

T.R
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

**A PERSONALIZED APPROACH ON MOLECULAR
MECHANISMS OF ALZHEIMER'S DISEASE:
MAPPING TRANSCRIPTOME ON PROTEIN INTERACTOME**

BETÜL CEYLAN
A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE
DEPARTMENT OF BIOENGINEERING
BIOINFORMATICS AND SYSTEMS BIOLOGY PROGRAMME

GEBZE
2023

T.R
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

**A PERSONALIZED APPROACH ON
MOLECULAR MECHANISMS OF
ALZHEIMER'S DISEASE: MAPPING
TRANSCRIPTOME ON PROTEIN
INTERACTOME**

BETÜL CEYLAN

**A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE
DEPARTMENT OF BIOENGINEERING
BIOINFORMATICS AND SYSTEMS BIOLOGY PROGRAMME**

**THESIS SUPERVISOR
PROF. DR. TUNAHAN ÇAKIR**

GEBZE

2023

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ALZHEİMER HASTALIĞININ MOLEKÜLER
MEKANİZMALARINA KİŞİSELLEŞTİRİLMİŞ
BİR YAKLAŞIM: TRANSKRİPTOM
VERİLERİNİ PROTEİN İNTERAKTOMUNA
HARİTALAMA

BETÜL CEYLAN
YÜKSEK LİSANS TEZİ
BİYOMÜHENDİSLİK ANABİLİM DALI
BİYOİNFORMATİK VE SİSTEM BİYOLOJİSİ PROGRAMI

DANIŞMANI
PROF. DR. TUNAHAN ÇAKIR

GEBZE
2023

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 07/02/2023 tarih ve 2023/11 sayılı kararıyla oluşturulan jüri tarafından 14/02/2023 tarihinde tez savunma sınavı yapılan Betül CEYLAN'ın tez çalışması BİYOMÜHENDİSLİK Anabilim Dalı Biyoinformatik ve Sistem Biyolojisi Programında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE

(TEZ DANIŞMANI) : Prof. Dr. Tunahan ÇAKIR

ÜYE

: Dr. Öğr. Üyesi Pınar PİR

ÜYE

: Prof. Dr. Kazım Yalçın ARĞA

ONAY

Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun

...../...../..... tarih ve/..... sayılı kararı.

SUMMARY

Alzheimer's disease is a long-term, progressive neurodegenerative disease that affects societies socially and economically, and there is no accurate treatment. The molecular mechanism of the disease, which is characterized by amyloid plaques and neurofibrillary tangles, has not yet been fully elucidated. The vast majority of research studies on the disease evaluate the data collectively, which obscures the molecular differences among patients due to the heterogeneous nature of the disease. In this study, a personalized approach was adopted and the transcriptome data of 403 Alzheimer's disease patients was analyzed by mapping it on the human protein interactome individually. To this end, protein interaction subnetworks for each patient were discovered by using two different algorithms, KeyPathwayMiner and DOMINO, and it was found that DOMINO is more effective in capturing genes and functional GO/pathway terms associated with the disease. It was observed that the disease mostly affects the pathways related to mitochondria, immune response, cell death and endoplasmic reticulum. Additionally, it was demonstrated that the personalized approach is significantly more efficient than bulk analysis and provides a more in-depth understanding of the molecular effects of Alzheimer's disease. Moreover, the patients were analyzed in five groups, in line with recently reported molecular subtypes of the disease based on the same transcriptomic dataset, and it was found that each subtype has its own characteristic profile.

Keywords: Alzheimer's disease, Subnetwork discovery, Personalized analysis, Transcriptome data, Disease subtypes

ÖZET

Alzheimer hastalığı, toplumları sosyal ve ekonomik olarak etkileyen, uzun süreli, ilerleyici bir nörodejeneratif hastalıktır ve kesin bir tedavisi yoktur. Amiloid plaklar ve nörofibriler yumaklarla karakterize olan hastalığın moleküler mekanizması henüz tam olarak aydınlatılamamıştır. Hastalıkla ilgili araştırmaların büyük çoğunluğu verileri toplu olarak değerlendirmekte, bu da hastalığın heterojen doğası nedeniyle hastalar arasında oluşan moleküler farklılıkları gizlemektedir. Bu çalışmada, kişiselleştirilmiş bir yaklaşım benimsenmiş ve 403 Alzheimer hastasının transkriptom verileri, insan protein interaktomu üzerinde ayrı ayrı haritalanarak analiz edilmiştir. Bu amaçla, KeyPathwayMiner ve DOMINO olmak üzere iki farklı algoritma kullanılarak her hasta için protein etkileşim alt ağları keşfedilmiş ve DOMINO'nun hastalıkla ilişkili genleri ve fonksiyonel GO/yolak terimlerini yakalamada daha etkili olduğu görülmüştür. Hastalığın en çok mitokondri, bağışıklık yanıtı, hücre ölümü ve endoplazmik retikulum ile ilgili yolları etkilediği gözlemlenmiştir. Ayrıca, kişiselleştirilmiş yaklaşımın toplu analizden önemli ölçüde daha verimli olduğu ve Alzheimer hastalığının moleküler etkilerinin daha derinlemesine anlaşılmasını sağladığı gösterilmiştir. Buna ek olarak, hastalar aynı transkriptomik veri setine dayalı olarak hastalığın yakın zamanda bildirilen moleküler alt tipleri doğrultusunda beş grupta analiz edilmiş ve her bir alt tipin kendine özgü karakteristik bir profile sahip olduğu tespit edilmiştir.

Anahtar Kelimeler: Alzheimer hastalığı, Alt-ağ keşfi, Kişiselleştirilmiş analiz, Transkriptom verisi, Hastalık alttipleri

ACKNOWLEDGEMENTS

I sincerely thank to my supervisor Prof Dr. Tunahan AKIR. His guidance and knowledge was great treasure for me to conduct this study. I also deeply thank to my friend, PhD candidate, Elif DÜZ for her support and kindness during this time. I would also like to thank the entire project team who contributed to this study, which is a part of the TÜBİTAK funded project (Project code: 120S824). Science and wisdom are the reflections of the distance traveled by man on his journey through the world. I should also like to thank all the scholars who have contributed to this long journey.

I would like to express my endless gratitude to my beloved parents, Zeynep and Abdul Münip CEYLAN, who taught me the real purpose of life, illuminated my way and have always been by my side with their love and trust. I am also grateful to my brothers Ömer and Bilal CEYLAN and their families and my dear sister Hilal CEYLAN, whose great support I always feel with me.

TABLE of CONTENTS

	<u>Page</u>
SUMMARY	v
ÖZET	vi
ACKNOWLEDGEMENTS	vii
TABLE of CONTENTS	viii
LIST of ABBREVIATIONS and ACRONYMS	x
LIST of FIGURES	xi
1. INTRODUCTION	1
2. BACKGROUND INFORMATION	3
2.1. Alzheimer's disease	3
2.1.1. Amyloid-beta plaques	3
2.1.2. Neurofibrillary Tangles	4
2.1.3. Neuroinflammation	5
2.2. Omics and Transcriptomics	6
2.3. Protein-Protein Interaction Network	10
2.4. Integrating Transcriptomic Data with PPI Network	11
2.4.1. KeyPathwayMiner	13
2.4.2. DOMINO	14
3. METHODS	16
3.1. Transcriptome Datasets	16
3.2. Statistical Analysis	16
3.3. Subnetwork Discovery from Protein-Protein Interaction Data	17

3.4. Enrichment Analysis	18
4. SAMPLE BASED SUBNETWORK ANALYSIS	19
4.1. Comparison of KPM and DOMINO results	19
4.2. Evaluation of sample based results	26
4.3. Comparison of sample based and bulk analysis results	30
4.4 Evaluation of results according to AD subtypes	33
5. CONCLUSION	39
REFERENCES	40
BIOGRAPHY	51

LIST of ABBREVIATIONS and ACRONYMS

<u>Abbreviations and Acronyms</u>	<u>Explanations</u>
ACO	: Ant Colony Optimization
AD	: Alzheimer's disease
AMI	: Active Module Identifier
A β	: Amyloid-beta
BiP	: Binding Immunoglobulin Protein
CNS	: Central Nervous System
COX	: Cytochrome C Oxidase
DAT	: Dopamine transporter
DEGs	: Differentially Expressed Genes
ER	: Endoplasmic Reticulum
HSPs	: Heat Shock Proteins
INES	: Individual Node ExceptionS
KPM	: KeyPathwayMiner
MAMs	: Mitochondria Associated ER Membranes
NFTs	: Neurofibrillary Tangles
PI3K	: Phosphatidylinositol 3-Kinase
PPI	: Protein Protein Interaction
PTMs	: Post-translational Modifications
ROS	: Reactive Oxygen Species
ROSMAP	: Religious Orders Study and Rush Memory and Aging Project
SUMO	: Small Ubiquitin-Like Modifier
TLRs	: Toll-Like Receptors
UBL	: Ubiquitin-Like Protein
UPS	: Ubiquitin-Proteasome System

LIST of FIGURES

<u>Figure No:</u>	<u>Page</u>
2.1: Proteolytic process of APP and A β plaque formation.	4
2.2: Formation of paired helical filaments and NFTs.	5
2.3: Comparison of microarray and RNA-Seq technologies.	8
2.4: Illustration of scale free network.	10
2.5: Schematic representation of subnetwork discovery process.	12
2.6: Schematic illustration of steps of DOMINO.	15
3.1: Human PPI network used in this study.	17
4.1: Boxplots of number of genes in the downregulated and upregulated subnetworks discovered by DOMINO and KPM.	20
4.2: Boxplots showing the range of the number of downregulated and upregulated DEGs for each AD patient from ROSMAP cohort.	21
4.3: Comparison of the number of occurrences of selected AD-related terms in the enrichment results of the downregulated subnetworks discovered with DOMINO and KPM.	23
4.4: Comparison of the number of occurrences of selected AD-related terms in the enrichment results of the upregulated subnetworks discovered with DOMINO and KPM.	24
4.5: For both algorithms, the number of intersections of the top 500 genes in the upregulated and downregulated networks with the AD-related gene lists.	25
4.6: Comparison of the prevalence rates of AD related terms in enrichment results of upregulated and downregulated subnetworks.	27
4.7: Interaction of the first 50 AD related genes most frequently involved in subnetworks. Color and size of genes is changing according to degree of nodes.	30
4.8: Comparison of prevalence rates of selected AD related terms in subtypes for downregulated and upregulated subnetworks.	34

4.9: Diagram that shows the alterations in AD subtypes in different cellular mechanism.

36



1. INTRODUCTION

Brain is the most intricate organ in the human body, and it is essential for carrying out both vital and cognitive functions. With the effect of ageing and various risk factors, different neurodegenerative diseases can occur. Alzheimer's disease (AD) is one of them. It is a long term, progressive neurodegenerative disease, and it is the most common type of dementia. It causes the death of brain cells and shrinkage (atrophy) of brain tissues, which can cause memory loss, cognitive decline, disorientation, behavioral disorders, and in more severe cases, death. It was first described by German neuropathologist and clinical psychiatrist Alois Alzheimer in 1906 [Hippius and Neundörfer, 2003]. According to the reports, the number of Alzheimer's patients is expected to exceed 100 million by 2050, with the effect of increasing life expectancy [Qiu et al., 2022]. Lifelong care and therapy requirements for these patients place a significant financial burden on the family and society. Therefore, uncovering the disease associated molecular mechanisms that were not yet fully elucidated would have a significant contribution to developing novel treatment options.

In order to understand the complex molecular mechanisms and reveal the heterogeneous nature of the diseases, various methods and data sources are employed. Omics approaches are commonly used for this purpose. Transcriptomics, one of the omics technologies, is widely used in disease diagnosis and for the elucidation of molecular mechanisms by enabling the extraction of important genes and pathways in complex systems[Lowe et al., 2017 a].

Although many studies have been conducted to understand the molecular mechanism of AD, its causes and complex mechanism still remain unclear. Due to the heterogeneous nature of the disease, putting all the data in the same pool and evaluating them as a whole will hinder the elucidation of missing key points. In this study, a personalized approach was adopted, and all 403 AD transcriptomic samples were evaluated individually to explore important genes and unknown mechanisms of disease pathology. To this end, each patient data was mapped separately on a human protein-protein interaction network to identify patient-specific subnetworks, which were

separately investigated by functional enrichment analysis. In further stages, Chapter 2 provides information about the disease pathology, as well as short information about the components that constitute the main elements of this study. In Chapter 3, the used methods and algorithms are covered, and in Chapter 4, the computational parts and the analysis of the findings are discussed. And the last chapter, Chapter 5, contains concluding remarks and recommendations for future studies.



2. BACKGROUND INFORMATION

2.1. Alzheimer's disease

Alzheimer's disease is a very complex and long-term disease. It is known that the pathological basis of the disease begins long before the clinical findings. In this silent phase of the disease, called the preclinical stage, abnormalities start at the cellular level and continue in a process that varies from patient to patient [Dubois et al., 2016]. Although disease process of each patient differs, there are some pathological hallmarks that characterize the disease in most of the cases.

2.1.1. Amyloid-beta plaques

In the pathophysiology of AD, there are several known physiological abnormalities in the brain tissues especially at hippocampus, entorhinal cortex, and neocortex. Extracellular amyloid-beta ($A\beta$) plaques are one of the hallmarks of AD. $A\beta$ peptides are products of proteolytic process of amyloid precursor protein (APP). APP is an integral membrane protein, which is expressed in many tissues but is particularly abundant in the synapses of neurons. It is crucial for different biological activities like intracellular transport, signaling, neuronal developments and homeostasis [Chen et al., 2017]. There are two pathways by which the APP can be processed through proteolysis: the amyloidogenic pathway and the non-amyloidogenic pathway. These processes are carried out by three different enzymes α -, β -, and γ -secretase. In the non-amyloidogenic pathway, which does not allow $A\beta$ accumulation, APP is first cleaved by α -secretase and forms larger soluble $APP\alpha$ peptide and smaller carboxy-terminal fragment (C83). The C83 part remaining in the membrane is then cleaved by γ -secretase and forms the short fragment P3. In the amyloidogenic pathway, APP is first cleaved by the β -secretase into 99 amino acids (C99) fragment that stays in the membrane and soluble $APP\beta$. The γ -secretase complex then cleaves the C99 fragment and generates predominantly $A\beta_{40}$ and more amyloidogenic $A\beta_{42}$ [LaFerla et al., 2007]. In AD cases it is cleaved by β -secretase

and γ -secretase dominantly and produces several types of A β peptides. Plaque formation occurs when clearance of A β is disrupted in several ways. In AD, accumulation of A β plaques causes serious metabolic disorders, neurotoxicity and eventually leads to neuronal cell death [Guo et al., 2020]. Figure 2.1 [Heppner et al., 2015] illustrates the amyloidogenic pathway, non-amyloidogenic pathway and formation of A β plaques.

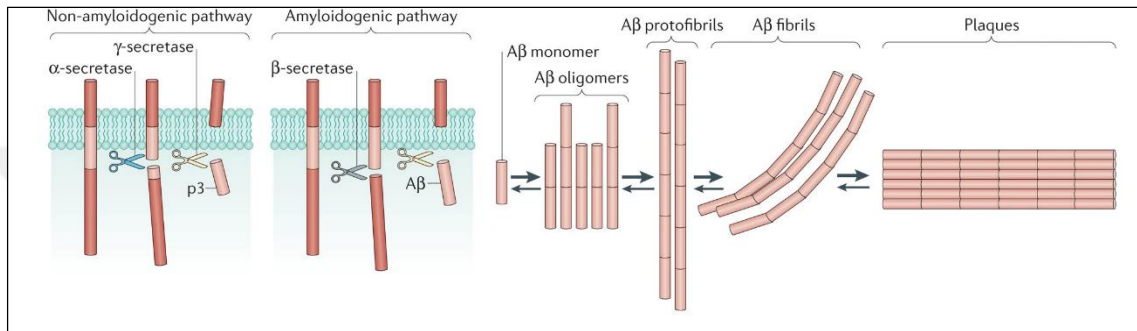


Figure 2.1: Proteolytic process of APP and A β plaque formation.

2.1.2. Neurofibrillary Tangles

Another hallmark of AD is intracellular neurofibrillary tangles (NFTs), which are formed by the aggregation of hyperphosphorylated tau proteins. Tau is a highly soluble, unfolded microtubule-associated protein that is encoded by the MAPT gene. It is essential for maintaining microtubule stability by tubulin assembly, and thus regulates normal functioning of neurons [Chong et al., 2018]. In the central nervous system, there are 6 different isoforms of tau as a consequence of alternative splicing of 8 exon regions. Mainly it has four domains, and alternative splicing occurs mostly in the N-terminal region and microtubule-binding domain. As a result, 3-repeat (3R) or 4-repeat (4R) isoforms of tau protein are produced and the 4R form has a stronger activation in microtubule assembly [Naseri et al., 2019]. The healthy adult brain contains almost equal amounts of the 3R and 4R isoforms. However, the 4R/3R ratio rises because of the alteration of this balance in tauopathy-associated neurodegenerative diseases [Chen et al., 2010].

