DISCOVERY OF GENES AND MOLECULAR PROCESSES THAT ARE IMPORTANT FOR THE PATHOGENESIS OF ALZHEIMER'S DISEASE

by

Alexander Kwakye

A Thesis submitted to the Faculty of the

Charles E. Schmidt College of Medicine

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

Florida Atlantic University

Boca Raton, FL

August 2020

DISCOVERY OF GENES AND MOLECULAR PROCESSES THAT ARE IMPORTANT FOR THE PATHOGENESIS OF ALZHEIMER'S DISEASE

by

Alexander Kwakye

This thesis was prepared under the direction of the candidate's thesis advisor, Dr. Zhongwei Li, Department of Biomedical Science, and has been approved by all members of the supervisory committee. It was submitted to the faculty of the Charles E. Schmidt College of Medicine and was accepted in partial fulfillment of the requirements for the degree of Master of Science.

SUPERVISORY COMMITTEE:

Zhongwei Li
Zhongwei Li (Jul 23, 2020 15:23 EDT)

Zhongwei Li, Ph.D.
Thesis Advisor

Lun-Ching Chang
Lun-Ching Chang (Jul 23, 2020 18:41 EDT)

Lun-Ching Chang, Ph.D.

Hanqi Zhuang
Hanqi Zhuang (Jul 23, 2020 21:41 EDT)

Hanqi Zhuang, Ph.D.

Janet Robishaw
Janet Robishaw (Jul 24, 2020 00:15 EDT)

Janet D. Robishaw, Ph.D.
Chair, Department of Biomedical Science

Phillip M. Boiselle
Phillip M. Boiselle (Jul 27, 2020 13:37 EDT)

Phillip M. Boiselle, M.D..
Dean, Charles E. Schmidt College of Medicine

Robert W. Stackman Jr., Ph.D.
Dean, Graduate College

July 27, 2020
Date

ACKNOWLEDGEMENTS

My profound gratitude goes to my advisor, Dr. Zhongwei Li, whose enormous guidance and support fostered my thoughts throughout this work. I am highly indebted to him for guiding me on how I can effectively organize and communicate my thoughts. Similarly, I deeply appreciate the guidance of my other advisory committee members: Dr. Lun-Ching Chang and Dr. Hanqi Zhuang. Their respective statistical and computational insights profoundly contributed to the success of this work. As well, insights from Dr. Gerda Breitwieser greatly influenced this work.

Many thanks to the staff and faculty of the graduate program in Biomedical Science for their dedication to training young scientists to advance science and society. I also thank Geri Mayer at the department of Biological Science, where I served as a teaching assistant, for her support. I cannot forget my colleague students who helped me to integrate into my new environment, thank you all.

Ultimately, I thank my family, especially my mother and my friends who have unconditionally supported me throughout my life.

ABSTRACT

Author:	Alexander Kwakye
Title:	Discovery of genes and molecular processes that are important for the pathogenesis of Alzheimer's disease
Institution:	Florida Atlantic University
Thesis Advisor:	Dr. Zhongwei Li
Degree:	Master of Science
Year:	2020

Alzheimer's Disease (AD) is a complex brain disorder that affects at least one in every ten persons aged 65 and above worldwide. The pathogenesis of this disorder remains elusive. In this work, we utilized a rich set of publicly available gene expression data to elucidate the genes and molecular processes that may underlie its pathogenesis. We developed a new ranking score to prioritize molecular pathways enriched in differentially expressed genes during AD. After applying our new ranking score, GO categories such as cotranslational protein targeting to membrane, SRP-dependent cotranslational protein targeting to membrane, and spliceosomal snRNP assembly were found to be significantly associated with AD. We also confirm the protein-protein interaction between APP, NPAS4 and ARNT2 and explain that this interaction could be implicated in AD. This interaction could serve as a theoretical framework for further analyses into the role of NPAS4 and other immediate-early genes in AD pathogenesis.

DISCOVERY OF GENES AND MOLECULAR PROCESSES THAT ARE IMPORTANT FOR THE PATHOGENESIS OF ALZHEIMER'S DISEASE

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Summary of research (Extended abstract)

Dementia is characterized by memory loss and reduction of cognitive capabilities leading to a reduced ability to perform routine tasks [1]. One common cause of dementia is Alzheimer's disease (AD), a disorder characterized by the accumulation of proteins such as beta amyloids and hyperphosphorylated tau [2]. AD could contribute to about 80% of dementia globally [3, 4]. In the United States, 6 million or more people are affected by AD [5] as at 2019; after age 65, 1 in every 10 persons may be affected [6]. This ratio is estimated to rise to 2 in every 10 persons if there is no effective interventions [7]. The prognosis could be in some states in the United States [8].

Elsewhere, the rates are not any lower although there are some variations based on geography [9]. The prevalence of AD in Europe is about 10% [10], while in Japanese populations it can be as low as 4% [11].

Several factors cause AD, some are environmental, others are genetic. The interaction between the genetic and environmental factors play key roles in the pathogenesis of the disorder. There are tools and approaches to examine how these factors lead to AD. For instance, genetic linkage mapping, genome wide association studies (GWAS), high-throughput microarray and sequencing approaches are some of the tools employed in interrogating the genetic factors responsible for AD.

The identification of some genes responsible for early onset AD (which is known to be an autosomal inherited trait) were done through linkage analyses [12]. Also, genome wide association studies have identified loci and genetic variants in the genome that increase a person's risk of developing late AD [13-22]. APP gene was discovered through genetic linkage analyses [12, 23, 24]. This gene codes for the amyloid precursor protein which is very key to amyloidosis–an overarching feature of AD [25].

Usually, these genetic factors interact with environmental factors that result in changes in how these genes are expressed. The first study to holistically examine how genes are differentially expressed in major brain regions affected by AD was [26]. Later, similar efforts ensued including [27] and [28]. The data from these studies have been deposited in the NCBI Gene Expression Omnibus (GEO) database to foster rediscovery studies.

By employing enrichment methods such as over-representation analyses and network-topology based analyses (NTA), this study reveals biological pathways and processes enriched in the genes that are differentially expressed from these studies. These biological pathways and processes are prioritized with a new ranking score. The biological processes and pathways ranked by this score enhance our understanding of the processes or pathways that are implicated in AD.

Through our analyses, we confirm that NPAS4, ARNT2 genes may be complicit in AD pathogenesis through their interactions with APP. We explain that this interaction presents a highly interconnected network between synaptic activities, gene transcription and amyloidosis, a network that can serve as a theoretical framework for further analyses into the role of NPAS4 and other immediate early genes in AD pathogenesis. Also, molecular processes such as cotranslational protein targeting to membrane, SRP-dependent

cotranslational protein targeting to membrane, protein targeting to ER, spliceosomal snRNP assembly, establishment of protein localization to endoplasmic reticulum pathways could play key roles in AD pathogenesis. To the best of our knowledge, this is the first comprehensive study to perform a systematic analyses of different publicly available gene expression data in Alzheimer's disease, and discuss the differentially expressed genes, biological processes and molecular functions that could cause the disorder.

1.2 Classification and characteristics of Alzheimer's disease (AD)

1.2.1 Classification of AD

AD can be classified as either early or late onset. Early onset, also called familial AD, is characterized by an autosomal dominant inheritance of one or more of a set of genes including APP, PSEN1 and PSEN2 genes [29]. Individuals with early onset AD experience the symptoms usually before age 65 [30]. Early onset AD contributes to less than 5% of all AD cases [31]. Late-onset or sporadic AD could cause more than 95% of all AD cases. This form of AD is not caused by direct inheritance of some specific genes but may be due to the interaction of environmental and genetic factors [32, 33]. One of well-studied genetic factors associated with late AD is the APOE gene [34, 35]. Early and late onset have some genetic underpinnings.

There are other features that early and late onset AD share like loss of memory, reduction in cognitive skills, as well as similar neuropathological features. Usually, the difference observed in these two forms of the disorder is in respect to the degree of the neuropathology, not the nature of the pathology [36, 37]. The difference is mostly attributed to the presence of a 'cognitive reserve' in early AD, as proposed by [38].

Cognitive reserve is the idea that individuals create a reservoir through engagement in higher cognitive activities, and this enables them to function normally in the face of a brain pathology. Early onset AD patients had milder symptoms due to the presence of large cognitive reserve, and vice versa. In section 1.2.2.1 Psychological features of Alzheimer's disease (AD), it is shown that episodic memory is critical to the formation of cognition and other mental states implicated during AD, and thus the varying degrees of 'cognitive reserve' observed by these studies could be as a result of differences in levels of damages to the episodic memory.

Due to the similarities between the two forms of AD, some authors have postulated that, perhaps, early and late onset AD may be just one entity and should be unified as one trait. These proposals were conceived when atypical cases of early onset AD started to surface [39-41]. Early onset AD is characterized by autosomal dominant inheritance of APP, PSEN1, PSEN2 genes, but these atypical cases did not exhibit this feature. Conversely, and rather interestingly, they all involved episodic memory dysfunction except the cases presented by [41]. Some have proposed newer ways of classifying the forms of AD. For instance, [42] argued that, because of its effect on the age of onset of AD, the presence or absence of APOE$\varepsilon 4$ should be used to classify AD. Others have also argued for the age threshold for differentiating AD to be altered [43]. In summary, AD may be just a single trait that affects the episodic memory, at its core.

1.2.2. Features of Alzheimer's disease

1.2.2.1 Psychological features of Alzheimer's disease (AD)

AD is characterized by loss of both episodic and semantic memory. Episodic memory is a system of memory that enables an individual to remember past events they took part in. It is characterized by subjectivity of time, autonoetic awareness and conception of a 'self' independent of the physical body occupying space [44]. Time flows continually in one direction. The unidirectional flow of time is violated when an individual ponders over events that happened in their past. The individual creates a mental reality in which they can go back and travel into the past. At that point, time is subjective to the individual, and they can travel as far back as possible, bending the unidirectional flow of time. To perceive that they are remembering past events, the individual needs to be aware of this time travel. This awareness has been termed auto-noesis (autonoetic awareness). Time travel requires a traveler, and therefore the individual creates a 'self' which can journey through the mental reality. These characteristics of episodic memory were described by Endel Tulving when he first coined the term in 1972 [45] (and discussed in [44]) to distinguish it from semantic memory.

Semantic memory is a collection or a body of concepts, the meanings and attributes linked to these concepts, facts about the world around an individual and an individual's beliefs [46]. Suppose that an individual encounter a cat for the first time. The cat possesses four limbs, two eyes, a tail, and fur. If the individual runs into the cat again, but this time, the tail and two limbs are colored differently, fur shortened, and so on. Upon another encounter, there are still more differences in some of the features of the cat. Over time, the individual abstracts a concept *'cat'* in their mind, such that should they encounter an object

with some or most features of '*cat*' such as an object with four limbs, two eyes and/or fur, they may be tempted to label that object as a cat. The decision to label the object as a cat is dependent on the weight given to the features of the concept *'cat'* during its conceptualization process, with more weight being given to features that were more constant, and lesser weight to more variable features. In a word, generalizable knowledge about the world around us are stored in the semantic memory, and this memory system is usually deprived of temporal and spatial context, unlike episodic memory.

To be sure, aside episodic and semantic memory, there are other memory systems [47, 48], their details of which are beyond the scope of this work, due to the fact that episodic and semantic memories are the systems predominantly damaged during AD [49-52]. To note, AD affects episodic and semantic memories to different extends, both spatially and temporally. Episodic memory is affected more severely and also at relatively early stages of the disease [53, 54], possibly because the hippocampal and para-hippocampal regions of the brain are the first regions to be affected by AD [55, 56]. These regions play key roles in episodic memory formation [57]. As the disease spreads to other regions of the brain like the temporal neocortex, semantic memory becomes affected [49, 58]. Because of the critical effects of AD on episodic, many patients with AD have been reported to suffer loss of autobiographical memory (ABM) or the ability to recollect recent personal experiences [59, 60]. The loss of ABM has led some authors to postulate that AD leads to the loss of a patient's 'self' to the extent that some authors even describe AD patients as 'living dead' [61].

The 'self' may be disrupted by AD, however, as [62] and others have cautioned, the loss of memory does not necessarily imply the loss of the 'self'. Considering (as has already

been discussed above and detailed in [44]) that the 'self' comes into play later in the formation of episodic memory, our attention rather should be directed to the earliest events of episodic memory formation, viz. the subjectivity of time and autonoetic awareness in efforts to understand the psychopathology of AD. Episodic memory is linked to cognition [63], emotions [64] and many other mental states that are impaired during AD. The centrality of episodic memory to these mental states makes the study of its features such as the ability to loop back time to recall past events (subjectivity of time) very key to understanding AD. Such studies can aid in designing more effective psychological tests to assess not only memory, but also other cognitive impairments that are implicated during AD. Also, the focus on the loss of 'self' that strips AD patients of their personhood will be reduced. Further inquiries of this phenomenon should be undertaken by students of memory and psychology as this work focuses on the neuropathology of AD.

1.2.2.2 Neuropathological features of Alzheimer's disease

There are many neuropathological changes that occur during AD. These changes were first reported by Alois Alzheimer [65-67] when he described the 'peculiar' cases of Auguste D. and Johan F [67]. The prominent neuropathological changes include the extracellular accumulation of amyloid plaques [68, 69] and/or the buildup of neurofibrillary tangles inside neurons [70, 71]. These neuropathological changes are associated with the memory loss and deterioration of cognitive skills observed during AD. Although there are no well-defined mechanisms that lead up to these features yet, there are overwhelming evidence showing that some genetic and environmental factors increase the risk or cause these neuropathological features.

1.2.2.2.1 BRAAK Staging of Alzheimer's disease

BRAAK staging of AD was developed [54, 72] to reduce the occurrence of false positives and negatives in AD staging and diagnosis. AD is characterized neuropathologically by the presence of β-amyloids deposits and neurofibrillary changes in regions of the brain. The density and patterns of β-amyloids deposits do not differ widely in different brain regions affected by AD or even between different individuals in AD populations. Therefore, the density and pattern of deposition of β-amyloids do not offer an effective means to distinguish the stages of AD and also AD from other dementia. This has led to the reliance on neurofibrillary changes to stage AD and differentiate AD from other dementia. Neurofibrillary changes include neuritic plaques, neurofibrillary tangles and neuritic threads. Neurofibrillary tangles and neuritic threads have regular distribution across different brain regions during AD. This observation was utilized to group AD into six(6) stages referred as BRAAK staging of AD [54].

BRAAK stages I and II are referred to as transentorhinal stages. These two stages are marked by neurofibrillary changes in the pre-α layer of the transentorhinal zone, which is a transitory zone from the entorhinal to temporal cortex. The main feature of this transitory zone is the superficial layer of the multipolar neurons emanating from the entorhinal cortex [73]. Stages III and IV, also referred as the limbic stage, feature a more pronounced neurofibrillary changes in the transentorhinal region, with an extension to the entorhinal region itself. Stages V and VI are characterized by an expansive deposition of the amyloid proteins in the whole neocortex, and therefore these stages are called isocortical stages.

This approach of grouping AD cases is in many ways better than the sole dependence on the deposition of β-amyloids to classify AD. β-amyloids are present in other dementia and

during normal aging, and therefore cases of non-AD dementia, as well as normally aging brains could be falsely assigned as AD and vice versa. Moreover, conventional staining of β-amyloids is mostly only reliable when there is large deposition of β-amyloids, which is present only during advanced stages of AD. These limitations present additional challenges in distinguishing preclinical stages of AD from symptomatic stages. To identify the molecular processes that lead to the onset of AD, it's important to study the preclinical and clinical stages of the disorder, making BRAAK staging a preferred choice in staging.

1.2.3 Environmental Factors that increase risk of AD Neuropathology

1.2.3.1 Aging

Aging is a major risk factor for developing AD [74, 75]. As an individual ages, their chances of developing AD increases [76], with the risk increasing significantly after 65 years [77]. Normal (brain) aging is described as advancing in age without any neuropathology. Normal (brain) aging is characterized by disturbance in the formation of episodic memory [78], accumulation of tau proteins [79] (Noting that [79] attributed this accumulation of tau proteins during normal aging to the APOE gene, a genetic risk factor for developing AD), cognitive decline [80], and other metabolic processes in the brain [81]. These features are also present in people with AD. This leads to a difficulty in differentiating normal (brain) aging from cases of AD [82], particularly the so-called preclinical cases of AD [83].

Susanne Wegmann and colleagues have demonstrated that there is an increasing spread of tau proteins, hallmark proteins in AD neuropathology, in old brains than young brains [84]. They showed that aging brains provide an environment which enhances about two-folds

transference of tau proteins through neurons comparing to young brains, thereby increasing the spread or migration of these toxic substances. It's challenging to pinpoint the exact factors that cause such a milieu in old brains that facilitate the spread of tau proteins, and by extension AD.

In an attempt to understand what factors lead to such an environment, Karl Herrup had argued that the continual exposure of the brain to injuries initiate cascade of events in the aging brain [85]. According to Herrup, "[his] hypothesis stipulates that some event—a physical head trauma, a major illness or infection, a vascular event (possibly so small as to be clinically undetectable), the metabolic stress associated with adult-onset diabetes, or even the stress associated with a major life event such as a death in the family…" could create an environment that aids the spread of these toxic proteins that cause AD pathology. These injuries could also be genetic mutations, but Herrup reasoned that such mutations need to interact with other aging-related processes before they can be expressed and help in the creation of such a milieu conducive for AD progression. The injury (whether genetic or environmental) initiate a response by the cells which continue even when the initiating injury has been resolved, leading to a cascade of inflammatory and cellular processes that ultimately result in the creation of an environment for the onset and/or progression of AD.

Herrup's hypothesis focuses on the response to an injury, rather than the injury itself. His proposal was enlightening although it failed to make the framework for understanding AD pathogenesis any simpler. Indeed, Herrup concluded that his model is "a complex web of interactions". To understand such a complex web of interactions that lead up to AD, a simpler model is needed.

In this work, a simpler model for understanding AD is proposed. As Herrup made clear, focusing on the age-dependency root of AD pathogenesis provides a better way to understand its pathogenesis. However, instead of centering the analyses on the numerous processes that occur in an age-related post-initiation of AD, it will be beneficial to refocus the inquiry of AD pathology on the well-regulated expression of genes that interact with these environmental factors.

In the central dogma of molecular biology, proposed by Francis Crick [86, 87], information flow from DNA to RNA then to proteins. Since its proposition, the central dogma has undergone scrutiny and debates and is now widely accepted by many molecular biologists to be fundamental to understand life. Most crucially, the subjection of the central dogma to critical investigations has revealed that the flow of information from DNA to RNA to protein is largely influenced by environmental factors. Perhaps, the contemporary surge in epigenetic studies [88] typifies our knowledge of the role of environmental factors on this flow of information. In AD for instance, [89] found aging related H4K16ac epigenetic changes near SNPs and other variants associated with AD, underscoring how environmental factors interact with genetic factors to cause AD.

AD is caused by injuries that alter informational flow from DNA to RNA to proteins, leading to alterations in biological pathways and processes which, presumably, ultimately affect the episodic memory system. These injuries can directly affect the DNA, in which case can lead to genetic mutations, such as mutations reported in APP gene. The injuries can also cause damage to the RNA or affect the proteins, some of which may regulate gene activities. All these affect the informational flow in the central dogma.

This work emphasizes the molecular changes caused by injury rather than immunologic response to them. The immunologic response to injury can be diverse and complicated, and usually difficult to study in vitro. To be clear though, the influence of age on AD depends largely on the response to the injury not the injury itself, because, as one ages, they lose their capacity to resolve the same magnitude of injury. Molecular changes to injuries are largely well-regulated, which means they are easier to study and model, while the immunologic response are more fluid– changing in responsiveness and becoming less effective as one ages [90]. The decrease in effectiveness in response to injuries explains why aging significantly increase the risk of getting AD.

