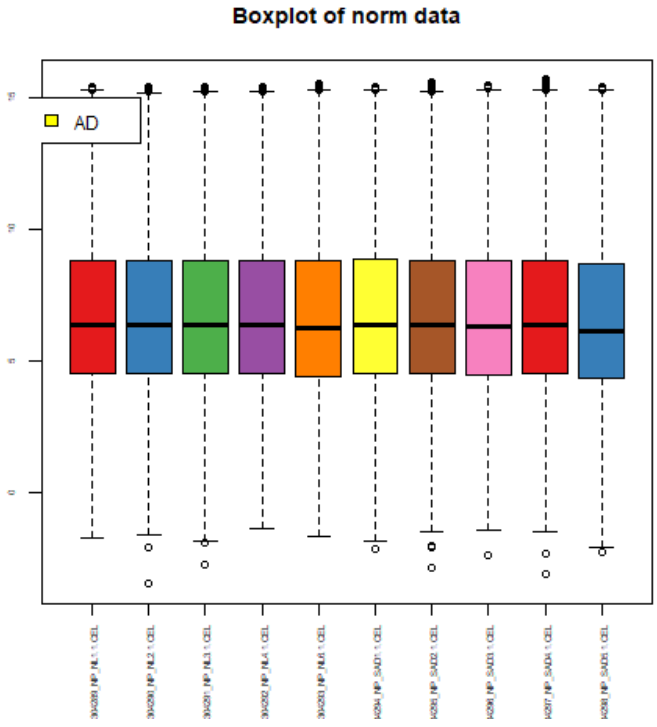

Figure 4.2: Boxplot of AD normalized data.

4.3.2 DEGs of AD

In this section, Revealing the identification and visualization of DEGs related to Alzheimer's disease, which is considered to be important for understanding the molecular mechanisms of critical insights into this neurodegenerative disorder.

Table 4.3 shows the identification of annotated DEGs related to AD using MANCOVA analysis. Among the 41,381 genes studied, 3,238 were found to be significant at $p < 0.05$. Among those, 836 had $p < 0.01$, 445 $p < 0.005$, 273 $p < 0.003$, and 192 $p < 0.002$. Although all these genes are valued significant, we reduced our list to just 192 genes whose p-value were < 0.002 that have to be further considered. As in the following table for any given probe ID, there must be an associated gene symbol, as well as a name exists. For some probe IDs no gene symbol or name could be assigned since no genomic sources include information on them.

Figure 4.3 illustrates the Visualization of Gene expression values of DEGs through Heatmap. This is a heatmap that is showing the expression levels of several genes across many samples, and it has applied hierarchical clustering. The color range used is from blue to red, which implies lower expression values, and it is only red that depicts higher expression values. At the top and left are the dendrograms, indicating clusters for samples and genes hence their similarities and differences in expression profiles. While the names of genes and sample IDs appear on the x- and yaxes, respectively, the plots are very similar. Blocks are clusters of genes exhibiting similarly expressed levels within a few groups of samples. This heatmap depicts a complex data set in order to further detail the expression patterns of genes. The clusters which result may represent the biological difference between the groups of samples, as embodied by these gene expression profiles.

Table 4.3: Differential Expressed Genes of AD

ID	Symbol	Name
1553179_at	ADAMTS19	ADAM metallopeptidase with thrombospondin type 1 motif 19
1553622_a_at	FSIP1	fibrous sheath interacting protein 1
1555993_at	CACNA1D	calcium voltage-gated channel subunit alpha1 D
1558819_at	NA	NA
1560023_x_at	NA	NA
1560490_at	FAT3	FAT atypical cadherin 3
1560743_a_at	NA	NA
1563071_at	ADAMTSL4-AS2	ADAMTSL4 antisense RNA 2
1563331_at	NA	NA
1564112_at	GARIN4	golgi associated RAB2 interactor family member 4
1564485_at	LINC00887	long intergenic non-protein coding RNA 887
1565681_s_at	DIP2C	disco interacting protein 2 homolog C
1566163_at	NA	NA
200771_at	LAMC1	laminin subunit gamma 1
201340_s_at	ENC1	ectodermal-neural cortex 1
201341_at	ENC1	ectodermal-neural cortex 1
202201_at	BLVRB	biliverdin reductase B
202442_at	AP3S1	adaptor-related protein complex 3 subunit sigma
202560_s_at	CHTOP	chromatin target of PRMT1
202752_x_at	SLC7A8	solute carrier family 7 member 8
202900_s_at	NUP88	nucleoporin 88
202912_at	ADM	adrenomedullin
202978_s_at	CREBZF	CREB/ATF bZIP transcription factor
203001_s_at	STMN2	stathmin 2
203020_at	RABGAP1L	RAB GTPase activating protein 1 like
203408_s_at	SATB1	SATB homeobox 1
203700_s_at	DIO2	iodothyronine deiodinase 2
203733_at	DEXI	Dexi homolog
203857_s_at	PDIA5	protein disulfide isomerase family A member 5
203928_x_at	MAPT	microtubule-associated protein tau
204024_at	OSGIN2	oxidative stress-induced growth inhibitor family member 2
204069_at	MEIS1	Meis homeobox 1
204195_s_at	PKNOX1	PBX/knotted 1 homeobox 1
204227_s_at	TK2	thymidine kinase 2
204321_at	NEO1	neogenin 1
204365_s_at	REEP1	receptor accessory protein 1
204646_at	DPYD	dihydro pyrimidine dehydrogenase

204820_s_at	NA	NA
204868_at	MRPL58	mitochondrial ribosomal protein L58
205081_at	CRIP1	cysteine-rich protein 1
205271_s_at	CDK20	cyclin-dependent kinase 20
205311_at	DDC	dopa decarboxylase
205669_at	NCAM2	neural cell adhesion molecule 2
206051_at	ELAVL4	ELAV-like RNA binding protein 4
206114_at	EPHA4	EPH receptor A4
206874_s_at	SLK	STE20 like kinase
207480_s_at	MEIS2	Meis homeobox 2
207725_at	POU4F2	POU class 4 homeobox 2
207952_at	IL5	interleukin 5
208051_s_at	PAIP1	poly(A) binding protein interacting protein 1

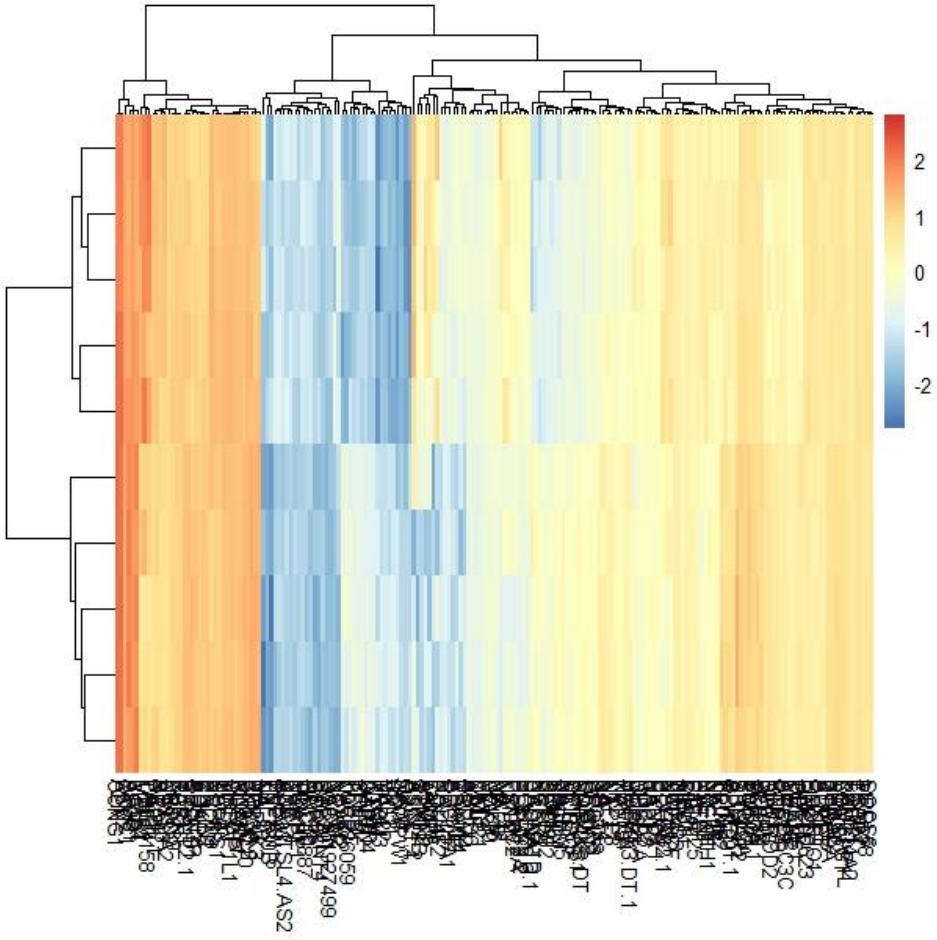

Figure 4.3: Visualization of Gene expression values of DEGs through Heatmap.

4.3.3 Multivariate Covariance of AD DEGs

The computation of the covariance between DEGs will give insight into their relations and interactions across the different samples. Covariance is a measure of how changes in the expression of one gene are related to the changes in the expression of another. We can see from the covariance matrix that genes have similar expression patterns an indication of co-regulation, common pathways, or shared biological processes.

Table 4.4: Symmetric Multivariate Covariance Matrix of the First 10 DE genes of AD. It is a covariance matrix and shows how the gene expression of one gene is interrelated with the gene expressions of others. ADAMTS19 and ADAMTSL4.AS2 is correlated at 0.89. The expression levels of these genes are extremely positively correlated. FSIP1 and LINC00887 (0.87) show the relationship of increasing or decreasing. FAT3 and 1560743_a_at show varying expression values that are either positive or negative but retain a significantly high coefficient of 0.92. The genes 1560743_a_at and FAT3 (0.92) have a very high positive covariance, which could indicate co-regulation or functional interrelation. CACNA1D and 1558819_at (-0.85) have a highly significant negative covariance, meaning their expression patterns are inversely correlated. GARIN4 and 1558819_at (-0.86): indicates that when GARIN4 is high in expression, 1558819_at is low, and vice-versa. GARIN4 and 1560743_a_at (-0.91) have high negative covariance thus, their expressions inversely correlate. Most of the gene pairs have some modest positive or negative covariance that shows the degree of their related expressions. Genes that show strong positive covariance might be regulated together or might be involved in similar biological processes, whereas genes that have a high negative covariance might be involved in a regulatory mechanism wherein the expression of one gene is negatively controlled by another.

Table 4.4: Symmetric Multivariate Covariance Matrix of First 10 DE genes

	ADAMTS19	FSIP1	CACNA1D	1558819_at	1560023_x_at	FAT3	1560743_a_at	ADAMTSL4.AS2	1563331_at	GARIN4
ADAMTS19	1.00	0.84	-0.77	0.64	0.61	0.76	0.73	0.89	0.74	-0.72
FSIP1	0.84	1.00	-0.70	0.61	0.62	0.77	0.64	0.81	0.77	-0.69
CACNA1D	-0.77	-0.70	1.00	-0.85	-0.73	-0.78	-0.80	-0.75	-0.71	0.67
1558819_at	0.64	0.61	-0.85	1.00	0.84	0.84	0.91	0.56	0.63	-0.86
1560023_x_at	0.61	0.62	-0.73	0.84	1.00	0.70	0.83	0.66	0.63	-0.81
FAT3	0.76	0.77	-0.78	0.84	0.70	1.00	0.92	0.74	0.66	-0.79
1560743_a_at	0.73	0.64	-0.80	0.91	0.83	0.92	1.00	0.75	0.73	-0.91
ADAMTSL4.AS2	0.89	0.81	-0.75	0.56	0.66	0.74	0.75	1.00	0.82	-0.68
1563331_at	0.74	0.77	-0.71	0.63	0.63	0.66	0.73	0.82	1.00	-0.79
GARIN4	-0.72	-0.69	0.67	-0.86	-0.81	-0.79	-0.91	-0.68	-0.79	1.00

4.3.4 Data Pre-Processing of Dementia

Preprocessing is considered a necessary step in gene expression analysis because it prepares the dataset to be free of inaccuracies for further, reliable analyses. Data preprocessing requires assessing the quality of raw data, normalization of the raw data, and the elimination of technical variability. Results of data preprocessing are provided in Figures 4.4 and 4.5 that include the raw data and normalized data, respectively.

Figure 4.4: Boxplot showing the raw data distribution across several samples grouped into two separate groups, "control" and "DM". Each box shows the interquartile range of the data with a horizontal line marking the median. The whiskers are extended up to 1.5 times the IQR and any point outside these whiskers are considered potential outliers. It is seen that the raw data distributions in both the control and DM groups do not have any apparent differences between them, either in central tendency or in variability. There is also a high consistency between the height of boxplots for each sample, hence, there is a considerable uniformity of data distribution for all samples. There are very few outliers, and none of them suggest an important deviation from the overall data pattern. This plot suggests that there does not appear to be a significant difference between the control and DM groups for the raw data and, hence, this would be an appropriate candidate for further analysis and normalization.

Figure 4.5 Boxplot of the normalized gene expression values distributed across the two conditions, control vs. DM. Each box plot corresponds to a unique gene or probe with a range of different expression values across the samples. A total of 22277 probe IDs were processed. The line inside the box shows the median expression value and the box itself shows an interquartile range. Whiskers extend showing the range of values outside the middle 50%. Points beyond whiskers show outliers. Graph showing the contrast in the expression levels for each gene or probe between two conditions and pointing out variations or outliers in the data.