Post-translational modifications (PTMs) in tau proteins lead to NFTs, and pathogenic tau aggregates promote synaptic and neuronal dysfunction by damaging

dendritic spines and impairing axonal functions [Tracy and Gan, 2018]. PTMs occur when an enzyme adds a chemical group, a protein, or a sugar to a specific residue of a target protein. This can happen either right away after the ribosome creates the protein or after protein folding and localization. Phosphorylation is the most commonly encountered form of PTM for tau, but tau also undergoes acetylation, ubiquitination, glycation, glycosylation, SUMOylation, methylation, oxidation, and nitration [Alquezar et al., 2021]. In AD brain, kinase and phosphatase activities are disrupted and tau protein is phosphorylated up to 2-3 times more than normal condition. This makes tau more likely to self-assemble, resulting in the formation of paired helical filaments and, as they build up, NFTs [Alonso et al., 2018]. As shown in Figure 2.2 [van Oostveen and de Lange, 2021], self-assembled tau cannot function properly and the stability of the microtubules is impaired. Consequently, the microtubules become dysfunctional and this negatively affects many neuronal functions.

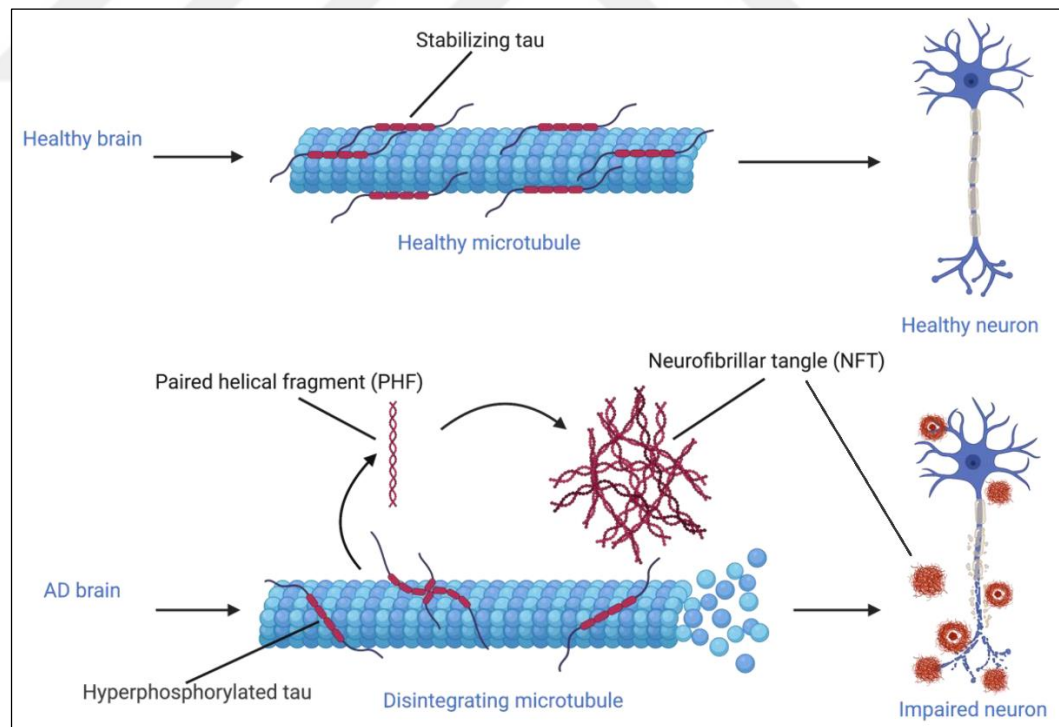


Figure 2.2: Formation of paired helical filaments and NFTs.

2.1.3. Neuroinflammation

The other major characteristic of disease is neuroinflammatory response. Neuroinflammatory response is a secondary response that occurs in the event of any injury or damage to the central nervous system (CNS). Glial cells, particularly microglia and astrocytes, concentrate in the area when cellular and molecular elements, such as cytokines and pattern recognition receptors, are activated in such a situation. The majority of the glial cells of the nervous system are astrocytes and they make up about 25% of the cerebral volume. They perform a wide range of tasks, such as promoting the development of neuronal synapses, blood-brain barrier formation and maintenance, neurotransmission, metabolic regulation, maintaining ion balance, and serving as a component of the "triple synapse" model of neurotransmission. On the other hand, microglial cells make up roughly 5-20% of the total population of glial cells in the CNS and can be found in the hippocampus and substantia nigra, as well as the brain, spinal cord, retina, and optic nerve. These cells are regarded as immune system representatives in the CNS because they can perform phagocytosis, release cytotoxic factors, and act as antigen-presenting cells [Morales et al., 2014]. Pathological alterations in AD state are detected by microglia and perivascular macrophages. They produce cytokines and chemokines that exacerbate inflammation and eventually result in neurodegeneration and neuronal loss [Heppner et al., 2015]. Like microglia, astrocytes also release cytokines after exposure to A β and misfolded proteins, hence they aggravate the neuroinflammatory response. In addition, astroglial atrophy, which occurs in AD state, leads to the destruction of synaptic transmission, which contributes to cognitive impairment [Heneka et al., 2015].

2.2. Omics and Transcriptomics

Omics technologies provide the opportunity to analyze the structure and function of a particular biological system at different levels. It consists of many fields such as genomics, transcriptomics, proteomics, metabolomics, and so on. The amount of high-throughput data produced in these fields is increasing exponentially with the advancement in technology. The first omics field to emerge is genomics, which allows for a comprehensive analysis by taking into account the entire genome rather than just

concentrating on specific variants or genes. Proteomics focuses on the quantity, modifications and interactions of peptides in biological systems, while metabolomics quantifies the products of cellular metabolic processes. Epigenomics, on the other hand, focuses on the genome-wide decoding of the reversible alterations of DNA or DNA associated proteins, such as histone acetylation or DNA methylation [Hasin et al., 2017].

Transcriptomics is another omics type that is most frequently used to investigate diseases. It is most frequently utilized alone or in combination with other omics data types [Athieniti and Spyrou, 2023]. All RNAs produced as a result of transcription constitute the source of the transcriptomic data. With the help of transcriptome analysis, we can comprehend how the genome expresses itself at the transcriptional level, learning about the structure of the genes, how they are regulated, how their products work, and how the genome changes over time [Dong and Chen, 2013]. There are two main techniques to produce transcriptomic data; microarrays and RNA-Seq. In microarray, the first emerging technique in mid 1990s, transcripts that are fluorescently marked bind to complementary probes arrayed on a chip through hybridization. Following this procedure, the fluorescence intensity in each probe region is measured to determine the abundance of the transcripts. In RNA-Seq, the isolated RNAs are fragmented and copied into double-stranded complementary DNA before being sequenced using high-throughput sequencing technology. Then, sequenced parts are aligned to the reference genome to reveal which parts of the genome are expressed [Lowe et al., 2017 b].

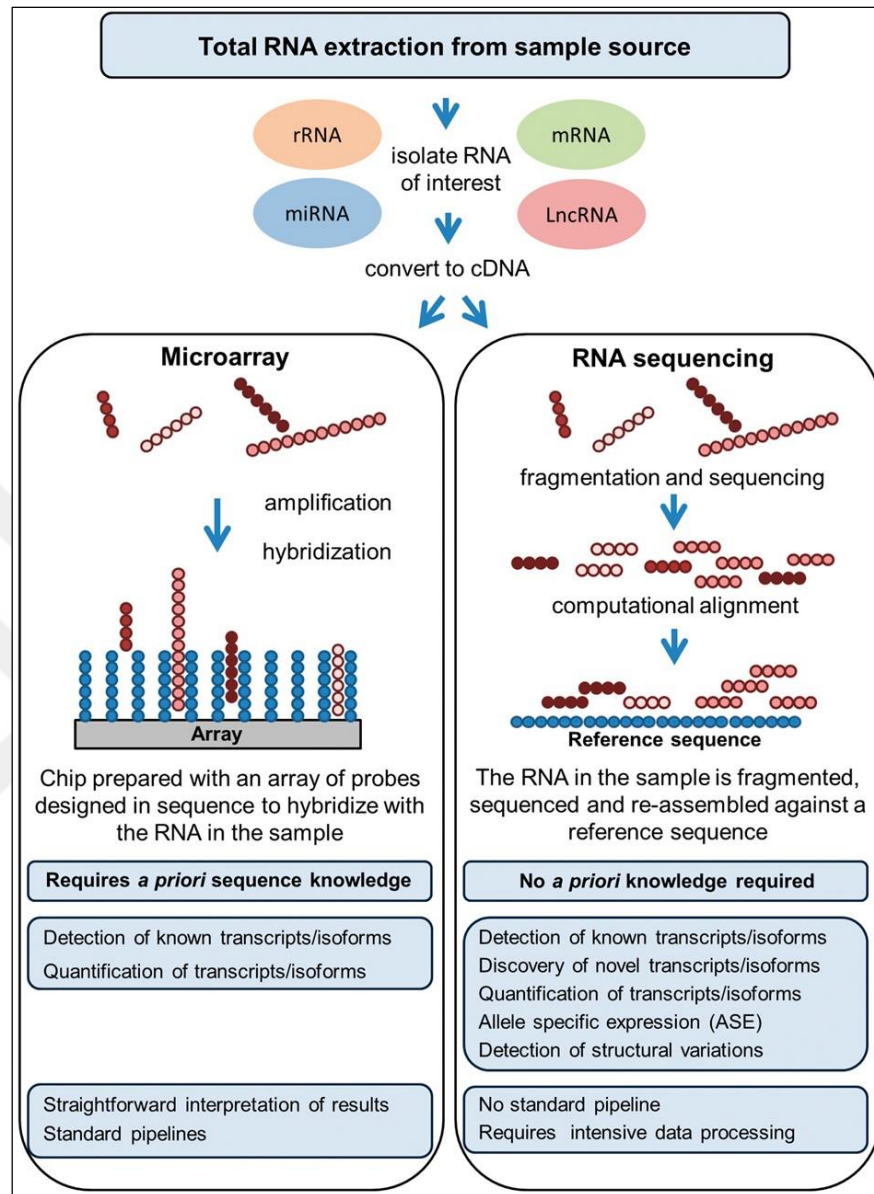


Figure 2.3: Comparison of microarray and RNA-Seq technologies.

Figure 2.3 [Manzoni et al., 2018] illustrates both methods roughly and gives some information on differences between them. Although it is still a developing technique, RNA-seq has become the dominant transcriptome technology because of the advantages that it offers. A large amount of transcriptome data created for different purposes is stored in private servers or public repositories and shared for common use. Some of the commonly used databases where transcriptome data can be accessed include Gene Expression Omnibus and ArrayExpress [Manzoni et al., 2018].

There are many studies in literature that employ transcriptome data to illuminate AD mechanisms. Williams and colleagues conducted a study with transcriptome data from prefrontal cortex. They revealed that synaptic function associated genes were downregulated while genes connected to immune response were upregulated. Genes that encode postsynaptic receptors, transporters, enzymes mediating glutamatergic and GABAergic transmission are also downregulated [Williams et al., 2021] . In another study, Li et al. performed a meta-analysis by combining six transcriptome datasets and found candidate biomarker genes for AD. These genes were substantially downregulated and strongly linked with neuropathology. Also, their pathway analysis showed that the most affected pathways were related to nitric oxide and reactive oxygen species, NFkB and mitochondrial dysfunction [Li et al., 2015 b].

Long term cohort studies Religious Orders Study (ROS) and Rush Memory and Aging Project (MAP) were initiated specifically to study aging and Alzheimer's disease in large cohorts. ROS was started in 1994 and its participants are catholic nuns, priests and brothers from all over the US, while MAP started in 1997 and enrolls older lay persons from Chicago area. Transcriptome data of the participants is also available. In AD studies, both cohorts are often used together and referred to as ROSMAP [Bennett et al., 2018]. The ROSMAP transcriptome dataset contains 165 controls and 404 AD patients, making it the most comprehensive AD-specific transcriptomic dataset. Therefore, it is commonly used in literature. In a study using three different datasets, including the ROSMAP, a significant downregulation of the neurosecretory protein VGF was observed in both frontal and temporal lobes of AD patients [Marques-Coelho et al., 2021]. Similarly, in a meta-analysis of three transcriptome datasets including ROSMAP, Wan and colleagues presented molecular models of the transcriptional state of the AD brain and identified five different consensus gene set clusters by examining 30 AD-associated coexpression modules [Wan et al., 2020]. In another study conducted using the ROSMAP dataset, it was determined that the development of AD in APOE4 allele carriers was significantly associated with vascular dysfunction. Also, they demonstrated that alterations in immune response have a significant impact on the development of AD in APOE4 carriers [McCorkindale et al., 2022].

2.3. Protein-Protein Interaction Network

It is known that there are approximately 20,000 protein-coding genes in the human genome. Each of these genes does not encode a single type of protein. As a result of different combinations of exon regions of the same gene with alternative splicing process, many protein isoforms are formed from one gene. Thus, this number rises to around 70,000 [Aebersold et al., 2018]. In order to sustain life and carry out cellular processes, these proteins have to interact with each other for different functionality, and this creates a gigantic network of interactions. Protein-protein Interaction (PPI) networks include all known protein interactions of an organism. Each organism has its own unique interaction network, and the size of these networks varies among them.

In a PPI network, the proteins are represented by nodes, their interactions between them are shown as edges and the total number of interactions a node has determines its degree. Like all biological networks, PPIs show a scale-free network feature. In scale free networks, a small fraction of nodes have high degrees, while the rest of the nodes have low degrees and the nodes with high interaction numbers are called as hub nodes. Figure 2.4 [Barabasi and Oltvai, 2004] illustrates a scale-free network example; the blue nodes have a higher degree than the other nodes, and they are the hub nodes.

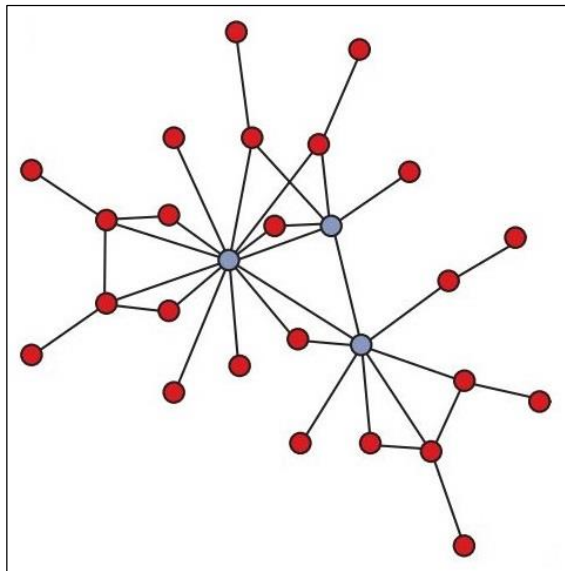


Figure 2.4: Illustration of scale free network.

There are several databases containing PPI network information that are important for bioinformatics studies. DIP [Xenarios et al., 2000], HPRD [Keshava Prasad et al., 2009], MINT [Licata et al., 2012], STRING [Szklarczyk et al., 2021], BIND [Bader et al., 2001] and BioGRID [Stark et al., 2006] are the most commonly used ones. The most recent version of BioGRID, 4.4.217 (release January 2023), has a total of 2,009,467 interactions and 87,157 proteins from various organisms that were assembled by curating 81,599 publications.

2.4. Integrating Transcriptomic Data with PPI Network

With advances in technology, the production of and access to omics data has become easier. However, analyzing what these data characterize constitutes the most important part of the scientific process. Transcriptomic data provide the information on gene activity at the transcript level, and the amount of RNA that is transcribed from a gene reflects the activity level of that gene. Transcriptomic profile of a biological system is determined on a gene basis using various statistical methods and the genes that exhibit differential expression from the normal condition are examined in further stages. Differentially expressed genes (DEGs) alone are often insufficient to identify complicated molecular interactions. At this point, subnetwork discovery tools, or, in other words, active module identifiers, are frequently employed in systems medicine to reveal potential molecular mechanisms of diseases. They mainly integrate the molecular profiles and the knowledge of biological networks with different strategies and algorithms. Figure 2.5 shows the integration of the PPI network and gene expression profiles. Thus, genes that are involved in the activated modules in case of diseases are revealed. This approach enables identification of DEGs that show physical interaction with each other.