By focusing on the injuries and the molecular changes they cause, the two classes of AD can be united as a single entity. Given the centrality of episodic memory to the mental states (such as cognition, emotions) that are disturbed during AD, understanding the molecular changes caused by injuries associated with AD, particularly their relationship to the memory system will provide insights into how the disorder begins. Viewed in this scope, cases described as 'typical' early onset AD are as a result of the inheritance of mutations in certain genes that adversely affect the episodic memory (or the memory system), while 'atypical' early onset AD are as results of the effect of other forms of injuries, other than genetic, on the episodic memory. Such non-genetic injuries will occur in early stage of a person's life and are likely to be very intense, otherwise they will be resolved by the body's response. As one ages, their responses to injuries (genetic or environmental) that could potentially affect episodic memory become weaker, leading to higher chances of getting AD. Late onset AD results from the interaction of environmental and genetic factors leading to changes in molecular processes and pathways that disrupt

episodic memory. Early onset familial AD and late onset (sporadic) therefore become one entity, characterized by the presence of injuries that affect the episodic memory.

In this work, this disruption of information flow is investigated to examine the molecular pathways or processes that are affected by injuries to cause AD. Before delving into this, other environmental factors, aside aging, as well as genetic factors that increase the risk of developing AD neuropathology are discussed.

1.2.3.2 Other Environmental factors that Increase Risk of AD Neuropathology

Apart from aging, there are other factors that increase the risk of developing AD. Perhaps, the most significant of these factors are comorbidities, particularly vascular disorders and some heritable diseases like Down's Syndrome. Incidence of vascular disorders like chronic hypertension in midlife likely predispose individuals to developing AD later in life [91]. Also, factors that increase risk of developing such vascular disorders, like being overweight or hypercholesterolemia, have predominantly been found to increase the risk of developing AD too [92, 93]. It is not clear why these factors lead to AD or how they contribute to the neuropathology of the disorder. Possibly, vascular disorders result in decreased oxy-perfusion to neurons, that eventually lead to the damage of nerve cells [94], ultimately to loss of memory. It could also be as a result of the damaging effects of vascular disorders to the blood-brain barrier which lead to influx of substances that have the potential to harm neurons into the brain. Even that, it is difficult to figure which substances cause injury to neurons, and specific molecular processes that eventually lead to AD. It is possible that in the presence of these substances, there still will be no AD unless the episodic memory is affected.

Another risk factor that has been associated with vascular disorders as well as AD is diet. Dietary habits have been associated with many other disorders. In AD, [95] found that an increased intake of saturated fat will increase the risk of developing AD, while vitamins may protect individuals from developing AD [96]. Because people eat these dietary components together, [97] wanted to examine how consumption of a combination of some so-called 'protective' and harmful dietary components influence the development of AD. They found that the 'protective' dietary patterns decrease risk of AD development.

Heritable diseases such as Down Syndrome also predispose individuals to developing AD. Down syndrome is caused by the presence of an extra copy of chromosome 21 (trisomy 21). Chromosome 21 contains the amyloid precursor protein (APP), which is overexpressed during trisomy 21. This overexpression of APP gene leads to accumulation of beta amyloid as patients with Down Syndrome age [98]. Down syndrome patients face decline in intellect leading to difficulty in learning, which can also increase their risk of developing AD.

1.2.4 Genetic factors that influence the neuropathology of AD

Early onset AD is known to be predominantly caused by an autosomal dominant inheritance of mutations in APP, PSEN1 and PSEN2 genes. APP gene is located on the short arm of chromosome 21 and codes for Amyloid Precursor Protein. Amyloid Precursor Protein is an integral membrane protein which can undergo proteolytic cleavage to form β-amyloid peptides secreted into the extracellular space [99, 100]. Mutations in the APP gene can lead to overproduction of the protein itself or its constituent β-amyloid peptides. β-amyloid peptides, particularly those with 42 amino acids, aggregate into amyloid fibrils

that constitute major portion of the senile plaques observed during AD. Despite this role of amyloid precursor protein in the pathology of AD, mutations in the APP gene contributes to less than 5% of all early onset AD.

One gene that is mutated in more than 80% of all cases of early onset AD is PSEN1. PSEN1 is located on chromosome 14 and codes for presenilin. Presenilin contains the active site of γ-secretase enzyme [101] which cleaves amyloid precursor protein at the intramembrane domain to produce β-amyloid peptides [102]. Because majority of the mutations in the PSEN1 gene are missense, it has been thought that loss of normal function of the protein leads to the increased proteolytic activity that cause β-amyloidosis [103]. Some studies have pointed to mutations in the gene as that of gain of function [104], leading to confusion on developing models to explain the pathogenesis of AD based on this gene. More recent studies support either partial or complete loss of function mechanism [105]. PSEN1 has been shown to play key roles in memory formation [106], and thereby loss of this function has been proposed to cause neurodegeneration in AD. There are even some evidence showing that mutations in the presenilin gene could lead to neurodegeneration in the absence of amyloid plaques [107]. Also, mutations in PSEN1 has been shown to present earlier onset of AD than APP and PSEN2 genes [108].

Mutations in PSEN2, an isoform of PSEN1 have been found to contribute to less than 5% cases of early onset AD. This gene, together with PSEN1 and APP genes are mostly found to be inherited by patients with early onset AD from their affected parents.

Another gene that has been predominantly found to be associated with AD, particularly late onset cases is APOE. This gene has different variants. The APOE$\varepsilon 4$ variant is widely known to increase the risk of developing AD, while the $\varepsilon 2$ provides some protection against

the disorder. Additionally, the $\varepsilon 4$ has been found to substantially affect the age of onset of AD with patients with two alleles of this variant at an increased risk of developing the disorder earlier.

Aside these genes, GWAS and high throughput gene expression approaches have identified many candidate genes, some of which may be involved in metabolic processes or pathways that are involved or lead to the phenotypes we observe during Alzheimer's disease. GWAS approaches led to the discovery of a number of risk alleles for AD [109]. Since the advent of GWAS, several dozens of risk loci that may be associated with AD have been identified. [16] reported the first GWAS in AD. This study identified 16 candidate SNPs associated with AD, including three markers which were in linkage disequilibrium with APOEe4 variant. Some of the other candidate SNPs identified by this study included GALP, GWA_14q32.13, LOC651924, PGBD1, TNK1. Since then several other GWASs in AD have been done [17-22] and identified many risk loci. These risk alleles were either found in or in linkage disequilibrium with genes such as TRPC4AP, LRAT, ATXN1, TGEN, FAM113B, CLU, GOLM1, PCDH11X, GAB2 and others.

Because these approaches usually yield big data, and the identification of many candidate genes or loci, it is difficult to specify which genes may be truly involved in causing AD. Most of these studies are also difficult to reproduce [110]. In other words, findings from these approaches may be stochastic, with little or no biological significance. To increase the power of discoveries by these approaches, it is tempting to increase the sample size, as [111] did in their analysis of late-onset AD patients. An increase in the sample size of a GWAS study may come up with new findings, such as in the case of [111]. However, this

does not rule out the possibility of the findings being stochastic. [111] conducted series of tests to determine the biological importance of their findings.

1.3 The role of gene expression in the pathophysiology of AD

One major way to access the biological significance of genetic changes in the course of AD is through gene expression analysis. Interaction of environmental and genetic factors results in variation in gene expression during Alzheimer's disease. Genes that are crucial to the pathophysiology of AD are more likely to be differentially expressed. For instance, [112] observed an increase in amyloid precursor protein (APP) mRNA in AD patients. This increase could result in an increased production of β-amyloids leading to more senile plaque formation. Conversely, overexpression of other genes such as TREM2, have been found to reduce β-amyloids loads in AD brain [113] by acting downstream of CD33 [114], another gene associated with AD [22]. Several other genes found to be associated with AD by different approaches are known to be differentially expressed [115, 116], suggesting that exploring how genes are differentially expressed in AD brains will be effective in finding pathways that are key to AD pathogenesis.

Exploring gene expression in the normal aging brain and AD brain is even more important because normal aging also alters the dynamics of gene expression. As a person advances in age, [117] found that there were age-specific differential changes in the gene expression, with the most significant changes occurring during the sixth and seventh decades of life. Interestingly, [117] noted that most of the changes occur in the superior central gyrus (SFG) and postcentral gyrus, not the entorhinal cortex (EC) or hippocampus (HIP). During AD, the EC and HIP are brain regions mostly severely affected, which could lead to greater

changes in gene expression. This shows that, although both normal brain aging and AD brain affect differential gene expression, but there are region specific and other differences that could be examined to reveal pathways that are implicated in AD, but not normal aging.

1.4 Analysis of Gene Expression in Alzheimer's Disease

There have been several efforts to describe the nature of gene expression in normal brains and AD brains in severely affected brain regions. [26] provided one of the earliest genome wide gene expression profiles of the six most important brain regions that are affected during AD. These brain regions include Entorhinal Cortex (EC), hippocampus (HIP), middle temporal gyrus (MTG), posterior cingulate cortex (PC), superior frontal gyrus (SFG), and primary visual cortex (VC). Changes in gene expression in these regions may play roles in age-related cognitive impairments or AD. To understand what factors may play key roles in the pathology of AD, [118], the same group that performed the study in [26], accessed how genes involved in metabolism are expressed in regions of the brain during AD. AD is characterized by reduction in glucose metabolism rates in many regions of the brain [119], which could indicate a general malfunction in metabolism in these brain regions [120]. This makes the assessment of the expression of genes involved in metabolism an important approach to understand AD pathology.

There has been evidence that certain viruses may be involved in AD [121, 122]. To examine the mechanism through which viruses may cause AD, [123] conducted a multi-omic analyses including gene expression data from [26].

Recent efforts have geared towards examining genome-wide expression patterns to understand the pathways that come into play during both preclinical or asymptomatic

stages and clinical stages of AD. The focus on preclinical stages of AD reflect the evidence that there are substantial neuropathological changes many years before the onset of clinical features of the disorder [124]. Preclinical cases of AD closely resemble normal aging brain as there are amyloidosis in both instances [83], presenting strain on distinguishing preclinical AD from normal brain aging. Since the patterns of gene expression in normal aging brain and AD (both preclinical and clinical) brain differ substantially, it is essential to explore gene expression to understand the pathology of AD. To this end, [27] performed gene expression profiling of brain regions mostly affected by AD as the disease progresses. Similarly, [28] profiled the transcriptomic changes in healthy, asymptomatic (preclinical) and symptomatic AD.

The gene expression dataset from these studies have been deposited in NCBI's Gene Expression Omnibus with GEO accession numbers GSE5281, GSE118553 and. This study employs these datasets to provide examine the pathology of AD.

1.5 Functional annotation of differentially expressed genes (DEGs)

1.5.1 Brief history of functional annotation

It is challenging to find the biological significance of hundreds to thousands of genes that are differentially expressed in phenotypic states, for example during typical or asymptomatic AD. It's tempting to select DEGs with the highest fold changes or lowest p-values (or those ranked highest according to the metric used in identifying DEGs), study their molecular functions and deduce that such molecular functions or processes drive the phenotype under consideration. This approach can be very laborious. Additionally, it can be spurious because there are many genes that may play key roles in molecular functions

or processes in a phenotypic state but may not be highly differentially expressed according to the metric used in DEG identification. Suppose, for instance, that gene A plays an essential role in molecular pathway M, which is disrupted to cause phenotype P. Under normal physiological condition, the concentration of A required for normal functioning of M is 10.0 u/L. Assuming that a 1.0 u/L increase or decrease of gene A will cause a disruption in P, while other genes in the same molecular pathway M, with the same normal physiological concentration of 10.0 u/L will need to have an increase or decrease of 4.0 u/L to cause disruption in P, then selecting only the highest differentially expressed genes will omit gene A. A more favorable approach will be to study the full list of DEGs, which will not be an easy task manually.

A lot of efforts have led to the formulations of mathematical or computational models to study full list of DEGs from microarray experiments. One of the earliest efforts involved Boolean networks, in which a gene is either ON or OFF during a specific phenotype. [125] showed that properties of Boolean networks such as self-organization, stability and global complexity are very similar to features exhibited by biological systems. Despite that, Boolean networks did not adequately account for some dynamics in genomic regulation or expression. For instance, the frequency and intensity of binding of a transcription factor to a part of a genome. The frequency or intensity of binding alter the rate or amount of differential expression. Additionally, there are enough evidence showing that certain messenger RNAs direct the production of many protein isoforms (alternative splicing). To solve this problem, [126] proposed the use of differential equations (the Linear Transcription Model) that served to describe the temporal and quantitative dynamics during the biological processes.

On the other hand, drawing on findings from studies such as [127], which observed that extrapolating biological significance from gene expression data will require adequate prior information on genes such as, which transcription factors bind to which genes, [128] suggested the employment of Bayesian networks in studying DEGs. In a multiply interacting genes, Bayesian networks assign probabilistic relationship between the genes, such that, qualitatively, these relationships are represented by conditional independence relations, while conditional probability distribution describe their quantitative relationship. Bayesian networks presented a more robust and objective estimation of networks between gene expression data.

Yet, Bayesian networks approach proposed by [128] inferred biological networks from a limited set of genes, unlike other approaches that employed hierarchical clustering [129], which utilize all DEGs. Hierarchical clustering groups genes into clusters according to their expression pattern, with no a priori knowledge on gene function. This method helped to infer the function of genes whose functions were unknown before, through gene co-expression.

Through inference from co-expression, biochemical experiments and other approaches, there have been a continuous growth in information on the function of genes and their interaction with other genes and proteins. A key product of such growth in information on genes, coupled with or resulting from the completion of the sequencing of many eukaryotic genomes, is the construction of the Gene Ontology database [130] and other knowledge-based databases. The Gene Ontology project sought to create a body of knowledge on eukaryotic genes that will enhance discovery. Other similar databases include the Kyoto Encyclopedia of Genes and Genomes (KEGG), Online Mendelian Inheritance of Man

(OMIM). Indeed, through these efforts, it has become easier to examine the biological significance of DEGs through various enrichment methods that employ computational methods to interpret gene lists by using terms in knowledge bases such as Gene Ontology. There have been several gene enrichment methods, and they have broadly been grouped into three generations.

1.5.2 First generation functional annotation methods

First Generation enrichment approaches borrowed the idea of hierarchical clustering to group genes into clusters, and then map them to predefined functional categories [131]. First Generation approaches utilized hypergeometric tests to find the statistical significance of the enrichment of genes in the clusters in the functional categories. They are also known as Over-Representation Analysis (OTA) [132-134]. Because they employ hypergeometric test (Fisher's Exact Test), their execution is simple and easy. However, first generation methods had some challenges. For instance, the independence of the gene expression assumed in the hypergeometric test is not easily realizable. Also, features in the sets of genes are grouped as either differentially expressed or not differentially expressed; a binarization that leads to loss of information.

1.5.3 Second generation functional annotation methods

Second Generation enrichment methods are typified by the well-known Gene Set Enrichment Analysis (GSEA) technique [135, 136]. GSEA ranks genes according to expression differences and hypothesize that if genes in a set happen to be at the top or bottom of the ranked list, they are associated with the phenotypes that characterize the

affected individuals. The association is measured by a non-parametric statistic, the Enrichment Score (ES), which mirror the extend of overrepresentation at the extremes of the ranked list. The significance of this association is measured by an empirical permutation test and further adjusted for false discovery rate.

1.5.4 Third generation functional annotation methods

A third method, which has been classed third generation enrichment analysis, was introduced by [137]. This approach is similar to GSEA, but it is more hypothesis driven compared to the more data driven GSEA. This algorithm integrates the topology of the network in calculating the enrichment score. This presents a more advanced measure of the complexity that characterize biological systems.

1.6 Hypothesis

Alzheimer's disease (AD) is a complex disorder caused by genetic and environmental factors. These genetic and environmental factors interact to cause changes in how genes are expressed, leading to alteration of metabolic processes and pathways that eventually cause AD. It is hypothesized that the alterations in the metabolic processes and pathways will disrupt episodic memory leading to the onset and progression of AD.

1.7 Specific objectives.

1. To characterize genes that are differentially expressed in AD affected brain regions. AD affects different brain regions differently, with limbic regions of the brain such

as entorhinal cortex (EC) and hippocampus (HIP) being severely affected, while regions such as cerebellum are relatively spared. This study identifies the genes that are commonly expressed differentially in both severely affected and less affected regions.

2. To identify metabolic processes and pathways that are enriched in these differentially expressed genes. AD is a complex disorder involving different metabolic processes. It is essential to identify the molecular processes and metabolic pathways that are enriched in the differentially expressed genes. This will provide better understanding of the biological significance of differentially expressed genes during AD pathogenesis.

3. To prioritize molecular processes and metabolic pathways significant to AD pathology using a novel ranking score. There are many genes that are differentially expressed, with many molecular pathways enriched in them. Prioritization with our novel ranking score helps to differentiate true and false positives biological processes that are important to AD.

CHAPTER 2: METHODS

2.1 Microarray profiling of differentially expressed genes

Microarrays are employed for the simultaneous profiling of thousands of genes in different phenotypes. There are two major platforms for microarray experiments; Affymetrix and Illumina. These platforms have driven discoveries in various diseases including AD. In AD, there have been many efforts from microarray experiments, that have enhanced the discovery of different pathways that may be important to its pathogenesis. Using statistical and computational approaches, this study seeks to deepen our current understanding of AD using data from such microarray experiments.

2.2 Statistical tests for identifying DEGs

Since the advent of microarray approaches, there have been debates on which statistical tests to be resorted to in identifying differentially expressed genes. Several methods ranging from simple approaches like fold change, t-test, Anova to more complex ones such as Bayesian models and non-parametric models have been proposed. Among these methods, t-test remains the most widely used due to its simple applicability and interpretability.

Critics of the application of t-test in identifying differentially expressed genes argue that there could be an asymmetry in the estimation of variances due to the presence of genes

with disproportionally low variances. Such genes will, in practice, have relatively large t-statistic and thereby falsely selected as differentially expressed. Also, they contend that, t-test yields low statistical power when applied to microarray data because microarray studies usually have few variables (low sample sizes) compared to observations (genes) [138]. Such criticisms particularly focus on the error variance incorporated in the computation of the t-statistic to assess its accuracy. The error variance is the residual variance that cannot be attributed to the phenotypic variables of interest. In this study, the *Normalize.Quantile* function from the LIMMA package was used to correct for such residual variance that may exist in the datasets.

Therefore, for its simplicity and interpretability, the t-test was used to identify genes that are differentially expressed between the phenotype groups in the datasets. That is, between controls and affected in GSE5281; typical Alzheimer's Disease (AD), asymptomatic AD (ASymAD) and Controls (C) for GSE118553; and BRAAK levels I-VI and controls in GSE131617. The Benjamini-Hochberg method (BH) for controlling false discovery in multiple testing [139] was employed to correct for multiple testing.

2.3 Software/ web platforms for performing functional annotation

There are many platforms that implement one or more of these enrichment methods to discover biological insights from DEGs generated from microarray platforms. One such platform is WebGestalt [140]. WebGestalt is a web-based application that integrates all three generations of gene set enrichment methods to retrieve important patterns in gene lists which may provide biological insights to reveal the pathogenesis of diseases.

Genes from various regions and levels of AD affection, in the different datasets (GSE5281, GSE118553 and GSE131617) were uploaded to the latest version of WebGestalt (2019) [141] and the pathways and biological functions that are enriched, as well as networks in the gene lists computed. For each study, the gene list in each region were analyzed with over-representation and network topology methods. GSEA was not included because the algorithm is very similar to network topology method, with the latter providing a more advanced measure of the complexity in the biological systems. For the over-representation analysis, the Kyoto Encyclopedia of Genes and Genomes (KEGG) was selected as the functional database to find the pathways. The Online Mendelian Inheritance of Man (OMIM) database was used to find the disorders that are associated with each gene list. The reference set for over-presentation analyses of GSE5281 was set to affy hg u133 plus 2, illumina humanht 12 v4 for GSE118553 and affy huex 1 0 st v2 for GSE131617. After the enrichment, the affinity propagation and weight cover options were both selected to reduce redundancy.

For the network topology analyses, the functional database was set to PPI-BIOGRID, and both network expansion and network retrieval and prioritization were used to find the top seeds from the gene lists with the highest random walk probability and top-ranking neighbors respectively. The highest ranked GO categories from each gene list were also recorded.

2.4 A novel ranking score for prioritizing gene ontology categories

A major aspect of enrichment analyses is the ranking of genes or attributes in a database based on their constituent genes. For instance, GSEA order DEGs into a list and determine if the genes in an attribute in an a priori database tend to have genes towards the bottom or

top of this list, in which case, they are considered enriched. The magnitude of enrichment is measured with the enrichment score (ES). Similarly, in the network topology analyses employed in this work, after using random walk method to find the paths of information flow between the genes, an enrichment score and a hypergeometric test statistic, to find the significance of this enrichment, were calculated for the set of genes. The p-value from the hypergeometric test was used to rank the attributes that were enriched in the DEGs. Here, a new method for ranking these attributes is devised. This method combines the enrichment score and the hypergeometric test statistic.