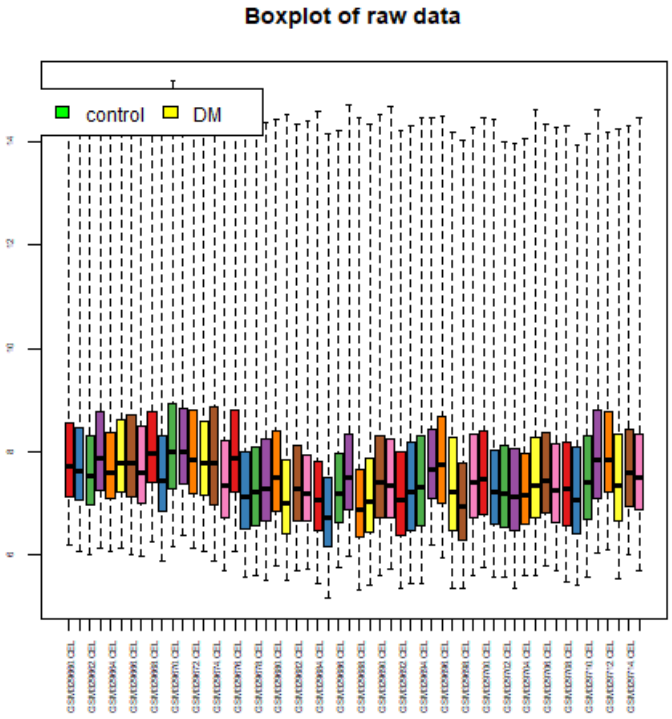

Figure 4.4: Boxplot of Dementia Raw Data

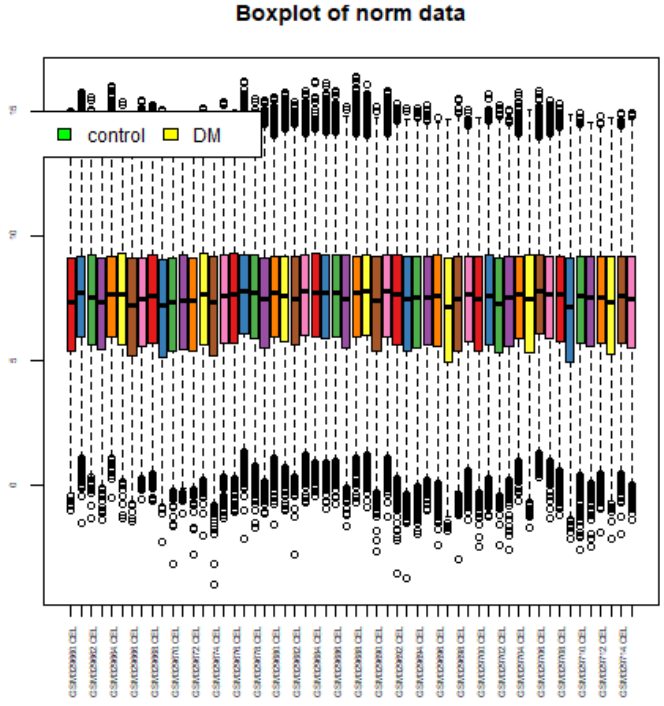

Figure 4.5: Boxplot of Dementia Normalized Data

4.3.5 DEGs of Dementia

In in this section, revealing the identification and visualization of DEGs related to Dementia, which is considered to be important for understanding the molecular mechanisms of critical insights into this neurodegenerative disorder.

Table 4.5 Annotated top 50 DEGs in DM using MANCOVA. Out of the total genes of 15038, 4608 genes have been found significant with a p-value less than 0.05. Among them, 2299 genes had a p-value below 0.01, 1662 genes below 0.005, 1299 genes below 0.003, and 1035 genes below 0.002. Although we acknowledge that all of these genes are important, we decided to study the 1035 genes with p-values less than 0.002 further. For each probe ID in this table, there is a corresponding gene symbol and gene name. There were some probe IDs that had no annotation to any gene symbol or gene name; therefore, they have been categorized as unknown genes because there was no information relevant to them from genomic resources. From a total of 1035 genes, the first 50 genes are presented in the table.

Figure 4.6 illustrates a heatmap that elucidates hierarchical clustering within gene expression data obtained from various samples. The accompanying color scale on the right denotes the extent of gene expression, with red signifying a high expression level reaching up to 4, blue indicating a low expression level descending to -4, and yellow representing the intermediate or moderate levels of expression. At the top and left of the dendrograms is shown hierarchical clustering of samples and genes. This clustering is an aggregation of samples and genes having similar expression patterns. Along the x-axis, it is the samples; the y-axis corresponds to genes.

The heatmap shows the presence of two distinct clusters for samples and genes with similar expression patterns. A strong cluster of genes on the left side shows mainly up-regulated expressions indicated by yellow to red across most of the samples, whereas another cluster on the right mainly expresses low levels of expression represented by yellow to blue. This might indicate that the gene groups listed above have differential expression across the samples and, by consequence, could be subjected to differential regulation or might have active biological pathways in these samples. The heatmap is essentially meant to portray the gene expression patterns and allow for the easy identification of co-expressed gene groups along with samples having similar expression profiles.

Table 4.5: Top 50 Differential Expressed Genes of Dementia

ID	Symbol	Name
1007_s_at	DDR1	discoidin domain receptor tyrosine kinase 1
1294_at	UBA7	ubiquitin-like modifier activating enzyme 7
200030_s_at	SLC25A3	solute carrier family 25 member 3
200064_at	HSP90AB1	heat shock protein 90 alpha family class B member 1
200092_s_at	RPL37	ribosomal protein L37
200595_s_at	EIF3A	eukaryotic translation initiation factor 3 subunit A
200613_at	AP2M1	adaptor-related protein complex 2 subunit mu 1
200621_at	CSRP1	cysteine and glycine-rich protein 1
200622_x_at	NA	NA
200638_s_at	YWHAZ	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta
200641_s_at	YWHAZ	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta
200642_at	SOD1	superoxide dismutase 1
200647_x_at	NA	NA
200653_s_at	NA	NA
200665_s_at	SPARC	secreted protein acidic and cysteine-rich
200677_at	PTTG1IP	PTTG1 interacting protein
200680_x_at	HMGB1	high mobility group box 1
200694_s_at	DDX24	DEAD-box helicase 24
200695_at	PPP2R1A	protein phosphatase 2 scaffold subunit Aalpha
200696_s_at	GSN	gelsolin
200708_at	GOT2	glutamic-oxaloacetic transaminase 2
200720_s_at	ACTR1A	actin-related protein 1A
200723_s_at	CAPRIN1	cell cycle-associated protein 1
200773_x_at	PTMA	prothymosin alpha
200780_x_at	GNAS	GNAS complex locus

200801_x_at	ACTB	actin beta
200802_at	SARS1	seryl-tRNA synthetase 1
200810_s_at	CIRBP	cold-inducible RNA binding protein
200811_at	CIRBP	cold-inducible RNA binding protein
200815_s_at	PAFAH1B1	platelet-activating factor acetylhydrolase 1b regulatory subunit 1
200820_at	PSMD8	proteasome 26S subunit, non-ATPase 8
200822_x_at	TPI1	triosephosphate isomerase 1
200825_s_at	HYOU1	hypoxia up-regulated 1
200832_s_at	SCD	stearoyl-CoA desaturase
200870_at	STRAP	serine/threonine kinase receptor-associated protein
200871_s_at	PSAP	prosaposin
200883_at	UQCRC2	ubiquinol-cytochrome c reductase core protein 2
200896_x_at	HDGF	heparin-binding growth factor
200912_s_at	EIF4A2	eukaryotic translation initiation factor 4A2
200927_s_at	RAB14	RAB14, member RAS oncogene family
200932_s_at	DCTN2	dynactin subunit 2
200937_s_at	RPL5	ribosomal protein L5
200954_at	ATP6V0C	ATPase H+ transporting V0 subunit c
200961_at	SEPHS2	selenophosphate synthetase 2
200972_at	TSPAN3	tetraspanin 3
200981_x_at	GNAS	GNAS complex locus
200982_s_at	ANXA6	annexin A6
200991_s_at	SNX17	sorting nexin 17
201014_s_at	PICS	phosphoribosylaminoimidazole carboxylase and phosphoribosylaminoimidazolesuccinocarboxamide synthase
201032_at	BLCAP	BLCAP apoptosis-inducing factor

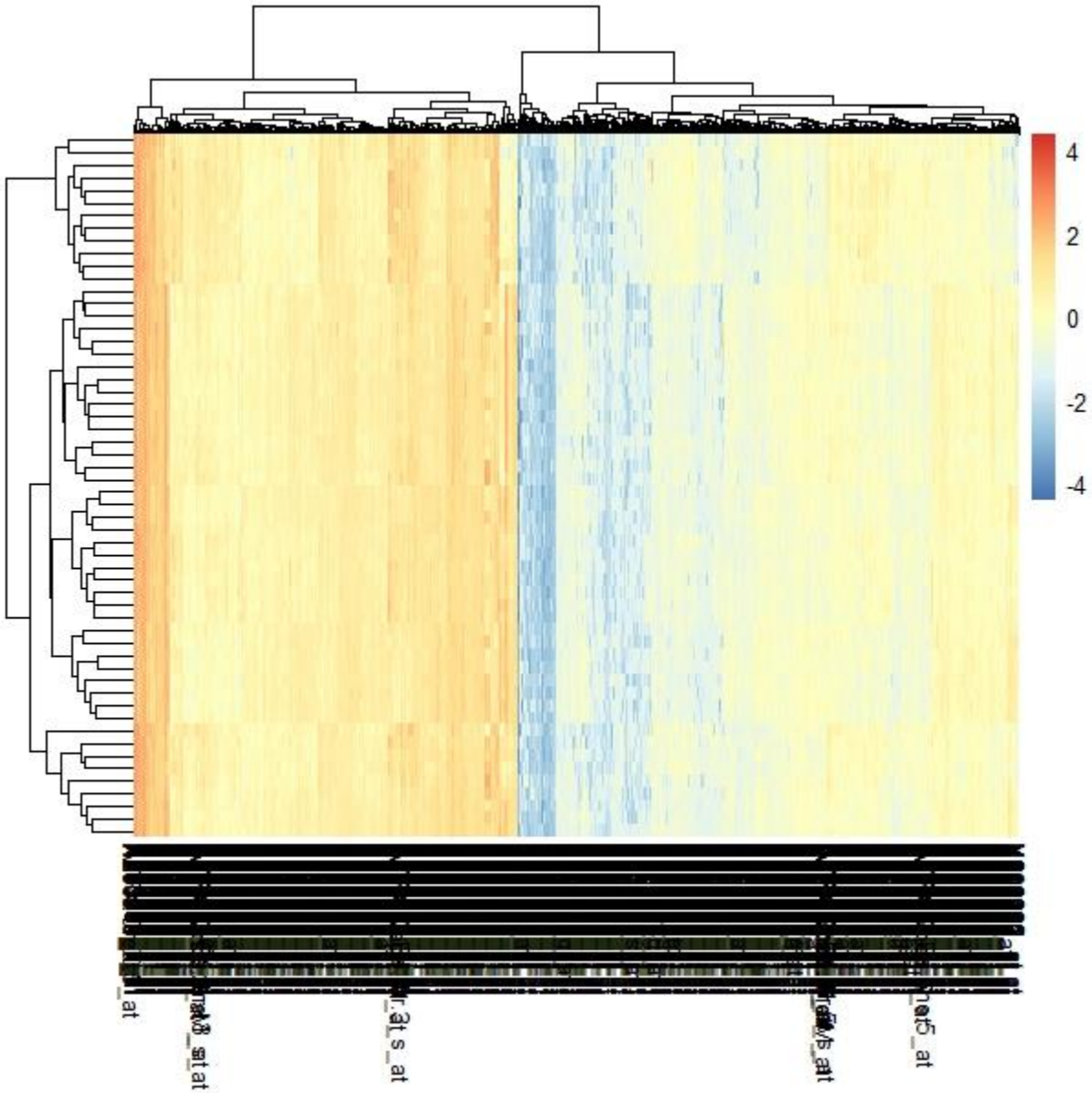

Figure 4.6: Visualization of Gene expression values of Dementia DEGs through Heatmap.

4.3.6 Multivariate Covariance of Dementia DEGs

This section analyzes the relations between the first 10 differentially expressed genes (DEGs) related to dementia from the investigation of a symmetric multivariate covariance matrix.

Table 4.6 illustrates the symmetric multivariate covariance matrix representing the correlations among the first 10 DE genes with relation to dementia. Every diagonal entry is always valued at 1.00 since these refer to the variance of each gene with respect to itself, which is one feature of a correlation matrix. Off-diagonal elements represent the covariance of two genes, showing that alterations in the expression level of one gene are related to similar changes in another. A few positive covariance connections exist, such as between DDR1 and CSRP1 (0.86), DDR1 and UBA7 (0.65), SLC25A3 and NA (0.66), HSP90AB1 and YWHAZ (0.72), and AP2M1 and NA (0.81). There is a negative covariance relationship between DDR1 and HSP90AB1 (-0.59), UBA7 and HSP90AB1 (-0.63), UBA7 and SLC25A3 (-0.45), and between SLC25A3 and CSRP1 (-0.38). Positive covariance at a higher value reflects a strong direct relationship among gene pairs, suggesting that if the expression level of one gene is increased, then the corresponding gene will also increase.

For example, between DDR1 and CSRP1, the positive covariance observed is as high as 0.86, implying that their expression patterns closely correlate with each other. Negative covariance was found between the genes, meaning a negative correlation in which an increase in the expression of one gene was inversely correlated with a decrease in the expression of the other. In this regard, the value of covariance obtained between DDR1 and HSP90AB1 was -0.59, hence establishing its negative correlation. This matrix gives deep insights into how these genes interact with one another in the context of dementia.

Table 4.6: Symmetric Multivariate Covariance Matrix of First 10 DE genes of Dementia

	DDR1	UBA7	SLC25A3	HSP90AB1	RPL37	EIF3A	AP2M1	CSRP1	NA	YWHAZ
DDR1	1.00	0.65	-0.26	-0.59	0.42	0.30	-0.37	0.86	-0.02	-0.33
UBA7	0.65	1.00	-0.45	-0.63	0.42	0.22	-0.50	0.59	-0.35	-0.49
SLC25A3	-0.26	-0.45	1.00	0.55	-0.27	-0.13	0.69	-0.38	0.66	0.64
HSP90AB1	-0.59	-0.63	0.55	1.00	-0.25	-0.06	0.67	-0.62	0.51	0.72
RPL37	0.42	0.42	-0.27	-0.25	1.00	0.15	-0.28	0.59	-0.23	-0.37
EIF3A	0.30	0.22	-0.13	-0.06	0.15	1.00	-0.16	0.21	-0.11	0.17
AP2M1	-0.37	-0.50	0.69	0.67	-0.28	-0.16	1.00	-0.43	0.81	0.80
CSRP1	0.86	0.59	-0.38	-0.62	0.59	0.21	-0.43	1.00	-0.16	-0.47
NA	-0.02	-0.35	0.66	0.51	-0.23	-0.11	0.81	-0.16	1.00	0.74
YWHAZ	-0.33	-0.49	0.64	0.72	-0.37	0.17	0.80	-0.47	0.74	1.00

4.4 GO & KEGG Pathway Enrichment Analysis of DEGs

In section 3.4, we performed the enrichment analysis of DEGs of both AD and Dementia Disease. It provides a context to the changes in gene expression observed with the specific biological function, pathway, or disease. DEGs can help identify key biological processes and pathways that are disrupted in AD and dementia. We compared the DEGs of both AD and Dementia to find common and unique pathways, processes, and functions that can give us in-depth details.

4.4.1 GO Enrichment Analysis of Differentially Expressed Genes

We visualized the GO Enrichment results of both AD and Dementia through dot plots and networks. Go enrichment analysis involves:

- Biological Process
- Molecular Function
- Cellular component

4.4.1.1 Biological Processes of DEGs in AD and Dementia

For the functional roles of the DEGs involved in AD, GO enrichment analysis was conducted to identify the most significantly biological processes associated with these genes.

Figure 4.7 illustrates a dot plot of the top 20 biological processes in AD. The x-axis represents Gene Ratio, which is the ratio of the genes in one biological process to the total number of genes that were used in the analysis. The y-axis shows the names of the specific biological processes. Each point represented in this illustration corresponds to a distinct biological process, with the dimensions of the point indicating the number of genes that are associated with that specific process. Points of greater size denote a higher gene count. The color gradient applied to these points reflects an adjusted p-value, which serves as an indicator of statistical significance. The presence of red indicates smaller p-values, which imply greater significance, whereas blue signifies larger p-values, suggesting a lesser degree of significance.