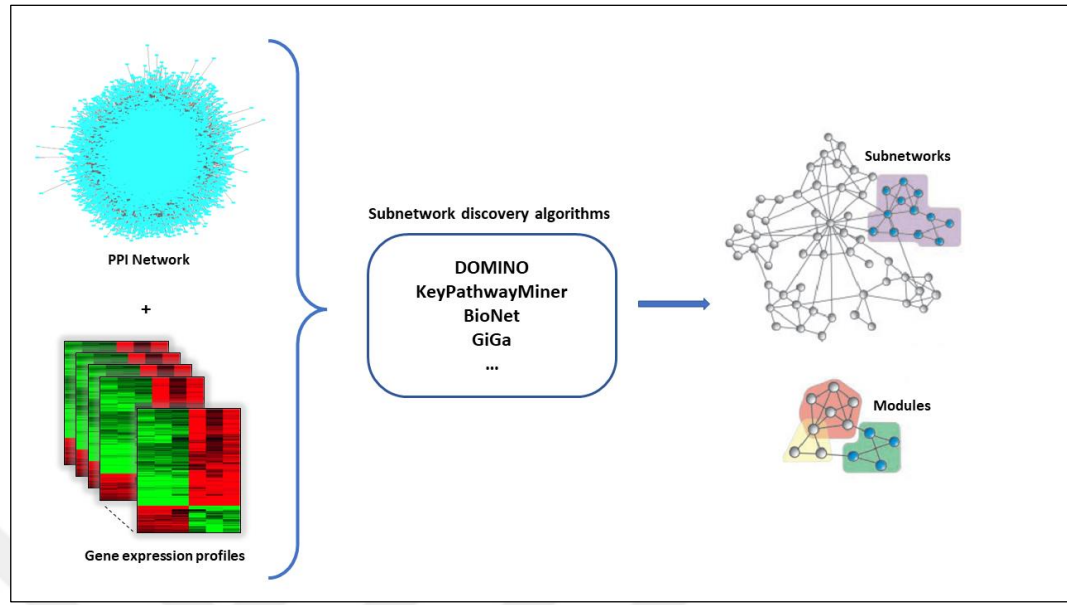


Figure 2.5: Schematic representation of subnetwork discovery process.

There are many different tools in the literature for subnetwork discovery, and they adopt different algorithms and strategies to find active modules that differ under changing conditions. Recently Lazareva et al. [Lazareva et al., 2021] conducted a study based on the randomization of networks using five different PPI networks. This study was conducted by using widely used eight different algorithms in the literature to measure whether active module identification methods actually benefit from PPI knowledge in real terms. Among ClustEx2 [Ding et al., 2018], COSINE [Ma et al., 2011], DIAMonD [Ghiassian et al., 2015], DOMINO [Levi et al., 2021], GiGA [Breitling et al., 2004], GXNA [Nacu et al., 2007], KeyPathwayMiner [Alcaraz et al., 2014] and GrandForest [Larsen et al., 2020] algorithms, DOMINO produced more meaningful modules and its performance was better than the others. When compared to other algorithms, it truly benefits from the PPI interaction information because it produces the best results. In another study, Batra et al. [Batra et al., 2017] compared the performance of seven algorithms, including KeyPathwayMiner (KPM), BioNet [Beisser et al., 2010], GiGA, GXNA, COSINE, DEGAS [Ulitsky et al., 2008], and PinnacleZ [Chuang et al., 2007], and discovered that KeyPathwayMiner and BioNet performed better than others. Based on these results, DOMINO and KeyPathwayMiner are used as two different subnetwork discovery tools in this study.

2.4.1. KeyPathwayMiner

KPM was first introduced in 2011, and over time, versions were updated till version 5.0. It works with both RNA-Seq and microarray data types and permits direct combination of various types of omics data. It offers different types of usage: web application module (KeyPathwayMinerWeb [List et al., 2016]) for easy use, standalone module for higher computational power, Cytoscape app module for advanced visualization and ease of use and KeyPathwayMineR [Mechteridis et al., 2022] for use in the R ecosystem. It offers two different strategies for subnetwork discovery; Individual Node ExceptionS (INES) allows the setting of the parameter K, which is the number of unchanged genes allowed in the discovered subnetwork and GLONE. It also allows to choose algorithm with 3 different options, GREEDY which is a faster option, Ant Colony Optimization (ACO), which is used in large networks and datasets and explores maximal subnetworks with a heuristic approach, and exact fixed-parameter tractability algorithm OPTIMAL. In addition, it also enables the usage of prior knowledge with the positive genes that are desired to be found in any case in the discovered subnetwork or negative lists that are not desired to be found at all. KPM mainly takes the PPI network and an indicator matrix where DEGs are represented as 1 and others as 0 as input. The run time varies depending on the size of the inputs, but the standalone version has the fastest processing power among the different types of usage.

Yurduseven et al. [Yurduseven et al., 2022] conducted a study using cell-type-specific transcriptome data to uncover biomarker candidates for multiple sclerosis and they used KPM as a subnetwork discovery tool. In another study, Emanetci and Çakır also used KPM for subnetwork discovery to elucidate the molecular mechanism underlying cognitive impairment and memory deficits [Emanetci and Çakır, 2021]. In another study aiming to illuminate the cancer dormancy mechanism, transcriptomic data was mapped to the PPI network using KPM, thus creating a dormancy-interaction network first time in the literature [Uzuner et al., 2021].

2.4.2. DOMINO

DOMINO is an active module identifier (AMI) algorithm that aims to eliminate non-specific over-represented GO terms. It was developed in response to the argument that existing AMI algorithms are not successful in finding content-specific functional terms and are not resistant to permutation of the dataset. As first step in finding active modules, DOMINO partitions the PPI network, which is given in SIF format, into slices using the Louvain modularity algorithm. This step is done only once for a PPI and this information can be used in all analyses if the same PPI is used. The slices that pass the hypergeometric test based on the number of DEGs on these slides are then identified as relevant slices. At this stage, the default FDR value is 0.3 and it can be changed if desired. Each determined relevant slice is divided into sub-slices by solving the prize collecting Steiner tree problem, and then these sub-slices are divided into putative active modules using the Newman-Girvan modularity algorithm. Finally, each putative active module is subjected to hypergeometric test for over-representation of DEGs to determine the final active modules. As a result, it generates multiple modules, which are biologically rich and functionally distinct. Figure 2.6 [Levi et al., 2021] shows each step that DOMINO performs in the AMI process.

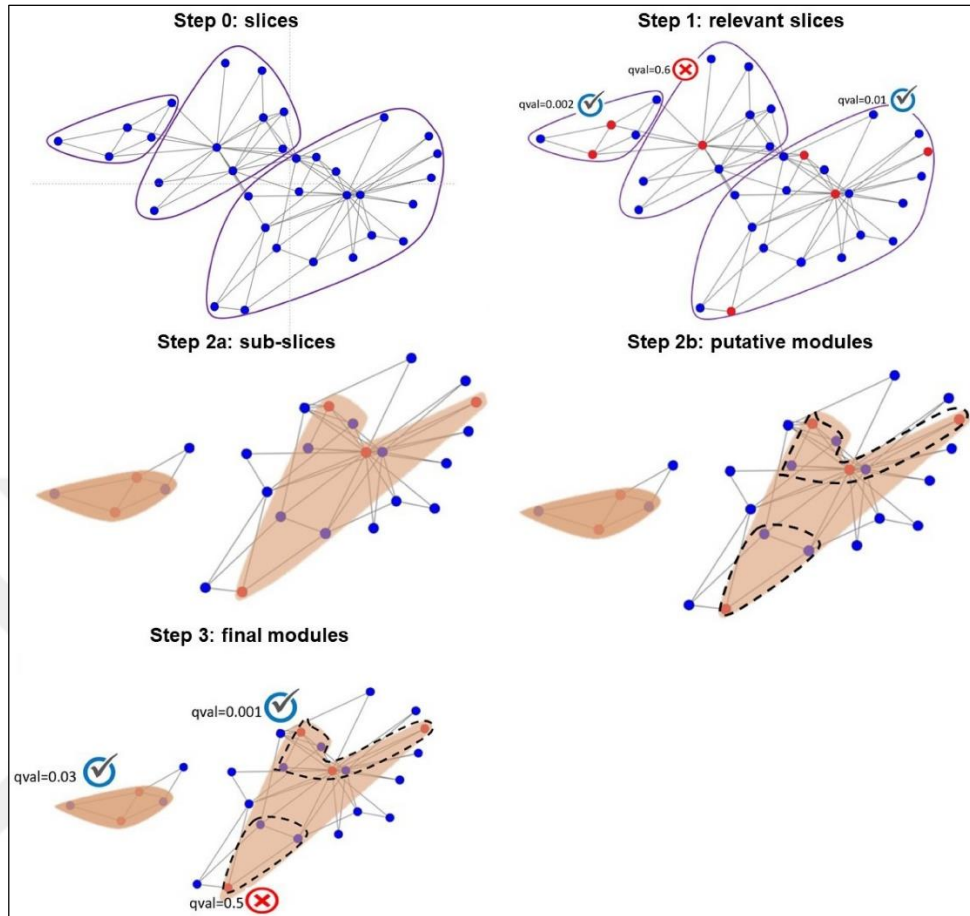


Figure 2.6: Schematic illustration of steps of DOMINO.

DOMINO accepts the DEGs list given as input only in Ensembl gene ID format and this is a disadvantage that prevents the use of other types of omics data. It works very fast regardless of the size of the PPI network or the number of DEGs, which is advantageous when multiple analyses are required. Different analyses with the exact same input can produce modules of different numbers and sizes, and this is due to the stochastic nature of the algorithms used in its design. It runs on Ubuntu and this can create limitations for some users. At this point, the recently developed DOMINO web [Levi et al., 2022] shows more user-friendly profile and offers advantages such as GO enrichment analysis, module visualization and output in different formats.

3. METHODS

3.1. Transcriptome Datasets

Religious Orders Study and Rush Memory and Aging Project (ROSMAP) [Bennett et al., 2018] RNA-seq data was used in this study. The samples are from dorsolateral prefrontal cortex of 569 participants, with 404 Alzheimer's disease patients (CERAD score 1 and 2, corresponding to definite and probable AD) and 165 controls (CERAD score 4, corresponding to no AD). Raw read counts were normalized by DESeq2 package [Love et al., 2014] in R. Principal component analysis was performed with the normalized data to detect outlier samples. Two samples were excluded from the cohort; one was from AD and the other was from the control group (sample IDs; 500_120515 and 380_120503). To eliminate confounding effects of age, sex, and PMI (postmortem interval), a mixed linear model was applied to log2-transformed normalized data.

3.2. Statistical Analysis

In this study, a personalized approach was adopted where each sample from the AD group was compared to the control group to detect differentially expressed genes (DEGs) in each patient. To this end, a ranking based approach was used where the patient value and all the values in the control group were combined and ranked for each gene. The patient gene was considered to be downregulated if it is within the top 5th percentile in the rank list while it was considered to be upregulated if it is within the bottom 5th percentile. Although it did not eliminate the selected genes most of the time, Benjamini-Hochberg corrected p-values were also calculated with one-sample t-test for each gene in each AD sample, and used this as additional cut-off (p-value<0.01) in identifying DEGs for each sample. In the bulk analysis where all AD samples were compared with all control samples, limma [Ritchie et al., 2015] package was used, and the cut-off value was chosen as 0.05 for Benjamini-Hochberg corrected p values.

3.3. Subnetwork Discovery from Protein-Protein Interaction Data

Human protein-protein interaction (PPI) network was downloaded from BioGRID [Stark et al., 2006] (version 4.4.207). Only human-human physical interactions were retained, self-loops and duplicated edges were removed from the network. In its final form, the PPI network included 19,634 proteins and 678,802 interactions. The final form of the PPI network used in this study is illustrated in Figure 3.1.

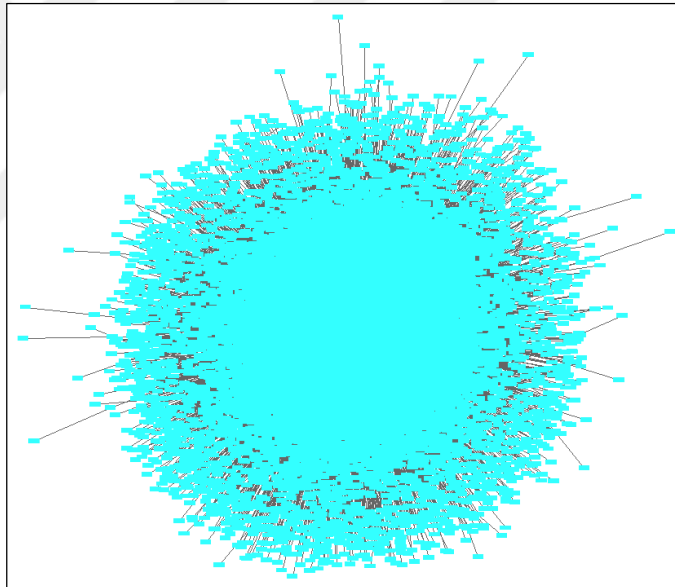


Figure 3.1: Human PPI network used in this study.

Subnetwork discovery was performed by two different algorithms: KeyPathwayMiner [Alcaraz et al., 2016] (KPM) and DOMINO [Levi et al., 2021]. For each patient, subnetworks were created for the upregulated and downregulated DEGs separately with both algorithms. For the KPM analysis, standalone version 5.0 was used with greedy algorithm and INES model. The maximum number of unchanged genes allowed in a discovered subnetwork was set to 8 through the parameter K. The

DOMINO analysis was performed with version 0.1.1 without changing the default parameters.

3.4. Enrichment Analysis

Functional enrichment analyses are performed to find out in which pathways the genes in the discovered subnetworks are functionally located. In the literature, there are many tools based on different strategies such as DAVID [Sherman et al., 2022], FunRich [Pathan et al., 2015], EasyGO [Zhou and Su, 2007], Enrichr [Chen et al., 2013], g:Profiler [Raudvere et al., 2019] to perform functional enrichment analysis. While doing this, they use different databases such as GeneOntology, Reactome, Panther, KEGG, WikiPathways, CORUM, Human Protein Atlas and many others created for different purposes [Reimand et al., 2019].

In this study, g:Profiler, a widely used enrichment analysis tool in the literature, was used as a functional enrichment analysis tool. Since a large number of analyses were required, the R package version of g:Profiler [Kolberg et al., 2020], which is more suitable for automation, was used. It can be easily integrated with the pipeline and provides fast results, which shortens the time of analysis and provides convenience. The genes in each discovered subnetwork were analyzed with default correction type (g:SCS) and 0.05 p-value cut-off. Since DOMINO produces multiple modules with differing size, enrichment analysis was applied to the modules with the size greater than or equal to 10, ignoring the smaller modules. Significantly enriched terms coming from the multiple modules of a given sample were combined for final analysis.

4. SAMPLE BASED SUBNETWORK ANALYSIS

4.1. Comparison of KPM and DOMINO results

In this study, it was aimed to elucidate the molecular mechanism of AD by adopting a personalized approach. RNA-seq data of 403 AD patients from ROSMAP cohort were used for this purpose. To reveal the affected mechanisms, subnetwork discoveries were performed through the KPM and DOMINO algorithms by using upregulated and downregulated DEGs separately for each patient. Functional enrichment analysis was applied to each subnetwork, and AD associated functional terms were used to assess the performance of KPM and DOMINO.

KPM discovered two separate subnetworks for each sample, one with upregulated DEGs and the other with downregulated DEGs, leading to a total of 806 subnetworks for 403 AD patients. DOMINO, on the other hand, did not discover any modules for the down-regulated DEGs of 23 patients and the up-regulated DEGs of 9 patients. There was only one patient with no modules. As a result of its design, DOMINO produces multiple distinct modules, each of which having a different number of genes. Genes from these discrete modules were combined when reporting the total number of genes in the subnetworks. Figure 4.1 shows boxplots of the number of genes in upregulated and downregulated subnetworks discovered by both DOMINO and KPM. It is observed that the number of nodes of subnetworks produced by KPM is considerably high compared to DOMINO, despite the same number of DEGs being used as input. The upregulated and downregulated DEG numbers for each AD patient are represented by the boxplot in Figure 4.2. The distribution of the number of genes included in the DOMINO subnetworks shows a more balanced profile closer to the normal distribution compared to KPM. The gene content of the KPM subnetworks, on the other hand, shows a more dispersed distribution.