Assuming there are **n** genes in a set of interesting genes **A** and **m** genes in a set of reference genes **B** in an a priori database;

Consider an attribute, **G**, such as Gene Ontology term, KEGG pathway or any biological process or function in this database;

If there are **k** genes from **A**, and **j** genes from **B** present in **G**;

The expected value of **k**, $\mathbf{k_e}$ will be;

$$k_e = \frac{n}{m} \times j$$

If $\boldsymbol{k}$ *is greater than* $\boldsymbol{k_e}$*, then G is said to be enriched in* ***A***. $\boldsymbol{k_e}$ *is the enrichment score* ***(ES)***

The magnitude of the enrichment is reflected in the enrichment score, and the statistical significance of this enrichment is computed using Fisher's Exact Test, given by:

$$P = \sum_{i=k}^{n} \frac{\binom{n}{i}\binom{m}{j+k \quad i}}{\binom{m+n}{j+k}}.$$

Using **p** and $\mathbf{k_e}$, the new ranking score, S is given by:

$$S = -\log(p) \times k_e$$

CHAPTER 3: RESULTS

3.1 Data acquisition and processing

The gene expression data sets from GSE5281, GSE118553 and GSE131617 were downloaded and parsed from NCBI's GEO using the Bioconductor package GEOquery [142]. These datasets contain DNA microarray data generated with Affymetrix and Illumina.

3.1.1 General description of datasets

The data in GSE5281 were generated from six brain regions: entorhinal cortex (EC), hippocampus (HIP), medial temporal gyrus (MTG), posterior cingulate (PC), superior frontal gyrus (SFG) and primary visual cortex (VCX), which are functionally relevant to Alzheimer's disease [26]. Expression profiling of affected and control samples were performed on Affymetrix U133 Plus 2.0 platform using cells obtained by Laser Capture Microscopy, as fully described by the authors [26].

As outlined in [28], GSE118553 comprises gene expression profiles of frontal cortex (FC), temporal cortex (TC), entorhinal cortex (EC) and cerebellum (CB) regions, generated using Illumina HT-12_V4 beadchips. The dataset measures the transcriptomic activity of

symptomatic (typical) and asymptomatic Alzheimer's patients in these four brain regions. Asymptomatic AD comprises a highly heterogenous group of people with increasing deposition of amyloids [143]. Some develop AD while others remain cognitively normal, making this a viable group to locate critical genes, metabolic processes or pathways that mediate the transition from an aged, cognitively normal brain to an AD brain. The samples in GSE118553 were selected according to guidelines provided by [144], which recommends the use of topographic methods such as BRAAK in staging Alzheimer's, particularly in research settings.

GSE131617 also measured gene expression data of patients which have been categorized into the BRAAK stages (I/II, III/IV, V/VI) of AD [27]. Patients in the BRAAK stages I/II, III/IV and V/VI respectively have low, intermediate, high probabilities that their dementia was due to Alzheimer's.

3.1.2 Reasons for selecting datasets GSE5281, GSE118553, GSE131617

These datasets were selected for several reasons. GSE5281 contains data from one of the first attempts to fully characterize the differentially expressed genes in AD relevant brain regions. This dataset has been used for many discoveries including a recent study by [145] to identify potential biomarkers for predicting AD incidence. This dataset did not specify the stage of AD of subjects, making it difficult to examine various processes that come into play as the disease progresses. As the disease is progressive in nature, such information will be vital in understanding the pathogenesis. GSE118553 and GSE131617 provided information on the stages of the subjects. Although GSE118553 and GSE131617 employed the BRAAK staging to group their study participants, GSE118553 had classified their

subjects as either belonging to control, asymptomatic or AD groups, while GSE131617 grouped their subjects into groups 0 (control), I-II, III-IV, V-VI. BRAAK stages I-IV are usually clinically asymptomatic cases of AD, while stages V-VI are symptomatic cases with clear neuropathological changes in affected brain regions. The results from these two later studies, which have a combined sample size of 614, are therefore very comparable. The large sample size, together with the close approximation of the findings from these datasets add power to the results of this current study.

Additionally, the datasets were generated with different microarray platforms (Affymetrix and Illumina arrays), and in the course of the past two decades, these platforms have witnessed tremendous improvement in their performances. Genes that are differentially expressed in microarray experiments designed with these different platforms are likely to be truly associated with AD.

Finally, these datasets were chosen because of the wide geographical span of the samples included in them. The samples in GSE5281 were generated from subjects predominantly from North America; subjects in GSE118553 were derived from the London Neurodegenerative Diseases Brain Bank; and that of GSE131617 were obtained from Japanese brain-donor bank, making our study inclusive of all global populations at high risk of developing the disease.

3.1.3 Challenges in analyzing data from different subject populations and array platforms

Effective cross platform and population analyses of gene expression dataset will be a considerable step in the study of complex diseases like AD. There have been some attempts

at cross-platform analyses [146, 147]. Despite these attempts with some fruitful results, care must be taken in designing such studies, especially genetic studies across different populations. There are inherent differences in alleles between individuals in the same population which result in stratification of the population. Population stratifications could lead to the observed inter-individual phenotypic differences, leading to false positives. Correcting such population stratifications convincingly reduced the false positive rates in some cases [148]. Population stratification can be due to population admixtures that happened in a very remote past or could be in a more recent cases of interbreeding. With the rise of cosmopolitanism, globalization and internationalism, there are likely to be more stratifications in the populations than ever, and therefore, it is imperative to consider this factor in designing cross-population studies.

Despite the success of RNA microarray approaches such as Affymetrix's spotted cDNA arrays or Illumina's bead-based arrays in studying the differential expression of genes under different conditions, there have long been well-documented challenges. Most commonly were the issues of technical errors with sample handling and data processing, which could lead to false variations in observations (technical variation). Additionally, there's extreme difficulty in collecting samples for microarray analyses which has resulted in wanting sample sizes in many gene expression studies. As [149] noted, the low sample sizes of microarray experiments, together with the large numbers of genes (usually few hundreds of samples and several thousands of genes) increase the odds of false discoveries, particularly when statistical or machine learning methods such as global feature selections are employed. Considerable caution must be taken when designing such cross-platform and population studies.

A very productive approach in cross-population and platform analyses, as this study will show, will be to design a robust pipeline and subject the data generated across platforms and populations to this same pipeline. In view of that, the downstream analyses in this study, including the processing techniques adopted as well as the prioritization technique developed, were performed on each dataset separately.

3.1.4 Distribution of samples in different phenotypes across different brain regions

In table 1, the samples in different phenotypic categories in all the datasets have been shown. GSE5281 had a total of 161 samples, while there were 401 and 213 samples in GSE118553 and GSE131617 respectively. In GSE131617, the samples in the phenotypes are the same across different brain regions because three brain regions were derived from each individual brain. Thus, the 213 total samples were extracted from 71 individual brains (persons). This isn't the case for GSE5281 and GSE118553. In GSE118553, some of the samples were duplicated. There were in fact about 384 unique samples in this dataset.

Table 1: The number of samples in different brain regions

	Samples	EC	TC*	FC*	CB	HIP	PC	VCX
GSE5281 (Total=161)								
	Control	13	12	11		13	13	12
	AD	10	18	24		10	9	17
GSE118553 (Total=401)								
	Control	24	31	23	22			
	ASymAD	37	52	40	38			
	Typical AD	37	32	33	32			
GSE131617 (Total=213)								
	Control	13	13	13				
	BRAAK I&II	20	20	20				
	BRAAK III&IV	19	19	19				
	BRAAK V&VI	19	19	19				

EC-Entorhinal Cortex
TC-Temporal Cortex
CB-Cerebellum

HIP-Hippocampus
PC-Posterior Cingulate Cortex
VCX-Visual Cortex
AD-Alzheimer's disease
AsymAD-Asymptomatic Alzheimer's disease
*For GSE5281 in the table, TC corresponds to MTG (Medial Temporal Gyrus) in the original dataset
Frontal Cortex
*For GSE5281 in the table, FC corresponds to SFG (superior frontal gyrus)

3.1.5 Pre-normalization processes

The probe sets were mapped to their corresponding gene names using the hgu95av2.db and illuminaHumanv1.db databases for Affymetrix (GSE5281) and Illumina (GSE118553) respectively. GSE131617 contained transcripts of 22,000, with each mapping to one gene, therefore, the gene conversion tool in the DAVID (v.6.8) software [150, 151] was employed in mapping the transcript IDs to their respective gene names. The similarity score (kappa) was set to greater than or equal to 3.5 and overlap to greater than or equal to 2. Only one transcript, with the highest similarity score (usually 1), per gene was reported. The raw signals from GSE5281 were transformed to log2. The *IQR.f* function (found in the scripts in Appendix E) was used to map multiple probe sets to their corresponding gene names. With the *IQR.f* function, when many probes map to one gene, the maximum inter-quartile range (75^{th} percentile) was reported. This resulted in total of approximately 22,000 mapped genes in the GSE5281 and GSE118553 datasets.

3.1.6 Normalization

An important step in gene expression analyses is normalization of the data to get rid of systematic variations, while preserving as much biological differences between the phenotypes of interest as possible. There are many proposed methods for normalization, such as global methods like quantile normalization (QN) [152]. QN removes variations in

statistical distributions, based on the assumption that there are equal variations in the phenotypes. In combined datasets, particularly in datasets generated by oligonucleotide arrays [153] , QN has been shown to correct undesirable variations [154]. Although QN is effective in removing such variations (mostly batch effects and other technical variations) [155], it could lead to loss of important biological variation of interest when the assumptions are not met [156]. Therefore, it is necessary, or even imperative to test the fundamental assumptions before applying QN.

Usually, graphical methods such as boxplots and density plots are utilized to display the statistical distributions of the datasets. But such visual representations can be misleading and subject to individual biases. Therefore, [157] proposed a method, *Quantro*, that provides a quantified estimate of the variations in the dataset, to aid in measuring the suitability of QN in transforming datasets. The *Quantro* package, implemented in R was used to find the variations within and between the phenotypes in the datasets. The summary is shown in table 2 below.

Table 2: Tests for global differences in the distributions of datasets

	GSE5281	**GSE118553**	**GSE131617**
nGroups	2	3	4
nTotSamples	161	401	213
nSamplesinGroups	87 74	167 134 100	39 60 57 57
anovaPval	0.42859	0.72886	0.66842
Variance within groups	0.05883946	0.003196329	0.005670415
Variance between groups	1.096589	0.02846446	0.002911345
quantroStat	18.63696	8.90536	0.51343

quantroPvalPerm	< 0.01	<0.01	0.77

nGroups-Number of groups
nTotSamples -Total number of samples in dataset
nSamplesinGroups -Total number of samples within each group compared in the datasets
quantroStat -Statistical test showing the variance between the different groups in a dataset

As shown in Table 2, ANOVA test showed there was no significant variations between the controls and affected samples in GSE5281. The variations within the individual phenotypes i.e. controls and affected is far lesser than between the two phenotypes. The *quantrostat* (18.63696), which was derived from both the variance between and within groups was significantly different between the phenotypes. This variation is further depicted in the boxplot of the distribution in Panel 1a of Figure 1. These statistics and plots indicate that QN can be used to transform the dataset in GSE5281 without loss of biological variations of interest. That is to say, the observed variations in the datasets are likely due to technical variations and should be corrected (normalized). As shown in the boxplot in Panel 1b, the data shows equal distribution post-normalization with *Normalize.Quantile* function from the Linear Models for Microarray Data (LIMMA) package [157] in R.

The quantrostat for GSE118553 suggests that there is significant variation between the phenotypes and therefore quantile normalization will be suitable for transforming this dataset too. This is supported by the boxplot in Panel 2a. After the *Normalize.Quantile* function from the LIMMA package was used to transform gene expression dataset in GSE118553, as shown in panel 2b of figure 1, the variations between the statistical distributions were corrected.

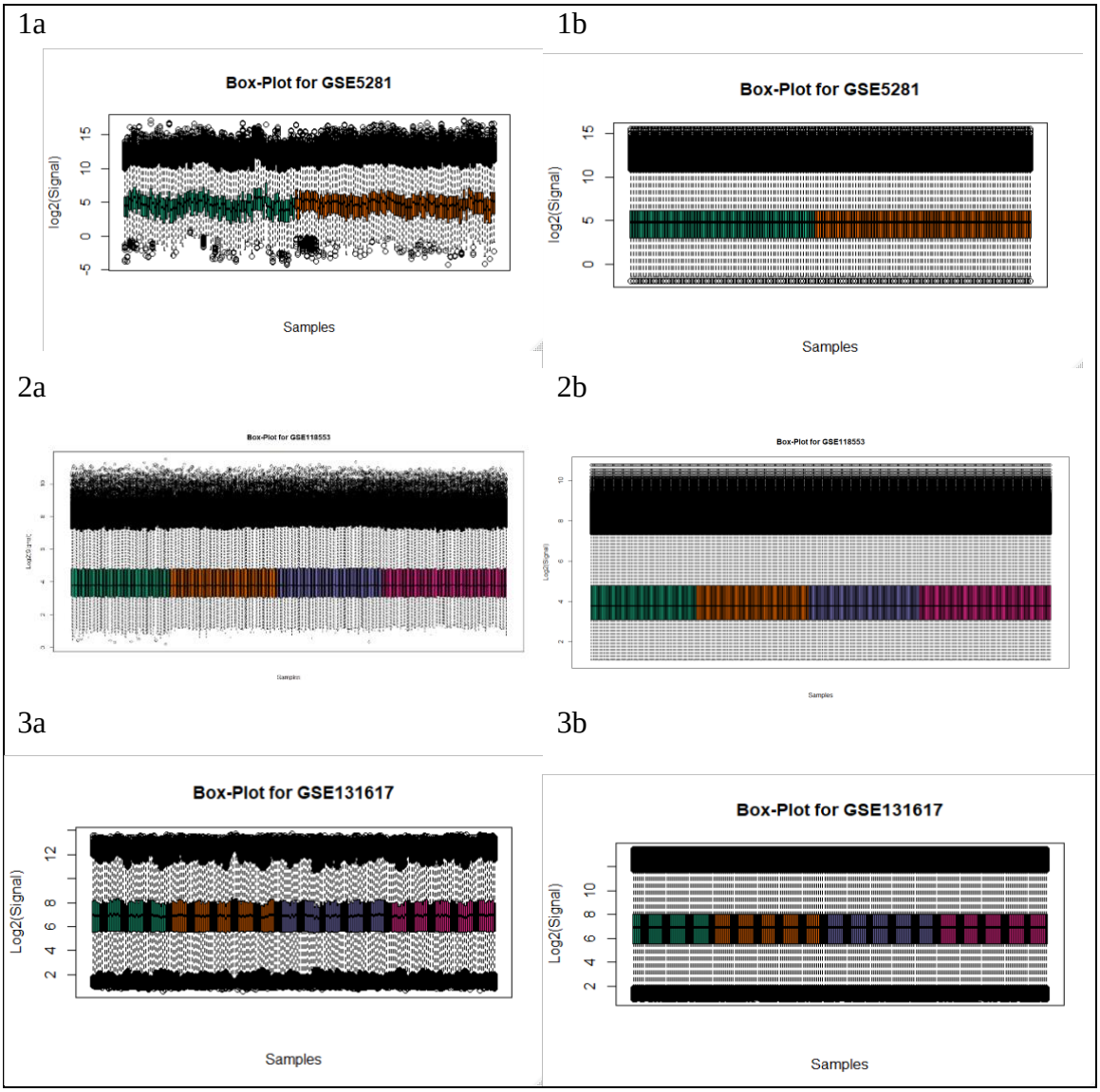

Figure 1: Statistical distribution of GSE5281, GSE118553 and GSE131617 before and after quantile normalization. Normalizaing with *Normalize.Quantile* function in the Limma Package, smoothens the probe levels in the datasets.

3.2 Identification of Differentially Expressed Genes (DEGs)

3.2.1 Differentially expressed genes (DEGs) from brain regions in datasets

The differentially expressed genes (DEGs) that were identified from the datasets using t-test and Benjamini-Hochberg (BH) correction have been shown in tables 3, 4 and 5. Table

3 shows the DEGs from GSE118553. Due to the reported high incidence of false positives in DNA microarray analyses, genes were identified at BH adjusted p-values of <0.01. For GSE131617 the BH adjusted p-values cut off showed zero (0) DEGs in all comparisons. This highly conservative cut-off ensures that there are minimal false positives in the DEGs.

In dataset GSE118553, when typical AD samples were compared with controls, there were 4065 DEGs in the EC region, 5190 in the TC region, 1534 in the FC region and 1029 in the CB region. After adjusting the p-values with BH, there were 2673, 2817, 64 and 25 DEGs in the EC, TC, FC and CB regions respectively. For asymptomatic AD (that is AsymAD vs. C), there were 715, 2083, 1028, 459 DEGs, with adjustment with BH resulting in 11, 197, 34 and 3 DEGs in the EC, TC, FC and CB regions respectively. These trends are in line with the results from [28]. As observed in this study and [28], overall there are more DEGs in the entorhinal cortex (EC) and temporal cortex (TC) regions than the other brain regions in typical (AD vs. C) and AD vs. AsympAD cases. However, there are more DEGs in the frontal cortex (FC) region than the EC in asymptomatic (AsymAD vs. C) AD cases. The DEGs in the cerebellum were comparatively low in the typical, asymptomatic and late cases of AD. For convenience, AD vs. AsympAD shall be referred as ‘late’ AD hereafter.

The DEGs derived from GSE5281 have been shown in table 4. Similar to what was observed in GSE118553, there are more DEGs in the EC region and TC (represented by the medial temporal gyrus, MTG in dataset GSE5281) regions than other regions. After adjusting the p-values with BH, there were 1095, 1850, 565 in the EC, TC and hippocampus (HIP) regions respectively. No DEG was identified in the VCX region after the adjustment. The number of DEGs identified in this study are relatively lower than what

[158] reported, most likely due to the more conservative cut-offs employed in identifying genes, as well as the different preprocessing methods used in this study.

In table 5, the DEGs that were identified at $p<0.01$ from the dataset GSE131617 align with the observations above that more DEGs are present in the EC and TC regions than other regions. When BH was applied, many of the DEGs did not meet the $p<0.01$ threshold. Importantly, [27] reported, after performing alt-splicing ANOVA, that there is alteration in the expression of the genes RELN, PTGS2, MYO5C, TRIL, DCHS2, GRB14, NPAS4 and PHYHD1 at almost all BRAAK stages in the three brain regions. Interestingly, the current study revealed similar results, although a different approaches of gene discovery have been used. For example, PHYHD1 was differentially expressed in all BRAAK stages in FC and TC regions, as well as in the III-IV and V-VI in EC. Also, DCHS2 was differentially expressed in BRAAK I-II, III-IV and V-VI levels in TC region, I-II, III-IV in EC region and II in FC region. The total number of genes in the EC in BRAAK level I-II is 620, higher than any other brain region. As described before (section 1.2.2.2.1 BRAAK Staging of Alzheimer's disease), BRAAK level I-II are characterized by changes in the entorhinal cortex.

The DEGs in GSE118553, GSE5281 and GSE131617 have been summarized in shown in Figure 2-Figure 7. For a full list of DEGs refer to appendix A.