Some of the key processes are "regulation of neuron projection development," "axonogenesis," and "synapse organization," which had a lower p-value and larger dots. Following the list were limb development and morphogenesis-related processes, although having a little higher p-values. In general, this plot shows that neuronal development- and morphogenesis-related processes dominated the gene set of AD. Gene ontology IDs with p-value and count are tabulated in Table

4.7.

Figure 4.8 depicts the top 20 biological process in Dementia ranked by their gene ratio, plotted along the x-axis. Biological process factorized appears along the y-axis along with axonogenesis, cell growth, and synapse organization, others. Different circles represent a process, and the size of these circles represents the number of genes a process contains. The shades of the circles represent the adjusted p-value, which is a measure of statistical significance. It can be noticed that 'axonogenesis' contains the highest ratio of genes; thus, it comprises the highest percentage of genes involved in other biological processes. Other processes significantly important are cell growth, regulation of neuron projection development, and calcium ion transport.

Smaller circles represent processes such as 'presynaptic endocytosis' that have low gene ratios and few genes involved. This also depicts the relative importance and relevance of these biological processes in a very clear manner. More prominent processes are represented by larger red circles, which indicate a high gene ratio along with a low adjusted p-value. The gene ontology IDs, along with their corresponding p-values and counts, are shown in Table 4.8.

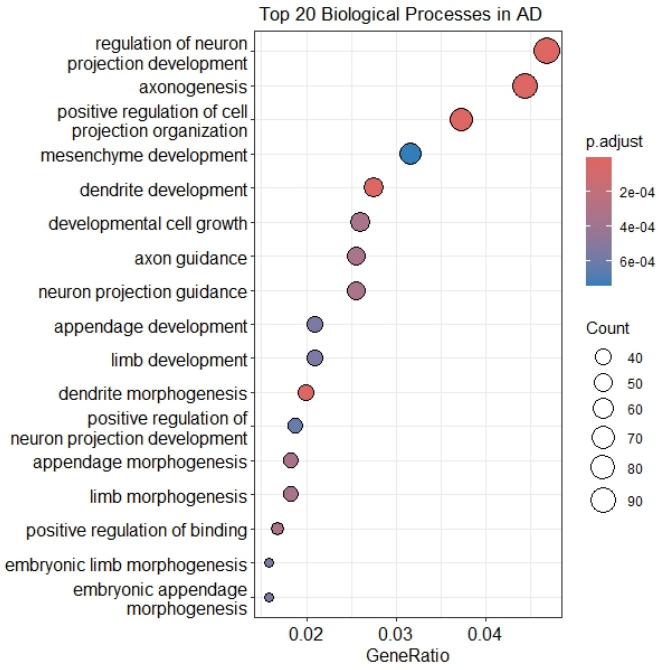

Figure 4.7: Dot plot of biological processes of DEGS in AD

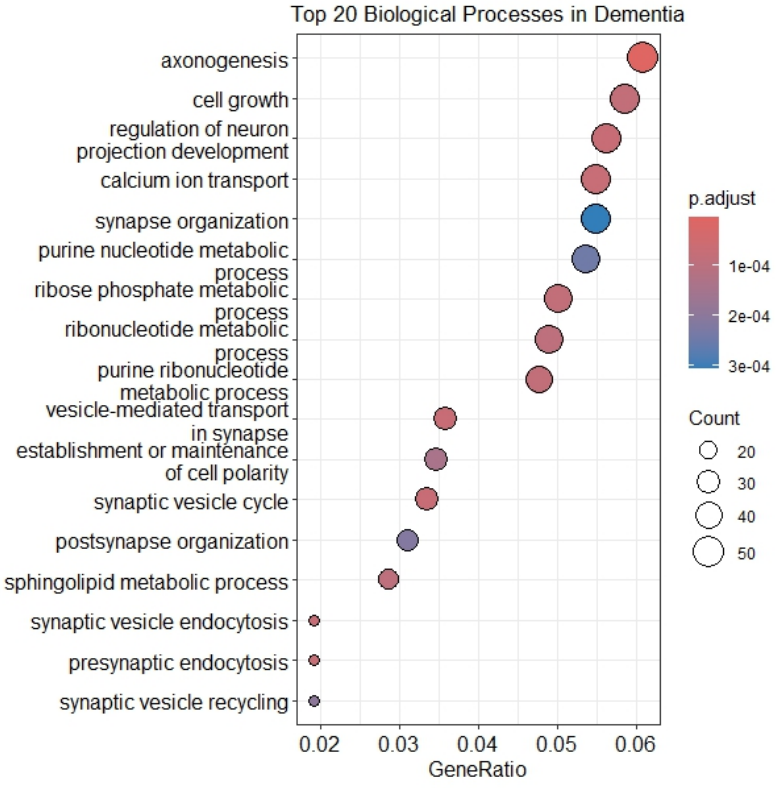

Figure 4.8: Dot plot of biological processes of DEGS in Dementia

Table 4.7 Gene Ontology IDs of Biological Processes in AD

ID	Description	p-value	Count
GO:0010975	regulation of neuron projection development	2.0351E-10	92
GO:0048813	dendrite morphogenesis	4.7274E-09	39
GO:0007409	axonogenesis	6.1809E-09	87
GO:0016358	dendrite development	1.1341E-08	54
GO:0048588	developmental cell growth	4.5729E-07	51
GO:0007411	axon guidance	4.9419E-07	50
GO:0097485	neuron projection guidance	4.9419E-07	50
GO:0051099	positive regulation of binding	5.0403E-07	33
GO:0035107	appendage morphogenesis	6.242E-07	36
GO:0035108	limb morphogenesis	6.242E-07	36
GO:0030326	embryonic limb morphogenesis	1.2661E-06	31
GO:0035113	embryonic appendage morphogenesis	1.2661E-06	31
GO:0048736	appendage development	1.3938E-06	41
GO:0060173	limb development	1.3938E-06	41
GO:0060485	mesenchyme development	2.1143E-06	62
GO:0050808	synapse organization	2.9214E-06	83
GO:0048762	mesenchymal cell differentiation	3.5076E-06	52

Table 4.8 Gene Ontology IDs of Biological Processes in Dementia

ID	Description	p-value	Count
GO:0007409	axonogenesis	7.04652E-10	51
GO:0099003	vesicle-mediated transport in synapse	4.55391E-08	30
GO:0099504	synaptic vesicle cycle	5.81933E-08	28
GO:0010975	regulation of neuron projection development	6.0916E-08	47
GO:0048488	synaptic vesicle endocytosis	6.41741E-08	16
GO:0140238	presynaptic endocytosis	7.9146E-08	16
GO:0006816	calcium ion transport	9.40442E-08	46
GO:0019693	ribose phosphate metabolic process	1.37057E-07	42
GO:0009150	purine ribonucleotide metabolic process	1.39522E-07	40
GO:0016049	cell growth	1.54989E-07	49
GO:0006665	sphingolipid metabolic process	2.11727E-07	24
GO:0009259	ribonucleotide metabolic process	2.12478E-07	41
GO:0007163	establishment or maintenance of cell polarity	3.45256E-07	29
GO:0036465	synaptic vesicle recycling	5.29477E-07	16
GO:0099173	postsynapse organization	6.13423E-07	26
GO:0006163	purine nucleotide metabolic process	7.43257E-07	45
GO:0050808	synapse organization	9.89188E-07	46
GO:0060052	neurofilament cytoskeleton organization	1.3575E-06	6
GO:0051924	regulation of calcium ion transport	2.53301E-06	29

4.4.1.2 Molecular Function of DEGs in AD and Dementia

This section highlights the top 20 molecular functions that relate to AD, underlined by significance and frequency of genes that reflect the ratio, count of genes, and adjusted p-values.

Figure 4.9 illustrates the top 20 molecular functions associated with Alzheimer's disease (AD), showing their importance through GeneRatio, count, and adjusted p-values. The y-axis lists functions such as "DNA-binding transcription factor binding," "GTPase regulator activity," and "nucleoside-triphosphatase regulator activity," while the x-axis represents the gene ratio, indicating the proportion of genes linked to each function.

"DNA-binding transcription factor binding" stands out with a high gene ratio and low adjusted p-value, indicating its significant association with AD. Other important functions include "GTPase regulator activity" and "nucleoside-triphosphatase regulator activity," both showing high gene ratios and adjusted p-values. Functions like "protein serine/threonine kinase activity" and "ubiquitin-like protein ligase binding" have medium gene ratios and p-values, suggesting moderate significance for AD. "Double-stranded RNA binding" has the lowest gene ratio and a high adjusted p-value, marking it as the least significant function in AD. Overall, the plot emphasizes that molecular functions related to DNA-binding transcription factors and GTPase regulation play major roles in AD pathogenesis. The gene ontology identifiers, p-values, and counts are detailed in Table 4.9.

Figure 4.10 displays the top 20 molecular functions associated with dementia, ranked by GeneRatio (the ratio of genes linked to each function compared to the total number of genes in the dataset). The gradient from red to blue indicates statistical significance, with red representing lower p-values (higher significance). The highest GeneRatio is found for "DNA-binding transcription factor binding," suggesting that many dementia-related genes are involved in this function. Other notable functions include "amide binding," "guanyl nucleotide binding," and "GTP binding." In contrast, functions like "structural constituent of postsynapse" and "adrenergic receptor binding" have lower GeneRatios and gene counts. This highlights the diversity of molecular functions related to dementia, with transcription factor binding, nucleotide binding, and transporter activity being particularly important. The color coding of p-values emphasizes the statistical significance of these functions, helping to identify critical pathways and interactions in dementia. The specific gene ontology identifiers, p-values, and gene counts are detailed in Table 4.10.

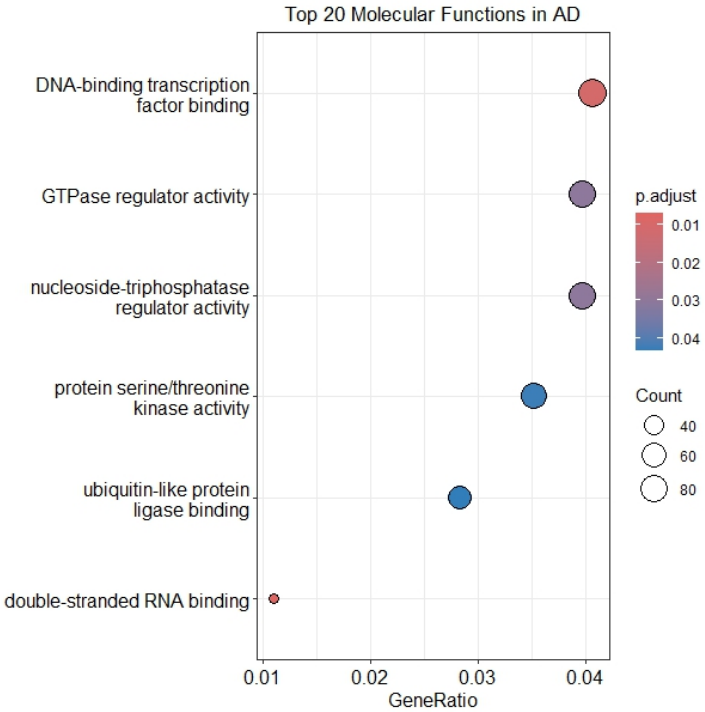

Figure 4.9: Dot plot of Molecular Function of DEGS in AD

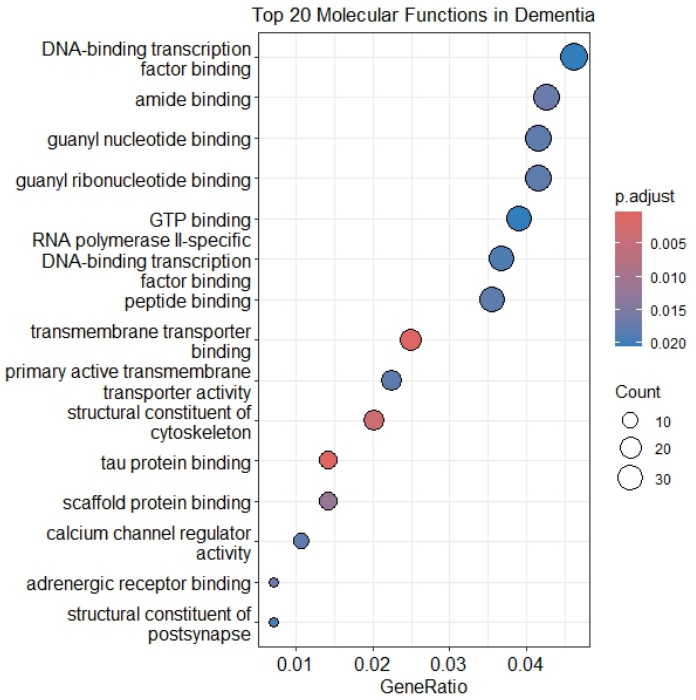

Figure 4.10: Dot plot of Molecular Function of DEGS in Dementia

Table 4.9 Gene Ontology IDs of Molecular Function in AD

ID	Description	p-value	Count
GO:0003725	double-stranded RNA binding	6.02398E-06	22
GO:0140297	DNA-binding transcription factor binding	1.81352E-05	82
GO:0030695	GTPase regulator activity	0.000104779	80
GO:0060589	nucleoside-triphosphatase regulator activity	0.000104779	80
GO:0004674	protein serine/threonine kinase activity	0.000184175	71
GO:0044389	ubiquitin-like protein ligase binding	0.000225109	57

Table 4.10 Gene Ontology IDs of Molecular Function in Dementia

ID	Description	p-value	Count
GO:0048156	tau protein binding	3.1261E-07	12
GO:0044325	transmembrane transporter binding	6.341E-07	21
GO:0005200	structural constituent of cytoskeleton	1.2529E-05	17
GO:0097110	scaffold protein binding	5.1953E-05	12
GO:0031690	adrenergic receptor binding	0.00010264	6
GO:0033218	amide binding	0.00010922	36
GO:0005246	calcium channel regulator activity	0.00016562	9
GO:0015399	primary active transmembrane transporter activity	0.00018	19
GO:0042277	peptide-binding	0.00020862	30
GO:0019001	guanyl nucleotide binding	0.00021195	35
GO:0032561	guanyl ribonucleotide binding	0.00021195	35
GO:0061629	RNA polymerase II-specific DNA-binding transcription factor binding	0.0002522	31
GO:0099186	structural constituent of postsynapse	0.00026676	6
GO:0140297	DNA-binding transcription factor binding	0.00032621	39
GO:0005525	GTP binding	0.00034974	33
GO:0009931	calcium-dependent protein serine/threonine kinase activity	0.00035273	6
GO:0010857	calcium-dependent protein kinase activity	0.00045893	6
GO:0003779	actin binding	0.00049496	36
GO:0019894	kinesin binding	0.00074715	8

4.4.1.3 Cellular Component of DEGs in AD and Dementia

This section highlight the top 20 molecular functions that relate to AD, underlined by significance and frequency of genes that reflect the ratio, count of genes, and adjusted p-values.