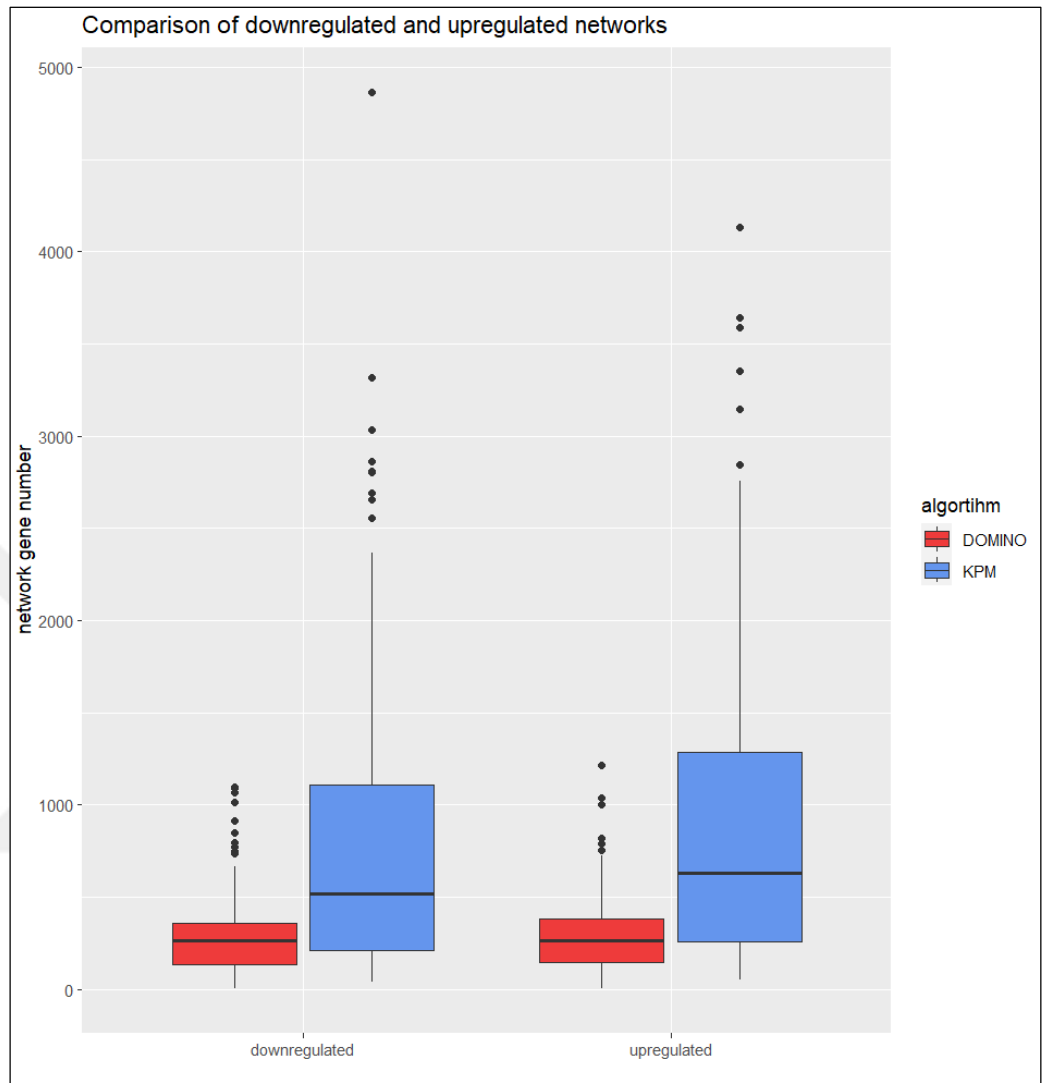


Figure 4.1: Boxplots of number of genes in the downregulated and upregulated subnetworks discovered by DOMINO and KPM.

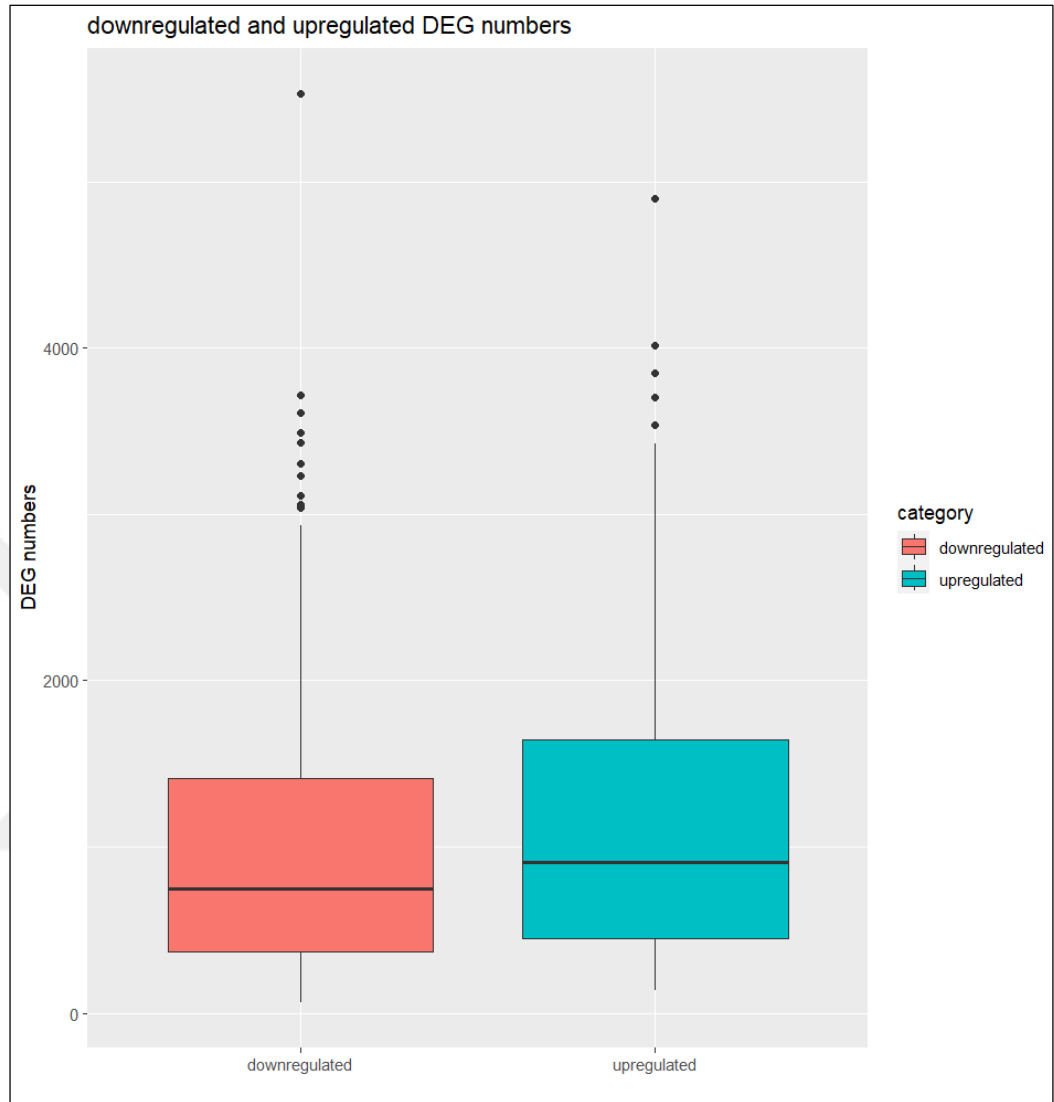


Figure 4.2: Boxplots showing the range of the number of downregulated and upregulated DEGs for each AD patient from ROSMAP cohort.

Next, comparison of the enrichment results was made for the two algorithms. To determine which pathways are associated with the subnetworks, an enrichment analysis was carried out using the genes of each subnetwork. Since KPM already produces a single subnetwork as an output, the genes of subnetworks were directly used in enrichment analysis. However, as DOMINO generates modules of varying numbers and sizes, its enrichment analyses were performed as described in Methods section. Combining the enrichment results of downregulated KPM subnetworks gives 6,332 unique terms while 7,560 unique terms were obtained from combination of the enriched

terms of upregulated subnetworks. These numbers were 15,451 and 16,286 respectively for DOMINO. A comparison was made through the enrichment results to find out which algorithm captures disease-related terms better. To this end, majority voting was applied to the enrichment results for each algorithm to identify the number of patients found to be significantly associated with a given term. To compare the algorithms, 20 terms that have been shown to be connected to AD were identified and represent the pathophysiology of the disease. While selecting these terms, a recent study by Morgan et al. [Morgan et al., 2022] was also used. Figure 4.3 and Figure 4.4 show these terms and the number of patients with these terms according to the enrichment results in downregulated and upregulated subnetworks. DOMINO captured the disease-related terms considerably better than KPM in almost all selected terms except the ‘apoptotic process’. Compared to KPM, DOMINO captured these terms in more patients, but it does so with subnetworks that are much smaller in size than KPM (Figure 4.1).

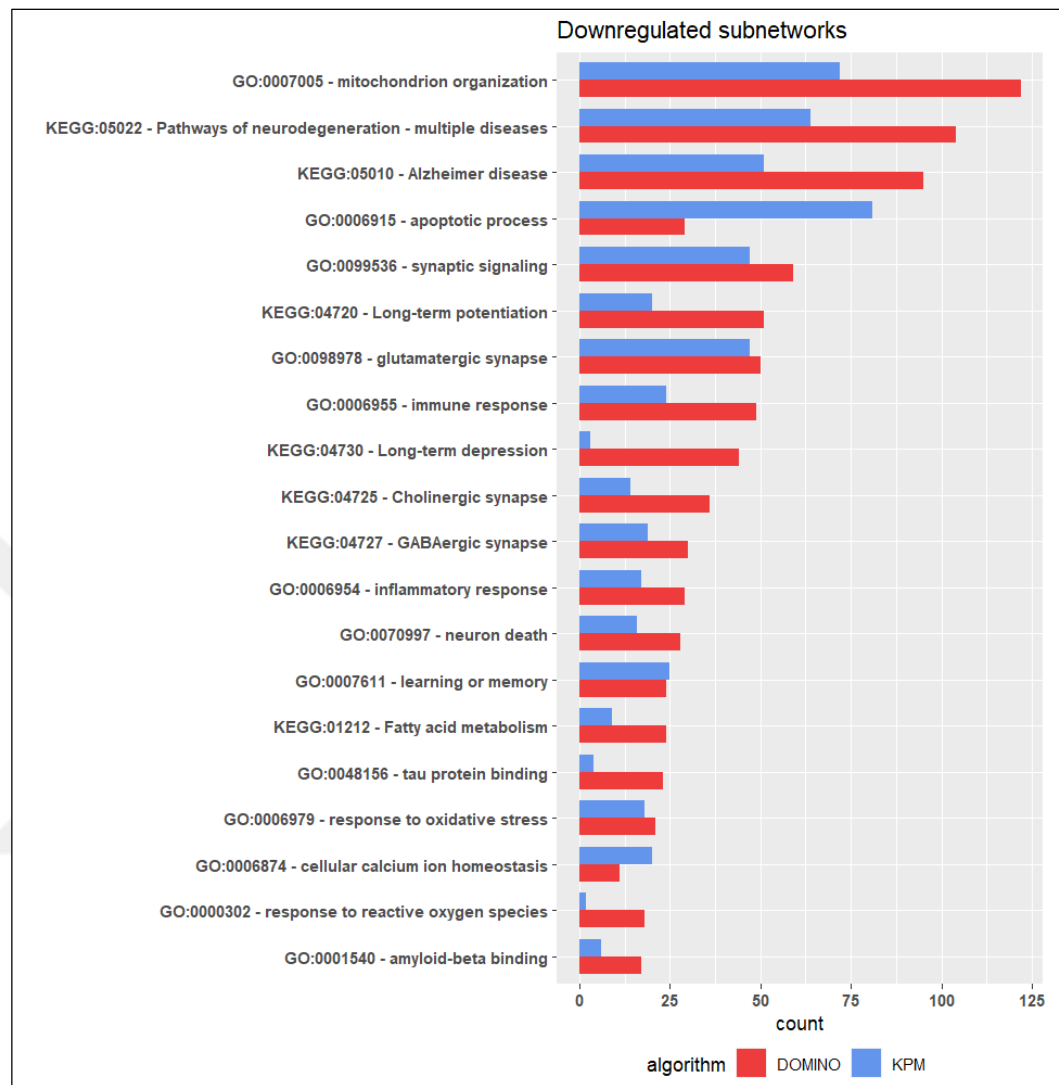


Figure 4.3: Comparison of the number of occurrences of selected AD-related terms in the enrichment results of the downregulated subnetworks discovered with DOMINO and KPM.

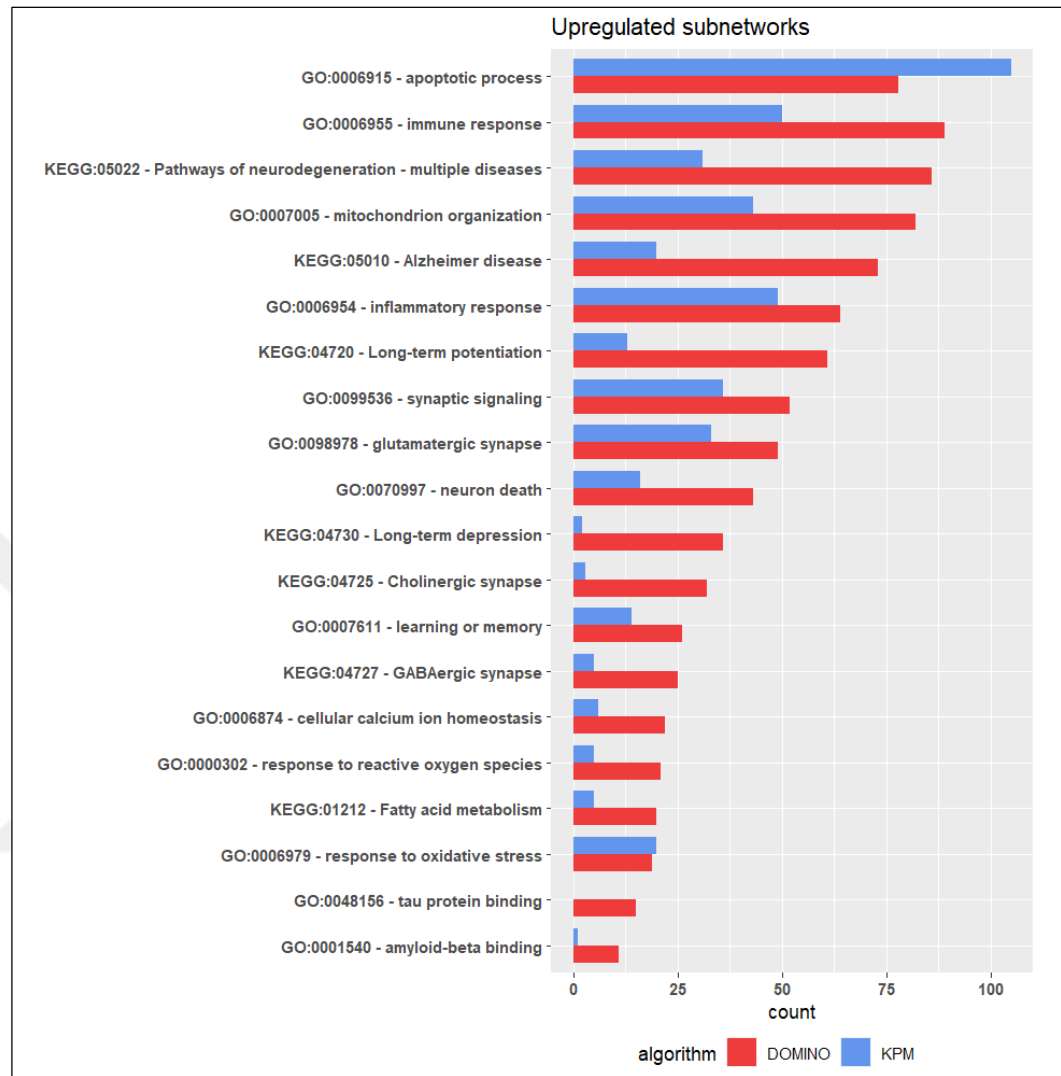


Figure 4.4: Comparison of the number of occurrences of selected AD-related terms in the enrichment results of the upregulated subnetworks discovered with DOMINO and KPM.

Another comparison was made to compare the performance of the two algorithms by identifying the number of known AD related genes in the discovered subnetworks. For this purpose, four different gene lists containing AD-associated genes were used: (i) A list compiled from three different GWAS studies that were conducted to identify AD risk genes [Andrews et al., 2020; Wightman et al., 2021; Bellenguez et al., 2022][Andrews et al., 2020; Wightman et al., 2021; Bellenguez et al., 2022], (ii) A list including the genes linked with the term "KEGG:05010 Alzheimer's disease", obtained via g:Profiler, (iii) A list created from the genes obtained by searching for the term

"Alzheimer's disease" in the GeneCards database. (iv) a manually curated list of AD associated genes, termed Alzgset, compiled by extensive scanning of 823 published genetic association studies in PubMed [Hu et al., 2017] . There are 117 genes in GWAS, 353 genes in KEGG, 191 genes in GeneCards and 430 genes in Alzgset. For both algorithms, the top 500 genes that were encountered in the discovered subnetworks were identified separately for upregulated and downregulated subnetworks, and their intersections with four different gene lists were determined (Figure 4.5). Genes with the same score as the 500th gene were also taken into account when calculating intersections. While the intersections with GWAS and GeneCards lists were very low in general, the highest intersections were observed with KEGG list. DOMINO generally has more intersections than KPM across all gene lists, and this is particularly evident in KEGG.