Table 3: Comparison of number of differentially expressed genes

	EC	**TC**	**FC**	**CB**	**HIP**	**PC**	**VCX**
GSE5281							
	*Raw (Adj)	Raw (Adj)	Raw (Adj)	Raw (Adj)	Raw (Adj)	Raw (Adj)	Raw (Adj)

AD vs. C	3205(1095)	3641(1850)	1639(566)		2213(243)	2321(494)	788(0)
GSE118553							
AD vs. C	4065(2673)	5190(2817)	1534(64)	1029(25)			
AD vs. AsymAD	4696(2346)	7537(1731)	1093(2)	683(3)			
AsymAD vs. C	715(11)	2083(197)	1028 (34)	459(3)			
GSE131617							
BRAAK_I-II vs. 0	620(0)	418(0)	447(0)				
BRAAK_III-IV vs. 0	429(0)	402(0)	374(0)				
BRAAK_V-VI vs. 0	1103(0)	559(0)	438(0)				

⋆Raw p-value (p<0.01)
⋆ Benjamini-Hochberg (BH) adjusted p-values (p<0.01)

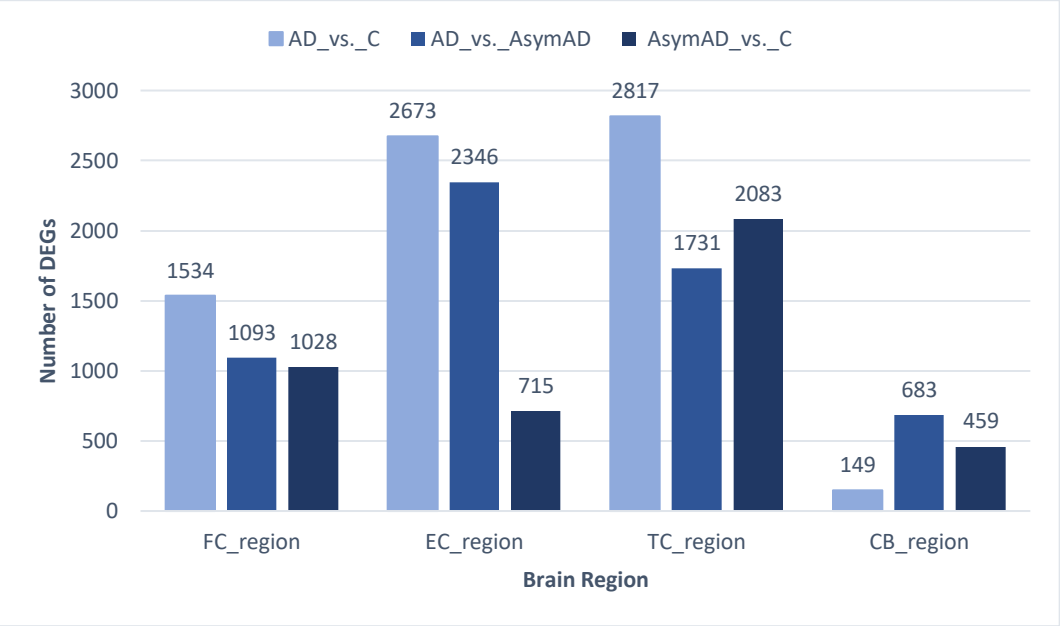

Figure 2:DEGs in different brain regions in dataset GSE118553. DEGs in GSE118553 at threshold of p< 0.01, after correcting for false discovery with Benjamini-Hochberg (BH) test.

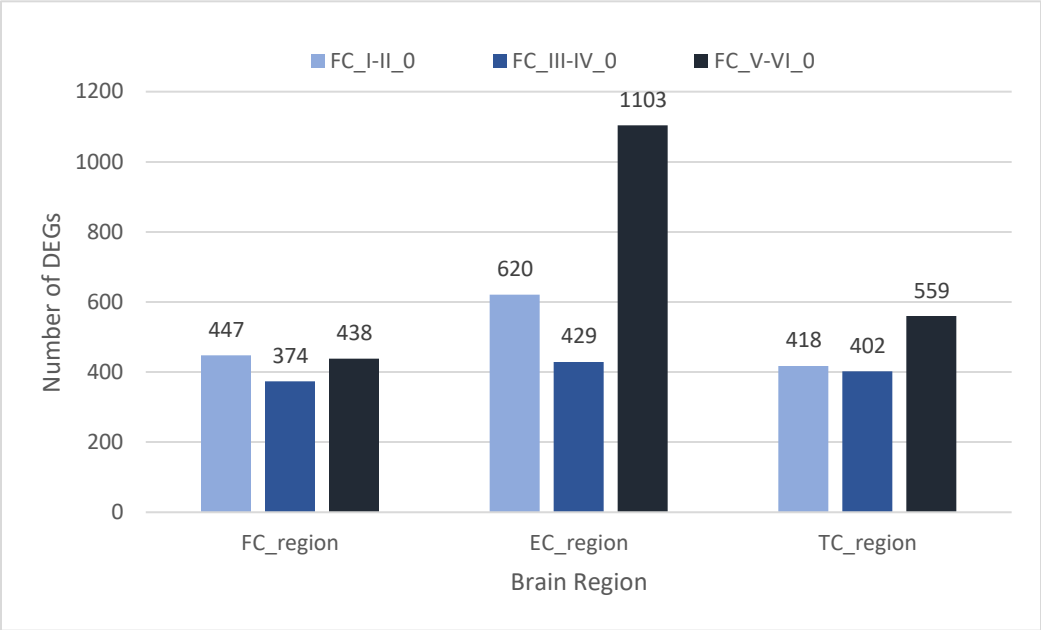

Figure 3: DEGs in different brain regions in dataset GSE131617.

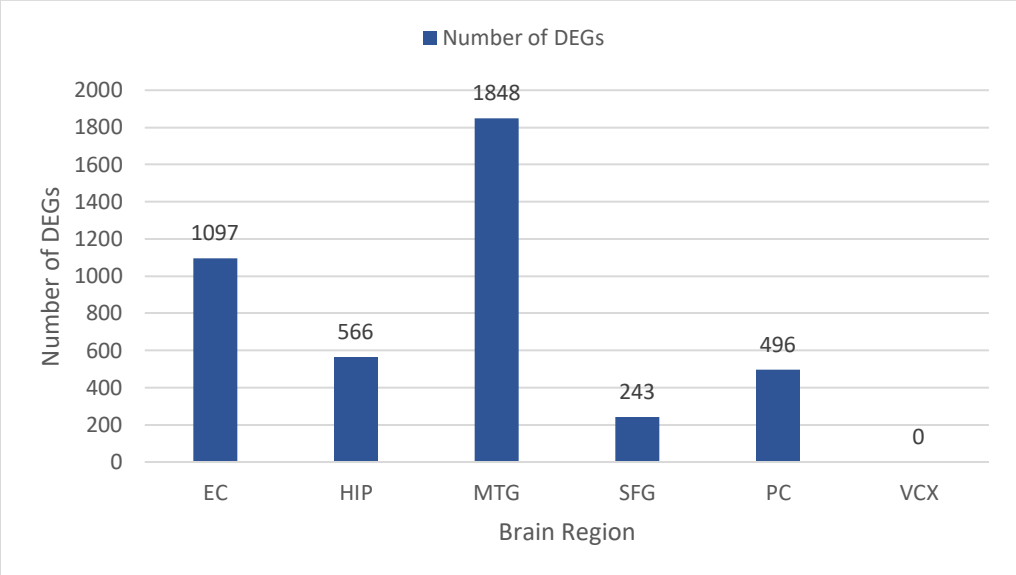

Figure 4: DEGs in different brain regions in dataset GSE5281. EC- entorhinal cortex, HIP-hippocampus, MTG-medial temporal gyrus, SFG-Superior frontal gyrus, PC-Posterior cingulate cortex VCX-visual cortex.

3.2.2 Intersecting differentially expressed genes at various levels of AD in different brain regions

To better understand which genes are differentially expressed in the three different studies, the intersection of DEGs in the different datasets, at different levels of AD and different brain regions were computed. An intersecting list could be made up of genes that are involved in biological processes implicated by AD, regardless of the region affected or stage of the disorder.

In Figure 5: : Differentially expressed genes in the entorhinal, frontal and temporal cortices of all three studies. 1. DEGs in Entorhinal Cortex (EC). SFG(superior frontal gyrus) and MTG(medial temporal gyrus) in the GSE5281 study correspond respectively to the frontal and temporal cortex in the other two studies

, the DEGs in the entorhinal, frontal and temporal cortices have been shown. These three brain regions are present in all three studies and are the most functionally affected by AD, together with hippocampus. The hippocampus however was only present in GSE5281 and therefore was not included in this comparison. In the entorhinal region, there are 429 genes that are differentially expressed in at least two of the three studies. There are eight (8) genes are present in all three (3) studies. These genes include SLC46A3, ADAMTSL1, PPP4C, ZC3H6, SEC62, TSSK2, KLHDC10, NDUFA10. There are 221 genes in the EC region in dataset GSE118553 and GSE131617, which is the highest in any two-study comparison. This observation corresponds to what was found in the FC region where there are 103 genes in the intersection between GSE131617 and GSE118553 compared to 23 and 7 genes in the other two-study comparisons. Out of the 626 genes found in at least two studies, 361 were found between GSE5281 and GSE118553 only. Fifteen (15) genes: GRPEL1,

HIST1H2AC, HIST1H2AB, PLP1, GALNT12, ORMDL2, SFT2D1, CASQ2, TRIL, SLC27A1, SLCO1A2, NFASC, OPTN, PCDH10, GABARAPL2 were found to be differentially expressed in all three (3) studies. Only one (1) gene, HIST1H4A was found in all three (3) studies in the FC region. The overlapping DEGs in different brain regions suggest that there are some genes that are fundamental to AD in affected brain regions.

Genes that are differentially expressed in typical, asymptomatic and late AD cases in the entorhinal (EC), frontal (FC), temporal (TC) and cerebellum (CB) were determined. As shown in Figure 6, in typical AD, there are 823 genes that are differentially expressed in both the EC and TC cortices, about 320 genes between the EC and FC cortices and only 29 between the EC and CB. There are disproportionately lower number of genes shared between the CB region and the other brain regions. For instance, there is only 15 genes between the CB and FC regions. This cuts across the other AD types. In asymptomatic AD, there are 61 genes common to CB and TC regions, compared to the 288 between the EC and TC or 197 between the EC and FC. In general, there are more intersecting genes when DEGs are compared among EC, TC and FC regions than comparing any of these regions with CB, reflecting the lower number of DEGs in the CB region or the disproportionately higher affection of the three brain regions compared to CB.

DEGs specific to BRAAK levels were compared in three brain regions: Entorhinal (EC), Frontal (FC) and Temporal Cortex (TC) and the results detailed in Figure 7. There were 57, 36 and 76 genes that were differentially expressed in all three (3) brain regions being compared in BRAAK levels I-II, III-IV and V-VI respectively. At BRAAK level I-II, there were only 16 genes shared between EC and FC, while there 104 genes in the BRAAK V-VI. Also, at BRAAK V-VI levels, there were 122 genes between EC and TC, while 39

were expressed in both FC and TC. This could indicate that as AD progresses, there is an increase in the number of genes that are differentially expressed.

The intersection of DEGs in various regions and AD levels show the high reproducibility of these genes generated from different studies, underscoring the significance of these genes to AD. Also, this intersectionality could indicate that some genes may be involved in some biological processes that are common in different brain regions and at different stages of the disorder.

For a full list of common DEGs in all studies, refer to Appendix B.

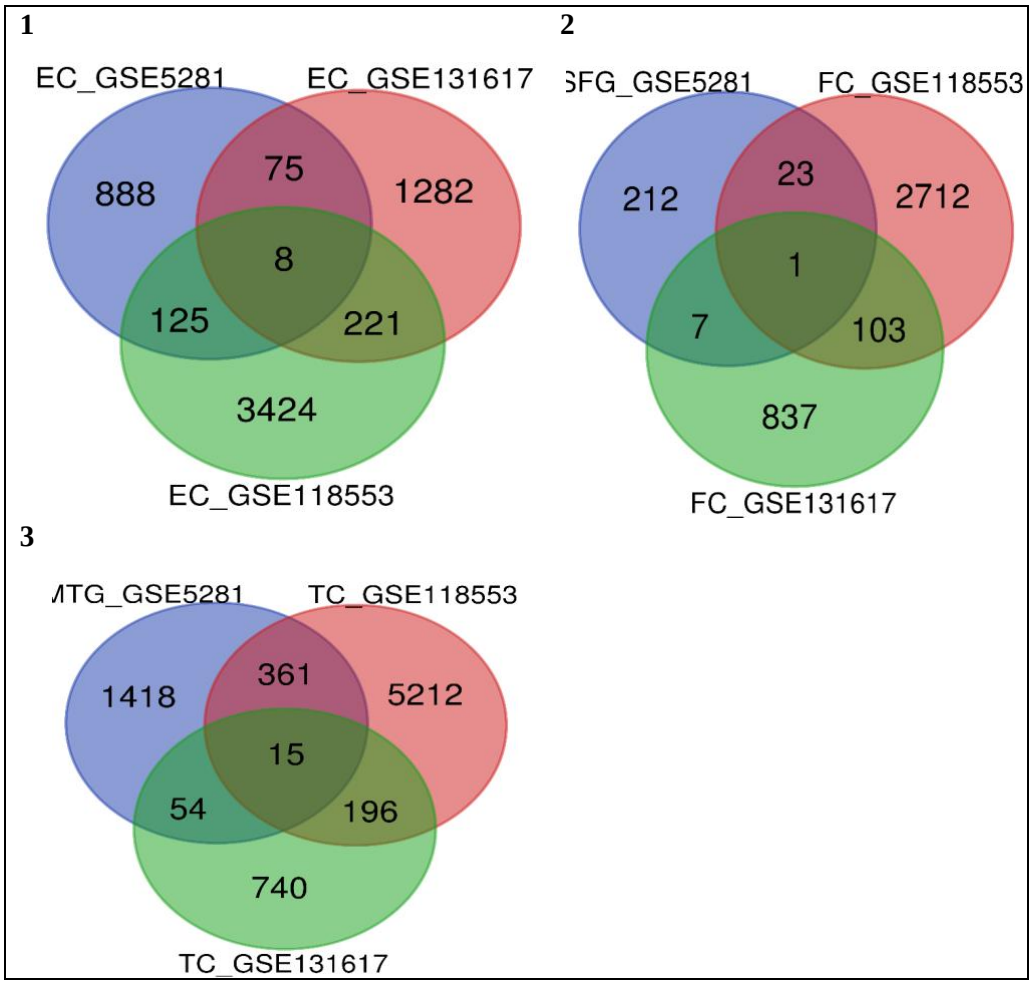

Figure 5: : Differentially expressed genes in the entorhinal, frontal and temporal cortices of all three studies. 1. DEGs in Entorhinal Cortex (EC). SFG(superior frontal gyrus) and MTG(medial temporal

gyrus) in the GSE5281 study correspond respectively to the frontal and temporal cortex in the other two studies

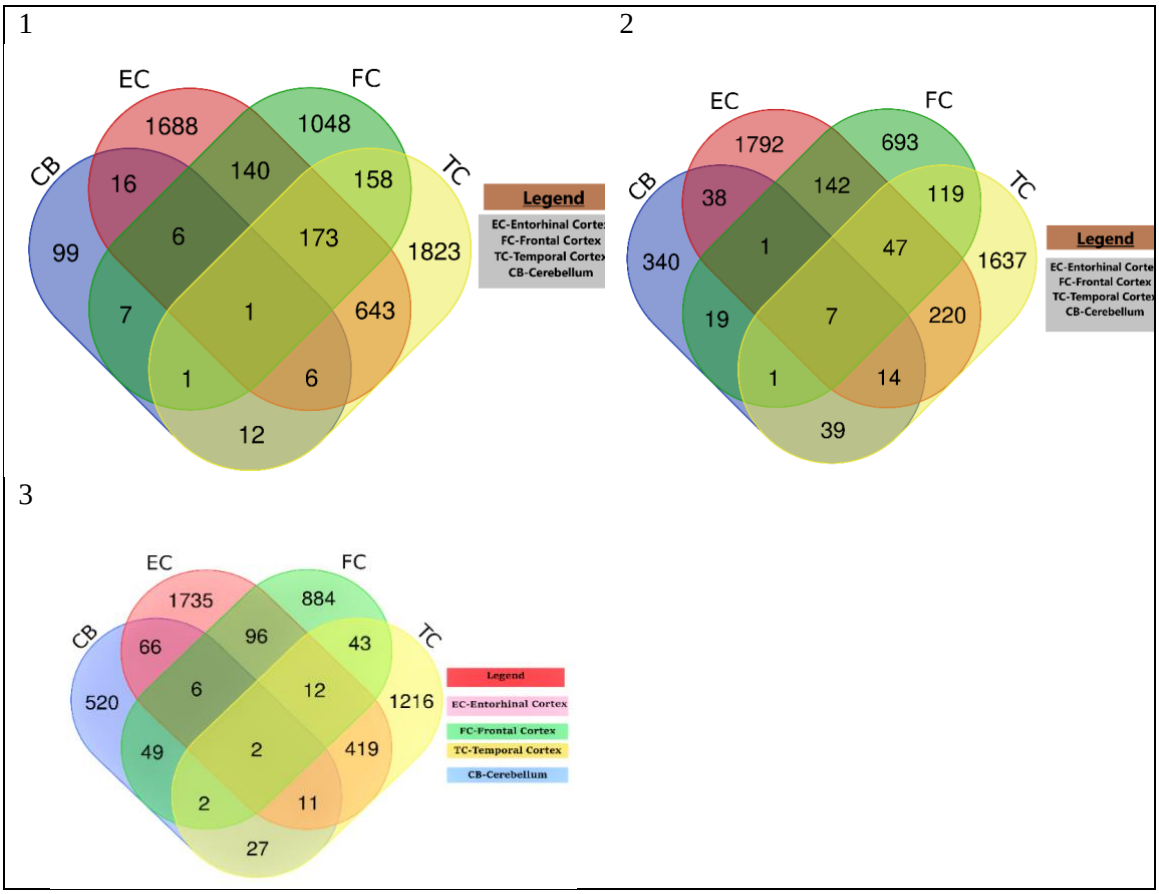

Figure 6: Differentially expressed genes (DEGs) in typical, asymptomatic and late AD cases in GSE118553. 1. DEGs in typical AD in four brain regions: entorhinal cortex (EC), temporal cortex (TC), frontal cortex (FC) and cerebellum (CB). 2. DEGs in asymptomatic AD in four brain regions.. 3. DEGs in 'late' AD in four brain regions.

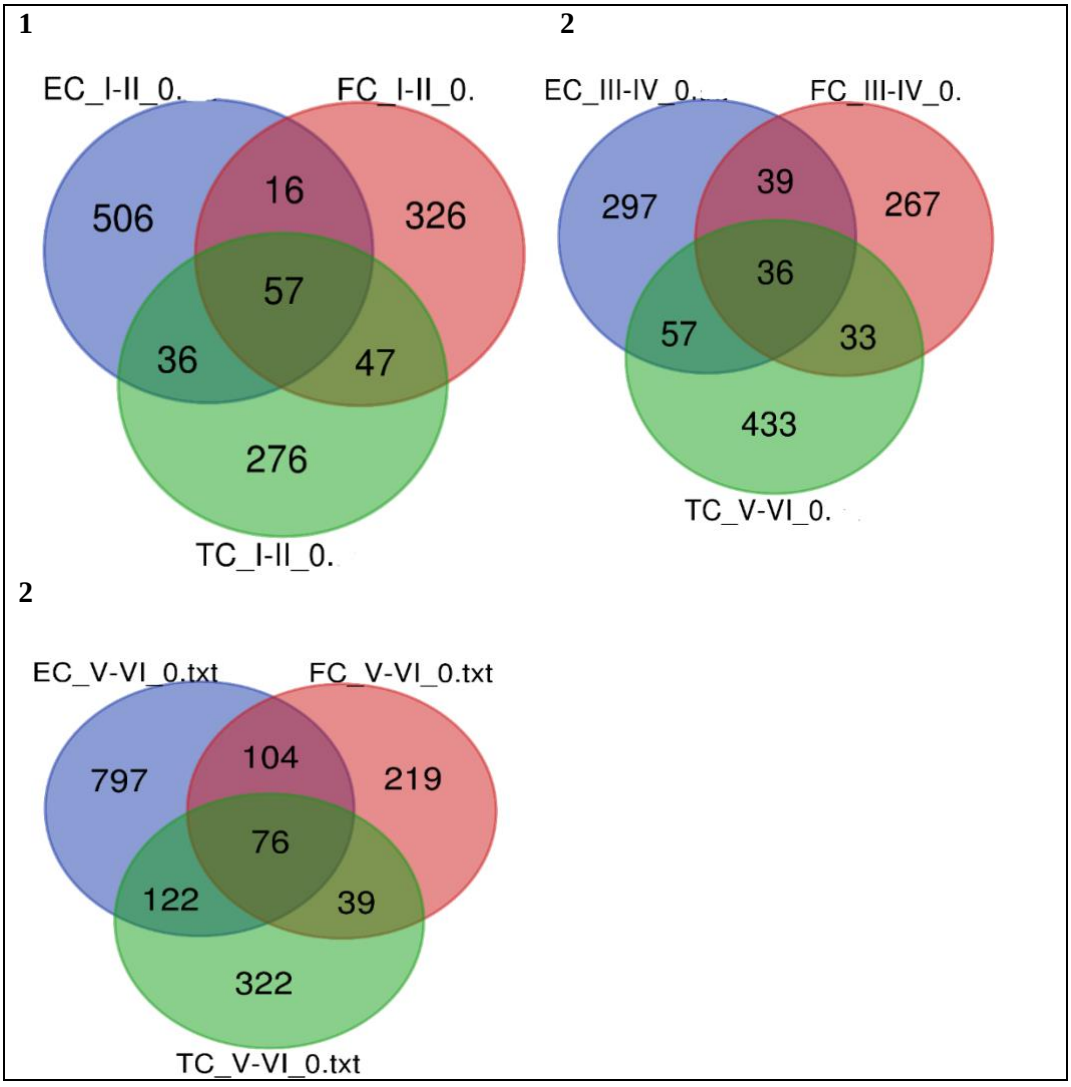

Figure 7: Differentially expressed genes in BRAAK levels I-VI in three brain regions in GSE131617. 1. DEGs in BRAAK levels I-II in different brain regions. 2. DEGs in BRAAK levels III-IV. 3. DEGs in BRAAK levels V-VI.