Figure 4.11: Dot plot of the top 20 cellular components implicated in AD. The x-axis is the GeneRatio, which here is the proportion of genes associated with that cellular component. The y-axis includes different cellular components such as neuronal cell body, nuclear envelope, glutamatergic synapse, and others. Each bubble is sized by the count of genes associated with the component; the larger the bubble, the higher the count. Each bubble is also colored on a gradient where the adjusted p-value is according to the key at the top left of the diagram: red and blue reflect higher and lower significance thresholds, respectively. The highest number of ratios of genes along with corresponding significance are observed with the "glutamatergic synapse" and "nuclear envelope". The "GABA-ergic synapse", "Schaffer collateral – CA1 synapse", etc., possess a significantly smaller number of ratios of genes and the significance attached. In the Table 4.11 the gene ontology IDs and corresponding p-values along with count for the mentioned IDs is represented.

Figure 4.12 depicts the top 20 cellular components of dementia. The x-axis represents the GeneRatio, indicating the ratio of relevant genes, and the y-axis indicates the designations of the components involved. A larger diameter of the dot indicates a greater gene amount. The neuronal cell body shows the highest GeneRatio and some of the largest dots, suggesting a great correlation with dementia. Other critical components include cell leading edge, transport vesicle, distal axon, neuron-to-neuron synapse, and cell-substrate junction; all of these have high GeneRatio along with high p-values, making them important. Most are in red, which denotes high significance. This figure stresses the significance of several neuronal and synaptic structures to dementia pathology, supported by significance in the p-values upon adjustment. Gene ontology IDs along with p-value and count is mentioned in Table 4.12.

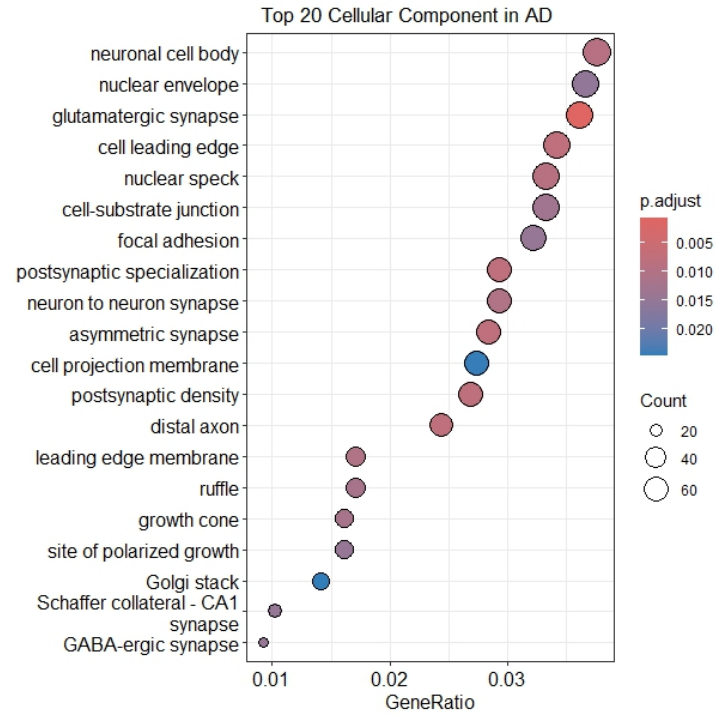

Figure 4.11: Dot plot of Cellular Component of DEGS in AD

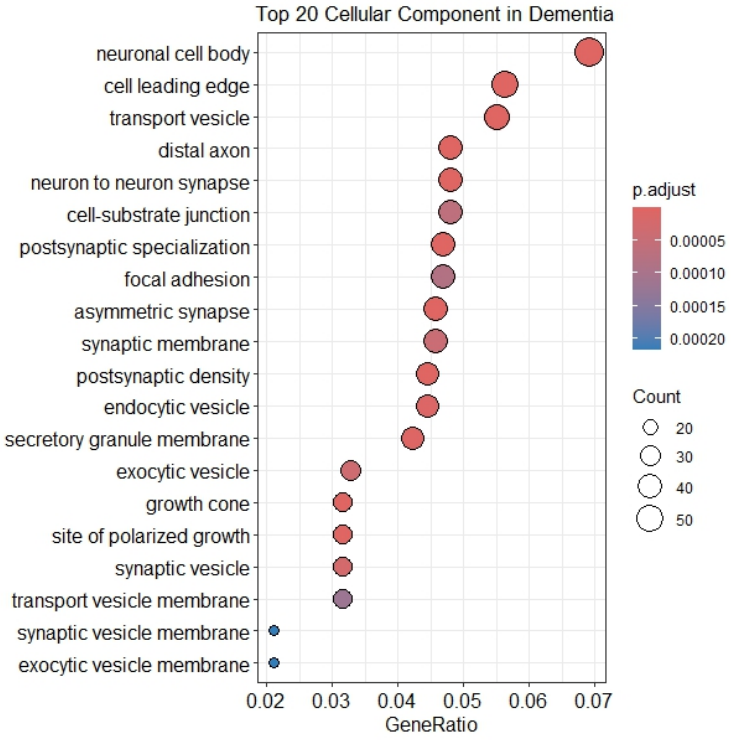

Figure 4.12: Dot plot of Cellular Component of DEGS in Dementia

Table 4.11 Gene ontology IDs of Cellular Component in AD

ID	Description	p-value	Count
GO:0098978	glutamatergic synapse	8.5279E-07	74
GO:0032279	asymmetric synapse	3.8329E-05	58
GO:0099572	postsynaptic specialization	3.8912E-05	60
GO:0031252	cell leading edge	4.4434E-05	70
GO:0150034	distal axon	5.3467E-05	50
GO:0014069	postsynaptic density	6.6297E-05	55
GO:0016607	nuclear speck	8.98E-05	68
GO:0043025	neuronal cell body	0.00010398	77
GO:0098984	neuron-to-neuron synapse	0.00014426	60
GO:0031256	leading-edge membrane	0.00014757	35
GO:0030426	growth cone	0.0002057	33
GO:0001726	ruffle	0.00020681	35
GO:0030055	cell-substrate junction	0.00024237	68
GO:0098982	GABA-ergic synapse	0.00033353	19
GO:0005925	focal adhesion	0.00035538	66
GO:0098685	Schaffer collateral - CA1 synapse	0.00035898	21
GO:0030427	site of polarized growth	0.00036062	33
GO:0005635	nuclear envelope	0.00039465	75
GO:0031253	cell projection membrane	0.00066819	56

Table 4.12 Gene ontology IDs of Cellular Component in Dementia

ID	Description	p-value	Count
GO:0043025	neuronal cell body	6.9094E-13	59
GO:0150034	distal axon	3.5134E-12	41
GO:0031252	cell leading edge	8.3251E-10	48
GO:0030426	growth cone	3.0533E-09	27
GO:0030133	transport vesicle	4.5551E-09	47
GO:0030427	site of polarized growth	5.8792E-09	27
GO:0014069	postsynaptic density	9.292E-09	38
GO:0032279	asymmetric synapse	1.1489E-08	39
GO:0099572	postsynaptic specialization	1.2797E-08	40
GO:0098984	neuron-to-neuron synapse	1.5193E-08	41
GO:0030667	secretory granule membrane	8.4321E-08	36
GO:0030139	endocytic vesicle	1.3334E-07	38
GO:0008021	synaptic vesicle	4.812E-07	27
GO:0070382	exocytic vesicle	8.1447E-07	28
GO:0097060	synaptic membrane	1.074E-06	39
GO:0030055	cell-substrate junction	1.6984E-06	41
GO:0005925	focal adhesion	2.3389E-06	40

Figure 4.13 shows the gene ontology networks in AD. **(A)** This figure shows biological process term networks in gene ontology, where nodes are reserved for different processes, while their sizes represent their values. Edges are used to signify relationships such as "is_a," "part_of," "positively_regulates," and "regulates.". The key concepts of neuron differentiation, neuron projection morphogenesis, and axon guidance are mutually interlinked, which puts emphasis on their critical nature in significance establishment. Specifically, the framework includes clusters indicating closely associated processes, that is, neuron development/differentiation. The network's hierarchical structure, characterized by overarching processes such as "biological regulation" and "developmental process," serves as a framework that encompasses more detailed processes, including "neuron projection guidance" and "dendrite morphogenesis." This structure facilitates the comprehension of the complex interactions and organization of biological processes, providing valuable insights into gene function and regulation. **(B)** This illustration represents a visual depiction of cellular components CC derived from Gene Ontology GO terms. Nodes represent the various cellular components in proportion to their significance. Edges symbolize "is_a" and "part_of" relations. Color of nodes has been depicted by the value of p-adjust; greater the shade of red, more the significance of the node. Highly significant nodes like "synapse," "nucleus," and "neuron projection" have high degrees of connectivity, which represents their very basic role in the structure of cellular architecture. Clusters centered on these hub nodes include related terms, including "postsynaptic density," "axons," and "nuclear speck." The modularity of the network structure indicates a hierarchical organization where general categories, like "membrane-bounded organelle," include more specific subunits, such as "glutamatergic synapse" and "nuclear speck," allowing for a better understanding of the connectivity and hierarchical organization of cellular components. **(C).** This figure shows a representation of the network of molecular functions as retrieved from gene ontology (GO) terms. In the network, the nodes stand for unique functions, their sizes in proportion to their strength, while the edges give out the "is_a" hierarchies in which nodes stand. Color filling up the nodes denotes the statistical significance extent by p-adjust value, thus it uses the red color for this figure:. Central nodes include "protein binding, ion binding, and purine ribonucleotide binding", accountable for their many well-cross-linked interactions.

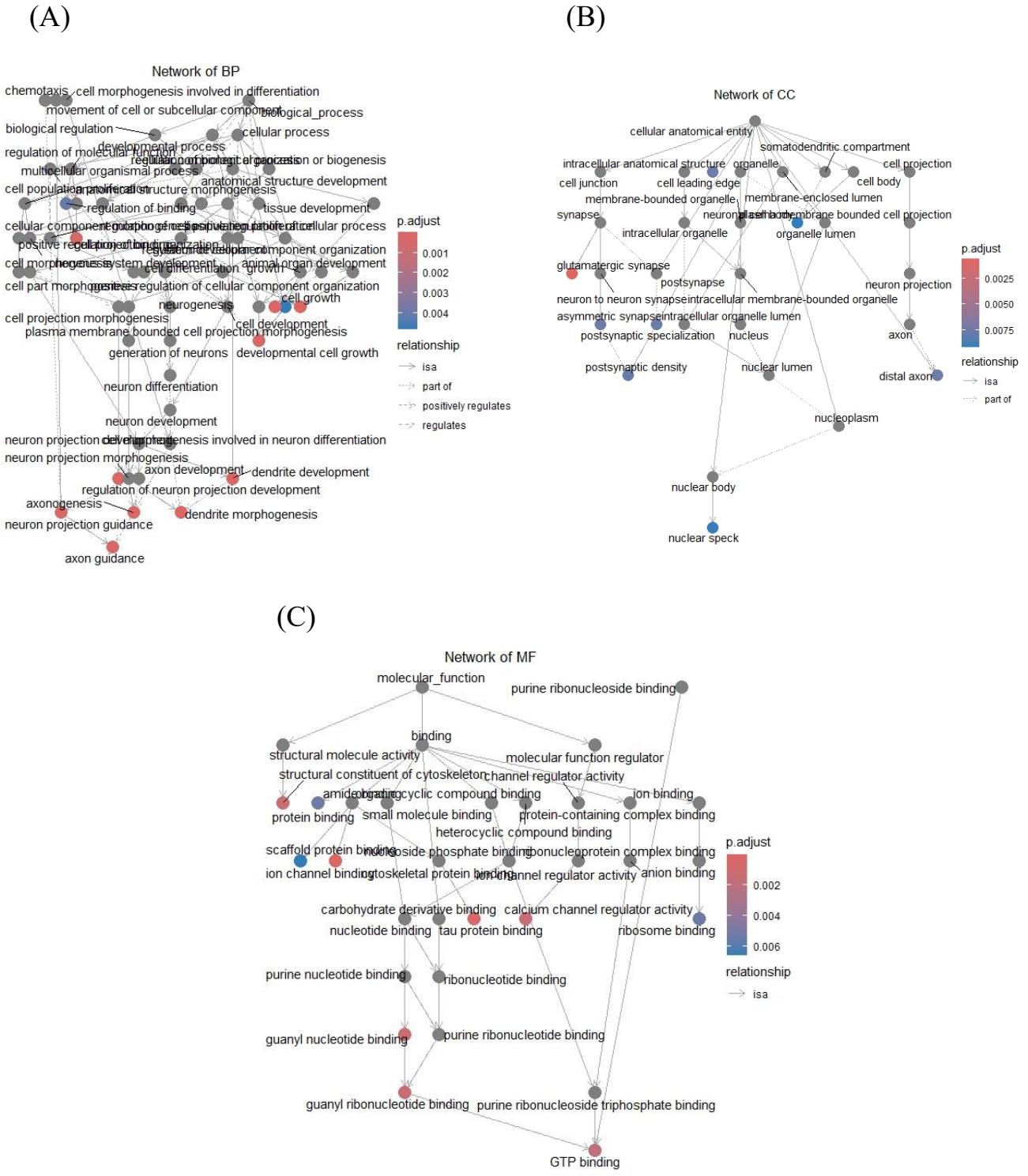

Figure 4.13: Network of Gene Ontologies in AD. (A) Network of Biological Process (B) Network of Cellular Component. (C) Network of Molecular Function

Figure 4.14 Gene ontology networks related to Dementia. **(A).** The network of interacting heterogeneous biological processes in a data set, which includes cell differentiation, transport mechanisms, developmental stages, and metabolic activities. Each node in this graph represents an individual biological process, and the edges represent different types of relationships between nodes, such as participating in the same process, regulatory interactions, or involvement in analogous biological functions. A gradient is used to indicate levels of significance; nodes that appear redder have a higher p-value, and nodes that appear greayer have a lower p-value. Core processes like "cell differentiation," "lipid metabolic process," and "membrane transport" were dominantly featured early in the list, which matches the expectation for core processes that significantly interact within the network. **(B).** The network figure depicts cellular components based on the biological dataset. Here, nodes are different cellular components, and edges represent relationships such as "is a" or "part of" that can be found between the cellular components. The color of the nodes indicates the values of the adjusted p-values, ranging from red to blue. In this case, red nodes mean lower p-values and thus greater significance, while blue nodes represent higher p-values, indicating reduced significance. Such things are the "distal axon," "growth cone," "transport vesicle," "intracellular organelle," and "cell junction." The network also refers to important structures and the association at the cell level: for example, elements in the "distal axon" and "growth cone" that will probably remain of crucial importance in a much richer biology. The general connectivity schema has the multilocation and multisite nature characteristic of the interactions between various elements within the cellular built-space. **(C).** The figure represents a graph of a network that was built from GO terminology on molecular functions. In this graph, the nodes represent the various functions scaled according to the importance of the function, while the edges between the nodes indicate the hierarchical "is_a" relationships. The color gradient of the nodes corresponds to the extent of the p-adjust value and therefore is indicative of the degree of statistical significance; for this case, red has been used. The core nodes are "protein binding," "ion binding," and "purine ribonucleotide binding" because of their high number and strongly established relationships. Sets of terms, like "purine ribonucleotide binding," in combination with more loosely linked terms that appear in the classes "molecular function regulator" and "structural molecule activity,"describe the specialized functions.