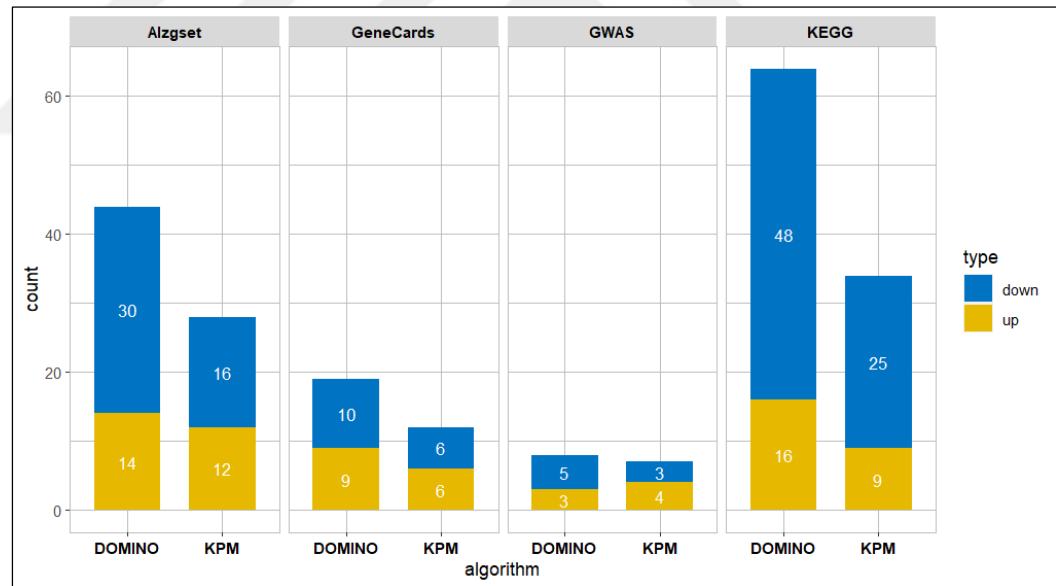


Figure 4.5: For both algorithms, the number of intersections of the top 500 genes in the upregulated and downregulated networks with the AD-related gene lists.

Based on these findings, it was decided to proceed with the results of DOMINO in further evaluations.

4.2. Evaluation of sample based results

The prevalence rate of enrichment terms and genes were calculated among patients in order to assess and draw conclusions from the large number of produced results. While doing this, the number of occurrences of a term or gene was divided by the total number of patient subnetworks and then multiplied by 100. In addition, terms associated with AD and reflecting different parts and functions of the neuronal cell were determined and the results were evaluated based on these terms. Based on the prevalence rates of each identified term in the upregulated and downregulated results, a heatmap was produced (Figure 4.6). The terms considered were grouped into seven categories as follows: mitochondria, immune response, cell death, calcium ion balance, synaptic activities, endoplasmic reticulum and fatty acid metabolism.

The first obvious difference is seen in the mitochondria-associated terms. Mitochondria is an essential component of neurons because they both produce ATP and buffer calcium ion concentration at the synaptic region [Cenini and Voos, 2019]. Impaired energy metabolism observed in AD state is an indication of disrupted mitochondrial activities. Moreover, damaged mitochondria have a higher potential of producing reactive oxygen species (ROS), and less efficient in ATP production. This increases oxidative stress, which further impairs mitochondrial functions [Wang et al., 2020]. Figure 4.6 shows that altered ATP related functions are primarily associated with the downregulated subnetworks. In the remainder part of the results, which cannot be shown in Figure 4.6 enrichment results of downregulated subnetworks contain more mitochondria-related terms than upregulated ones, and their prevalence rates are higher. This indicates that there is a suppression of genes that control the mitochondrial activity in AD. The two genes mostly encountered in the downregulated subnetworks are TBRG4 and COX4I1. Mitochondrial localized protein TBRG4 (transforming growth factor beta-regulator 4) is a member of FASTK family. They are involved in mitochondrial RNA processing, regulation of the mitochondrial mRNA stability and

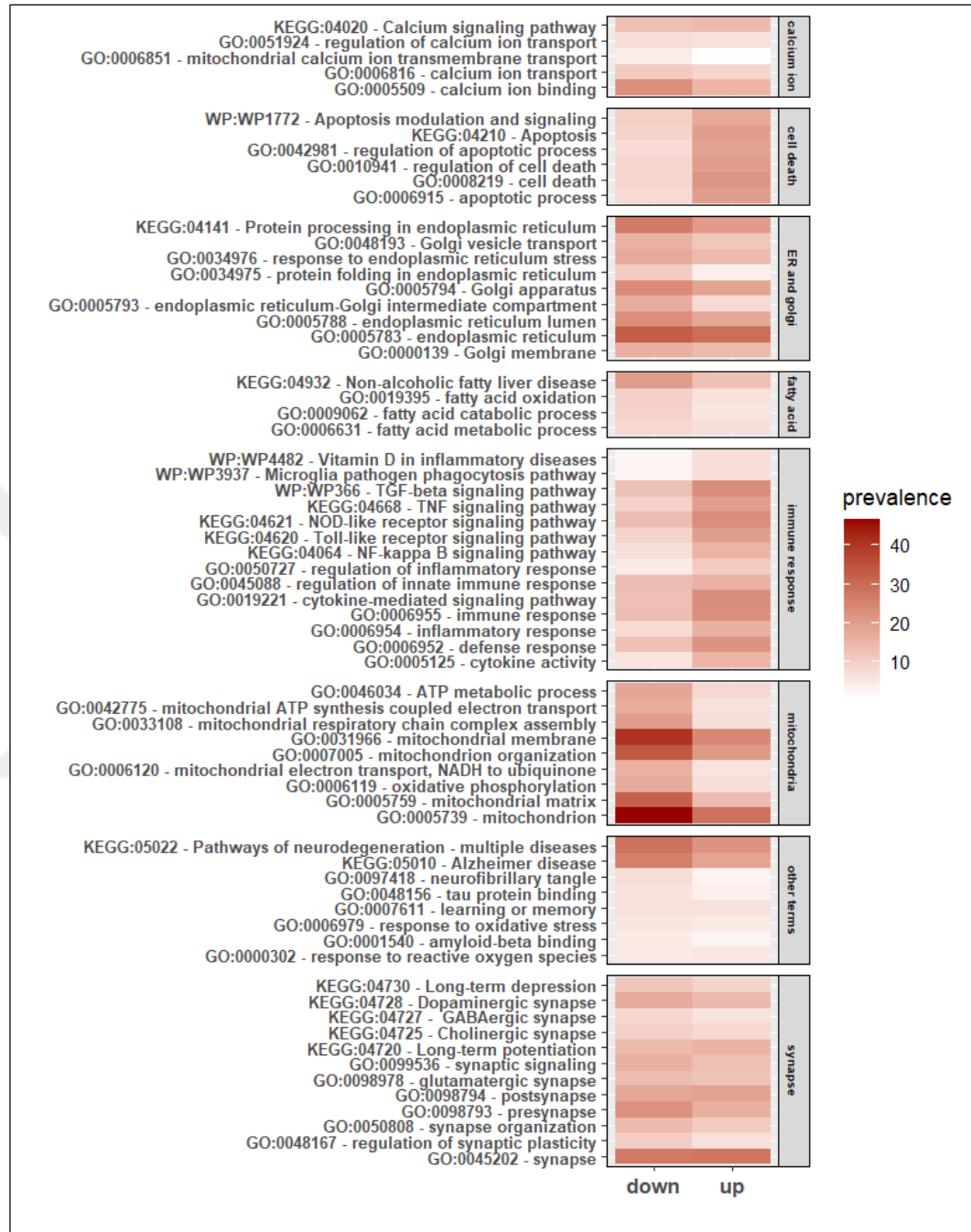


Figure 4.6: Comparison of the prevalence rates of AD related terms in enrichment results of upregulated and downregulated subnetworks.

modulation of mitochondrial respiration. TBRG4 deficiency promotes apoptosis and elevates the generation of ROS [Zhang et al., 2022]. Mitochondrial cytochrome c oxidase (COX, complex IV) is involved in the mitochondrial respiratory chain as the final electron acceptor. It has 13 subunits, one of them COX4 and most common isoform

of COX4 is COX4I1. Its deficiency elevates ROS levels and impairs ATP production [Abu-Libdeh et al., 2017] .

Immune response is another notable issue. Its related terms seem to be more prevalent in upregulated subnetworks than downregulated ones. In AD case, accumulated or misfolded proteins in the extracellular matrix like A β bind to pattern recognition receptors on microglia and astrocyte and activate them. This activation is predominantly initiated by toll-like receptors (TLRs) [Kumar, 2019]. Activation of different types of TLRs trigger cytokine signaling networks, causing the release of various cytokines like TNF- α , IL6, TGF- β , which have proinflammatory effect [Momtazmanesh et al., 2020]. The predominance of pathways related to these mechanisms in upregulated subnetworks is consistent with disease pathology. Another striking difference is seen in the terms related to cell death. It is a highly controlled process required for healthy growth and tissue homeostasis. In AD, a high rate of neuronal death is seen in various parts of the brain, especially in the entorhinal cortex, hippocampus, amygdala and cerebral cortex [Goel et al., 2022]. When the terms of upregulated and downregulated networks are compared, it is observed that the terms related to cell death take more place in upregulated networks. Endoplasmic reticulum (ER) and Golgi apparatus associated pathways are also among the pathways where the differences were observed. ER is a membranous organelle with very basic functions such as folding of protein molecules, transport of synthesized proteins to the Golgi with vesicles, regulation of calcium homeostasis in the cell and so on [Li et al., 2015 a]. In the ER, a subpopulation of cytosolic proteins, integral membrane proteins, and secretory proteins are all synthesized. Ribosomes on the cytosolic face of the ER produce proteins that are taken up into the ER lumen for processing [Schwarz and Blower, 2016]. Before being exported from the ER, proteins are folded, assembled, and undergo posttranslational modifications. Misfolded proteins are either corrected in the ER or, if not corrected, are discovered by the ER and sent to the cytosol for degradation [Almanza et al., 2019]. Buildup of misfolded proteins and disruption of intracellular Ca²⁺ homeostasis cause ER stress, resulting in neuronal dysfunction eventually cell death [Hashimoto and Saido, 2018]. Our results show that the terms “ER lumen”, “protein processing in the ER”, “response to ER stress”, “protein folding in the ER” are

dominantly take place in down-regulated subnetworks and are consistent with AD pathology.

The alterations that occurred in the disease state were examined with the comparison of the enrichment results. However, identifying candidate genes that can be shown as targets is quite important in terms of revealing mechanism of the disease and drug development. To this end, the genes in the 4 lists of AD associated genes mentioned in the previous section were ranked according to their prevalence in the networks obtained by combining the up and downregulated subnetworks, and the interaction map was created by selecting the first 50 genes (Figure 4.7). Here, it is seen that HSPA5 and HSP90AA1 from the heat shock protein family in the first row. In third place is the COX4I1 gene, the most common isoform of one of the subunits of the Cox involved in the mitochondrial respiratory chain.

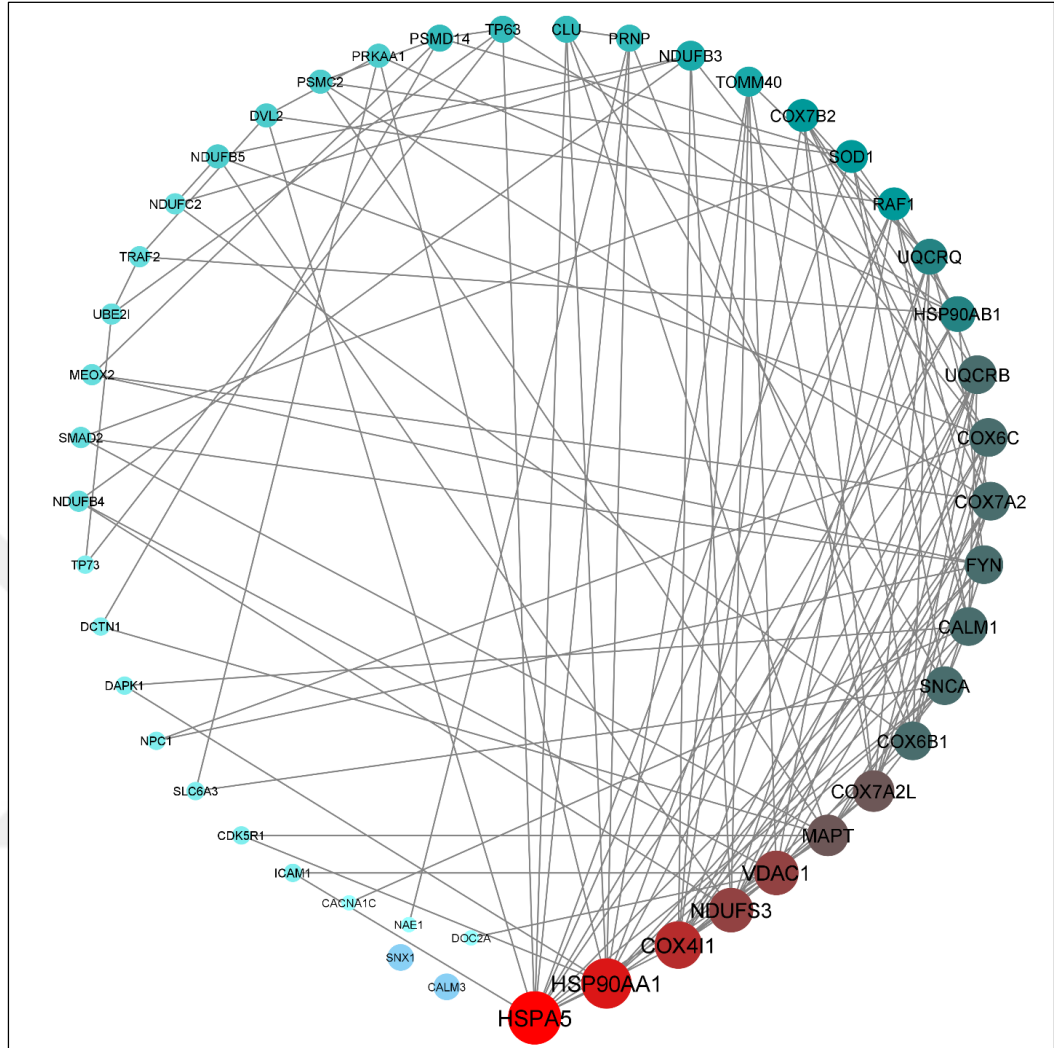


Figure 4.7: Interaction of the first 50 AD related genes most frequently involved in subnetworks. Color and size of genes is changing according to degree of nodes.

4.3. Comparison of sample based and bulk analysis results

To understand whether the sample-based approach has an advantage over the bulk analysis, which is often used in the literature, a bulk analysis was performed by comparing all control samples with all AD samples and obtained 2073 DEGs. Then, as in the sample based analysis, subnetwork discovery was performed with these DEGs by using DOMINO, and enrichment analysis was applied to the genes in the discovered subnetwork. DOMINO and g:Profiler analyses were performed using the same parameters as in the sample based analysis.

The enriched terms for the bulk subnetwork included many terms related to mitochondrial functions, respiration, ATP synthesis, which are very important for AD. Likewise, these terms are frequently encountered in the sample-based analysis results too. Looking at the terms related to synaptic functions, which are also important for AD pathology, there are only few terms in bulk result and these are related with “presynapse”, “synapse”, “postsynapse” and “synaptic vesicle”. On the other hand, terms that allow us to learn more and in-depth information about synapses are found only in the sample-based results. Another striking difference is that inflammation or immune system related terms, which has a central place in AD pathology, is not identified in the bulk analysis enrichment results. In the sample-based enrichment results, there are many terms related to inflammation and immune response. The terms related to the endoplasmic reticulum (ER) also differ between the two approaches. The bulk analysis results have only a few terms from Reactome Database reflecting ER - golgi transmission, on the other hand sample based results contains many terms related to endoplasmic reticulum and its functional processes. In addition, there are a limited number of terms in the bulk analysis enrichment results that reflect the mechanism of cell death, and these are only terms related to the autophagy mechanism. However, as it is known, the predominant cell death mechanisms in AD are apoptosis and necrosis. These mechanisms are triggered by the activation of caspase cascade, TNF and TLR pathways [Chi et al., 2018]. But in the findings of the bulk analysis, there is no term referring to these mechanisms. Another issue where differences are captured is the terms related to the SMAD protein. These proteins are activated by TGF- β binding to cell wall receptors and then move from the cytoplasm to the nucleus to activate or suppress transcription [Attisano and Tuen Lee-Hoeflich, 2001] . Abnormal localization of phosphorylated SMAD proteins in AD affects TGF- β signaling pathway, impairing the neuroprotective functions of this mechanism [Yang and Xu, 2022] . While the terms related to SMAD were observed in the sample based results, no terms were found in the bulk analysis results.