3.3 Over-representation analyses (ORA) of differentially expressed genes (DEGs)

ORA was employed to analyze the DEGs in GSE5281, GSE118553 and GSE131617 to reveal pathways and genetic disorders that are enriched in the DEGs. As described in the methods, ORA employs hierarchical clustering to group sets of genes and map them to predefined functional categories in a reference database. For pathways, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used as the reference, while

the Online Mendelian Inheritance in Man (OMIM) was used as reference for finding the genetic disorders associated with the DEGs. Some genetic disorders are associated with AD. The aim of the ORA analyses with the OMIM database is to find these disorders.

In each search, all the DEGs identified in each comparison in the brain regions in the GSE5281, GSE118553, GSE131617 were used for the ORA. Sets of DEGs with significantly over-represented pathways have been shown in the diagrams below. A pathway is taken to be significantly over-represented at a -log10FDR ≥ 1.3 (FDR ≤ 0.05). As shown in figure 8, in the TC (MTG) of GSE5281, only neuroactive ligand-receptor interaction was significantly over-represented. A full list of all significant pathways and disorders associated with all the set of DEGs in the three studies have been displayed in table 5 and 6.

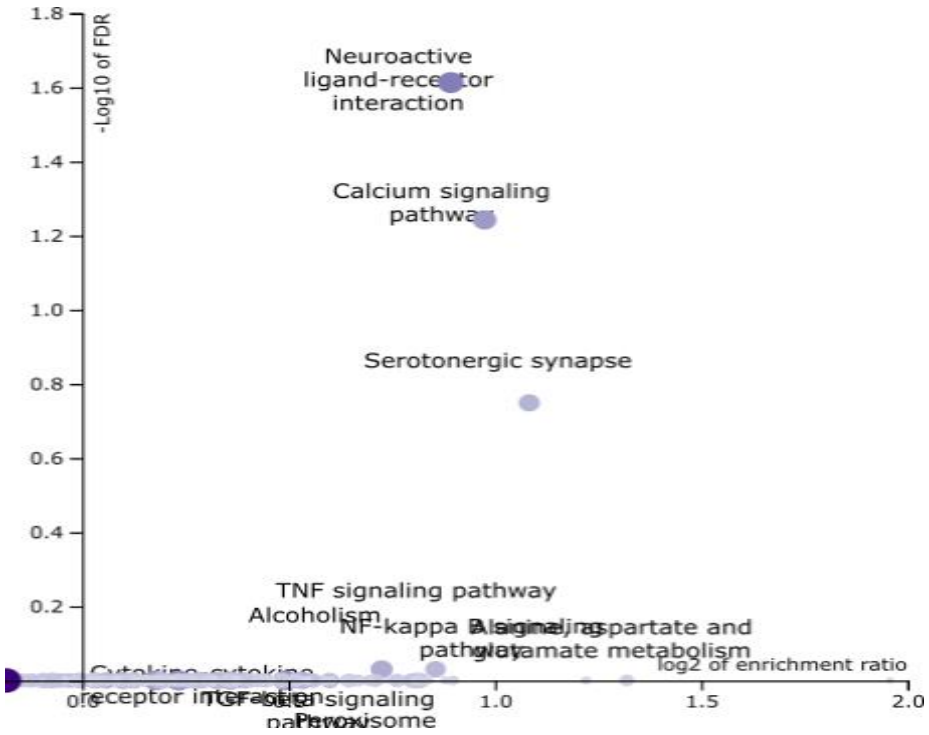

Figure 8: Over-representation analysis of DEGs in the MTG of GSE5281 using KEGG database as reference. Neuroactive ligand-receptor interaction was significantly enriched in this set of DEGs**.**

Table 4: Pathways enriched at FDR ≤0.05 in gene sets from all studies

Study	Brain Region	Pathways	Enrichment ratio	FDR	P-value
GSE5281 AD vs C	EC	None			
	FC	None			
	TC	Neuroactive ligand-receptor interaction	1.8573	2.44E-3	7.52E-5
	HIP	None			
	PC	None			
	VCX	None			
GSE118553 AD vs C	EC	None			
	TC	None			
	FC	None			
GSE118553 AsymAD vs C	EC	None			
	TC	None			
	FC				
GSE118553 AD vs AsymAD	EC	None			
	TC	None			
	FC	None			
GSE131617 BRAAK I-II vs C	EC	None			
	TC	None			
	FC	None			
	CB	None			
GSE131617 BRAAK III-IV vs C	EC	Ribosome	3.6484	9.9E-4	3.06E-6
	TC	Spliceosome	4.2813	0.0064125	1.98E-5
		Ribosome	3.6937	2.79E-3	1.72E-4
	FC	C-type lectin receptor signaling pathway	5.6332	0.00035616	1.10E-6
		Phosphatidylinositol signaling system	4.6570	1.92E-2	1.19E-4
	CB	None			
GSE131617 BRAAK V-VI vs C	EC	Proteasome	3.6920	4.14E-2	2.56E-04
		Spliceosome	2.6932	1.9E-3	5.97E-5
	TC	None			
	FC	None			
	CB	None			

Table 5: Genetic disorders enriched at FDR ≤0.05 in gene sets from all studies

Study	Brain Region	Genetic Disorders	Enrichment ratio	FDR	P-value
GSE5281 AD vs C	EC	Esophageal canceresophageal squamous cell carcinoma	69.663	0.0083494	0.00030924
		Tracheoesophageal fistula with or without esophageal atresia	30.734	0.0056050	0.00010380
		Leigh syndrome	29.026	0.031682	0.0019926
		Colorectal cancer	26.794	0.031682	0.00234468
	FC	None			
	TC	Non-insulin dependent diabetes mellitus	28.282	0.021287	0.0019710
		Susceptibility to type 1 HIV	21.212	0.0014639	0.000027109
		Susceptibility to myocardial infarction	21.212	0.0055723	0.00030957
	HIP	None			
	PC	None			
	VCX	None			
GSE118553 AD vs C	EC	Susceptibility to malaria	15.909	0.0023193	0.000085899
	TC	Essential hypertension	19.353	0.0011498	0.000035930
		Rheumatoid arthritis	19.353	0.0048805	0.00038129
		Susceptibility and resistance to malaria	13.661	0.0025898	0.00016186
		Acute myeloid leukemia	13.195	0.0011498	0.000027424
	FC	Schizophrenia	24.171	0.012790	0.00021739
GSE118553 AsymAD vs C	EC	Macular degeneration, age-related	55.836	0.00081424	0.000012723
		Susceptibility to type 1 HIV	16.824	0.020923	0.00065383
	TC	Alcohol dependence	45.907	0.0015065	0.000023539
	FC	None			
GSE118553 AD vs AsymAD	EC	Atopic IgE responsiveness	37.013	0.0033355	0.000052116
	TC	None			
	FC	None			
GSE131617 BRAAK I-II vs C	EC	None			
	TC	None			
	FC	None			
GSE131617 BRAAK III-IV vs C	EC	Susceptibility and resistance to malaria	87.088	0.00023466	0.00000366
	TC	None			
	FC	None			
GSE131617 BRAAK V-VI vs C	EC	Acute myeloid leukemia	20.706	0.0076696	0.00035951
	TC	None			
	FC	None			

3.4 Network Topology Analyses (NTA) of differentially expressed genes in the entorhinal cortex of all studies

Network topology analyses (NTA) was employed to find the biological processes that are enriched in the DEGs. The PPI Biological General Repository for Interaction Datasets (BioGRID) database was selected as the functional database. This database has close to 2 million well-curated interactions and is widely used in many studies. In the WebGestalt software, the option for network retrieval and prioritization was used to find the GO categories in each brain region and AD levels in various studies. To reduce the chances of introducing false positives, GOs with p-values of <0.01 were included for each set of genes. In cases where there are more than 500 GOs with p-values <0.01, the top 500 GOs were selected.

This analysis focused on the EC because it is the area of the brain that is severely affected during AD. NTA was employed to identify the GO categories enriched in AD vs. C for EC of GSE5281; AD vs. C, AsymAD vs. C and AD vs. AsymAD of EC in the GSE118553; BRAAK I-II vs. 0, BRAAK III-IV vs. 0, BRAAK V-VI vs. 0 of EC in GSE131617 gene sets. The full list of GO categories derived from this analysis can be found in Appendix B.

Network topology analysis ranks the GO categories according to the p-values derived from hypergeometric test. P-value is indicative of the probability that an identified GO category is truly enriched in the gene set. However, the p-value alone is insufficient to rank the GO categories, therefore, the enrichment ratio will also be considered to rank the GOs. Enrichment ratio is suggestive of the degree of enrichment of a GO category in a gene set.

Table 7 shows the top GO categories in the DEGs from the entorhinal cortex of GSE5281 with cellular protein metabolic process being the topmost category. (For a full list of GO

categories derived from NTA, refer to Appendix B). This category has Gene Ontology ID (GoId) GO:0044267. GoId is a unique alphanumeric identification assigned to each functional category in the Gene Ontology database. Each functional category, with a unique GoId has a specific number of genes (its size), with each gene having been experimentally determined to play a role in that category. The size of GO:0044267 for instance is 4630. From the list of interesting genes (1085) that were inputted into the WebGestalt software, 735 seeds were selected in the subnetwork during the process. It is expected that approximately 128 of the 735 seeds overlap the 4630 genes in GO:0044267 by random if the GO were not enriched. However, there were 176 genes that overlapped this GO category, indicating that this GO category is enriched in the interesting genes. The enrichment ratio was 1.37, which shows a modest enrichment although the p-value is highly significant. In other cases, such as GO:2000058, the enrichment ratio is 3.5 with significant p-value. We reasoned that a combined ranking score derived from both enrichment ratio and p-value is useful. The enrichment ratio and the hypergeometric test are very key to discover the functional categories in gene sets. In section

3.5 Prioritization of GO categories from NTA with a novel ranking score, a new ranking score (fully described in methods) based on these two key parameters is used to re-rank the GO categories.

Table 6: GO categories enriched in the entorhinal cortex (EC) of GSE5281

3.5 Prioritization of GO categories from NTA with a novel ranking score

GoId	Description	Size	Overlap	Expect	Enrichment Ratio	P Value	FDR
GO:0044267	cellular protein metabolic process	4630	176	128.78	1.37	3.41E-07	5.4E-3
GO:0006464	cellular protein modification process	3699	142	102.89	1.38	6.85E-06	3.6E-2
GO:0036211	protein modification process	3699	142	102.89	1.38	6.85E-06	3.6E-2
GO:0019538	protein metabolic process	5124	184	142.52	1.29	9.13E-06	3.6E-2
GO:0070647	protein modification by small protein conjugation or removal	977	49	27.18	1.8	4.12E-05	8.3E-2
GO:2000058	regulation of ubiquitin-dependent protein catabolic process	143	14	3.98	3.52	4.51E-05	8.3E-2
GO:0032434	regulation of proteasomal ubiquitin-dependent protein catabolic process	117	12	3.25	3.69	1.00E-04	1.32E-1
GO:0032092	positive regulation of protein binding	86	10	2.39	4.18	1.33E-04	1.44E-1
GO:0032446	protein modification by small protein conjugation	761	39	21.17	1.84	1.69E-04	1.54E-1
GO:0060078	regulation of postsynaptic membrane potential	124	12	3.45	3.48	1.75E-04	1.54E-1
GO:0061024	membrane organization	709	37	19.72	1.88	1.75E-04	0.154008
GO:0032435	negative regulation of proteasomal ubiquitin-dependent protein catabolic process	32	6	0.89	6.74	2.19E-04	1.64E-1
GO:2000059	negative regulation of ubiquitin-dependent protein catabolic process	45	7	1.25	5.59	2.23E-04	1.64E-1
GO:0034138	toll-like receptor 3 signaling pathway	21	5	0.58	8.56	2.29E-04	1.64E-1
GO:1903362	regulation of cellular protein catabolic process	230	17	6.4	2.66	2.46E-04	1.69E-1
GO:0051099	positive regulation of binding	170	14	4.73	2.96	2.87E-04	1.89E-1
GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	374	23	10.4	2.21	3.30E-04	1.97E-1
GO:1903363	negative regulation of cellular protein catabolic process	79	9	2.2	4.1	3.35E-04	1.97E-1
GO:0051098	regulation of binding	351	22	9.76	2.25	3.42E-04	1.97E-1

3.5.1 Ranking of GO categories with novel ranking score

The novel ranking score is applied to GO functional categories enriched in the EC region of all the three studies from NTA analyses. GO categories with FDR ≥ 0.20 were excluded

from this analysis to decrease the chances of including spurious results. AsymAD vs. C in the EC region of GSE118553 was not included in this analysis because all the GO categories had FDR ≥ 0.20.

It has been shown previously that the enrichment ratio and the p-value from the hypergeometric test are important parameters to find GO categories enriched in sets of interesting genes. The novel ranking score developed in this study employs these parameters. As detailed in the methods, the ranking score, S, uses the negative logarithm of the p-value and the enrichment ratio to rank the GO categories. The negative logarithm of the p-values and the enrichment ratio have almost the same weights on the ranking score. Although these parameters contribute almost equally to deriving the new ranking score, the maximal GO category will have the lowest p-value and the highest enrichment score.

As shown in Figure 9- Figure 11, after applying the new ranking score to the GO categories in the EC region of all the three studies, the highly ranked GOs were distinct from those with low scores. Between these two marked groups, there are GOs that appear to be a transitional zone. The highly ranked GO categories have high enrichment scores and low p-values are therefore less likely to be falsely associated with AD. While the low ranked with low enrichment scores and high p-values could be false positives.

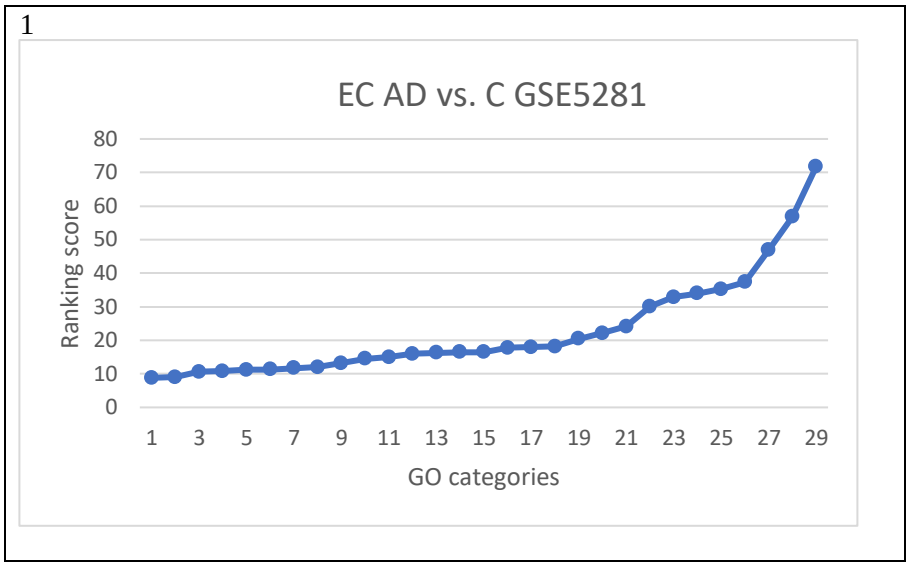

Figure 9: GO categories in the entorhinal cortex (EC) in GSE5281

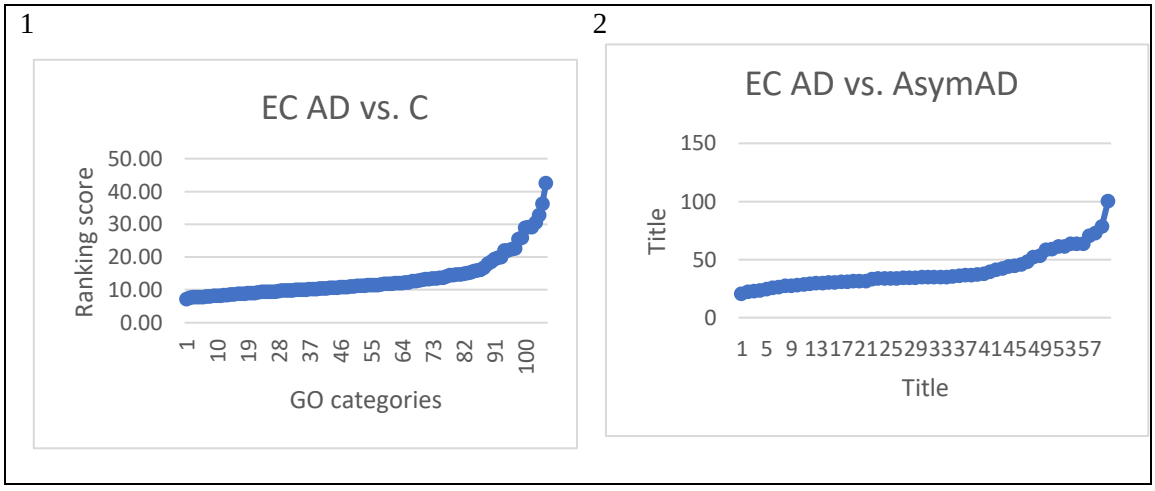

Figure 10: GO categories in the AD vs. C, AD vs. AsymAD in the entorhinal cortex (EC) in GSE118553.