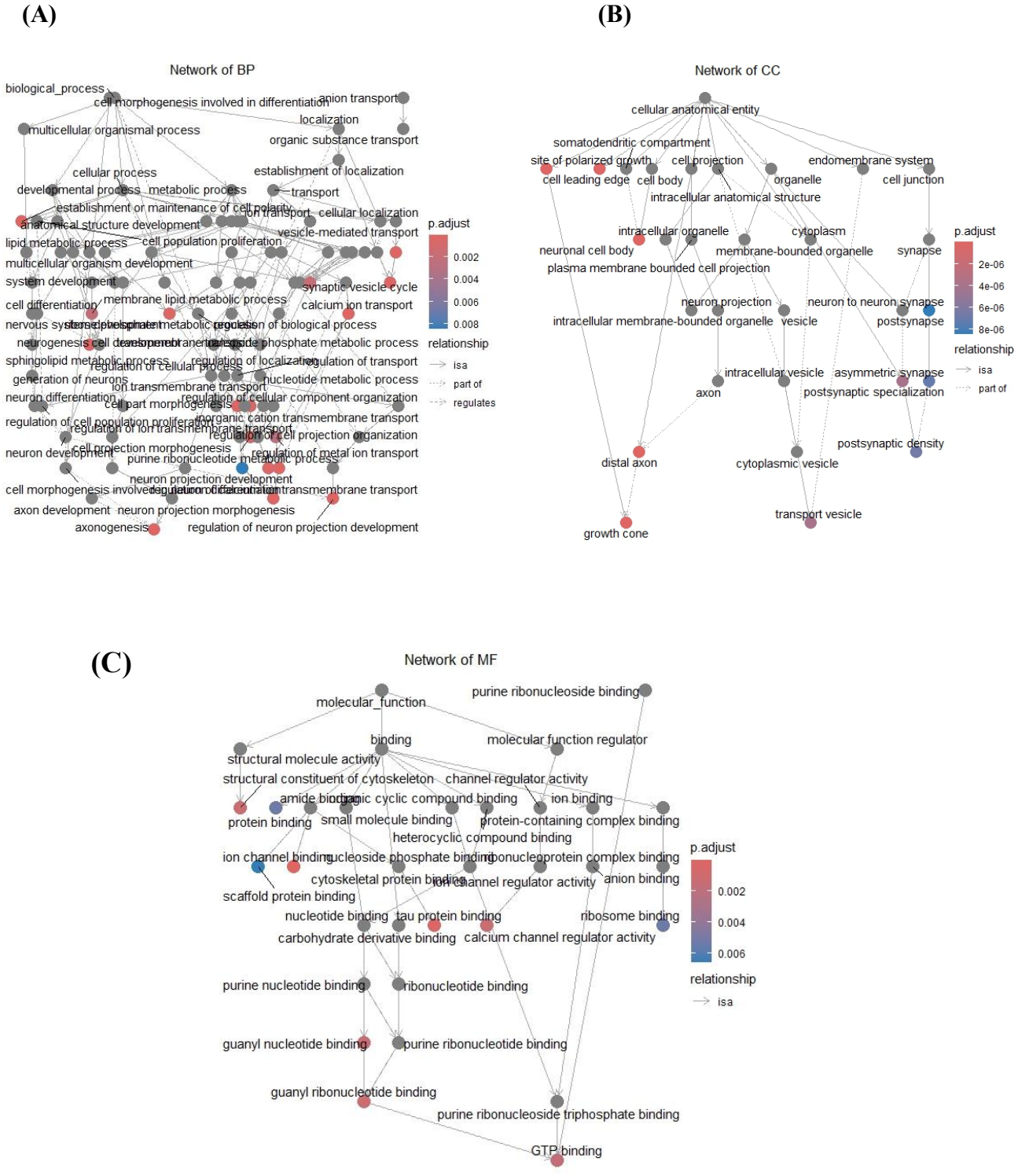

Figure 4.14: Network of Gene Ontologies in Dementia. (A) Network of Biological Process (B) Network of Cellular Component. (C) Network of Molecular Function

4.4.2 KEGG Pathway Enrichment Analysis

In section 3.4.2, we visualized the KEGG Enrichment results of both AD and Dementia through a Bar plot.

Figure 4.15 shows the KEGG Pathways in AD. The bar graph showing the number of genes of various KEGG pathways for a DEGS regarding AD. Each bar of this graph is representing one unique pathway and the height of the bar indicates how many genes are associated with it. The order of the pathways from top to bottom is according to count, with the p53 signaling pathway being the smallest and the MAPK signaling pathway being the largest. The color of the bar is provided by the adjusted p-value (p.adjust); it goes from red for small p.adjust values (more significant) to blue for big p.adjust values less significant. The pathways are colored red, such as p53 signaling pathway, Hippo signaling pathway, and hepatocellular carcinoma, which indicate that the adjusted p-value is lower, which in turn means higher significance. Other pathways like Cushing syndrome, chronic myeloid leukemia, and propanoate metabolism are shown in blue, meaning that they have a higher adjusted p-value, and hence lower significance.

A comprehensive enumeration of highly significant pathways characterized by high gene counts and low adjusted p-values encompasses the MAPK signaling pathway, proteoglycans in cancer, and the TGF-beta signaling pathway, which may be underlined as potentially critical within the context of Alzheimer's disease. Other notable pathways represented are malignancies such as pancreatic, gastric, breast, and hepatocellular carcinoma along with endocrine resistance and chronic myeloid leukemia. These thus point to the complex interaction of AD with these biological processes. KEGG IDs with p-value and count are as shown in Table 4.13.

Figure 4.16 illustrates a bar graph representing several KEGG pathways associated with dementia, organized hierarchically based on the significance of the adjusted p-values. Amongst identified conditions, neurodegenerative disorders prominently rank as the conditions include conditions like Huntington's disease, Prion disease, Parkinson's disease, and Amyotrophic lateral sclerosis which account for a strong relationship of the conditions with dementia; another notable pathways being in the list include Salmonella infection, Sphingolipid signaling pathway, Sphingolipid metabolism. There are also those in the subcategories of cellular and systemic functions, including "Fc gamma R-mediated phagocytosis," "Chemical carcinogenesis reactive oxygen species," "Apoptosis," "Thermogenesis," and "Oxytocin signaling pathway." Metabolic and cellular pathways end with "Oxidative phosphorylation," "Endocytosis," "Non-alcoholic fatty liver disease," and "Diabetic cardiomyopathy," which explain that the condition of dementia is very complicated and multifaceted. The gradient color spectrum from red to blue indicates p.adjust, where red means it is highly significant with a p.adjust of 0.001 and blue indicates lesser significance with a p.adjust of 0.003. Such a presentation will bring out all relevant biological pathways that are related to the pathology of dementia. It may prove helpful in further studies for the selection of potential therapeutic targets. Table 4.14. KEGG ids along with corresponding p-values and counts.

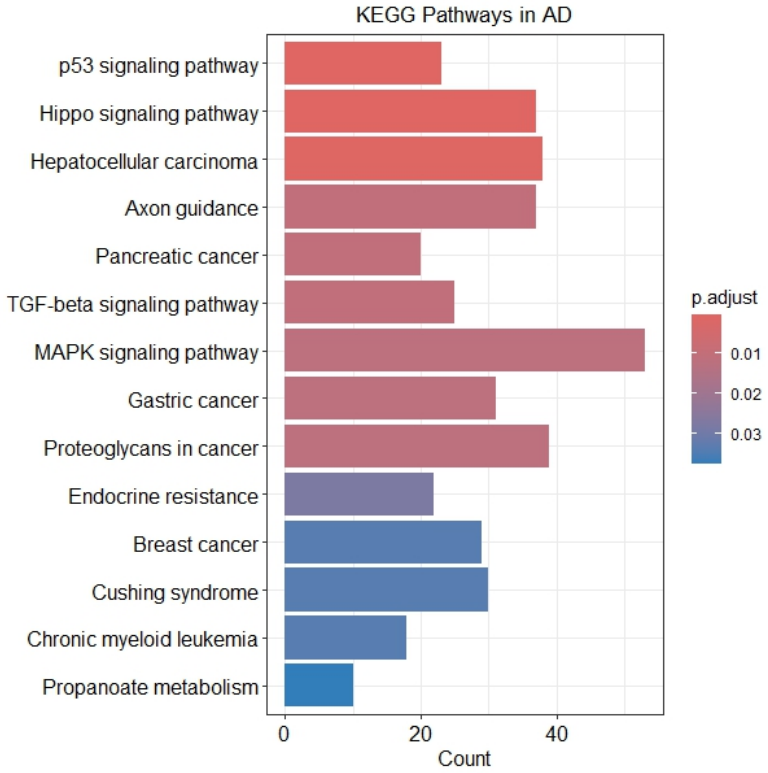

Figure 4.15: Bar plot of KEGG pathways in AD

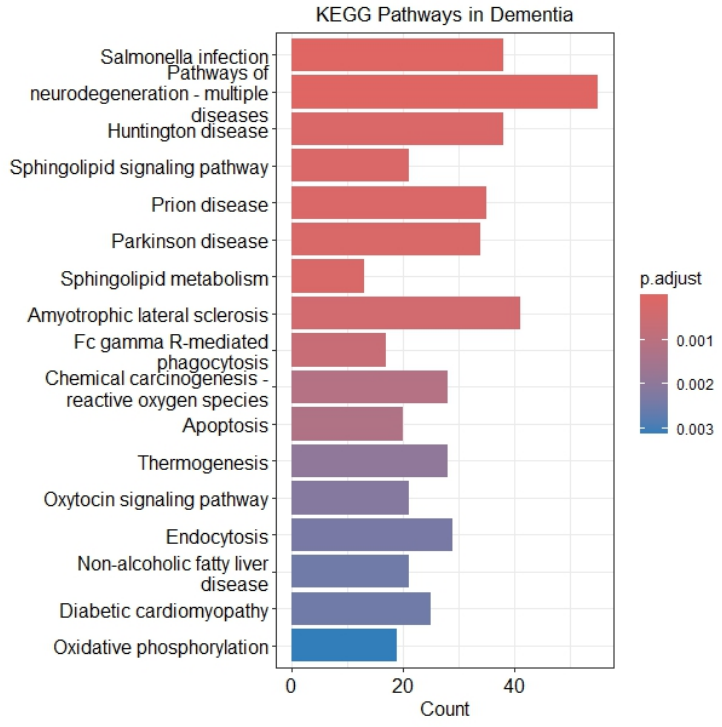

Figure 4.16: Bar plot of KEGG pathways in Dementia.

Table 4.13: KEGG IDs of Pathways involved in AD

ID	Description	p-value	Count
hsa04115	p53 signaling pathway	2.9329E-06	23
hsa04390	Hippo signaling pathway	4.6099E-06	37
hsa05225	Hepatocellular carcinoma	9.6345E-06	38
hsa04360	Axon guidance	0.00014249	37
hsa05212	Pancreatic cancer	0.00014459	20
hsa04350	TGF-beta signaling pathway	0.00021385	25
hsa04010	MAPK signaling pathway	0.00031235	53
hsa05226	Gastric cancer	0.00031525	31
hsa05205	Proteoglycans in cancer	0.00040228	39
hsa01522	Endocrine resistance	0.00078345	22
hsa05224	Breast cancer	0.00119137	29
hsa05220	Chronic myeloid leukemia	0.00119369	18
hsa04934	Cushing syndrome	0.00135988	30
hsa00640	Propanoate metabolism	0.00164336	10

Table 4.14: KEGG IDs of Pathways involved in AD

ID	Description	p-value	Count
hsa05132	Salmonella infection	7.2795E-09	38
hsa05022	Pathways of neurodegeneration - multiple diseases	1.1001E-07	55
hsa05016	Huntington disease	2.112E-06	38
hsa04071	Sphingolipid signaling pathway	2.261E-06	21
hsa05020	Prion disease	2.7235E-06	35
hsa05012	Parkinson disease	4.1019E-06	34
hsa00600	Sphingolipid metabolism	4.5156E-06	13
hsa05014	Amyotrophic lateral sclerosis	1.0606E-05	41
hsa04666	Fc gamma R-mediated phagocytosis	1.8777E-05	17
hsa05208	Chemical carcinogenesis - reactive oxygen species	3.6935E-05	28
hsa04210	Apoptosis	4.3717E-05	20
hsa04714	Thermogenesis	7.4575E-05	28
hsa04921	Oxytocin signaling pathway	8.7842E-05	21
hsa04144	Endocytosis	0.00010427	29
hsa04932	Non-alcoholic fatty liver disease	0.00011637	21
hsa05415	Diabetic cardiomyopathy	0.0001241	25
hsa00190	Oxidative phosphorylation	0.0001661	19
hsa04145	Phagosome	0.00032662	20
hsa04216	Ferroptosis	0.00035011	9

4.5 Identification of Hub Genes

In section 3.5, we identified the hub genes in AD and Dementia through the LASSO (Least Absolute Shrinkage and Selection Operator) Algorithm. LASSO is a statistical technique whose main purpose is feature selection and regularization of models.

Figure 4.17 shows the LASSO regression plot of expression of the top 10 hub genes in AD conditions. Coefficients are a function of the regularization parameter, Lambda, on a log scale. Each line corresponds to a gene: as Lambda increases, the coefficient for each gene shrinks toward zero, meaning it has been regularized. Notably, at alpha 0.1, the genes FAM114A2, DNAJC10, SOCS6, MRPL4P1L1, MED11, TCF3, E4F1, NUP88, RBSN, and SCAF11 show relatively higher coefficients even with higher Lambda values, suggesting their relevance to AD. Color of genes represents particular genes of interest; on the right-hand side is a legend where names are associated with their corresponding colors. The Hub genes along with their coefficients are as listed in Table 4.14.

Figure 4.18 shows the Lasso regression plot that includes coefficient profiles of the top 10 genes ordered by their order in Dementia as a function of the regularization parameter log Lambda. The x-axis represents the log-transformed regularization parameter lambda, and the y-axis represents the coefficients for genes. As one moves left to right along this plot, more genes' coefficients are shrunk to zero as regularization increases. It identifies the genes at alpha 0.7 that are at the intersection with IGHMBP2, CCHCR1, ENTPD4.1, THBS3, FKBP11, NA.21, SOCS6, RPL5, WWOX, and DKC1, among these, IGHMBP2 is significantly positively, and DKC1 significantly negatively associated with dementia at the lower Lambda values, and thus likely of importance as potential biomarkers or therapeutic targets in dementia. Hub genes along with coefficients are mentioned in Table 4.15.