Protein homeostasis is essential for normal functioning of the cells. Ubiquitin-proteasome system (UPS) and the ubiquitin-like protein (UBL) conjugation pathways are crucial for maintaining this homeostasis [Bedford et al., 2011] . Ubiquitin is a

member of UBL family and it is an important component of the UPS, which degrades more than 80% of normal or misfolded cellular proteins [Gadhav et al., 2016] . In the AD brain, abnormal UPS activity, ubiquitin accumulated plaques and tangles are seen. Also proteasomal activity decreases especially in hippocampus [Oddo, 2008] . Although there are many terms related to these mechanisms and proteins in sample based results, there is no term in bulk analysis. Another UBL where the difference is observed between two approaches is small ubiquitin-like modifier (SUMO), and its related terms. SUMO is a group of proteins that are attached to or detached from proteins during post translational modifications, and this phenomenon is called SUMOylation. It plays important roles in many cellular processes and regulation of biological functions of proteins [Chen et al., 2021] . In AD, dysregulations occur in SUMOylation [Lee et al., 2013], but no terms reflecting these alterations are found in the bulk analysis result, while SUMO and related terms are found in the sample based enrichment results. The same observation holds also for NEDD8, another UBL member whose activity is deregulated in AD [Chen et al., 2012] .

The phosphatidylinositol 3-kinase (PI3K)/AKT signaling pathway is another pathway where the differences are noticeable between bulk and sample-based results. The PI3K/AKT signaling pathway regulates a wide range of processes in the brain, including motility, intracellular trafficking, cell proliferation, growth, differentiation, and more complex procedures like dendrite and axon extension. It also has essential role in maintaining synaptic plasticity and memory process [Razani et al., 2021] . An abnormality in PI3K/AKT signaling can increase the activity of GSK-3 β , which is a crucial molecule in the downstream of the pathway, and cause hyperphosphorylation of Tau, which causes the formation of NFTs [Kitagishi et al., 2014] . Neurons can be protected against A β induced toxicity by activating PI3K/AKT signaling pathway [Long et al., 2021] . The terms reflecting these pathways and the proteins that are central for a thorough understanding of the molecular mechanism underlying AD are present in the sample-based results, but they are absent in the results of the bulk analysis. These altogether show that the sample based approach can reveal different aspects of each patient, and the unobtrusive points in the bulk analysis are made visible with the sample-based analysis.

4.4 Evaluation of results according to AD subtypes

It is known fact that AD is a heterogeneous disease and has subtypes. Therefore, not all patients show the same pathological findings, and the disease progresses by different mechanisms [Jellinger, 2022]. Evaluation of the results by ignoring the subtypes of the disease will be an obstacle to revealing the molecular differences between the subtypes. Therefore, patient-centered personalized results were also evaluated by referring to the subtyping suggested in a recent study by Neff et al. [Neff et al., 2022] In that study, they used the Mount Sinai Brain Bank dataset to subgroup AD patients and validate these results with the ROSMAP dataset, which also used in this study. They identified 5 molecular subtypes which called *blue*, *green*, *red*, *turquoise* and *yellow* under three major classes. There are 123 patients in the blue group, 96 patients in green, 80 in red, 56 in turquoise, and 35 in yellow group. In this section, the 390 ROSMAP samples used by Neff et al. in their study were evaluated to exploit subtype information. In order to show differences between groups, prevalence rates of terms and genes for each group were calculated separately for upregulated and downregulated subnetworks.

Figure 4.8 illustrates how the prevalence rates of the terms differ in the subgroups. This shows that the results are consistent with the fact that there are subtypes of the disease. Looking at downregulated side of Figure 4.8, it is seen that the patients in the *blue* group have a low prevalence rate in terms related to immune response and cell death. However, the prevalence rates of the terms related with mitochondria, synapse and calcium are high compared to the other groups. In upregulated side of Figure 4.8, *blue* group shows very high prevalence rates in terms related to cell death and immune system. There is also a slight upregulation of fatty acid metabolism. High levels of oxidative stress and ongoing bioenergetic disruption activate the apoptotic pathway, impair mitophagy, and ultimately cause the death of neuronal cells [Lanzillotta et al., 2019], which explains why there is an increase in immune response and cell death while

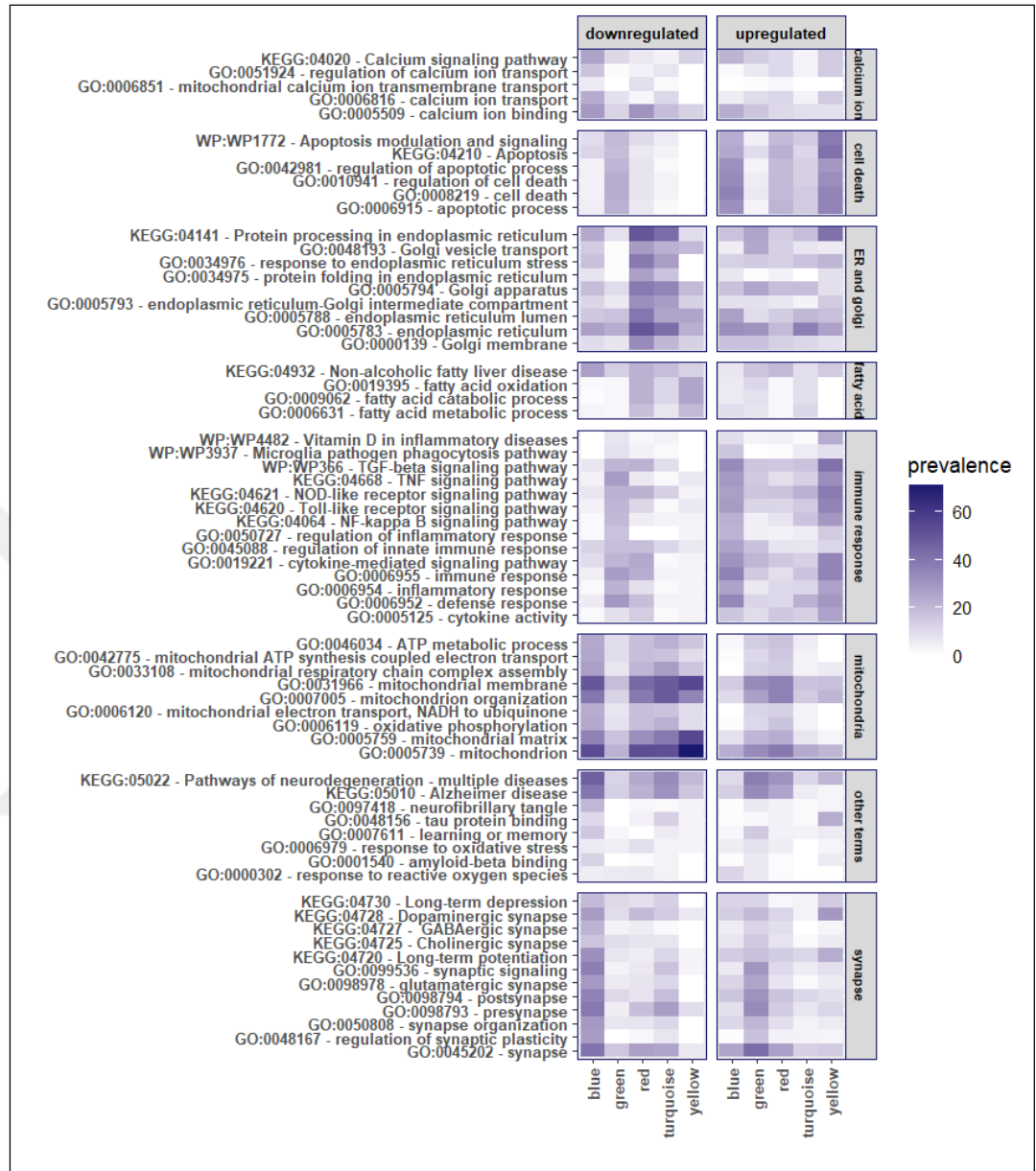


Figure 4.8: Comparison of prevalence rates of selected AD related terms in subtypes for downregulated and upregulated subnetworks

there is a decrease in mitochondrial activity. In the *green* group, it was seen that cell death and immune response, which do not have high prevalence rates in other groups, are downregulated. Moreover, unlike the other groups, the rates of synapse related terms in the *green* group are higher in the upregulated subnetwork results. At the same time, the prevalence rate of terms related to mitochondria is high. In *red* group while endoplasmic reticulum and fatty acid metabolism related terms are downregulated, cell

death related terms are upregulated. Poly unsaturated fatty acids play critical roles in brain development and maintenance, and their deficiency or abnormality is associated with neurological problems such as AD, bipolar disorders and major depression [Zhu et al., 2019]. Linoleic acid, docosahexaenoic acid and eicosapentaenoic acid levels are decreased in AD brain. On the other hand, proinflammatory eicosanoid pathways are upregulated and arachidonic acid levels are increased. At the same time, endocannabinoids and their receptors, which are crucial for the proper functioning of the central nervous system (CNS) and synaptic plasticity, are also decreased [Chew et al., 2020]. Although fatty acid metabolism does not affected in each group, these results show that it plays an active role in some subtypes of the disease.

In the *turquoise* group, endoplasmic reticulum, mitochondria and synapse related terms were downregulated, while cell death and immune response related terms were upregulated, similar to the majority of the groups. Close physical contact exists between mitochondria and ER subcompartments. These contact sides called as mitochondria associated ER membranes (MAMs) play very critical role in calcium homeostasis, lipid synthesis, apoptotic signaling and autophagy. This region is affected by AD pathogenesis. In the study with the APP_{SWE/LON} mouse model, upregulation of MAM-associated proteins was observed, and changes were detected in this region before plaque formation [Hedskog et al., 2013; Yang et al., 2020]. In the *yellow* group, it was observed that the terms associated with fatty acids and mitochondria are downregulated whereas the terms associated with immune response and cell death are upregulated.

As shown in the Figure 4.9, the cell death mechanism was up-regulated in 4 of 5 subtypes, while it was down-regulated only in the green group. In some AD cases, while no apoptotic morphology was observed in any part of the brain, it was observed that the cells were swollen and DNA fragmentation increased[Chi et al., 2018]. With the exception of the red group, cell death and immune response appear to act in together. This is evidence that both mechanisms trigger or activate each other.

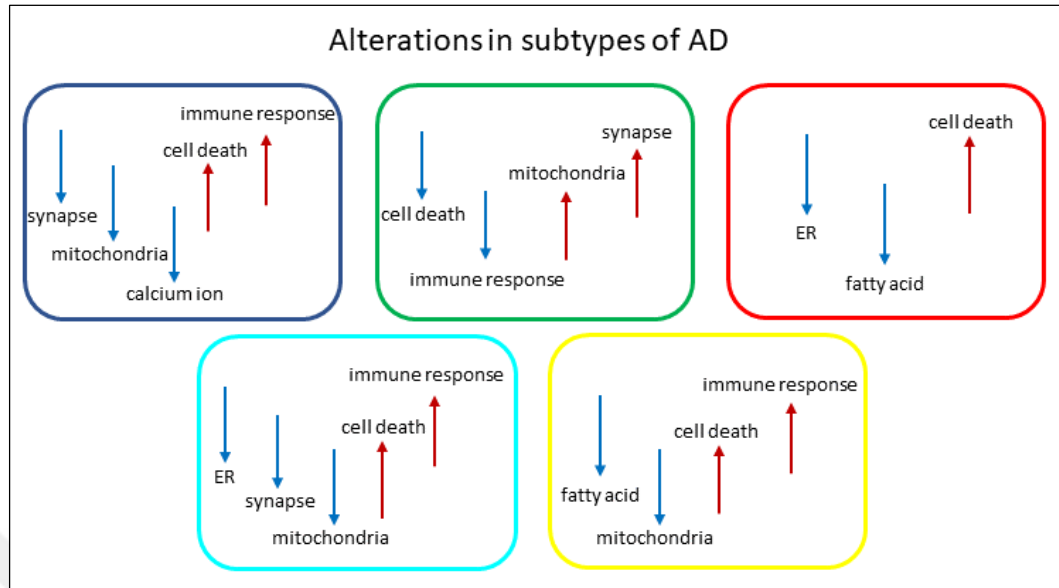


Figure 4.9: Diagram that shows the alterations in AD subtypes in different cellular mechanism.

The evaluation of the enrichment results revealed differences between the disease subtypes on the basis of biological processes and cellular functions. However, in order to contribute to the elucidation of the molecular mechanism of the disease, the genetic alterations that cause these differences should also be revealed. For this purpose, first the prevalence rates of the genes in the subnetworks of each subtype were determined separately, just as it was done for the enrichment results. Next, previously used AD associated genes were determined in each subtype, and the genes that had the highest prevalence rate were selected when compared to the other groups. For *blue* group, the SLC6A3 gene which is present in 23% of the subnetworks, is the most prominent gene. Dopamine transporter (DAT) protein is encoded by SLC6A3, and it has often been associated with neuropsychiatric diseases. It is one of the main regulators of synaptic transmission of dopamine, which plays a very essential role in many brain functions [Reith et al., 2022] . Abnormality in regulations of dopamine levels is associated with depression, bipolar disorder, Parkinson's disease. DAT is mainly localized in the presynaptic region of dopaminergic neurons as a membrane protein [Aykaç and Sehirli, 2020] . Since low levels of dopaminergic neurotransmitters are linked to AD [Pan et al., 2019], the function and significance of this gene can be further examined in terms of AD. Another gene is SNCA, which is found in 21% of subnetworks in the *blue* group.

Although the function of α -synuclein encoded by SNCA is not exactly known, it is considered to be involved in the regulation of dopamine release and transport, the fibrillation of tau, neuroprotection by decreasing caspase activity [Siddiqui et al., 2016] . Despite extensive research linking the SNCA gene to Parkinson's disease, it has also been found that AD patients have significantly higher levels of SNCA in their cerebrospinal fluid than healthy controls, suggesting that SNCA may be a risk factor for AD [Wang et al., 2016] . For the *green* group, PSMC5 and PSMB6 genes, which encode subunits of the proteasome complex, draw attention with 13% prevalence rate in subnetworks. As mentioned earlier, the proteasome is a large, highly complex protein that is responsible for protein degradation. It has two main subcomplexes; 20S catalytic core particle, and 19S regulatory particle at one or both end. While PSMC5 encodes the Rpt6 protein in the ATPase subunit of the 19S subcomplex, PSMB6 encodes the β 1 protein in the β type subunit of the 20S subcomplex [Tanaka, 2009] . In a study by Drummond et al., it was observed that phosphorylated tau has a significant interaction with PSMC5 and some other 19S subcomplex proteins [Drummond et al., 2020] . In another study, it was observed that the PSMC5 gene was significantly overexpressed in AD patients with the APOE4 allele compared to wild type [Conejero-Goldberg et al., 2011] . The first AD related gene identified in the *red* group, with a prevalence rate of 38%, is HSPA5. It encodes the binding immunoglobulin protein (BiP), which is a member of the HSP70 chaperon family. BiP is a protein that is localized in the ER lumen and can initiate the unfolded protein response when the amount of unfolded/misfolded protein in the ER exceeds the capacity of the protein folding machinery [Wang et al., 2017] . In a study by Montibeller and de Belleruche, it was observed that the HSPA5 gene expression is elevated in the temporal cortex of AD patients [Montibeller and de Belleruche, 2018] . In another study, it was reported that the accumulation of A β oligomers upregulates HSPA5 and may lead to ER stress [Soejima et al., 2013] . The other notable gene is TP73, which has a 29% prevalence rate in the red group and encodes the p73 protein. It is a transcription factor that is essential for neuronal development, stem cell renewal, differentiation, cell death, and proliferation [Killick et al., 2011] . In their study with p73 transgenic mice, Wetzel et al. found decreased motor and cognitive functions, brain atrophy and neuronal degeneration. They

also determined that p73 haploinsufficiency causes neuronal degeneration and tangles containing phosphorylated tau [Wetzel et al., 2008] . However, there are studies that disprove this argument [Vardarajan et al., 2013] . MYH8 is the first AD associated gene with a prevalence of 20% in the *turquoise* group. Actin-based motor protein MYH8 is involved in the contraction of skeletal muscles as well as cell adhesion and migration mediated by integrins [Park et al., 2020] . In some studies, it was observed that the MYH8 gene was downregulated in AD cases and was associated with AD [Taguchi et al., 2005; Cacabelos et al., 2014] . Another gene that stands out in terms of prevalence rate is the famous AD related gene PSEN1 that has been studied extensively. The presenilin-1 protein encoded by the PSEN1 gene is a member of the γ -secretase complex and plays an important role in the cleavage of the APP protein, which has a central place in A β plaque formation [Bagaria et al., 2022] . It is well known that this gene's mutations contribute to autosomal dominant early onset Alzheimer's disease [Lanoiselée et al., 2017] . However, it is not predominant in all disease subtypes, suggesting that it may not be the defining feature in all AD cases. HSP90AA1 and HSP90AB1 genes draw attention in the *turquoise* group with a 29% prevalence rate, unlike the other groups. Stressful situations like heat shock, ischemia, heavy metals, and hypoxia trigger the synthesis of heat shock proteins (HSPs). They are divided into several families, and the molecular chaperone HSP90s is one of them. They are able to target specific proteins for degradation, suppress protein aggregation, and solubilize protein aggregates. HSP90 has two types, α and β , and HSP90 α , including HSP90AA1 and HSP90AB1, is sensitive to stressful situations [Ou et al., 2014] . In a study with peripheral blood and brain transcriptomic data, it was determined that HSP90AA1 was differentially expressed and showed a negative correlation with α and β -secretase activity [Qian et al., 2022] . HSP family genes are also present in the first rows of the rank, which motivates further study of these genes in an effort to uncover previously unknown aspects of the disease.