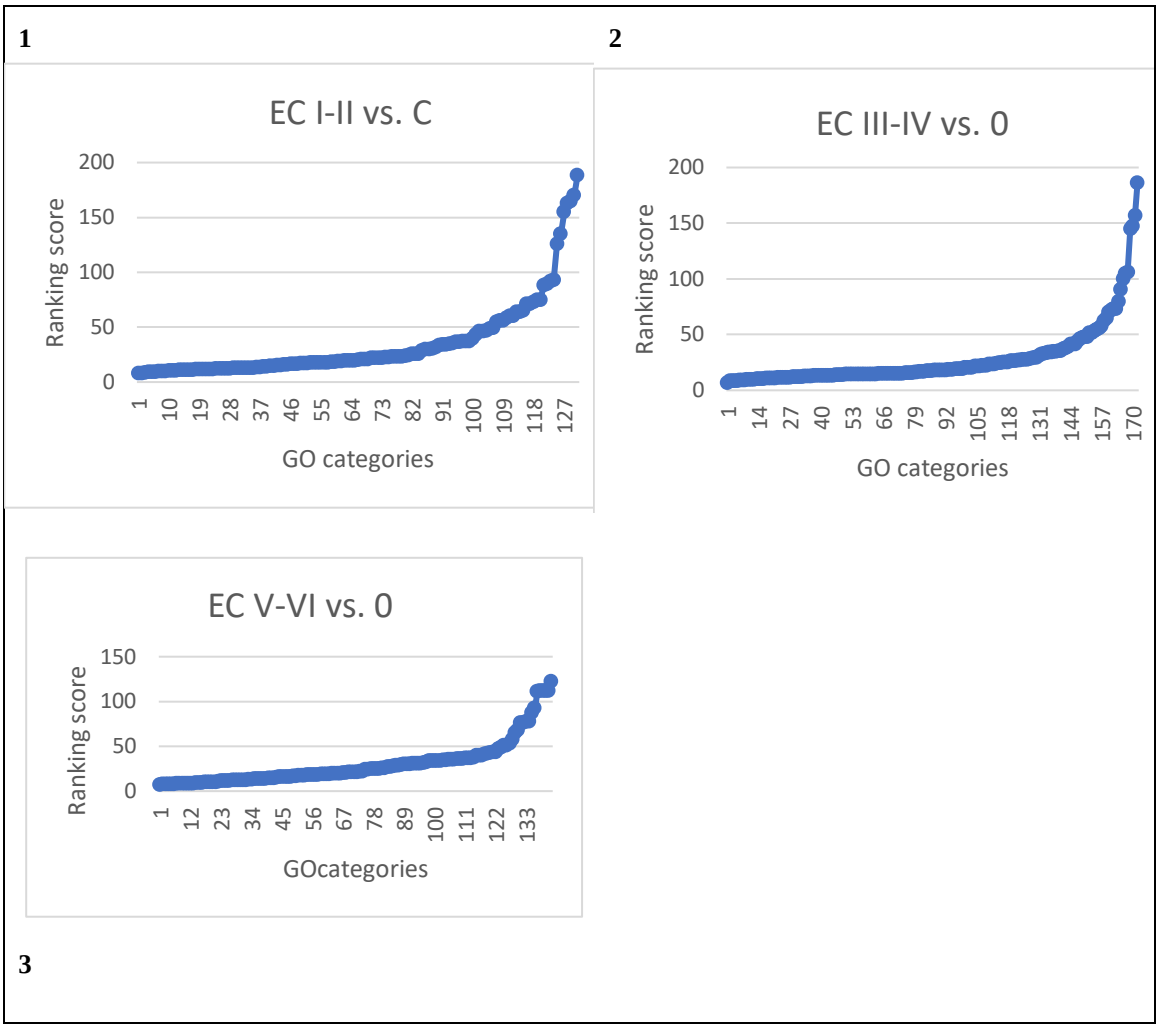

Figure 11 GO categories in the BRAAK I-II vs. 0, BRAAK III-IV vs. 0, BRAAK V-VI vs. 0 for EC in the GSE131617

3.5.2 GO categories enriched in Entorhinal cortex (EC) according to the new ranking score

After normalizing the ranking scores with the maximum score of each set, the GO categories above 60th percentile were selected and have been shown in

Table 7: Ranked GO categories in the AD vs. C of the entorhinal cortex of GSE118553

GoId	Description	Size	Overlap	Expect	Enrichment Ratio	P Value	FDR	Ranking score	Normalized score

GO:0021680	cerebellar Purkinje cell layer development	21	8	1.68	4.75	1.32E-04	7.0E-2	42.42	100.00
GO:0030730	sequestering of triglyceride	15	6	1.20	4.99	6.99E-04	1.3E-1	36.23	85.41
GO:0051123	RNA polymerase II preinitiation complex assembly	24	8	1.93	4.16	3.84E-04	1.0E-1	32.68	77.04
GO:0061028	establishment of endothelial barrier	44	12	3.53	3.40	1.26E-04	7.0E-2	30.53	71.96
GO:0009110	vitamin biosynthetic process	21	7	1.68	4.16	8.94E-04	1.6E-1	29.17	68.76
GO:0030865	cortical cytoskeleton organization	45	12	3.61	3.32	1.60E-04	7.0E-2	29.06	68.49
GO:0030866	cortical actin cytoskeleton organization	40	11	3.21	3.43	2.22E-04	8.0E-2	28.84	67.99
GO:0060260	regulation of transcription initiation from RNA polymerase II promoter	27	8	2.17	3.69	9.33E-04	1.6E-1	25.77	60.75
GO:0001885	endothelial cell development	59	14	4.73	2.96	1.82E-04	7.0E-2	25.47	60.05

7 -**Error! Reference source not found.**. In each set, GO categories that were above the 60$_{th}$ percentile were highly ranked by the ranking score. The full list of GO categories after re-ranking can be found in appendix D.

In the AD vs. C of the entorhinal cortex of GSE118553, cerebellar Purkinje cell layer development and sequestering of triglyceride were the topmost two GO categories, while cortical actin cytoskeleton organization and cortical cytoskeleton organization were the highly enriched pathways in AD vs. AsymAD.

Cotranslational protein targeting to membrane, SRP-dependent cotranslational protein targeting to membrane were the ranked highest in BRAAK I-II vs. C in the entorhinal cortex (EC) of GSE131617; decidualization and regulation of RNA polymerase II regulatory region sequence-specific DNA binding in BRAAK III-IV vs. C; regulation of

telomerase RNA localization to Cajal body and RNA localization to Cajal body in the BRAAK V-VI vs. C

For the entorhinal cortex in GSE5281, toll-like receptor 3 signaling pathway and negative regulation of proteasomal ubiquitin-dependent protein catabolic process were the top 2 GO categories.

Table 7: Ranked GO categories in the AD vs. C of the entorhinal cortex of GSE118553

GoId	Description	Size	Overlap	Expect	Enrichment Ratio	P Value	FDR	Ranking score	Normalized score
GO:0021680	cerebellar Purkinje cell layer development	21	8	1.68	4.75	1.32E-04	7.0E-2	42.42	100.00
GO:0030730	sequestering of triglyceride	15	6	1.20	4.99	6.99E-04	1.3E-1	36.23	85.41
GO:0051123	RNA polymerase II preinitiation complex assembly	24	8	1.93	4.16	3.84E-04	1.0E-1	32.68	77.04
GO:0061028	establishment of endothelial barrier	44	12	3.53	3.40	1.26E-04	7.0E-2	30.53	71.96
GO:0009110	vitamin biosynthetic process	21	7	1.68	4.16	8.94E-04	1.6E-1	29.17	68.76
GO:0030865	cortical cytoskeleton organization	45	12	3.61	3.32	1.60E-04	7.0E-2	29.06	68.49
GO:0030866	cortical actin cytoskeleton organization	40	11	3.21	3.43	2.22E-04	8.0E-2	28.84	67.99
GO:0060260	regulation of transcription initiation from RNA polymerase II promoter	27	8	2.17	3.69	9.33E-04	1.6E-1	25.77	60.75
GO:0001885	endothelial cell development	59	14	4.73	2.96	1.82E-04	7.0E-2	25.47	60.05

Table 8: Ranked GO categories in the AD vs. AsymAD of the entorhinal cortex of GSE118553

GoId	Description	Size	Overlap	Expect	Enrichment Ratio	P-Value	FDR	Ranking score	Normalized score

GO:0030866	cortical actin cytoskeleton organization	40	11	2.81	3.92	6.76 E-05	5.0E-2	37.59	100.00
GO:0030865	cortical cytoskeleton organization	45	11	3.16	3.48	2.15 E-04	8.0E-2	29.39	78.18
GO:0006695	cholesterol biosynthetic process	65	14	4.57	3.07	1.35 E-04	8.0E-2	27.32	72.68
GO:1902653	secondary alcohol biosynthetic process	66	14	4.64	3.02	1.60 E-04	8.0E-2	26.40	70.21
GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	119	21	8.36	2.51	7.56 E-05	5.0E-2	23.84	63.42
GO:0009057	macromolecule catabolic process	1263	136	88.71	1.53	1.84 E-07	1.50 E-3	23.77	63.24
GO:0009057	macromolecule catabolic process	1263	136	88.71	1.53	1.84 E-07	1.50 E-3	23.77	63.24
GO:0070972	protein localization to endoplasmic reticulum	136	23	9.55	2.41	6.97 E-05	5.0E-2	23.05	61.30
GO:0016126	sterol biosynthetic process	70	14	4.92	2.85	3.08 E-04	1.0E-1	23.02	61.24

Table 9: Ranked GO categories in BRAAK I-II vs. C in the entorhinal cortex (EC) of GSE131617

GoId	**Description**	**Size**	**Overlap**	**Expect**	**Enrichment Ratio**	**P Value**	**FDR**	**Ranking score**	**Normalized score**
GO:0006613	cotranslational protein targeting to membrane	98	16	1.96	8.16	9.49 E-11	6.39 E-07	188.22	100.00
GO:0006614	SRP-dependent cotranslational protein targeting to membrane	94	15	1.88	7.97	5.18 E-10	1.49 E-06	170.44	90.55

GO:0045047	protein targeting to ER	106	16	2.12	7.54	3.20E-10	1.27E-06	164.85	87.58
GO:0000387	spliceosomal snRNP assembly	36	8	0.72	11.10	4.35E-07	2.15E-04	162.61	86.39
GO:0072599	establishment of protein localization to endoplasmic reticulum	110	16	2.20	7.27	5.63E-10	1.49E-06	154.75	82.22
GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	119	16	2.38	6.72	1.85E-09	3.06E-06	135.06	71.75
GO:0070972	protein localization to endoplasmic reticulum	136	17	2.72	6.24	1.80E-09	3.06E-06	125.73	66.80

Table 10: Ranked GO categories in BRAAK III-IV vs. C in the entorhinal cortex (EC) of GSE131617

GoId	Description	Size	Overlap	Expect	Enrichment Ratio	P Value	FDR	Ranking score	Normalized score
GO:0046697	decidualization	17	4	0.21	18.82	4.95E-05	1.0E-2	186.57	100.00
GO:1903025	regulation of RNA polymerase II regulatory region sequence-specific DNA binding	12	3	0.15	20.00	3.89E-04	6.0E-2	157.01	84.15
GO:0048025	negative regulation of mRNA splicing, via spliceosome	20	4	0.25	16.00	9.79E-05	2.0E-2	147.67	79.15
GO:0050686	negative regulation of mRNA processing	29	5	0.36	13.79	2.69E-05	1.0E-2	145.12	77.78

Table 11: Ranked GO categories in BRAAK V-VI vs. C in the entorhinal cortex (EC) of GSE131617

GoId	Description	Size	Overlap	Expect	Enrichment Ratio	P Value	FDR	Ranking score	Normalized score
GO:1904872	regulation of telomerase RNA localization to Cajal body	18	7	0.72	9.73	3.40E-06	1.6E-3	122.52	100.00

GO:0090670	RNA localization to Cajal body	19	7	0.76	9.22	5.20E-06	2.1E-3	112.16	91.54
GO:0090671	telomerase RNA localization to Cajal body	19	7	0.76	9.22	5.20E-06	2.1E-3	112.16	91.54
GO:0090672	telomerase RNA localization	19	7	0.76	9.22	5.20E-06	2.1E-3	112.16	91.54
GO:0090685	RNA localization to nucleus	19	7	0.76	9.22	5.20E-06	2.1E-3	112.16	91.54
GO:1904874	positive regulation of telomerase RNA localization to Cajal body	15	6	0.60	10.01	1.46E-05	4.93 E-3	111.44	90.95
GO:0008380	RNA splicing	402	51	16.07	3.17	1.98E-13	3.13E-09	92.85	75.78
GO:0022618	ribonucleoprotein complex assembly	222	33	8.87	3.72	6.21E-11	1.23E-07	87.41	71.34
GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	318	41	12.71	3.23	3.03E-11	1.15E-07	78.13	63.77
GO:0000398	mRNA splicing, via spliceosome	318	41	12.71	3.23	3.03E-11	1.15E-07	78.13	63.77
GO:0071826	ribonucleoprotein complex subunit organization	236	33	9.43	3.50	3.23E-10	5.08E-07	76.46	62.40
GO:0000375	RNA splicing, via transesterification reactions	321	41	12.83	3.20	4.09E-11	1.15E-07	76.44	62.39

Table 12: Ranked GO categories in the AD vs. C of the entorhinal cortex of GSE5281

GoId	**Description**	**Size**	**Overlap**	**Expect**	**Enrichment Ratio**	**P Value**	**FDR**	**Ranking score**	**Normalized Score**
GO:0034138	toll-like receptor 3 signaling pathway	21	5	0.58	8.56	2.29E-04	1.6E-1	71.75	100.00
GO:0032435	negative regulation of proteasomal ubiquitin-dependent protein catabolic process	32	6	0.89	6.74	2.19E-04	1.6E-1	56.79	79.16

GO:2 0000 59	negative regulation of ubiquitin-dependent protein catabolic process	4 5	7	1.2 5	5.59	2.2 3E-04	1.6 E-1	47.00	65.51

3.6 Protein-protein interaction of differentially expressed genes (DEGs)

The relationship between the products of genes is an important aspect of molecular biology. In network topology analyses, the probability of random walk method finds the relationship between the seeds in the proteins encoded by the inputted gene lists and their neighbors. These interactions could be vital in AD pathogenesis. The BioGRID database that was used as a reference in the network topology analyses contains millions of interactions of products encoded by genes. Taking advantage of the probability of random walk method, the interaction or relationship between selected proteins encoded by the DEGs were performed. In the original study that generated the data for GSE131617, protein-protein interaction (PPI) was analyzed using a set of selected proteins. Here we show that by using a set of re-selected proteins based on our analysis, a substantially significant relationship between PPI data and AD pathogenesis can be identified.

3.6.1 Protein-protein interaction of PTGS2, TRIL, DCHS2, GRB14, NPAS4, MYO5C and PHYHD1

BRAAK staging is very effective in classifying cases of AD. [27] showed that 8 genes: RELN, PTGS2, MYO5C, TRIL, DCHS2, GRB14, NPAS4 and PHYHD1 are associated with the BRAAK stage of AD. The authors also performed PPI using the BioGRID database 3.2.104 (September 2013) that included these 8 genes. Our analyses confirm that

7 of these genes are differentially expressed in the EC region during AD. In the EC region, RELN and GRB14 were found in only BRAAK V-VI; PTGS2 and PHYHD1 were found in both BRAAK III-IV and V-VI; DCHS2 in I-II and III-IV; while MYO5C, TRIL, NPAS4 and were found to be differentially expressed in all BRAAK levels. The pattern of differential expression were also similar to that found by [27]. NPAS4 for instance was downregulated in all BRAAK levels. A PPI using an updated version (update 2019) of the BioGRID database [159] was therefore performed for these 7 genes present in the EC region of the GSE131617 dataset. The results shown in figure 9 is similar to the PPI performed by [27] on their list of genes. The top-ranking neighbors from this PPI are COP25, APP, POT1, HNRNPLL, MAGED1, ARNT, ARNT2, ARNTL, TRIM25 and KRAS. It can be observed from this PPI that APP interacts with genes such as PTGS2, PHYHD1, and GRB14 through EGFR. APP also interacts with ARNT2, a transcription factor which contains the highly conserved PAS domain and found to dimerize with the immediate-early gene, NPAS4 which also contains the PAS domain. It will be shown later that this interaction between NPAS4 and ARNT2 may be crucial to the pathogenesis of AD. Additionally, the interaction between other genes such as PTGS2 and AD could be important to AD pathogenesis.

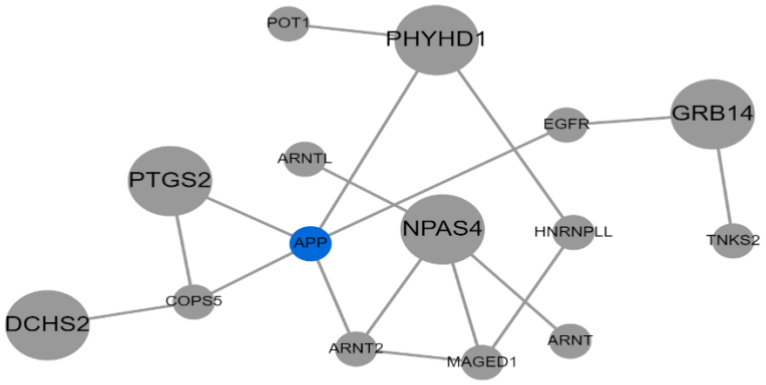

Figure 12: Network of protein-protein interaction of DEGs in this study and GSE131617, GSE118553 (PTGS2, TRIL, DCHS2, GRB14, NPAS4 and PHYHD1). The nodes represent proteins in the expanded sub-network and the edges indicate the connection and flow of information between the proteins. The bigger nodes are the seeds and the smaller nodes are their top neighbors.

3.6.2 Protein-protein interaction of differentially expressed genes in the entorhinal cortex of GSE131617

There are 119 genes common to all the three BRAAK levels in the EC of GSE131617 (full list of genes in the appendix). These genes were used for PPI and the results have been shown in figure 16. The top-ranking neighbors from this PPI included TRIM25, UBC, ELAVL1, HSP90AA1, CD7, AKAP5, APP, EFTUD2, MTNR1A, ABL1.

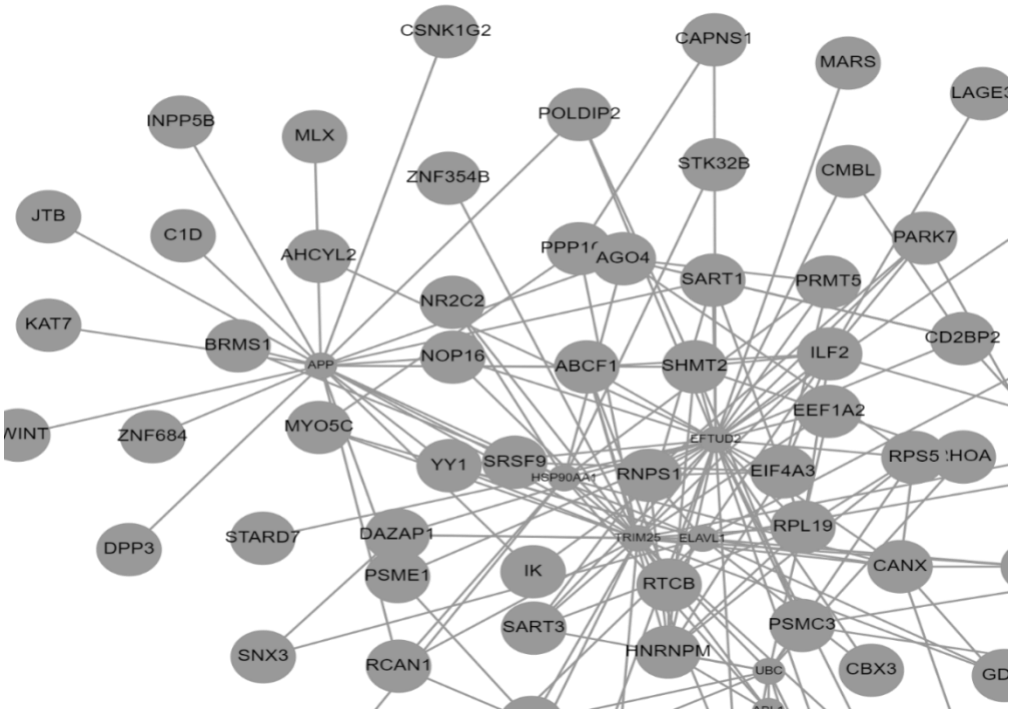

Figure 13: Protein-protein interaction of differentially expressed genes in the entorhinal cortex of GSE131617 (zoomed in to show genes that interact with APP)

3.6.3 Protein-protein interaction of differentially expressed genes in the entorhinal cortex of GSE118553

There are 139 genes common to all types of AD in the EC region of GSE118553 (full list of genes found in the appendix). The results from this PPI has been shown in figure 17. The top ranking neighbors from this PPI included APP, NXF1, ELAVL1, TRIM25, KRAS, LMNA, KRT40, FUS , C6orf211, TRIM15.