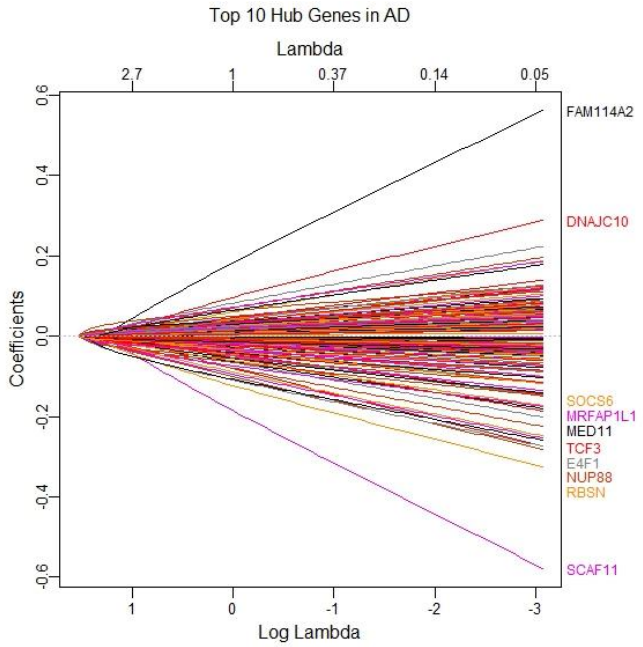

Figure 4.17: Top 10 Hub Genes of AD

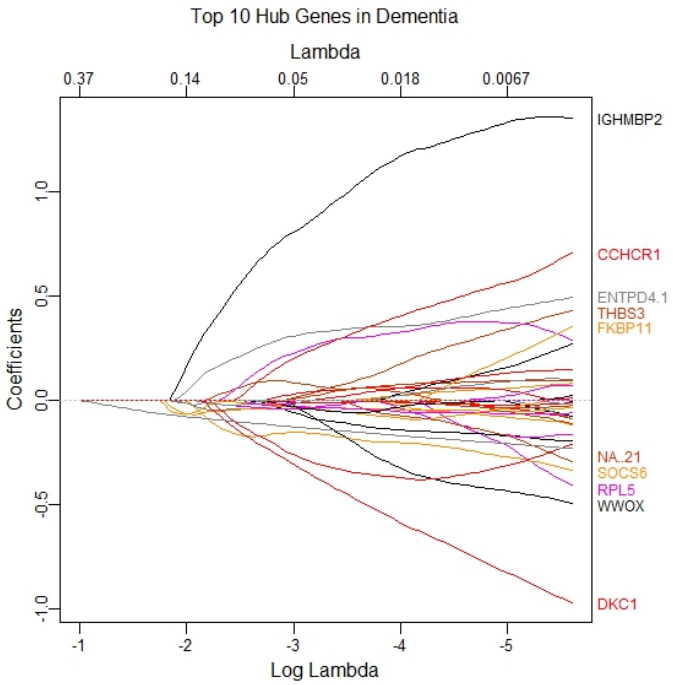

Figure 4.18: Top 10 Hub Genes of Dementia

Table 4.15: Coefficients of AD hub genes.

Gene	Coefficients
FAM114A2	0.465765
DNAJC10	0.306105
SOCS6	-0.21614
MRFAP1L1	-0.26009
MED11	-0.23804
TCF3	-0.25149
E4F1	-0.22901
NUP88	-0.24014
RBSN	-0.29771
SCAF11	-0.48783

Table 4.16: Coefficients of Dementia hub genes.

Gene	Coefficients
IGHMBP2	0.871737
CCHCR1	0.435546
ENTPD4.1	0.504457
THBS3	0.074929
FKBP11	0.038264
NA.21	-0.27334
SOCS6	-0.09019
RPL5	-0.88993
WWOX	-0.060786
DKC1	-0.15655

4.5.1 Cross-validation of Lasso Regression

This section plots the coefficient distribution for AD and Dementia. The plot shows the process of selecting the appropriate value of the regularization parameter (λ) to optimize between model complexity and prediction ability.

Figure 4.19 is the distribution plot of the coefficients for AD, and the lowest value of lambda_min is 0.006. That is shown as the farthest left vertical dashed line on the plot, which aligns with the optimal amount of regularization strength that can minimize binomial deviance. This value of λ will bring the best-fitting model that provides a balance between bias and variance. The right-most vertical dashed line gives another, more parsimonious model: within one standard error of minimum deviance, it gives a much sparser model with a lot fewer features. The mean binomial deviance goes in red dots in the plot; the error bars trace its variability. The number of nonzero coefficients at the top of each plot indicates a compromise between model complexity and the quality of fit. These plots help one to find proper λ; this is how deviance minimized is traded for the simplicity of the model.

Figure 4.20 shows a plot of binomial deviance versus Log(λ) (lambda) linked with Dementia, which exhibits a minimum value of lambda_min of 0.003. The x-axis is the logarithm of the regularization parameter λ, and the y-axis is the binomial deviance, which acts as a measure of the fit of the model. Red dots represent the average deviance for each value of λ, and grey bars represent the error (noise) in those estimates. The value of λ = 0.003 is associated with a log(λ) value that is roughly situated between –2 and –3 along the x-axis. This region is characterized by low and relatively stable deviance, indicating that the performance is high and complexity is effectively balanced. For example, in the context of dementia research, selecting an appropriate λ for a model means that it must have uncovered strong signals within the data that are either genetic signals or some clinical markers of the condition being in question, yet is overfitting-free. The selected value of λ must be sufficiently large to mitigate the risk of overfitting, while concurrently being small enough to ensure that the model demonstrates effective generalization to new data—an essential aspect for producing accurate predictions regarding dementia risk or its progression. Examination of the plot indicates that the chosen value of 0.003 for λ appears justifiable, as it is situated within the area characterized by low binomial deviance, thereby suggesting that this model performs effectively.

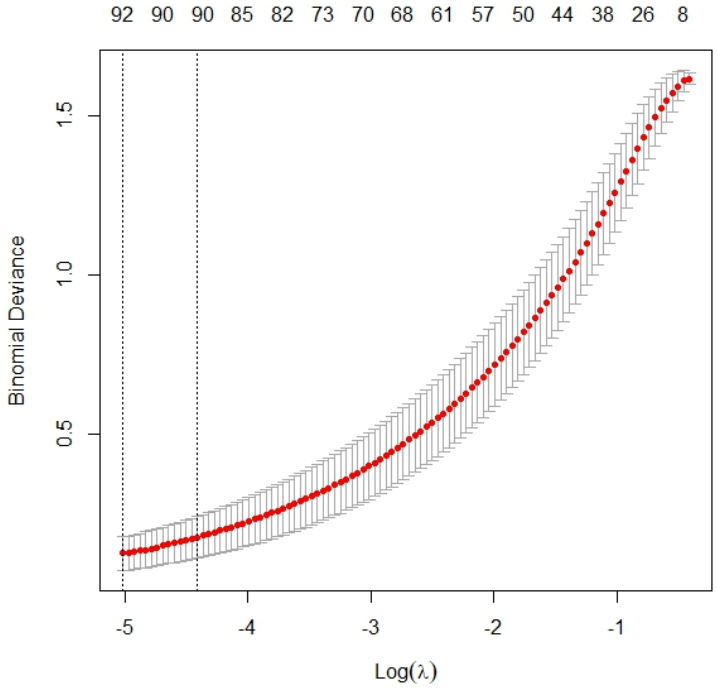

Figure 4.19: coefficient distribution plots for log(lambda) sequence in AD

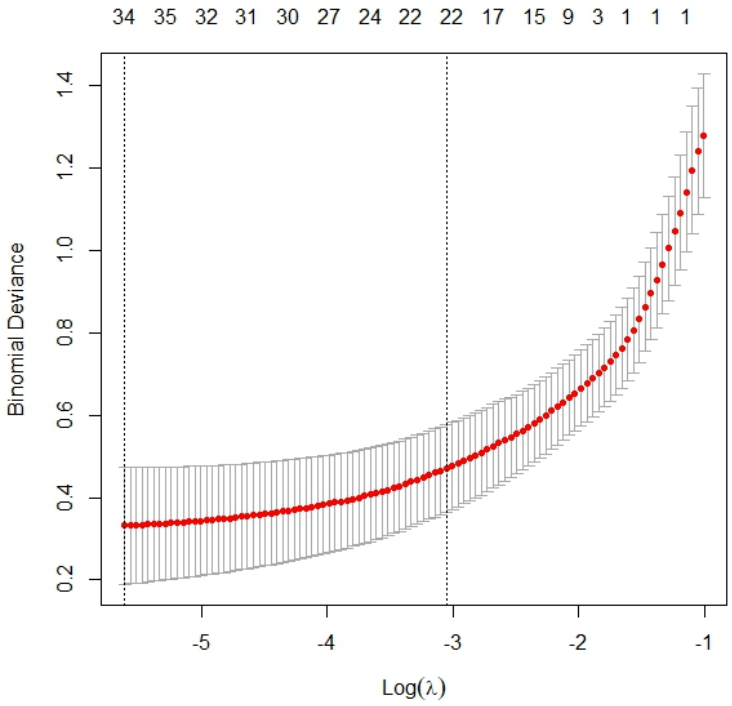

Figure 4.20: coefficient distribution plots for log(lambda) sequence in Dementia

4.6 PPI of Hub Genes

In section 3.5.3, we visualized the identified hub genes through the STRING Database and also found the common genes of AD and Dementia through the visualization of the PPI network.

Figure 4.21 shows PPI network of AD Hub genes, likely downloaded from a database like STRING. Nodes are proteins, with gene or protein names as labels, and edges connecting nodes represent known or predicted interactions. The different colored lines indicate whether evidence exists for the interaction type: experimental data, co-expression, and database annotations. Part of this protein-protein interaction network has implicated cytokine receptors and the JAK-STAT signaling pathway in regulation in SOCS6. Since SOCS6 is associating with IL1R1, PIK3R1, and MAPT, it suggests that it acts like a modulator in immune responses and cellular signaling processes. Through such functions, SOCS6 should be assumed to act as an inhibitory regulator of cytokine signaling, thereby controlling overwhelming inflammatory responses. SOCS6 is shown to reside in the network to indicate how it will also play the role in controlling precise signaling pathway and, more important, by interacting with other proteins signalling.

Figure 4.22 Demonstrates the network of interactions for Dementia Hub genes. These different nodes of the proteins might demonstrate the kind of interaction that involves a protein with its counterpart on forming an interaction or relationship represented by the lines and therefore different proteins such as RPL5, PPP2CB, and FBXO38. The different edge colors represent differences in the nature of evidence for interaction; for example, some are experimental evidence and others are database or co-expression evidence. This general structure suggests the proteins are connected in some biological context, perhaps by a pathway or process. The major nodes are SOCS6, RPL5, and VAV2, which suggests that there are at least some 'hub' proteins that are critical in interactions.

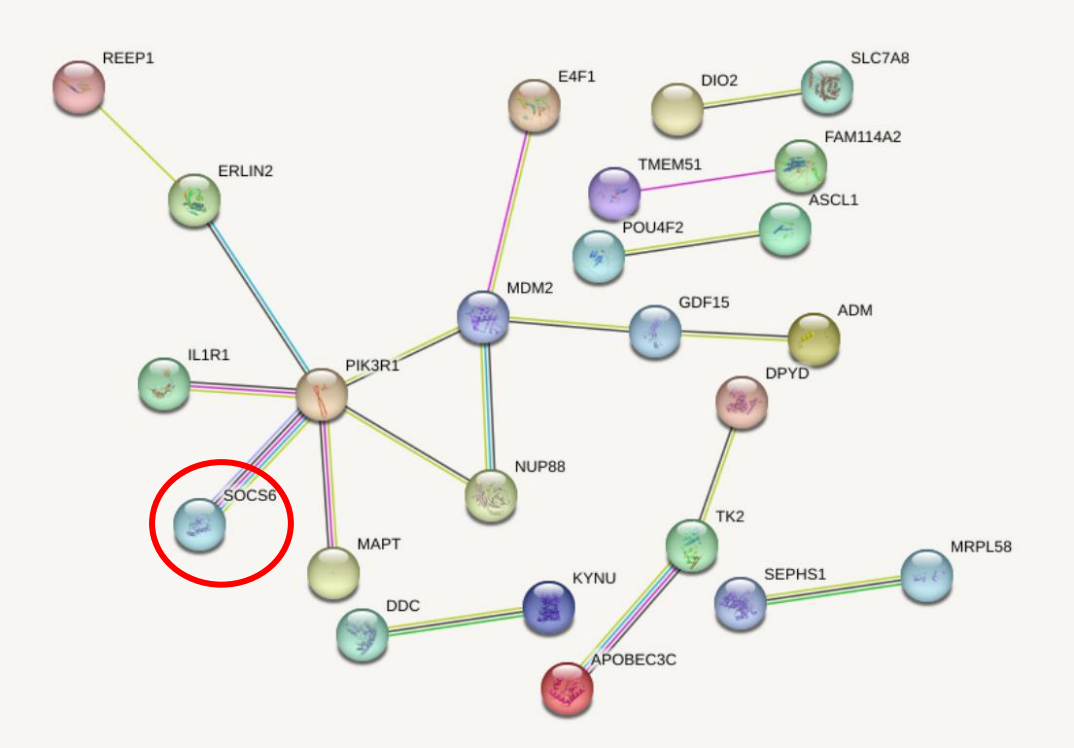

Figure 4.21: PPI Network of AD Hub Genes

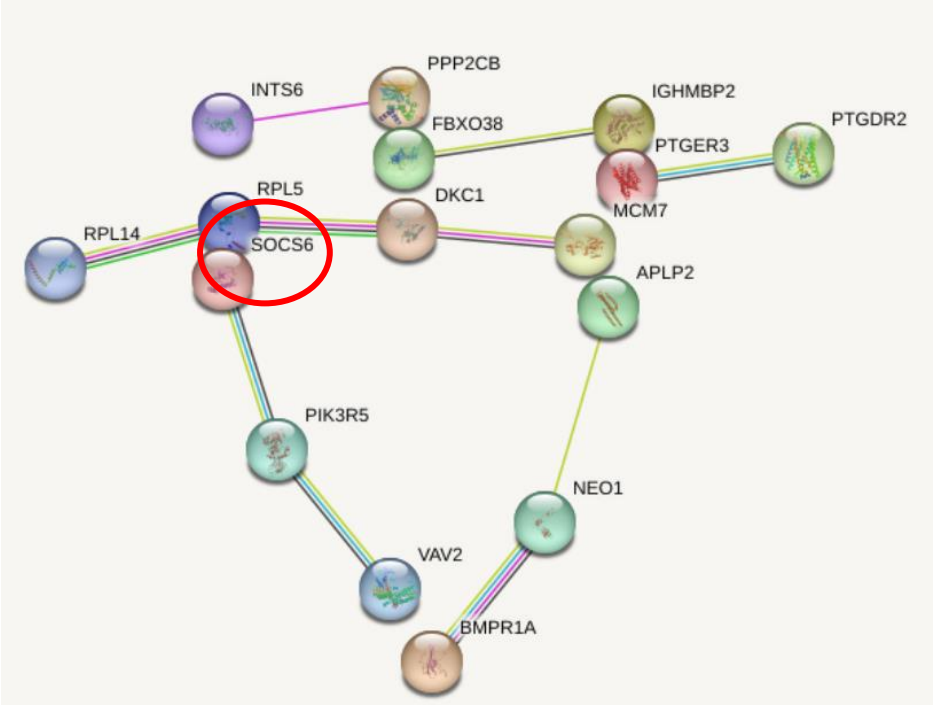

Figure 4.22: PPI Network of Dementia Hub Genes

4.7 Common Hub Gene in AD and Dementia

In the 3.5.2 section, we identified common hub genes from dementia and AD hub genes through intersect function in R.