5. CONCLUSION

The disease, described by Alois Alzheimer in 1906, has been the subject of many research since then. Despite these studies conducted from different perspectives, the cause of the disease and the underlying molecular mechanisms are not yet fully understood. Although there are many studies in the literature with transcriptomic data, the majority of these studies handled the data of the patients collectively and thus could not reveal the differences caused by the heterogeneous nature of the disease.

In this study, a personalized approach was adopted, and each patient was compared with the whole control group, leading to patient specific profiles in terms of genes and pathways. Instead of considering only DEGs in the evaluations, subnetwork discoveries were performed, and KPM and DOMINO algorithms were employed for this purpose. From these two algorithms, DOMINO was found to perform significantly better than KPM in capturing disease related genes and terms. As a result of the evaluation of sample based results, it was observed that there was a significant differentiation in terms related to mitochondrial activities, immune response, cell death mechanism and endoplasmic reticulum in AD state. Additionally, it was demonstrated that the sample-based approach produced significantly more detailed results than the bulk analysis, indicating that a personalized approach is a much more effective way to decipher previously undiscovered features of the disease. It was shown that the disease does not progress with the same process in every case and there are differences between subtypes. In addition, several AD-related genes that are common in the subnetworks of the identified disease subtypes were highlighted.

REFERENCES

- Abu-Libdeh B., Douiev L., Amro S., Shahrour M., Ta-Shma A., Miller C., Elpeleg O., Saada A., (2017), "Mutation in the COX4I1 gene is associated with short stature, poor weight gain and increased chromosomal breaks, simulating Fanconi anemia", *European Journal of Human Genetics.*, 25, 1142–1146.
- Aebersold R., Agar J.N., Amster I.J., Baker M.S., Bertozzi C.R., Boja E.S., Costello C.E., Cravatt B.F., Fenselau C., Garcia B.A., Ge Y., Gunawardena J., Hendrickson R.C., Hergenrother P.J., Huber C.G., Ivanov A.R., Jensen O.N., Jewett M.C., Kelleher N.L., Kiessling L.L., Krogan N.J., Zhang B. et al., (2018), "How many human proteoforms are there?", *Nature Chemical Biology.*, 14, 206–214.
- Alcaraz N., Pauling J., Batra R., Barbosa E., Junge A., Christensen A.G.L., Azevedo V., Ditzel H.J., Baumbach J., (2014), "KeyPathwayMiner 4.0: condition-specific pathway analysis by combining multiple omics studies and networks with Cytoscape", *BMC systems biology.*, 8, 1–6.
- Alcaraz N., List M., Dissing-Hansen M., Rehmsmeier M., Tan Q., Mollenhauer J., Ditzel H.J., Baumbach J., (2016), "Robust de novo pathway enrichment with KeyPathwayMiner 5", *F1000Research.*, 5.
- Almanza A., Carlesso A., Chintla C., Creedican S., Doultzinos D., Leuzzi B., Luís A., McCarthy N., Montibeller L., More S., (2019), "Endoplasmic reticulum stress signalling—from basic mechanisms to clinical applications", *The FEBS journal.*, 286, 241–278.
- Alonso A.D., Cohen L.S., Corbo C., Morozova V., ElIdrissi A., Phillips G., Kleiman F.E., (2018), "Hyperphosphorylation of tau associates with changes in its function beyond microtubule stability", *Frontiers in cellular neuroscience.*, 12, 338.
- Alquezar C., Arya S., Kao A.W., (2021), "Tau post-translational modifications: dynamic transformers of tau function, degradation, and aggregation", *Frontiers in Neurology.*, 11, 595532.
- Andrews S.J., Fulton-Howard B., Goate A., (2020), "Interpretation of risk loci from genome-wide association studies of Alzheimer's disease", *The Lancet Neurology.*, 19, 326–335.
- Athieniti E., Spyrou G.M., (2023), "A guide to multi-omics data collection and integration for translational medicine", *Computational and Structural Biotechnology Journal.*, 21, 134–149.
- Attisano L., Tuen Lee-Hoeflich S., (2001), "The Smads", *Genome Biology.*, 2, reviews3010.1.

Aykaç A., Sehirli A.O., (2020), "The role of the SLC transporters protein in the neurodegenerative disorders", *Clinical Psychopharmacology and Neuroscience.*, 18, 174.

Bader G.D., Donaldson I., Wolting C., Ouellette B.F.F., Pawson T., Hogue C.W. v, (2001), "BIND—The Biomolecular Interaction Network Database", *Nucleic Acids Research.*, 29, 242–245.

Bagaria J., Bagyinszky E., An S.S.A., (2022), "Genetics, Functions, and Clinical Impact of Presenilin-1 (PSEN1) Gene", *International Journal of Molecular Sciences.*, 23, 10970.

Barabasi A.L., Oltvai Z.N., (2004), "Network biology: understanding the cell's functional organization", *Nature reviews genetics.*, 5, 101–113.

Batra R., Alcaraz N., Gitzhofer K., Pauling J., Ditzel H.J., Hellmuth M., Baumbach J., List M., (2017), "On the performance of de novo pathway enrichment", *npj Systems Biology and Applications.*, 3, 6.

Bedford L., Lowe J., Dick L.R., Mayer R.J., Brownell J.E., (2011), "Ubiquitin-like protein conjugation and the ubiquitin–proteasome system as drug targets", *Nature Reviews Drug Discovery.*, 10, 29–46.

Beisser D., Klau G.W., Dandekar T., Müller T., Dittrich M.T., (2010), "BioNet: an R-Package for the functional analysis of biological networks", *Bioinformatics.*, 26, 1129–1130.

Bellenguez C., Küçükali F., Jansen I.E., Kleideidam L., Moreno-Grau S., Amin N., Naj A.C., Campos-Martin R., Grenier-Boley B., Andrade V., Holmans P.A., Boland A., Damotte V., van der Lee S.J., Costa M.R., Kuulasmaa T., Yang Q., de Rojas I., Bis J.C., Yaqub A., Prokic I., Sánchez-Arjona M.B. et al., (2022), "New insights into the genetic etiology of Alzheimer's disease and related dementias", *Nature Genetics.*, 54, 412–436.

Bennett D.A., Buchman A.S., Boyle P.A., Barnes L.L., Wilson R.S., Schneider J.A., (2018), "Religious Orders Study and Rush Memory and Aging Project", *Journal of Alzheimer's Disease.*

Breitling R., Amtmann A., Herzyk P., (2004), "Graph-based iterative Group Analysis enhances microarray interpretation", *BMC bioinformatics.*, 5, 1–10.

Cacabelos R., Cacabelos P., Torrellas C., (2014), "Personalized medicine of Alzheimer's disease", *Handbook of pharmacogenomics and stratified medicine.*, 563.

Cenini G., Voos W., (2019), "Mitochondria as potential targets in Alzheimer disease therapy: an update", *Frontiers in pharmacology.*, 10, 902.

Chen E.Y., Tan C.M., Kou Y., Duan Q., Wang Z., Meirelles G.V., Clark N.R., Ma'ayan A., (2013), "Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool", *BMC bioinformatics.*, 14, 1–14.

Chen G. fang, Xu T. hai, Yan Y., Zhou Y. ren, Jiang Y., Melcher K., Xu H.E., (2017), "Amyloid beta: structure, biology and structure-based therapeutic development", *Acta Pharmacologica Sinica*, 38, 1205–1235.

Chen S., Townsend K., Goldberg T.E., Davies P., Conejero-Goldberg C., (2010), "MAPT isoforms: differential transcriptional profiles related to 3R and 4R splice variants", *Journal of Alzheimer's Disease*, 22, 1313–1329.

Chen X., Zhang Y., Wang Q., Qin Y., Yang X., Xing Z., Shen Y., Wu H., Qi Y., (2021), "The function of SUMOylation and its crucial roles in the development of neurological diseases", *The FASEB Journal*, 35, e21510.

Chen Y., Neve R.L., Liu H., (2012), "Neddylation dysfunction in Alzheimer's disease", *Journal of cellular and molecular medicine*, 16, 2583–2591.

Chew H., Solomon V.A., Fonteh A.N., (2020), "Involvement of lipids in Alzheimer's disease pathology and potential therapies", *Frontiers in physiology*, 11, 598.

Chi H., Chang H.Y., Sang T.K., (2018), "Neuronal cell death mechanisms in major neurodegenerative diseases", *International journal of molecular sciences*, 19, 3082.

Chong F.P., Ng K.Y., Koh R.Y., Chye S.M., (2018), "Tau Proteins and Tauopathies in Alzheimer's Disease", *Cellular and Molecular Neurobiology*, 38, 965–980.

Chuang H., Lee E., Liu Y., Lee D., Ideker T., (2007), "Network-based classification of breast cancer metastasis", *Molecular systems biology*, 3, 140.

Conejero-Goldberg C., Hyde T.M., Chen S., Dreses-Werringloer U., Herman M.M., Kleinman J.E., Davies P., Goldberg T.E., (2011), "Molecular signatures in post-mortem brain tissue of younger individuals at high risk for Alzheimer's disease as based on APOE genotype", *Molecular Psychiatry*, 16, 836–847.

Ding Z., Guo W., Gu J., (2018), "ClustEx2: gene module identification using density-based network hierarchical clustering", 2018 Chinese Automation Congress (CAC), IEEE, pp. 2407–2412.

Dong Z., Chen Y., (2013), "Transcriptomics: Advances and approaches", *Science China Life Sciences*, 56, 960–967.

Drummond E., Pires G., MacMurray C., Askenazi M., Nayak S., Bourdon M., Safar J., Ueberheide B., Wisniewski T., (2020), "Phosphorylated tau interactome in the human Alzheimer's disease brain", *Brain*, 143, 2803–2817.

Dubois B., Hampel H., Feldman H.H., Scheltens P., Aisen P., Andrieu S., Bakardjian H., Benali H., Bertram L., Blennow K., (2016), "Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria", *Alzheimer's & Dementia*, 12, 292–323.

Emanetci E., Çakır T., (2021), "Network-Based Analysis of Cognitive Impairment and Memory Deficits from Transcriptome Data", *Journal of Molecular Neuroscience.*, 71, 2415–2428.

Gadhav K., Bolshette N., Ahire A., Pardeshi R., Thakur K., Trandafir C., Istrate A., Ahmed S., Lahkar M., Muresanu D.F., (2016), "The ubiquitin proteasomal system: a potential target for the management of Alzheimer's disease", *Journal of cellular and molecular medicine.*, 20, 1392–1407.

Ghiassian S.D., Menche J., Barabási A.L., (2015), "A DIseAse MOdule Detection (DIAMOND) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome", *PLoS computational biology.*, 11, e1004120.

Goel P., Chakrabarti S., Goel K., Bhutani K., Chopra T., Bali S., (2022), "Neuronal cell death mechanisms in Alzheimer's disease: An insight", *Frontiers in Molecular Neuroscience.*, 15.

Guo T., Zhang D., Zeng Y., Huang T.Y., Xu H., Zhao Y., (2020), "Molecular and cellular mechanisms underlying the pathogenesis of Alzheimer's disease", *Molecular neurodegeneration.*, 15, 1–37.

Hashimoto S., Saido T.C., (2018), "Critical review: involvement of endoplasmic reticulum stress in the aetiology of Alzheimer's disease", *Open biology.*, 8, 180024.

Hasin Y., Seldin M., Lusi A., (2017), "Multi-omics approaches to disease", *Genome Biology.*, 18, 83.

Hedskog L., Pinho C.M., Filadi R., Rönnbäck A., Hertwig L., Wiehager B., Larssen P., Gellhaar S., Sandebring A., Westerlund M., (2013), "Modulation of the endoplasmic reticulum–mitochondria interface in Alzheimer's disease and related models", *Proceedings of the National Academy of Sciences.*, 110, 7916–7921.

Heneka M.T., Carson M.J., Khoury J. el, Landreth G.E., Brosseron F., Feinstein D.L., Jacobs A.H., Wyss-Coray T., Vitorica J., Ransohoff R.M., Herrup K., Frautschy S.A., Finsen B., Brown G.C., Verkhratsky A., Yamanaka K., Koistinaho J., Latz E., Halle A., Petzold G.C., Town T., Morgan D., Shinohara M.L., Perry V.H., Holmes C., Bazan N.G., Brooks D.J., Hunot S., Joseph B., Deigendesch N., Garaschuk O., Boddeke E., Dinarello C.A., Breitner J.C., Cole G.M., Golenbock D.T., Kummer M.P., (2015), "Neuroinflammation in Alzheimer's disease", *The Lancet Neurology.*, 14, 388–405.

Heppner F.L., Ransohoff R.M., Becher B., (2015), "Immune attack: the role of inflammation in Alzheimer disease", *Nature Reviews Neuroscience.*, 16, 358–372.

Hippius H., Neundörfer G., (2003), "The discovery of Alzheimer's disease", *Dialogues in Clinical Neuroscience.*, 5, 101–108.

Hu Y.S., Xin J., Hu Y., Zhang L., Wang J., (2017), "Analyzing the genes related to Alzheimer's disease via a network and pathway-based approach", *Alzheimer's Research & Therapy*, 9, 29.

Jellinger K.A., (2022), "Recent update on the heterogeneity of the Alzheimer's disease spectrum", *Journal of Neural Transmission*, 129, 1–24.

Keshava Prasad T.S., Goel R., Kandasamy K., Keerthikumar S., Kumar S., Mathivanan S., Telikicherla D., Raju R., Shafreen B., Venugopal A., Balakrishnan L., Marimuthu A., Banerjee S., Somanathan D.S., Sebastian A., Rani S., Ray S., Harrys Kishore C.J., Kanth S., Ahmed M., Kashyap M.K., Mohmood R., Ramachandra Y.L., Krishna V., Rahiman B.A., Mohan S., Ranganathan P., Ramabadran S., Chaerkady R., Pandey A., (2009), "Human Protein Reference Database—2009 update", *Nucleic Acids Research*, 37, D767–D772.

Killick R., Niklison-Chirou M., Tomasini R., Bano D., Rufini A., Grespi F., Velletri T., Tucci P., Sayan B.S., Conforti F., Gallagher E., Nicotera P., Mak T.W., Melino G., Knight R.A., Agostini M., (2011), "p73: A Multifunctional Protein in Neurobiology", *Molecular Neurobiology*, 43, 139–146.

Kitagishi Y., Nakanishi A., Ogura Y., Matsuda S., (2014), "Dietary regulation of PI3K/AKT/GSK-3 β pathway in Alzheimer's disease", *Alzheimer's Research & Therapy*, 6, 35.

Kolberg L., Raudvere U., Kuzmin I., Vilo J., Peterson H., (2020), "gprofiler2--an R package for gene list functional enrichment analysis and namespace conversion toolset g: Profiler", *F1000Research*, 9.

Kumar V., (2019), "Toll-like receptors in the pathogenesis of neuroinflammation", *Journal of Neuroimmunology*, 332, 16–30.

LaFerla F.M., Green K.N., Oddo S., (2007), "Intracellular amyloid- β in Alzheimer's disease", *Nature Reviews Neuroscience*, 8, 499–509.