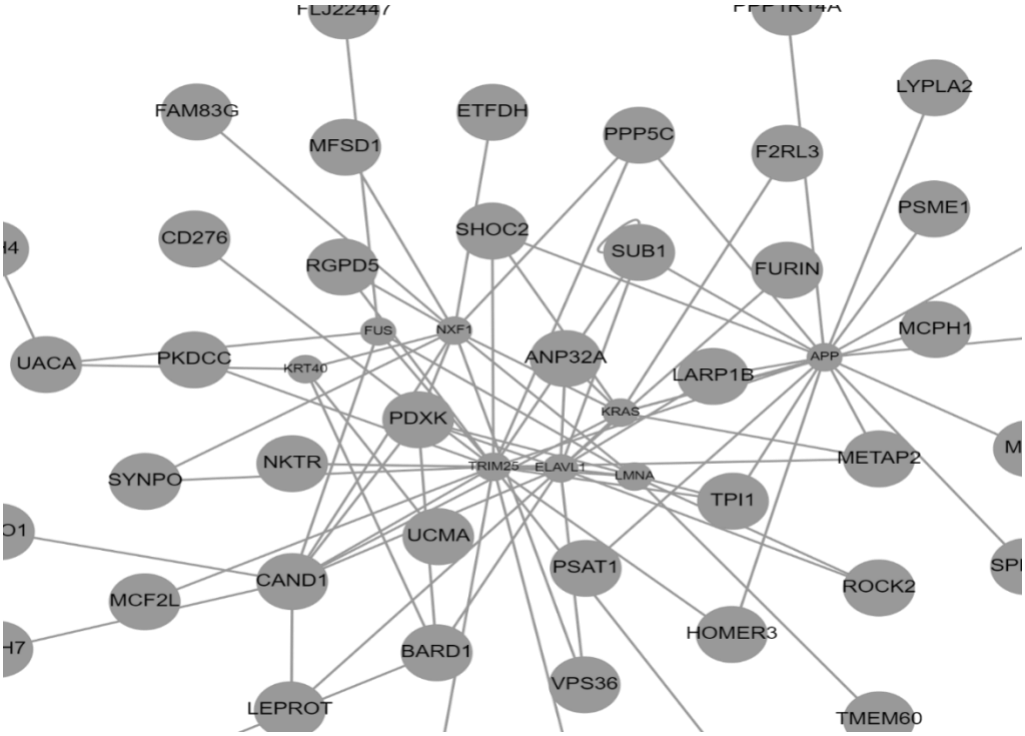

Figure 14: Protein-protein interaction of differentially expressed genes in the entorhinal cortex of GSE118553 (zoomed in to show genes that interact with APP).

3.6.4. Protein-protein interaction of differentially expressed genes in the entorhinal cortex of all three studies

It was observed that eight (8) genes differentially expressed in all three (3) studies. These genes included: SLC46A3, ADAMTSL1, PPP4C, ZC3H6, SEC62, TSSK2, KLHDC10, NDUFA10. A PPI was performed for the proteins encoded by these 8 genes. It was observed that APP interacts with TSSK, and also with ADAMTSL1 through FHL2 and SLC46A3 through LGALS3.

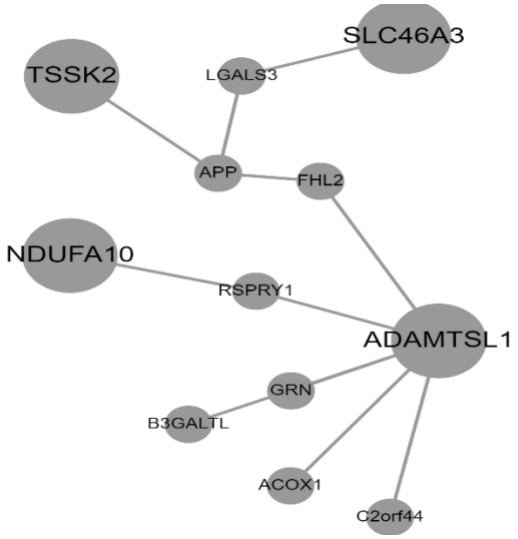

Figure 15: Protein-protein interaction of differentially expressed genes in the entorhinal cortex of all three(3) studies

CHAPTER 4: DISCUSSION

This work sought to identify the genes and molecular processes that are central to Alzheimer's Disease (AD) pathogenesis. Genes that are differentially expressed in various levels of AD in brain regions in AD were analyzed. Functional annotation methods, specifically, over-representation analysis (ORA) and network topology analyses (NTA) were employed to examine the biological significance of the list of differentially expressed genes to AD. The Gene Ontology (GO) categories revealed by the NTA were prioritized with a new ranking score that we developed.

ORA showed that GO categories such as, neuroactive ligand-receptor, ribosome and phosphatidylinositol signaling system could be involved in AD. When our novel ranking score was employed to prioritize the GO categories enriched in these genes derived from NTA, GO categories such as cotranslational protein targeting to membrane, SRP-dependent cotranslational protein targeting to membrane, protein targeting to ER, spliceosomal snRNP assembly, establishment of protein localization to endoplasmic reticulum, nuclear-transcribed mRNA catabolic process, nonsense-mediated decay, decidualization, negative regulation of mRNA splicing, via spliceosome and negative regulation of mRNA processing were significantly ranked high. PPI showed that the proteins encoded by the DEGs interact with APP, TRIM25 and other genes that have previously been reported to play key roles in the pathogenesis of AD, confirming previous PPIs.

4.1 Differentially expressed genes from three datasets

The number of DEGs found in the GSE5281 after correcting for multiple testing with Benjamini-Hochberg were 1095, 1850, 566, 243, 494 and 0 for the EC, TC(MTG), FC(SFG), HIP, PC and VCX regions respectively. These DEGs are lower than what was observed by [158], which used the same dataset.

In GSE118553, there were higher number of DEGs in the AD vs. C and AD vs. AsymAD (2673, 2346) respectively, than was observed by [28]. The number of DEGs in the AsymAD vs. C were however almost the same (11 in this study and 19 in [28]). There were similar number of DEGs in AD vs. AsymAD of the TC region (1731 in this study and 1546 in [28]). For AD vs. C, we reported 2817 DEGs, while [28] had 1517. In the FC we observed much lower DEGs than [28] did. For instance, there are only 64 DEGs in AD vs. C in this study, while 398 were identified by [28]. Both studies identified relatively lower number of DEGs in the CB region, compared to other brain regions.

For GSE131617, we identified 620, 429 and 1103; 418, 402 and 559; 447, 374 and 438 DEGs in the BRAAK I-II, III-IV and V-VI levels in the EC, TC and FC regions respectively. It was observed that the 8 genes that [27] found to be associated with the progression of AD from one BRAAK stage to another were also found in DEGs found in this study. These genes included RELN, PTGS2, MYO5C, TRIL, DCHS2, GRB14, NPAS4 and PHYHD1. Our analyses confirm that these genes are differentially expressed in AD. In the EC in particular, RELN and GRB14 were found in only BRAAK V-VI, PTGS2 and PHYHD1 were found in both BRAAK III-IV and V-VI, DCHS2 in I-II and III-IV, while MYO5C, TRIL, NPAS4 and were found to be differentially expressed in all

BRAAK levels. The pattern of differential expression were also similar to that found by [27]. NPAS4 for instance was downregulated in all BRAAK levels.

4.2 Genes that could play roles in the pathogenesis of AD

Amyloid Precursor protein, encoded by the APP gene, is central to AD pathology. The amyloid precursor protein is a single pass transmembrane protein, which can be cleaved to form amyloidogenic fragments by β or γ-secretase. This protein interacts with many other proteins including the secretases that cleave it into various fragments, and many other proteins yet to be discovered. The APP gene itself undergoes mosaicism in neurons that result from somatic recombination [160]. The many interactions of the protein and the mosaicism of the gene contribute to the complexity associated with AD, which is known to be caused, mostly by alterations to the amyloid precursor protein or its functions. Through our systematic approach, hundreds to thousands of genes were found to be expressed differentially at different levels of AD affection in different brain regions, the proteins encoded by some of which were found, through PPI, to interact with APP.

In the EC region of all three studies, there are eight (8) genes that were differentially expressed. These included SLC46A3, ADAMTSL1, PPP4C, ZC3H6, SEC62, TSSK2, KLHDC10, NDUFA10. It was observed that APP interacts with TSSK, and also with ADAMTSL1 through FHL2 and SLC46A3 through LGALS3.

In the EC region of GSE131617, it was observed that RELN, GRB14, PTGS2, PHYHD1, MYO5C, TRIL, NPAS4 from the eight genes that [27] found to be associated with BRAAK stages of AD were differentially expressed. PTGS2 codes for cyclooxygenase-2 (COX2) enzyme which converts arachidonic acid into prostaglandins that function in inflammatory

response [161] and also serve as target for nonsteroidal anti-inflammatory drugs (NSAIDS). Polymorphisms in PTGS2 has been shown to be associated with AD [162-164]. [163] observed that the rs20417 polymorphism in PTGS2 led to a decreased risk of developing AD. This rs20417 polymorphism could be the contributing factor in reducing the risk of developing AD in certain epidemiological studies, not the anti-inflammatory activity of NSAIDs as these studies [165, 166] have purported. To note, all clinical trials with NSAIDs have failed [167, 168]. As [169] made clear in their review, it's very likely that the effect of NSAIDs on AD is independent of the anti-inflammatory activity of COX enzymes. Since the epidemiological studies did not measure the presence or absence of the rs20417 polymorphism, it's unclear if this polymorphism was masking the observed effect of NSAID in AD patients.

TRIL gene codes for a TLR4 interactor with leucine-rich repeats, a crucial accessory molecule in the Toll-like receptor 4 (TLR4) complex pathway [170]. Toll-like receptors (TLRs) are involved in the identification of microbial products in the innate immune system. TLRs are localized either on the surface of the membrane or within some endosomes. TLR4 is localized on the membranous surface where it detects liposaccharides from microbes. A protein-protein interaction (PPI) by [171] pointed out that TRIL may be a hub gene in a molecular function that's crucial to AD. A more clearer interaction was observed by [27], which found TRIL connecting to the GRB14 pathway through TLR4. In that PPI, the GRB14 pathway is connected to APP gene through the IGF1R gene.

We performed a similar PPI as [27] with the proteins encoded by the seven genes RELN, GRB14, PTGS2, PHYHD1, MYO5C, TRIL, NPAS4 genes using network topology analyses and the result is shown in Figure 12, confirming what [27] found. This figure

indicates that TRIL does not interact with any of the inputted genes, or other genes known to be involved in AD, after network expansion. This is possibly due to the absence of TLR4 in the subnetwork, indicating that TRL4 may be required for TRIL to exert any influence on AD pathogenesis, as suggested by [27]. [27] focused their discussion of their PPI on the relationship between APP and PHYHD1. In this work, we focus on the interaction between APP, ARNT2 and NPAS4 to propose a hypothesis that could underlie the pathogenesis of AD.

The protein-protein interaction in Figure 12 shows that the transcriptional regulator, NPAS4, could play a major role in AD pathogenesis, although it's role may not be through direct interactions with APP. NPAS4, expressed predominantly in the neurons, belongs to the immediate-early class of genes; factors that are induced by external stimuli and regulate transcriptional activities [172]. In neurons, immediate-early genes are involved in regulating neuroplasticity and memory formation [173, 174]. NPAS4 regulates the development of GABAergic synapses [175], which helps in maintaining excitation/inhibition balance during neuronal activity.

A recent study suggested that APP may interact with NPAS4 in modulating inhibitory synaptic transmission [176, 177]. When the authors knocked out APP, NPAS4 was downregulated. They also observed that the interaction between APP and NPAS4 is likely to be through the soluble ectodomain, and not the APP intracellular domain (AICD). After treating mice models with the soluble APP ectodomain, only APP+/+ mice showed transcriptional effect on NPAS4, and they suggested that the soluble APP ectodomain require an APP holoprotein to exert its effect on NPAS4. It's possible that the ectodomain of APP may not require an APP holoprotein *per se*, but APP may interact with another

factor such as ARNT2 that enhance its interaction with NPAS4, as suggested PPI in figure 16. ARNT2 is a transcription factor containing the highly conserved PAS domain that is also present in NPAS4 and other immediate-early genes and is selectively expressed in the brain [178]. Genes with the PAS domain have been shown to exhibit heterodimerization [179]. In fact, the dimerization of NPAS4 and ARNT2 has been reported to necessitate the function of NPAS4 [180].

NPAS4 is neuroprotective, promoting neuronal survival when activated through nuclear calcium signals induced by synaptic excitation [181]. NPAS4 is implicated in conditions that negatively affect memory or neuronal health such as neurocognitive deficits resulting from tumors [182] or glucocorticoid induced age-related memory impairment [183]. Considering that in postnatal life, NPAS4 and other immediate-early genes are involved in activity-regulated synapse development and cognition [184], it's not surprising that dysregulation of NPAS4 or other immediate-early genes may be involved in neurodegeneration.

Specifically, NPAS4 may be complicit in AD pathogenesis through its interactions with APP. Amyloidogenic processing of APP occurs during the endocytic trafficking of APP [185] (See [186] for a review on APP processing and trafficking). Perhaps, NPAS4 (or through its dimerization with ARNT2) cause an increased transcription of APP gene, leading to an increased APP available for amyloidogenic processing by β or γ-secretase in the endosomes. Aberrations in any of these steps could lead to increased production of beta-amyloids that characterize AD. For instance, mutations in the secretases such as BACE1 that cleave APP during its endocytic trafficking may result in abnormal production of beta amyloids. Essentially, the interaction between NPAS4 and APP is an earlier event

to the amyloidogenic processes that are traditionally known to be associated with AD pathogenesis. An even earlier event will be the synaptic activities that result in the transcription of NPAS4 or other immediate-early genes. This interconnection between the synapse, immediate-early genes and APP may be central to AD pathogenesis.

The top neighbors from the protein-protein interaction shown in **Error! Reference source not found.**included APP, NXF1, ELAVL1, TRIM25, EGFR, KRAS, LMNA, KRT40, FUS, C6orf211, TRIM15, UBC, HSP90AA1, CD7, AKAP5, EFTUD2, MTNR1A, ABL1.

APP interactions with other genes in the protein-protein network shown in **Error! Reference source not found.**16 have also been well-studied. Polymorphisms in the EGFR gene has been shown to protect against the risk of AD [187]. EGFR binds to the AICD derived from cleavage of APP and is involved in adult neurogenesis [188]. For A-type Lamin (LMNA) gene, [189] showed that it is disproportionately increased during late stages of AD. To be sure, [190] has even demonstrated that AD could be a neurodegenerative laminopathy. Indeed, there are several theories to explain the pathogenesis of AD, but most of these theories explain a part of the complex network that drive AD.

A theory that encompasses the significant neuronal activities driving AD is proposed. Synaptic activities such as excitation/inhibition lead to transcription of NPAS4. The activities of NPAS4 (and by extension other activity-dependent immediate-early genes) such as its dimerization with ARNT2, cause it to interact and enhance the transcription of APP gene or the functional activity of protein encoded by this gene. An increase in transcription leads to a surge in nascent APP available for cleavage by β or γ-secretases during the endocytic shuttling of APP, leading to an increased production of beta-amyloids.

As one ages, the capacity to clear beta-amyloids reduce, therefore perturbations to any of these steps that lead to an increased production of beta-amyloids will result in accumulation of these proteins, which is characteristic of AD. Although disturbance to any of these steps could result in AD, the interaction between NPAS4, ARNT2, and APP may be fundamental to the pathogenesis of AD. This is because the NPAS4-ARNT2-APP interaction serves as an interconnection between environmental factors such as neuroinflammation, depression or acute head injury and the phenotype of AD, that is, amyloidosis. It has been established that NPAS4 dimerize with ARNT2, however, the interaction between NPAS4-ARNT2 and APP isn't well-studied, with the only evidence being this computational study and the study by [27] and [176].

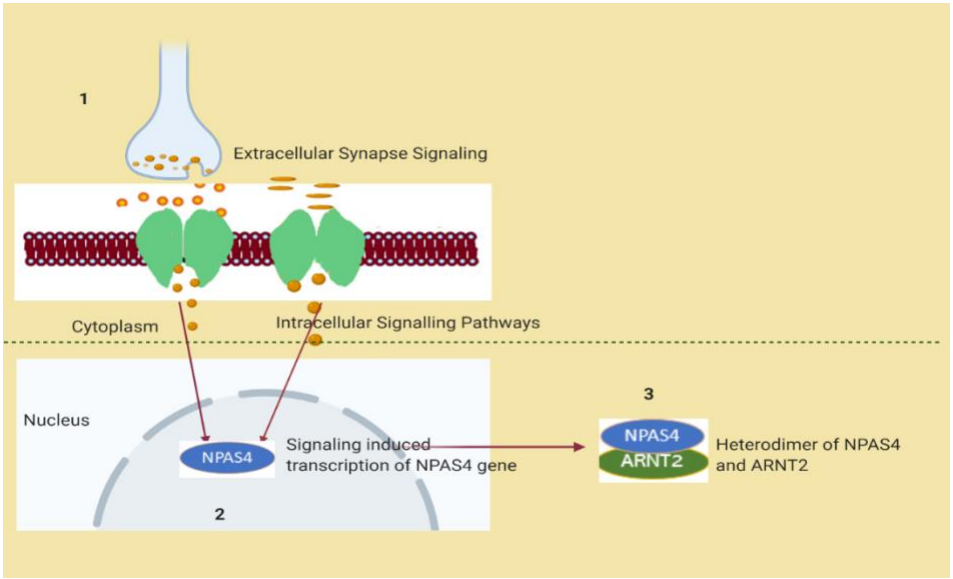

Figure 16: Synaptic activity induced transcription of NPAS4. Synaptic activities are transmitted to the nucleus through signaling pathways where NPAS4 transcription is induced. NPAS4 dimerize with ARNT2, activating it to interact with APP and presumably causing its transcription or enhancing its function.

Memory impairment is a hallmark of many brain disorders including AD. Memory can be episodic or semantic, although there are other systems of classifying memory, like short-term vs long-term. Episodic memory is a collection of spatial and temporal contexts of an individual. Semantic memory consists of generalizable knowledge and concepts such as an individual's beliefs. Regardless of the system used to classify memory, it's been established that memory is characterized by a time-dependent consolidation facilitated by neuroplasticity [191]. There are overwhelming evidence that transcription plays a major role in consolidation of memory [174]. Immediate-early genes code for transcription factors that regulate memory formation and consolidation [192]. Upon retrieval of memory such as when one recalls an event or concept, memory becomes labile, and needs to be reconsolidated [193]. Again, immediate-early genes play key roles in memory reconsolidation [194, 195]. Immediate-early genes are transcribed in response to neuronal activities [184], and regulate the transcription of other genes involved in excitatory and inhibitory synapses [196]. Immediate-early genes are involved in synapse development, plasticity, and memory formation.

Although the specific physiological role of APP is still under investigation, APP may be involved in neurogenesis [197]. APP has also been shown to play roles in synaptic activity and neuronal plasticity [198]. There are evidence indicating that APP may be involved in the regulation of transcription of other genes. For instance, the intracellular domain of APP has been found to regulate transcription of APP itself [199]. The soluble ectodomain may regulate the transcription of NPAS4, leading to the an increase in inhibitory synapses [176]. It can be hypothesized that during memory formation (including consolidation and reconsolidation), synaptic activities lead to an increased transcription of NPAS4, which

dimerize with ARNT2, and enhance its interaction with APP to promote synapse development. Disruption of these processes could lead to abnormal production of beta-amyloids, which in turn affect synaptic activities, therefore acting as feedback loop to the NPAS4-APP interaction. This dysregulation could result in neuronal cell death.

4.2 GO categories enriched in differentially expressed genes (DEGs) during AD

GO categories consist of biological processes, molecular functions and cell components. In this study, GO categories such as neuroactive ligand-receptor, ribosome, spliceosome, phosphatidylinositol signaling system were found to be enriched in the differentially expressed genes using over-representation analyses (ORA). When network topology analyses (NTA) was used to find GO categories enriched in the gene sets, and the output further ranked by a new ranking score we developed, we found certain GO categories to be significantly enriched. In the EC of GSE5281 for instance, toll-like receptor 3 signaling pathway and negative regulation of proteasomal ubiquitin-dependent protein catabolic process were significantly enriched.

There are some common or similar GO categories found from both ORA and NTA analyses. The spliceosome was found by ORA and negative regulation of RNA splicing through the spliceosome was found to be significantly associated with AD. Also, the proteasome was found by ORA, while negative regulation of proteasomal ubiquitin-dependent protein catabolic process was revealed through ORA.

4.2.1 GO categories from over-representation analyses (ORA)

ORA showed that neuroactive ligand-receptor, ribosome, spliceosome, phosphatidylinositol signaling system and C-type lectin receptor signaling GO categories could be involved from the KEGG database could be important for the pathogenesis of AD.