Figure 4.23 shows Venn diagram indicating the common core gene that connects two varied conditions-Alzheimer's disease and dementia. The right-hand side orange circle shows the unique genes of Alzheimer's disease containing 91 genes labeled "AD". Another circle, on the right side, is filled in blue and named "DM," with genes associated uniquely with Dementia within count 27. The smaller overlap between the two circles represents one common hub gene that overlaps in both conditions. There is one gene for the presence of SOCS6 in both conditions. The given diagram depicts the overlapping genetics between AD and DM wherein SOCS6 can act as a significant component for both diseases. Furthermore, the SOCS6 gene is also among the top 10 genes in both conditions. SOCS6 coefficient value in AD is -0.21614 and in Dementia is -0.09019. This can be said that SOCS6 is negatively correlated with AD and Dementia but according to the coefficient value AD shows a strong negative association.

4.8 AUC Curve of Common Gene

In the 3.5.4 section, we calculate the AUC value of common genes in both conditions (AD and Dementia).

Figure 4.24 shows the performance of the binary classification system for both conditions, Alzheimer's Disease and Dementia, on the ROC curve for the common gene SOCS6. Here, the classifier for AD is the blue line with an AUC of 1.0. This is a perfect sensitivity and specificity meaning that the model completely differentiates between the presence and absence of AD. Instead, the red line implies that performance is related to DM with an AUC of 0.81, meaning that it can classify fairly well but still has imperfections. Because the value is close to 1, it discriminates the positive cases from negative cases to a fair extent. The x-axis in this receiver operating characteristic curve represents specificity and the y-axis sensitivity. The plot suggests that the SOCS6 gene classifier is very good for AD but only reasonably good for DM, with some scope for improvement.

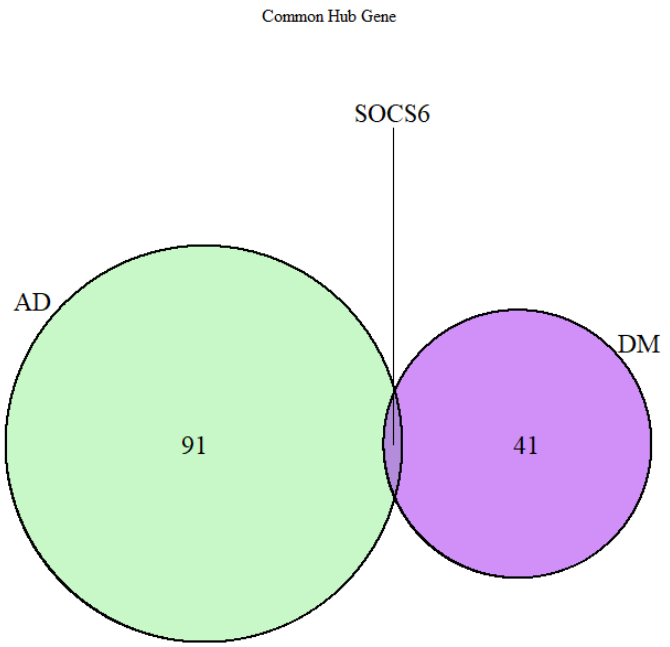

Figure 4.23: Common Hub Gene in AD and Dementia

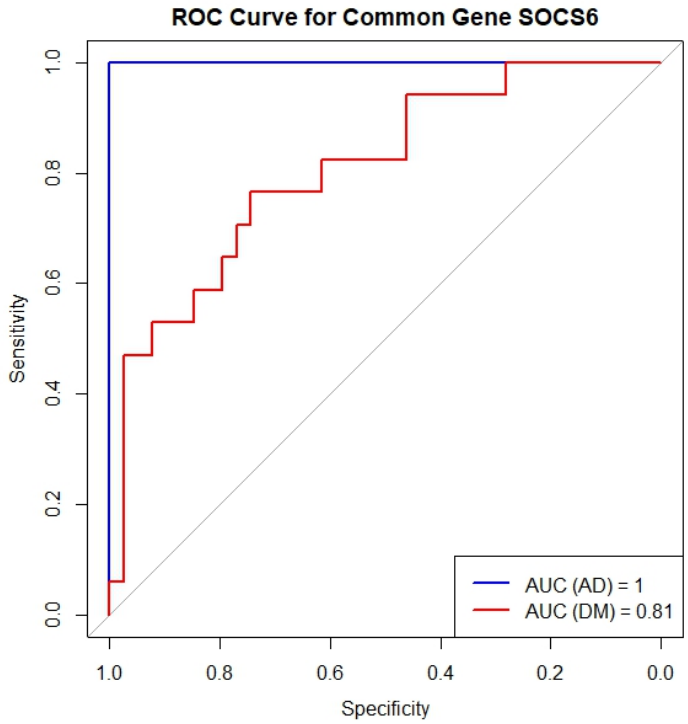

Figure 4.24: AUC Curve of SOCS6 Gene.

4.9 Identification of Functional Role of SOCS6

In section 3.5, we analyze the identified common hub gene in the signaling pathway which is involved in neuroinflammatory disease.

4.9.1 JAK/STAT Pathway

The JAK/STAT pathway is a major regulator of immune responses and inflammation, issues directly related to AD and other neurodegenerative dementias. In this regard, this pathway has been implicated as misregulated in AD pathogenesis, wherein it contributes to neuroinflammation and neuronal damage. In AD, overactive JAK/STAT signaling results in increased production of the pro-inflammatory cytokines that further activate the inflammatory environment in the brain. This might account for chronic neuroinflammation, leading to neuronal injury and death, which underlies the characteristic decline in cognition in patients with dementia [68]. Moreover, JAK/STAT pathway involvement in glial cell activity further underlines its importance in neurodegenerative diseases and therefore stays as a potential therapeutic target for intervention against inflammation and neuronal protection in AD and dementia [69].

The participation of JAK2 in different inflammatory processes, and its activation in response to cytokine signaling, may yield neuroinflammation and neuronal damage typical of AD. Thus, dysregulation of JAK2 and its downstream effects on the STAT3 pathway may contribute much to the development of neurodegenerative diseases, making JAK2 one of the key targets for possible therapeutic interventions [70]. It has been supposed that STAT3 plays a central role in mediating inflammatory responses, and activation has been shown in some neuroinflammatory diseases. Its dysregulation has been related to AD pathogenesis through the promotion of proinflammatory cytokine production, reactive gliosis, and neuronal damage. This could be achieved by targeting STAT3 and its upstream regulators, such as JAK kinases, which may be one of the potential modulators for curbing neuroinflammation and alleviating AD and other dementia.

SOCS6 is an inhibitor of excess activation of this pathway by inducing degradation of JAKs and inhibiting STATs. SOCS6 plays a protective role in controlling inflammatory signaling through the JAK/STAT pathway, and its dysfunction may contribute to the pathological inflammation and neurodegeneration observed in AD and dementia.

In the context of AD and dementia, SOCS6 modulates neuroinflammation and neuronal health. The increase in the levels of pro-inflammatory cytokines in AD could result in an enhanced activation of the JAK/STAT pathway responsible for neurodegeneration. SOCS6 will act to temper this effect by controlling the duration and strength of the inflammation response. Altered activity/expression of SOCS6 may thus lead to uncontrolled JAK/STAT activation, which heightens inflammation and neuronal damage.

4.10 Discussion

Alzheimer's disease (AD) and dementia are two of the most prevalent neurodegenerative disorders in the world, whose molecular basis is only partly understood and whose therapeutic interventions have been marked by modest efficacy to prevent the worsening of the disease or fully reverse the symptoms.This study attempted to bridge these gaps by exploring the common hub genes to understand the pathophysiological mechanisms of these diseases. Using a computational methodology with differential gene expression profiling, Multivariate Analysis of Covariance (MANCOVA), and Least Absolute Shrinkage and Selection Operator (LASSO) regression, we identified 192 DEGs in AD and 1,035 DEGs in dementia. Functional enrichment analysis revealed DEGs are closely associated with key biological processes in neurodegeneration such as neuron projection development, axonogenesis, and synapse organization. These findings emphasize synaptic dysfunction and neuronal structural changes as the main protagonists of AD and dementia pathogenesis and stress the significance of specific molecular interventions. Our results highlight synaptic integrity as a determinant of neurodegeneration, in line with previous work that has shown synaptic dysfunction to be a leading cause of cognitive impairment. Functional enrichment analysis further revealed shared molecular mechanisms, including DNA-binding transcription factor activity and GTPase regulation, suggesting common disruptions in gene regulation and intracellular signaling pathways. Among the key hub genes identified, APP, PSEN1, PSEN2, MAPT, APOE, and TREM2 were reaffirmed as critical molecular regulators in both AD and dementia. These genes play essential roles in amyloid-beta clearance, tau protein aggregation, and neuroinflammatory responses, all of which are hallmark features of neurodegenerative pathophysiology). Notably, SOCS6 emerged as the only shared hub gene between AD and dementia, demonstrating strong negative correlations (-0.21614 in AD and -0.09019 in dementia) and exceptional predictive accuracy (AUC = 1.0 for AD, AUC = 0.81 for dementia). These findings suggest that SOCS6 may

serve as a potential biomarker and therapeutic target, particularly due to its involvement in modulating neuroinflammatory pathways. In addition, dementia-specific genes such as IGHMBP2 and DKC1 were identified, illustrating the heterogeneity of dementia subtypes and reiterating the need for individualized therapeutic approaches. Previous research has mostly dealt with disease-specific pathways, limiting their potential to identify common molecular targets associated with both AD and dementia. Our research addresses this gap by highlighting convergent genetic and molecular characteristics, supporting the theory that these neurodegenerative diseases have similar etiologic processes, including immune system dysfunctions, dysregulated intracellular signaling, and aberrant synaptic function. Our results imply that in order to effectively slow the progression of the disease, treatment must also target convergent molecular pathways, even if conventional methods have focused on the amyloid-beta and tau protein aggregation theories. Furthermore, the use of machine learning approaches significantly enhanced the predictive potential of identified hub genes, confirming the utility of computational approaches in demystifying complex biological information. These findings open the door to the integration of multi-gene signatures into prediction models, thereby enabling the development of personalized treatment regimens for neurodegenerative disorders. Overall, our research on the shared biological pathways between AD and dementia has identified important hub genes and pathways that may be therapeutic targets. The discovery of SOCS6 as a common hub gene with positive predictive ability highlights its potential as a biomarker and therapeutic target. Furthermore, the identification of dementia-specific genes highlights the need to establish specific treatments for a range of neurodegenerative diseases.

SOCS6 is a negative regulator of the JAK/STAT signaling pathway to prevent excessive inflammation. Since dysregulation of JAK/STAT is involved in neuroinflammation and neurodegenerative diseases, the loss or abnormal expression of SOCS6 may lead to uncontrolled pathway activation and subsequent neuronal damage [67].

To counteract this effect, inhibition of JAK2 and STAT3 with inhibitory molecules is a treatment strategy to enhance SOCS6 activity to re-establish balance to the pathway and reduce neuroinflammation. It is of particular significance in situations where SOCS6 expression is suboptimal, defective, or unable to completely control JAK/STAT activity in pathologic conditions.

The JAK/STAT pathway plays a critical role in immune response and inflammation, which is a significant feature of the pathogenesis of AD. JAK2 and STAT3 dysregulation leads to an overexpression of pro-inflammatory cytokines and, hence, exacerbates neuroinflammation and neuronal damage [71]. SOCS6, as a negative regulator of the pathway, plays a crucial role in dampening excessive inflammatory signaling. The present study focuses on the significance of targeting JAK2 and STAT3 as a therapeutic strategy.

Study Limitations and Future Directions

Despite its findings, this study has several limitations. This study is based on publicly available transcriptome data that is susceptible to sample bias and that may be unable to reflect important regulatory pathways involved in neurodegeneration. In addition, the findings are computational predictions whose biological relevance has to be tested experimentally to establish their importance. Future studies will aim to verify these findings in larger, independent cohorts employing multi omics approaches, including proteomics and metabolomics, in an attempt to surmount these limitations and drive a more robust understanding of the etiologies of disease. Determining the specific roles of such important hub genes such as SOCS6 in AD.

This study identifies SOCS6 as a critical hub gene linked to dementia and Alzheimer's disease (AD) and proposes its possible function as a therapeutic target and biomarker. By combining cutting-edge computational techniques with machine learning technologies, this study clarifies the critical role of SOCS6 in the molecular pathophysiology of neurodegenerative diseases. The information provided here establishes a strong foundation for future studies toward the development of targeted therapeutic strategies, which may ultimately improve clinical outcomes and the quality of life in neurodegenerative disease patients.

Chapter 5

References

5.0 Biblography

[1] D. A. Başer and M. Satilmiş, "Dementia," in *Chronic Disease Follow-Ups for Adults in Primary Care*, 2022. doi: 10.1177/0091217416636579.

[2] J. W. Błaszczyk, "Pathogenesis of Dementia," *International Journal of Molecular Sciences*. 2023. doi: 10.3390/ijms24010543.

[3] M. Pawlowski, A. Johnen, and T. Duning, "Young onset dementia," *Nervenarzt*, 2020, doi: 10.1007/s00115-020-00967-0.

[4] X. P. Mfene and B. J. Pillay, "Dementia-Friendly Communities: Exploring Terms used to Describe Dementia, Attitudes and Reactions Towards People with Dementia in Ilembe District, South Africa," *Dementia*, 2023, doi: 10.1177/14713012231165112.

[5] F. Boller and M. M. Forbes, "History of dementia and dementia in history: An overview," *Journal of the Neurological Sciences*. 1998. doi: 10.1016/S0022-510X(98)00128-2.

[6] F. Assal, "History of Dementia," *Frontiers of Neurology and Neuroscience*. 2019. doi: 10.1159/000494959.

[7] Y. Wang, H. Xu, Z. Geng, G. Geng, and F. Zhang, "Dementia and the history of disease in older adults in community," *BMC Public Health*, 2023, doi: 10.1186/s12889-023-16494-x.

[8] D. J. Oh *et al.*, "Parental history of dementia and the risk of dementia: A cross-sectional analysis of a global collaborative study," *Psychiatry Clinical. Neuroscience.*, 2023, doi: 10.1111/pcn.13561.

[9] J. C. Morris and E. M. McDade, "Alzheimer Disease," *CONTINUUM Lifelong Learning in Neurology*. 2022. doi: 10.1212/CON.0000000000001131.

[10] M. Vally and F. Kathrada, "Understanding alzheimer disease," *South African Fam. Pract.*, 2019, doi: 10.4102/safp.v61i2.4999.

[11] D. S. Knopman *et al.*, "Alzheimer disease," *Nat. Rev. Dis. Prim.*, 2021, doi: 10.1038/s41572-021-00269-y.

[12] J. M. Long and D. M. Holtzman, "Alzheimer Disease: An Update on Pathobiology and

Treatment Strategies," *Cell*. 2019. doi: 10.1016/j.cell.2019.09.001.

[13] I. Aillaud and S. A. Funke, "Tau Aggregation Inhibiting Peptides as Potential Therapeutics for Alzheimer Disease," *Cellular and Molecular Neurobiology*. 2023. doi: 10.1007/s10571-022-01230-7.