Lanoiselée H.M., Nicolas G., Wallon D., Rovelet-Lecrux A., Lacour M., Rousseau S., Richard A.C., Pasquier F., Rollin-Sillaire A., Martinaud O., (2017), "APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease: A genetic screening study of familial and sporadic cases", *PLoS medicine*, 14, e1002270.

Lanzillotta C., di Domenico F., Perluigi M., Butterfield D.A., (2019), "Targeting Mitochondria in Alzheimer Disease: Rationale and Perspectives", *CNS Drugs*, 33, 957–969.

Larsen S.J., Schmidt H.H.H.W., Baumbach J., (2020), "De novo and supervised endophenotyping using network-guided ensemble learning", *Systems Medicine*, 3, 8–21.

- Lazareva O., Baumbach J., List M., Blumenthal D.B., (2021), "On the limits of active module identification", *Briefings in Bioinformatics.*, 22, bbab066.
- Lee L., Sakurai M., Matsuzaki S., Arancio O., Fraser P., (2013), "SUMO and Alzheimer's Disease", *NeuroMolecular Medicine.*, 15, 720–736.
- Levi H., Elkon R., Shamir R., (2021), "DOMINO: a network-based active module identification algorithm with reduced rate of false calls", *Molecular systems biology.*, 17, e9593.
- Levi H., Rahmanian N., Elkon R., Shamir R., (2022), "The DOMINO web-server for active module identification analysis", *Bioinformatics.*, 38, 2364–2366.
- Li J.Q., Yu J.T., Jiang T., Tan L., (2015)a, "Endoplasmic Reticulum Dysfunction in Alzheimer's Disease", *Molecular Neurobiology.*, 51, 383–395.
- Li X., Long J., He T., Belshaw R., Scott J., (2015)b, "Integrated genomic approaches identify major pathways and upstream regulators in late onset Alzheimer's disease", *Scientific Reports.*, 5, 12393.
- Licata L., Briganti L., Peluso D., Perfetto L., Iannuccelli M., Galeota E., Sacco F., Palma A., Nardoza A.P., Santonico E., Castagnoli L., Cesareni G., (2012), "MINT, the molecular interaction database: 2012 update", *Nucleic Acids Research.*, 40, D857–D861.
- List M., Alcaraz N., Dissing-Hansen M., Ditzel H.J., Mollenhauer J., Baumbach J., (2016), "KeyPathwayMinerWeb: online multi-omics network enrichment", *Nucleic Acids Research.*, 44, W98–W104.
- Long H.Z., Cheng Y., Zhou Z.W., Luo H.Y., Wen D.D., Gao L.C., (2021), "PI3K/AKT signal pathway: a target of natural products in the prevention and treatment of Alzheimer's disease and Parkinson's disease", *Frontiers in pharmacology.*, 12, 648636.
- Love M., Anders S., Huber W., (2014), "Differential analysis of count data—the DESeq2 package", *Genome Biol.*, 15, 10–1186.
- Lowe R., Shirley N., Bleackley M., Dolan S., Shafee T., (2017)a, "Transcriptomics technologies", *PLoS computational biology.*, 13, e1005457.
- Lowe R., Shirley N., Bleackley M., Dolan S., Shafee T., (2017)b, "Transcriptomics technologies", *PLoS computational biology.*, 13, e1005457.
- Ma H., Schadt E.E., Kaplan L.M., Zhao H., (2011), "COSINE: COndition-SpecIfic sub-NEtwork identification using a global optimization method", *Bioinformatics.*, 27, 1290–1298.

Manzoni C., Kia D.A., Vandrovcova J., Hardy J., Wood N.W., Lewis P.A., Ferrari R., (2018), "Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences", *Briefings in Bioinformatics.*, 19, 286–302.

Marques-Coelho D., Iohan L. da C.C., Melo de Farias A.R., Flaig A., Letournel F., Martin-Négrier M.L., Chapon F., Faisant M., Godfraind C., Maurage C.A., Deramecourt V., Duchesne M., Meyronnet D., Streichenberger N., de Paula A.M., Rigau V., Vandenbos-Burel F., Duyckaerts C., Seilhean D., Milin S., Chiforeanu D.C., Laquerrière A., Marguet F., Lannes B., Lambert J.C., Costa M.R., Network T.B.N.N., (2021), "Differential transcript usage unravels gene expression alterations in Alzheimer's disease human brains", *npj Aging and Mechanisms of Disease.*, 7, 2.

McCorkindale A.N., Mundell H.D., Guennewig B., Sutherland G.T., (2022), "Vascular Dysfunction Is Central to Alzheimer's Disease Pathogenesis in APOE e4 Carriers", *International journal of molecular sciences.*, 23, 7106.

Mechteridis K., Lauber M., Baumbach J., List M., (2022), "KeyPathwayMineR: de novo pathway enrichment in the r ecosystem", *Frontiers in Genetics.*, 2844.

Momtazmanesh S., Perry G., Rezaei N., (2020), "Toll-like receptors in Alzheimer's disease", *Journal of Neuroimmunology.*, 348, 577362.

Montibeller L., de Belleruche J., (2018), "Amyotrophic lateral sclerosis (ALS) and Alzheimer's disease (AD) are characterised by differential activation of ER stress pathways: focus on UPR target genes", *Cell Stress and Chaperones.*, 23, 897–912.

Morales I., Guzmán-Martínez L., Cerda-Troncoso C., Fariás G.A., Maccioni R.B., (2014), "Neuroinflammation in the pathogenesis of Alzheimer's disease. A rational framework for the search of novel therapeutic approaches", *Frontiers in cellular neuroscience.*, 8, 112.

Morgan S.L., Naderi P., Koler K., Pita-Juarez Y., Prokopenko D., Vlachos I.S., Tanzi R.E., Bertram L., Hide W.A., (2022), "Most Pathways Can Be Related to the Pathogenesis of Alzheimer's Disease", *Frontiers in Aging Neuroscience.*, 14.

Nacu Ș., Critchley-Thorne R., Lee P., Holmes S., (2007), "Gene expression network analysis and applications to immunology", *Bioinformatics.*, 23, 850–858.

Naseri N.N., Wang H., Guo J., Sharma M., Luo W., (2019), "The complexity of tau in Alzheimer's disease", *Neuroscience Letters.*, 705, 183–194.

Neff R.A., Wang M., Vatansever S., Guo L., Ming C., Wang Q., Wang E., Horgusluoglu-Moloch E., Song W. min, Li A., Castranio E.L., TCW J., Ho L., Goate A., Fossati V., Noggle S., Gandy S., Ehrlich M.E., Katsel P., Schadt E., Cai D., Brennand K.J., Haroutunian V., Zhang B., (2022), "Molecular subtyping of Alzheimer's disease using RNA sequencing data reveals novel mechanisms and targets", *Science Advances.*, 7, eabb5398.

Oddo S., (2008), "The ubiquitin-proteasome system in Alzheimer's disease", *Journal of cellular and molecular medicine.*, 12, 363–373.

Ou J.R., Tan M.S., Xie A.M., Yu J.T., Tan L., (2014), "Heat Shock Protein 90 in Alzheimer's Disease", (Chang, R. C. C., Ed.)*BioMed Research International.*, 2014, 796869.

Pan X., Kaminga A.C., Wen S.W., Wu X., Acheampong K., Liu A., (2019), "Dopamine and dopamine receptors in Alzheimer's disease: a systematic review and network meta-analysis", *Frontiers in aging neuroscience.*, 11, 175.

Park H., Kim D., Kim D., Park J., Koh Y., Yoon S.S., (2020), "Truncation of MYH8 tail in AML: a novel prognostic marker with increase cell migration and epithelial–mesenchymal transition utilizing RAF/MAPK pathway", *Carcinogenesis.*, 41, 817–827.

Pathan M., Keerthikumar S., Ang C., Gangoda L., Quek C.Y.J., Williamson N.A., Mouradov D., Sieber O.M., Simpson R.J., Salim A., (2015), "FunRich: An open access standalone functional enrichment and interaction network analysis tool", *Proteomics.*, 15, 2597–2601.

Qian X. hang, Liu X. li, Tang H. dong, (2022), "Integrating peripheral blood and brain transcriptomics to identify immunological features associated with Alzheimer's disease in mild cognitive impairment patients", *Frontiers in Immunology.*, 13.

Qiu C., Kivipelto M., von Strauss E., (2022), "Epidemiology of Alzheimer's disease: occurrence, determinants, and strategies toward intervention", *Dialogues in clinical neuroscience.*

Raudvere U., Kolberg L., Kuzmin I., Arak T., Adler P., Peterson H., Vilo J., (2019), "g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)", *Nucleic acids research.*, 47, W191–W198.

Razani E., Pourbagheri-Sigaroodi A., Safaroghli-Azar A., Zoghi A., Shanaki-Bavarsad M., Bashash D., (2021), "The PI3K/Akt signaling axis in Alzheimer's disease: a valuable target to stimulate or suppress?", *Cell Stress and Chaperones.*, 26, 871–887.

Reimand J., Isserlin R., Voisin V., Kucera M., Tannus-Lopes C., Rostamianfar A., Wadi L., Meyer M., Wong J., Xu C., Merico D., Bader G.D., (2019), "Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap", *Nature Protocols.*, 14, 482–517.

Reith M.E.A., Kortagere S., Wiers C.E., Sun H., Kurian M.A., Galli A., Volkow N.D., Lin Z., (2022), "The dopamine transporter gene SLC6A3: multidisease risks", *Molecular Psychiatry.*, 27, 1031–1046.

Ritchie M.E., Phipson B., Wu D., Hu Y., Law C.W., Shi W., Smyth G.K., (2015), "limma powers differential expression analyses for RNA-sequencing and microarray studies", *Nucleic Acids Research.*, 43, e47–e47.

Schwarz D.S., Blower M.D., (2016), "The endoplasmic reticulum: structure, function and response to cellular signaling", *Cellular and molecular life sciences.*, 73, 79–94.

Sherman B.T., Hao M., Qiu J., Jiao X., Baseler M.W., Lane H.C., Imamichi T., Chang W., (2022), "DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update)", *Nucleic acids research.*, 50, W216–W221.

Siddiqui I.J., Pervaiz N., Abbasi A.A., (2016), "The Parkinson Disease gene SNCA: Evolutionary and structural insights with pathological implication", *Scientific Reports.*, 6, 24475.

Soejima N., Ohyagi Y., Nakamura N., Himeno E., M Iinuma K., Sakae N., Yamasaki R., Tabira T., Murakami K., Irie K., (2013), "Intracellular accumulation of toxic turn amyloid- β is associated with endoplasmic reticulum stress in Alzheimer's disease", *Current Alzheimer Research.*, 10, 11–20.

Stark C., Breitkreutz B.J., Reguly T., Boucher L., Breitkreutz A., Tyers M., (2006), "BioGRID: a general repository for interaction datasets", *Nucleic acids research.*, 34, D535–D539.

Szklarczyk D., Gable A.L., Nastou K.C., Lyon D., Kirsch R., Pyysalo S., Doncheva N.T., Legeay M., Fang T., Bork P., Jensen L.J., von Mering C., (2021), "The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets", *Nucleic Acids Research.*, 49, D605–D612.

Taguchi K., Yamagata H.D., Zhong W., Kamino K., Akatsu H., Hata R., Yamamoto T., Kosaka K., Takeda M., Kondo I., (2005), "Identification of hippocampus-related candidate genes for Alzheimer's disease", *Annals of neurology.*, 57, 585–588.

Tanaka K., (2009), "The proteasome: overview of structure and functions", *Proceedings of the Japan Academy, Series B.*, 85, 12–36.

Tracy T.E., Gan L., (2018), "Tau-mediated synaptic and neuronal dysfunction in neurodegenerative disease", *Current Opinion in Neurobiology.*, 51, 134–138.

Ulitsky I., Karp R.M., Shamir R., (2008), "Detecting disease-specific dysregulated pathways via analysis of clinical expression profiles", *Annual International Conference on Research in Computational Molecular Biology.*, Springer, pp. 347–359.

Uzuner D., Akkoç Y., Peker N., Pir P., Gözüaık D., Çakır T., (2021), "Transcriptional landscape of cellular networks reveal interactions driving the dormancy mechanisms in cancer", *Scientific reports.*, 11, 1–10.

van Oostveen W.M., de Lange E.C.M., (2021), "Imaging techniques in Alzheimer's disease: a review of applications in early diagnosis and longitudinal monitoring", *International journal of molecular sciences.*, 22, 2110.

Vardarajan B., Vergote D., Tissir F., Logue M., Yang J., Daude N., Ando K., Rogaeva E., Lee J., Cheng R., Brion J.P., Ghani M., Shi B., Baldwin C.T., Kar S., Mayeux R., Fraser P., Goffinet A.M., George-Hyslop P.S., Farrer L.A., Westaway D., (2013), "Role of p73 in Alzheimer disease: lack of association in mouse models or in human cohorts", *Molecular Neurodegeneration.*, 8, 10.

Wan Y.W., Al-Ouran R., Mangleburg C.G., Perumal T.M., Lee T. v, Allison K., Swarup V., Funk C.C., Gaiteri C., Allen M., (2020), "Meta-analysis of the Alzheimer's disease human brain transcriptome and functional dissection in mouse models", *Cell reports.*, 32.

Wang J., Lee J., Liem D., Ping P., (2017), "HSPA5 Gene encoding Hsp70 chaperone BiP in the endoplasmic reticulum", *Gene.*, 618, 14–23.

Wang Q., Tian Q., Song X., Liu Y., Li W., (2016), "SNCA gene polymorphism may contribute to an increased risk of Alzheimer's disease", *Journal of clinical laboratory analysis.*, 30, 1092–1099.

Wang W., Zhao F., Ma X., Perry G., Zhu X., (2020), "Mitochondria dysfunction in the pathogenesis of Alzheimer's disease: recent advances", *Molecular Neurodegeneration.*, 15, 30.

Wetzel M.K., Naska S., Laliberté C.L., Rymar V. v, Fujitani M., Biernaskie J.A., Cole C.J., Lerch J.P., Spring S., Wang S.H., Frankland P.W., Henkelman R.M., Josselyn S.A., Sadikot A.F., Miller F.D., Kaplan D.R., (2008), "p73 Regulates Neurodegeneration and Phospho-Tau Accumulation during Aging and Alzheimer's Disease", *Neuron.*, 59, 708–721.

Wightman D.P., Jansen I.E., Savage J.E., Shadrin A.A., Bahrami S., Holland D., Rongve A., Børte S., Winsvold B.S., Drange O.K., Martinsen A.E., Skogholt A.H., Willer C., Bråthen G., Bosnes I., Nielsen J.B., Fritsche L.G., Thomas L.F., Pedersen L.M., Gabrielsen M.E., Johnsen M.B., Team 23andMe Research et al., (2021), "A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease", *Nature Genetics.*, 53, 1276–1282.

Williams J.B., Cao Q., Yan Z., (2021), "Transcriptomic analysis of human brains with Alzheimer's disease reveals the altered expression of synaptic genes linked to cognitive deficits", *Brain communications.*, 3, fcab123.

Xenarios I., Rice D.W., Salwinski L., Baron M.K., Marcotte E.M., Eisenberg D., (2000), "DIP: the Database of Interacting Proteins", *Nucleic Acids Research.*, 28, 289–291.

Yang C., Xu P., (2022), "The role of transforming growth factor β 1 /Smad pathway in Alzheimer's disease inflammation pathology", *Molecular Biology Reports.*

Yang M., Li C., Yang S., Xiao Y., Xiong X., Chen W., Zhao H., Zhang Q., Han Y., Sun L., (2020), "Mitochondria-associated ER membranes—the origin site of autophagy", *Frontiers in cell and developmental biology.*, 8, 595.

Yurduseven K., Babal Y.K., Celik E., Kerman B.E., Kurnaz I.A., (2022), "Multiple Sclerosis Biomarker Candidates Revealed by Cell-Type-Specific Interactome Analysis", *OMICS: A Journal of Integrative Biology.*, 26, 305–317.

Zhang H., Wong J.P., Ni G., Cano P., Dittmer D.P., Damania B., (2022), "Mitochondrial protein, TBRG4, modulates KSHV and EBV reactivation from latency", *PLoS pathogens.*, 18, e1010990.

Zhou X., Su Z., (2007), "EasyGO: Gene Ontology-based annotation and functional enrichment analysis tool for agronomical species", *BMC genomics.*, 8, 1–4.

Zhu T.B., Zhang Z., Luo P., Wang S.S., Peng Y., Chu S.F., Chen N.H., (2019), "Lipid metabolism in Alzheimer's disease", *Brain Research Bulletin.*, 144, 68–74.

BIOGRAPHY

Betül CEYLAN completed her undergraduate degree at Fatih University, Department of Genetics and Bioengineering in 2015. She is an M.Sc. student in Bioinformatics and Systems Biology Program under Bioengineering Department at Gebze Technical University Graduate School of Natural and Applied Sciences.