[14] D. S. Knopman, R. C. Petersen, and C. R. Jack, "A brief history of 'Alzheimer disease,'" *Neurology*, 2019, doi: 10.1212/wnl.0000000000007583.

[15] D. S. Knopman, R. C. Petersen, and C. R. Jack, "A brief history of 'Alzheimer disease': Multiple meanings separated by a common name," *Neurology*. 2019. doi: 10.1212/WNL.0000000000007583.

[16] K. A. Jellinger, K. Seppi, G. K. Wenning, and W. Poewe, "Impact of coexistent Alzheimer pathology on the natural history of Parkinson's disease," *Journal of Neural Transmission.*, 2002, doi: 10.1007/s007020200027.

[17] G. Li *et al.*, "The validity of the family history method for identifying Alzheimer disease," *Arch. Neurol.*, 1997, doi: 10.1001/archneur.1997.00550170104021.

[18] M. N. Shakir and B. N. Dugger, "Advances in Deep Neuropathological Phenotyping of Alzheimer Disease: Past, Present, and Future," *Journal of Neuropathology and Experimental Neurology*. 2022. doi: 10.1093/jnen/nlab122.

[19] C. Moreno Cervantes *et al.*, "Factors associated with mixed dementia vs Alzheimer disease in elderly Mexican adults," *Neurol. (English Ed.*, 2017, doi: 10.1016/j.nrleng.2015.12.003.

[20] C. R. Jack *et al.*, "Prevalence of Biologically vs Clinically Defined Alzheimer Spectrum Entities Using the National Institute on Aging-Alzheimer's Association Research Framework," *JAMA Neurol.*, 2019, doi: 10.1001/jamaneurol.2019.1971.

[21] E. Han *et al.*, "Severe Hypoglycemia Increases Dementia Risk and Related Mortality: A Nationwide, Population-based Cohort Study," *Journal of Clinical. Endocrinology & Metabolism.*, 2022, doi: 10.1210/clinem/dgab860.

[22] R. N. Taddei *et al.*, "Tau Oligomer-Containing Synapse Elimination by Microglia and Astrocytes in Alzheimer Disease," *JAMA Neurol.*, 2023, doi: 10.1001/jamaneurol.2023.3530.

[23] A. Adali, A. Yirün, B. Koçer Gümüşel, and P. Erkekoğlu, "The possible effects of biological agents on the development of Alzheimer's disease," *Ankara Universitesi Eczacilik Fakultesi Dergisi*. 2020. doi: 10.33483/jfpau.523804.

[24] S. Bylund, L. B. Von Kobyletzki, M. Svalstedt, and Å. Svensson, "Prevalence and incidence of atopic dermatitis: A systematic review," *Acta Dermato-Venereologica*. 2020. doi: 10.2340/00015555-3510.

[25] M. S. Baek *et al.*, "Annual Trends in the Incidence and Prevalence of Alzheimer's Disease in South Korea: A Nationwide Cohort Study," *Frontier of Neurology.*, 2022, doi: 10.3389/fneur.2022.883549.

[26] K. Trevisan, R. Cristina-Pereira, D. Silva-Amaral, and T. A. Aversi-Ferreira, "Theories of aging and the prevalence of Alzheimer's disease," *BioMed Research International*. 2019. doi: 10.1155/2019/9171424.

[27] K. Chidwick *et al.*, "Prevalence, incidence and management of atopic dermatitis in Australian general practice using routinely collected data from MedicineInsight," *Australas. Journal of Dermatology*, 2020, doi: 10.1111/ajd.13268.

[28] H. Huque *et al.*, "Could Country-Level Factors Explain Sex Differences in Dementia Incidence and Prevalence? A Systematic Review and Meta-Analysis," *Journal of Alzheimer's Disease*. 2022. doi: 10.3233/JAD-220724.

[29] B. Dubois, C. A. F. von Arnim, N. Burnie, S. Bozeat, and J. Cummings, "Biomarkers in Alzheimer's disease: role in early and differential diagnosis and recognition of atypical variants," *Alzheimer's Research and Therapy*. 2023. doi: 10.1186/s13195-023-01314-6.

[30] B. S. Appleby, D. Nacopoulos, N. Milano, K. Zhong, and J. L. Cummings, "A review: Treatment of Alzheimer's disease discovered in repurposed agents," *Dementia and Geriatric Cognitive Disorders*. 2013. doi: 10.1159/000345791.

[31] H. Shikanai, T. Shindo, S. Hiraide, S. Kimura, K. Iizuka, and T. Izumi, "New drug treatment strategies for AD/HD," *Proc. Annu. Meet. Japanese Pharmacol. Soc.*, 2022, doi: 10.1254/jpssuppl.96.0_1-b-w03-2.

[32] X. Gu *et al.*, "Hub Genes, Diagnostic Model, and Predicted Drugs Related to Iron

Metabolism in Alzheimer's Disease," *Frontier of Aging Neuroscience*, 2022, doi: 10.3389/fnagi.2022.949083.

[33] F. Zhang, R. jia Zhong, C. Cheng, S. Li, and W. dong Le, "New therapeutics beyond amyloid-β and tau for the treatment of Alzheimer's disease," *Acta Pharmacologica Sinica*. 2021. doi: 10.1038/s41401-020-00565-5.

[34] D. A. Moreira, S. D. Santos, V. Leiro, and A. P. Pêgo, "Dendrimers and Derivatives as Multifunctional Nanotherapeutics for Alzheimer's Disease," *Pharmaceutics*. 2023. doi: 10.3390/pharmaceutics15041054.

[35] N. Reich and C. Hölscher, "The neuroprotective effects of glucagon-like peptide 1 in Alzheimer's and Parkinson's disease: An in-depth review," *Frontiers in Neuroscience*. 2022. doi: 10.3389/fnins.2022.970925.

[36] F. Guo *et al.*, "Applications of Carbon Dots for the Treatment of Alzheimer's Disease," *International Journal of Nanomedicine*. 2022. doi: 10.2147/IJN.S388030.

[37] N. M. Rozumna, V. V. Hanzha, and E. A. Lukyanetz, "Memantine protects the cultured rat hippocampal neurons treated by NMDA and amyloid β1–42," *Frontier of Neuroscience.*, 2023, doi: 10.3389/fnins.2023.1269664.

[38] C. Zhang *et al.*, "Discovery of novel dual RAGE/SERT inhibitors for the potential treatment of the comorbidity of Alzheimer's disease and depression," *Eur. J. Med. Chem.*, 2022, doi: 10.1016/j.ejmech.2022.114347.

[39] M. W. Lin, Y. H. Chen, H. Ben Yang, C. C. Lin, and S. Y. Hung, "Galantamine Inhibits Aβ1–42-Induced Neurotoxicity by Enhancing α7nAChR Expression as a Cargo Carrier for LC3 Binding and Aβ1–42 Engulfment During Autophagic Degradation," *Neurotherapeutics*, 2020, doi: 10.1007/s13311-019-00803-7.

[40] R. A. Abdel-Aal, O. A. Hussein, R. G. Elsaady, and L. A. Abdelzaher, "Naproxen as a potential candidate for promoting rivastigmine anti-Alzheimer activity against aluminum chloride-prompted Alzheimer's-like disease in rats; neurogenesis and apoptosis modulation as a possible underlying mechanism," *Eurpeon. Journal of Pharmacology.*, 2022, doi: 10.1016/j.ejphar.2021.174695.

[41] B. C. Tang, Y. T. Wang, and J. Ren, "Basic information about memantine and its treatment of Alzheimer's disease and other clinical applications," *Ibrain*. 2023. doi: 10.1002/ibra.12098.

[42] P. M. Salkovskis, "Chapter 7 Cognitive–behavioural therapy," in *Understanding Counselling and Psychotherapy*, 2010.

[43] M. Bleibel, A. El Cheikh, N. S. Sadier, and L. Abou-Abbas, "The effect of music therapy on cognitive functions in patients with Alzheimer's disease: a systematic review of randomized controlled trials," *Alzheimer's Research and Therapy*. 2023. doi: 10.1186/s13195-023-01214-9.

[44] J. C. Millán-Calenti, L. Lorenzo-López, B. Alonso-Búa, C. de Labra, I. González-Abraldes, and A. Maseda, "Optimal nonpharmacological management of agitation in alzheimer's disease: Challenges and solutions," *Clin. Interv. Aging*, 2016, doi: 10.2147/CIA.S69484.

[45] D. Zhang *et al.*, "Bioinformatics identification of potential biomarkers and therapeutic targets for ischemic stroke and vascular dementia," *Exp. Gerontol.*, 2024, doi: 10.1016/j.exger.2024.112374.

[46] J. Xu *et al.*, "Identification of key genes and signaling pathways associated with dementia with Lewy bodies and Parkinson's disease dementia using bioinformatics," *Frontier of Neurology*, 2023, doi: 10.3389/fneur.2023.1029370.

[47] J. Shu, W. Wei, and L. Zhang, "Identification of Molecular Signatures and Candidate Drugs in Vascular Dementia by Bioinformatics Analyses," *Front. Mol. Neurosci.*, 2022, doi: 10.3389/fnmol.2022.751044.

[48] H. Gui, Q. Gong, J. Jiang, M. Liu, and H. Li, "Identification of the Hub Genes in Alzheimer's Disease," *Comput. Math. Methods Med.*, 2021, doi: 10.1155/2021/6329041.

[49] Q. Jing *et al.*, "A Comprehensive Analysis Identified Hub Genes and Associated Drugs in Alzheimer's Disease," *Biomed Res. Int.*, 2021, doi: 10.1155/2021/8893553.

[50] L. Wang, H. Qian, and L. Wang, "Identification of Key Genes Related with Alzheimer's Disease Treatment Through Bioinformatics Analysis," *Journal of Biological Life Sciences.*, 2018, doi: 10.5296/jbls.v9i1.12474.

[51] B. Vastrad, C. Vastrad, and C. Nilaya, "Identification of Key Pathways and Genes in Dementia via Integrated Bioinformatics Analysis," *bioRxiv*, 2021.

[52] Y. Cheng, M. Sun, F. Wang, X. Geng, and F. Wang, "Identification of Hub Genes Related to Alzheimer's Disease and Major Depressive Disorder," *Am. Journal of Alzheimers. Disorder and Other Dementia.*, 2021, doi: 10.1177/15333175211046123.

[53] C. Fenoglio, E. Scarpini, M. Serpente, and D. Galimberti, "Role of Genetics and Epigenetics in the Pathogenesis of Alzheimer's Disease and Frontotemporal Dementia," *Journal of Alzheimer's Disease*. 2018. doi: 10.3233/JAD-170702.

[54] H. Stocker *et al.*, "Alzheimer's polygenic risk scores, APOE, Alzheimer's disease risk, and dementia-related blood biomarker levels in a population-based cohort study followed over 17 years," *Alzheimer's Research Therapy*, 2023, doi: 10.1186/s13195-023-01277-8.

[55] S. Silvaieh *et al.*, "Comprehensive genetic screening of early-onset dementia patients in an Austrian cohort-suggesting new disease-contributing genes," *Human Genomics*, 2023, doi: 10.1186/s40246-023-00499-z.

[56] L. Mathioudakis *et al.*, "Study of Alzheimer's disease- and frontotemporal dementia-associated genes in the Cretan Aging Cohort," *Neurobiology Aging*, 2023, doi: 10.1016/j.neurobiolaging.2022.07.002.

[57] T. Guo, D. Zhang, Y. Zeng, T. Y. Huang, H. Xu, and Y. Zhao, "Molecular and cellular mechanisms underlying the pathogenesis of Alzheimer's disease," *Molecular Neurodegeneration*. 2020. doi: 10.1186/s13024-020-00391-7.

[58] G. K. Smyth *et al.*, "RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR," *F1000Research*, 2018, doi: 10.12688/f1000research.9005.3.

[59] G. R. Brown *et al.*, "Gene: A gene-centered information resource at NCBI," *Nucleic Acids Res.*, 2015, doi: 10.1093/nar/gku1055.

[60] T. Barrett *et al.*, "NCBI GEO: Archive for functional genomics data sets - Update," *Nucleic Acids Res.*, 2013, doi: 10.1093/nar/gks1193.

[61] T. Barrett *et al.*, "NCBI GEO: Mining tens of millions of expression profiles - Database and tools update," *Nucleic Acids Reearch.*, 2007, doi: 10.1093/nar/gkl887.

[62] M. Khalid, S. Khan, J. Ahmad, and M. Shaheryar, "Multivariate Covariance using Principal Component Analysis for Reconstruction of Bidirected Gene Regulatory Networks," in *Proceedings - 2017 International Conference on Frontiers of Information Technology, FIT 2017*, 2017. doi: 10.1109/FIT.2017.00048.

[63] M. Khalid, S. Khan, J. Ahmad, and M. Shaheryar, "Identification of self-regulatory network motifs in reverse engineering gene regulatory networks using microarray gene expression data," *IET System. Biology.*, 2019, doi: 10.1049/iet-syb.2018.5001.

[64] J. Piñero *et al.*, "The DisGeNET knowledge platform for disease genomics: 2019 update," *Nucleic Acids Res.*, 2020, doi: 10.1093/nar/gkz1021.

[65] M. Safran *et al.*, "The GeneCards Suite," in *Practical Guide to Life Science Databases*, 2022. doi: 10.1007/978-981-16-5812-9_2.

[66] A. P. Davis, T. C. Wiegers, R. J. Johnson, D. Sciaky, J. Wiegers, and C. J. Mattingly, "Comparative Toxicogenomics Database (CTD): update 2023," *Nucleic Acids Reearch.*, 2023, doi: 10.1093/nar/gkac833.

[67] G. Wang, X. Shen, X. Song, N. Wang, X. Wo, and Y. Gao, "Protective mechanism of gold nanoparticles on human neural stem cells injured by β-amyloid protein through miR-21–5p/SOCS6 pathway," *Neurotoxicology*, 2023, doi: 10.1016/j.neuro.2022.12.011.

[68] S. Gupta, K. Mishra, A. Surolia, and K. Banerjee, "Suppressor of cytokine Signalling-6 promotes neurite outgrowth via JAK2/STAT5-Mediated signalling pathway, involving negative feedback inhibition," *PLoS One*, 2011, doi: 10.1371/journal.pone.0026674.

[69] G. C. Wikumpriya, M. W. S. Prabhatha, J. Lee, and C. H. Kim, "Molecular Cloning and Gene Expression of Type I Suppressors of Cytokine Signaling 6 and 7 (SOCS6 and SOCS7) in Whiteleg Shrimp (Litopenaeus vannamei)," *Fishes*, 2023, doi: 10.3390/fishes8080416.

[70] Y. Tian *et al.*, "Identification of diagnostic signatures associated with immune infiltration in Alzheimer's disease by integrating bioinformatic analysis and machine-learning strategies," *Frontier Aging Neuroscence.*, 2022, doi: 10.3389/fnagi.2022.919614.

[71] W. Li *et al.*, "Folic Acid Alters Methylation Profile of JAK-STAT and Long-Term Depression Signaling Pathways in Alzheimer's Disease Models," *Molecular Neurobiology*, 2016, doi: 10.1007/s12035-015-9556-9.