Neuroactive ligand-receptor interaction has recently been found in several studies to be associated with AD. This pathway was found in a protein-protein interaction network constructed from differentially expressed genes associated with AD [200]. Also, the recent genetic network by [201] found this pathway in AD patients. There are multiple other studies that have identified this pathway [202-204], although majority of these were *in silico* studies. This mounting evidence warrants further biochemical experiments to elucidate the role of this pathway in AD pathogenesis.

Our analyses show that other GO categories such as ribosomes and phosphatidylinositol signaling system may be implicated during AD. There are evidence showing that ribosomal dysfunction may be an earlier event before the onset of AD [23]. Other studies have suggested that the entire protein synthesis machinery, including the ribosome could be altered during AD [24]. For phosphatidylinositol signaling system, several studies have noted that impairment in this biological process to be associated with AD [205, 206].

Our analyses with ORA show, for the first time that, C-type lectin receptor signaling pathway could be involved in AD. C-type lectin receptors are pattern recognition receptors that play crucial roles in immune response.

4.2.2 GO categories from network topology analyses (NTA) after prioritizing with new ranking score

After prioritizing the GO categories with our ranking score, it was found that SRP-dependent cotranslational protein targeting to membrane was significantly associated with AD. For a nascent protein, SRP-dependent cotranslational protein targeting to membrane is an important pathway in its trafficking in the cell. The dynamism between the signal-recognition particle (SRP) and its receptor helps in the delivery of nascent proteins to their destination membrane [207]. It was recently shown by [208] that differentially expressed genes in AD patients were enriched with SRP-dependent cotranslational protein targeting to membrane pathway. Similarly, [209] found that SRP-dependent protein targeting was enriched in their hub genes after performing a transcriptomic and lipidomic profiles of AD brains. AD is characterized by the intracellular accumulation of aggregated tau proteins. Protein aggregation impairs the dynamics of the endoplasmic reticulum leading to the accumulation of non-translocated proteins in neurodegenerative disorders [210]. These findings indicate that SRP-dependent protein targeting pathway could play significant roles in AD, and therefore should be further studied. In the GO tree, SRP-dependent cotranslational protein targeting is an offspring of protein targeting to ER, and establishment of protein localization to endoplasmic reticulum.

The small nuclear ribonucleoprotein particles (snRNPs) play important role in the spliceosome, the complex machinery that performs the excision and ligation of introns from transcribed mRNA. The five major snRNPs include the U1, U2, U4, U5 and U6 [211]. Disruption in U1 snRNP has been found to be associated to early stages of AD [212]. Downregulation of the spliceosomal subunits could contribute to AD [213]. As [214] described, mis-localization of the subunits in the spliceosomal complex, particularly the

U1 subunit, could result in neuronal cell cycle reentry. Cell cycle reentry occurs when the machinery involved in cell division enter non-dividing cells such as the neurons. There is overwhelming evidence that this could mediate AD [215]. Other evidence suggests that disturbance of the spliceosomal complex by tau proteins could exacerbate AD [216].

Our results indicate that the nuclear-transcribed mRNA catabolic process, nonsense-mediated decay is enriched in the dataset. There is evidence that this pathway could be involved in certain neurodegenerative disorders such as Parkinson's disease [217], however there has been no evidence of this pathway being associated with AD, except for [218]. This study is the first to show that nuclear-transcribed mRNA catabolic process, nonsense-mediated decay could play a key role in AD pathogenesis.

We find through our results that decidualization, the transformation of endometrial cells in preparation for pregnancy, is associated with AD. The relationship between decidualization and AD is not clear. Through decidualization, pro-inflammatory mediators such as IL-33 are released [219]. Certain pro-inflammatory mediators could break down the blood-brain barrier [220], suggesting that decidualization could result in blood-brain barrier disturbance that could possibly lead to AD.

The prioritization of GO categories with our ranking score showed novel pathways such as corticalactin cytoskeleton organization, vitamin biosynthetic process, endothelial cell development to be associated with AD. Such pathways could be studied further for developing therapeutic and prognostic interventions to bring down the incidence of AD.

4.3 Significance of ranking in gene set functional annotation

In gene set enrichment analyses, [221] showed that the choice of parameters to rank the genes significantly influence the results from the analyses. It was proposed by [222] that improving the statistical power of gene set enrichment analyses could be done by splitting the dataset and averaging the enrichment scores. Although these studies focused on GSEA, they indicate the need to improve ranking during the performance of such analyses as well as the results from them. One of the few studies to undertake a comprehensive, systematic analyses of the sensitivity, prioritization and specificity of different methods of gene set enrichment analyses was [223]. The authors found that the methods that performed well in prioritizing the targeted pathways did not show any high performance with sensitivity, and vice versa. Prioritization and sensitivity are two very significant aspects of gene set enrichment. Prioritization ranking pathways that are relevant to the phenotype to the top, while sensitivity means there are small p-values associated with such pathways. The specificity is the chance of finding true positive pathways. Usually, the p-values are used in ranking pathways.

For each pathway, the enrichment ratio shows how much the pathways are enriched in gene sets. We combined the p-values and the enrichment ratios of the pathways and their p-values to derive a new way of ranking the pathways. Interestingly, we find that in AD, the pathways that ranked close to the top had very low FDR, suggesting that are truly associated with the phenotype (AD). This simple ranking score can be applied in most gene set enrichment analyses methods to improve the prioritization of the identified pathways.

CHAPTER 5: CONCLUSION

To the best of our knowledge, this is the first comprehensive study to examine the molecular functions and biological processes that may be implicated during AD. This study utilized publicly available gene expression data to rediscover genes and molecular processes from gene ontological databases that may be important for the pathogenesis of AD. In particular, this study confirms that genes such as PTGS2, DCHS2, GRB14, NPAS4 and PHYHD1 interact with APP during AD pathogenesis. This study also developed a new method for prioritizing Gene Ontology (GO) categories that are enriched in sets of genes. The ranking score is simple to implement and can be adopted to discover processes that may be involved in the pathogenesis of complex disorders.

APPENDICES

Appendix A: Data for the identification of differentially expressed genes and common DEGs from all three studies.

Appendix B: Full list of GO categories.

Appendix C: Full list of pathways and genetic disorders found through ORA.

Appendix D: Full list of GO categories with FDR ≤ 0.20 used for re-ranking GO categories from NTA.

Appendix E: Scripts for parsing data and performing other analyses.

Appendix A

The full list of differentially expressed genes and the common genes expressed in brain regions and different AD levels have been deposited in github and can be found at https://github.com/alexis263/Processed-data-for-identifying-DEGs-and-common-DEGs-in-all-studies.git .

Appendix B

The full list of GO categories in AD vs. C for EC; AD vs. C, AsymAD vs. C and AD vs. AsymAD for EC in the GSE118553; BRAAK I-II vs. 0, BRAAK III-IV vs. 0, BRAAK V-VI vs. 0 for EC in the GSE131617 https://github.com/alexis263/List-of-GO-categories-.git

Appendix C

The full list of pathways and disorders enriched in the datasets found through over-representation analyses by the WebGestalt software can be found at https://github.com/alexis263/Pathways-retrieved-from-ORA-in-WebGestalt.git.

Appendix D

Full list of GO categories with FDR ≤ 0.20 that were used for re-ranking

https://github.com/alexis263/GO-categories-with-FDR-0.20-

Appendix E

The scripts for parsing data and performing other analyses. Are deposited in github https://github.com/alexis263/Scripts-for-discovery-of-genes-and-molecular-processes-that-are-important-for-alzheimer-s-disease-pa.git

REFERENCES

1. Zivetz L: **The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines**, vol. 1: World Health Organization; 1992.
2. Alonso AdC, Grundke-Iqbal I, Iqbal K: **Alzheimer's disease hyperphosphorylated tau sequesters normal tau into tangles of filaments and disassembles microtubules**. *Nature medicine* 1996, **2**(7):783-787.
3. Abubakar I, Tillmann T, Banerjee A: **Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013**. *Lancet* 2015, **385**(9963):117-171.
4. Barker WW, Luis CA, Kashuba A, Luis M, Harwood DG, Loewenstein D, Waters C, Jimison P, Shepherd E, Sevush S: **Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank**. *Alzheimer Disease & Associated Disorders* 2002, **16**(4):203-212.
5. Association As: **2019 Alzheimer's disease facts and figures**. *Alzheimer's & Dementia* 2019, **15**(3):321-387.
6. [https://www.census.gov/data/tables/2014/demo/popproj/2014-summary-tables.html]
7. Hebert LE, Weuve J, Scherr PA, Evans DA: **Alzheimer disease in the United States (2010–2050) estimated using the 2010 census**. *Neurology* 2013, **80**(19):1778-1783.
8. Weuve J, Hebert LE, Scherr PA, Evans DA: **Prevalence of Alzheimer disease in US states**. *Epidemiology* 2015, **26**(1):e4-e6.
9. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP: **The global prevalence of dementia: a systematic review and metaanalysis**. *Alzheimer's & dementia* 2013, **9**(1):63-75. e62.
10. Niu H, Álvarez-Álvarez I, Guillén-Grima F, Aguinaga-Ontoso I: **Prevalence and incidence of Alzheimer's disease in Europe: A meta-analysis**. *Neurología (English Edition)* 2017, **32**(8):523-532.
11. Montgomery W, Ueda K, Jorgensen M, Stathis S, Cheng Y, Nakamura T: **Epidemiology, associated burden, and current clinical practice for the**

diagnosis and management of Alzheimer's disease in Japan. *ClinicoEconomics and outcomes research: CEOR* 2018, **10**:13.

12. Mullan M, Crawford F, Axelman K, Houlden H, Lilius L, Winblad B, Lannfelt L: **A pathogenic mutation for probable Alzheimer's disease in the APP gene at the N–terminus of β–amyloid**. *Nature genetics* 1992, **1**(5):345-347.

13. Lee JH, Mayeux R, Mayo D, Mo J, Santana V, Williamson J, Flaquer A, Ciappa A, Rondon H, Estevez P: **Fine mapping of 10q and 18q for familial Alzheimer's disease in Caribbean Hispanics**. *Molecular psychiatry* 2004, **9**(11):1042.

14. Sillén A, Andrade J, Lilius L, Forsell C, Axelman K, Odeberg J, Winblad B, Graff C: **Expanded high-resolution genetic study of 109 Swedish families with Alzheimer's disease**. *European Journal of Human Genetics* 2008, **16**(2):202.

15. Olson JM, Goddard KA, Dudek DM: **A second locus for very-late-onset Alzheimer disease: a genome scan reveals linkage to 20p and epistasis between 20p and the amyloid precursor protein region**. *The American Journal of Human Genetics* 2002, **71**(1):154-161.

16. Grupe A, Abraham R, Li Y, Rowland C, Hollingworth P, Morgan A, Jehu L, Segurado R, Stone D, Schadt E: **Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants**. *Human molecular genetics* 2007, **16**(8):865-873.

17. Coon KD, Myers AJ, Craig DW, Webster JA, Pearson JV, Lince DH, Zismann VL, Beach TG, Leung D, Bryden L: **A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease**. *The Journal of clinical psychiatry* 2007, **68**(4):613-618.

18. Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, Zismann VL, Joshipura KD, Pearson JV, Hu-Lince D, Huentelman MJ: **GAB2 alleles modify Alzheimer's risk in APOE ε4 carriers**. *Neuron* 2007, **54**(5):713-720.

19. Li H, Wetten S, Li L, Jean PLS, Upmanyu R, Surh L, Hosford D, Barnes MR, Briley JD, Borrie M: **Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease**. *Archives of neurology* 2008, **65**(1):45-53.

20. Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, Georgieva L, Dowzell K, Cichon S, Hillmer AM, O'Donovan MC: **A genome-wide association study for late-onset Alzheimer's disease using DNA pooling**. *BMC medical genomics* 2008, **1**(1):44.

21. Carrasquillo MM, Zou F, Pankratz VS, Wilcox SL, Ma L, Walker LP, Younkin SG, Younkin CS, Younkin LH, Bisceglio GD: **Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease**. *Nature genetics* 2009, **41**(2):192.

22. Bertram L, Lange C, Mullin K, Parkinson M, Hsiao M, Hogan MF, Schjeide BM, Hooli B, DiVito J, Ionita I: **Genome-wide association analysis reveals putative**

Alzheimer's disease susceptibility loci in addition to APOE. *The American Journal of Human Genetics* 2008, **83**(5):623-632.

23. Goate A, Chartier-Harlin M-C, Mullan M, Brown J, Crawford F, Fidani L, Giuffra L, Haynes A, Irving N, James L: **Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease**. *Nature* 1991, **349**(6311):704-706.

24. Kamino K, Orr HT, Payami H, Wijsman EM, Alonso ME, Pulst SM, Anderson L, O'dahl S, Nemens E, White JA: **Linkage and mutational analysis of familial Alzheimer disease kindreds for the APP gene region**. *American journal of human genetics* 1992, **51**(5):998.

25. Ghiso J, Frangione B: **Amyloidosis and Alzheimer's disease**. *Advanced drug delivery reviews* 2002, **54**(12):1539-1551.

26. Liang WS, Dunckley T, Beach TG, Grover A, Mastroeni D, Walker DG, Caselli RJ, Kukull WA, McKeel D, Morris JC: **Gene expression profiles in anatomically and functionally distinct regions of the normal aged human brain**. *Physiological genomics* 2007, **28**(3):311-322.

27. Miyashita A, Hatsuta H, Kikuchi M, Nakaya A, Saito Y, Tsukie T, Hara N, Ogishima S, Kitamura N, Akazawa K: **Genes associated with the progression of neurofibrillary tangles in Alzheimer's disease**. *Translational psychiatry* 2014, **4**(6):e396.

28. Patel H, Hodges AK, Curtis C, Lee SH, Troakes C, Dobson RJ, Newhouse SJ: **Transcriptomic analysis of probable asymptomatic and symptomatic alzheimer brains**. *Brain, behavior, and immunity* 2019, **80**:644-656.

29. Campion D, Dumanchin C, Hannequin D, Dubois B, Belliard S, Puel M, Thomas-Anterion C, Michon A, Martin C, Charbonnier F: **Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum**. *The American Journal of Human Genetics* 1999, **65**(3):664-670.

30. Mendez MF: **Early-onset Alzheimer disease**. *Neurologic clinics* 2017, **35**(2):263-281.

31. [https://www.nia.nih.gov/health/early-onset-alzheimers-disease-resource-list]

32. Chouraki V, Seshadri S: **Genetics of Alzheimer's disease**. In: *Advances in genetics. Volume 87*, edn.: Elsevier; 2014: 245-294.

33. Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, Jun G, DeStefano AL, Bis JC, Beecham GW: **Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease**. *Nature genetics* 2013, **45**(12):1452.

34. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small G, Roses AD, Haines J, Pericak-Vance MA: **Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families**. *Science* 1993, **261**(5123):921-923.

35. Saunders AM, Strittmatter WJ, Schmechel D, George-Hyslop PS, Pericak-Vance M, Joo S, Rosi B, Gusella J, Crapper-MacLachlan D, Alberts M: **Association of apolipoprotein E allele ϵ4 with late-onset familial and sporadic Alzheimer's disease**. *Neurology* 1993, **43**(8):1467-1467.

36. Marshall GA, Fairbanks LA, Tekin S, Vinters HV, Cummings JL: **Early-onset Alzheimer's disease is associated with greater pathologic burden**. *Journal of geriatric psychiatry and neurology* 2007, **20**(1):29-33.

37. Bigio EH, Hynan L, Sontag E, Satumtira S, White III C: **Synapse loss is greater in presenile than senile onset Alzheimer disease: implications for the cognitive reserve hypothesis**. *Neuropathology and applied neurobiology* 2002, **28**(3):218-227.

38. Katzman R, Terry R, DeTeresa R, Brown T, Davies P, Fuld P, Renbing X, Peck A: **Clinical, pathological, and neurochemical changes in dementia: a subgroup with preserved mental status and numerous neocortical plaques**. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 1988, **23**(2):138-144.

39. Aikawa H, Suzuki K, Iwasaki Y, Iizuka R: **Atypical Alzheimer's disease with spastic paresis and ataxia**. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 1985, **17**(3):297-300.

40. Balasa M, Gelpi E, Antonell A, Rey M, Sanchez-Valle R, Molinuevo J, Llado A: **Clinical features and APOE genotype of pathologically proven early-onset Alzheimer disease**. *Neurology* 2011, **76**(20):1720-1725.

41. Koedam EL, Lauffer V, van der Vlies AE, van der Flier WM, Scheltens P, Pijnenburg YA: **Early-versus late-onset Alzheimer's disease: more than age alone**. *Journal of Alzheimer's Disease* 2010, **19**(4):1401-1408.

42. van der Flier WM, Pijnenburg YA, Fox NC, Scheltens P: **Early-onset versus late-onset Alzheimer's disease: the case of the missing APOE ε4 allele**. *The Lancet Neurology* 2011, **10**(3):280-288.

43. Palasí A, Gutiérrez-Iglesias B, Alegret M, Pujadas F, Olabarrieta M, Liébana D, Quintana M, Álvarez-Sabín J, Boada M: **Differentiated clinical presentation of early and late-onset Alzheimer's disease: is 65 years of age providing a reliable threshold?** *Journal of neurology* 2015, **262**(5):1238-1246.

44. Tulving E: **Episodic memory: From mind to brain**. *Annual review of psychology* 2002, **53**(1):1-25.

45. Tulving E: **Episodic and semantic memory**. *Organization of memory* 1972, **1**:381-403.

46. Yee E, Jones MN, McRae K: **Semantic memory**. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* 2018, **3**:1-38.

47. Thompson RF, Kim JJ: **Memory systems in the brain and localization of a memory**. *Proceedings of the national academy of sciences* 1996, **93**(24):13438-13444.

48. Schacter DL, Tulving E: **Memory systems**. In.: Cambridge, MA: MIT Press; 1994.

49. Hodges JR, Patterson K: **Is semantic memory consistently impaired early in the course of Alzheimer's disease? Neuroanatomical and diagnostic implications**. *Neuropsychologia* 1995, **33**(4):441-459.

50. Nebes RD: **Semantic memory in Alzheimer's disease**. *Psychological bulletin* 1989, **106**(3):377.

51. Bäckman L, Jones S, Berger AK, Laukka EJ, Small B: **Multiple cognitive deficits during the transition to Alzheimer's disease**. *Journal of internal medicine* 2004, **256**(3):195-204.

52. Hodges JR: **Memory in the dementias**. *The Oxford handbook of memory* 2000:441-459.

53. Weintraub S, Wicklund AH, Salmon DP: **The neuropsychological profile of Alzheimer disease**. *Cold Spring Harbor perspectives in medicine* 2012, **2**(4):a006171.

54. Braak H, Braak E: **Neuropathological stageing of Alzheimer-related changes**. *Acta neuropathologica* 1991, **82**(4):239-259.

55. Eichenbaum H: **Prefrontal–hippocampal interactions in episodic memory**. *Nature Reviews Neuroscience* 2017, **18**(9):547.

56. Budson AE, Price BH: **Memory dysfunction**. *New England Journal of Medicine* 2005, **352**(7):692-699.

57. Moscovitch M, Cabeza R, Winocur G, Nadel L: **Episodic memory and beyond: the hippocampus and neocortex in transformation**. *Annual review of psychology* 2016, **67**:105-134.

58. Rogers SL, Friedman RB: **The underlying mechanisms of semantic memory loss in Alzheimer's disease and semantic dementia**. *Neuropsychologia* 2008, **46**(1):12-21.

59. Grilli MD, Wank AA, Bercel JJ, Ryan L: **Evidence for reduced autobiographical memory episodic specificity in cognitively normal middle-aged and older individuals at increased risk for Alzheimer's disease dementia**. *Journal of the International Neuropsychological Society* 2018, **24**(10):1073-1083.

60. Kirk M, Berntsen D: **The life span distribution of autobiographical memory in Alzheimer's disease**. *Neuropsychology* 2018, **32**(8):906.

61. George DR: **Overcoming the social death of dementia through language**. *The Lancet* 2010, **376**(9741):586-587.

62. Strikwerda-Brown C, Grilli MD, Andrews-Hanna J, Irish M: **"All is not lost"– Rethinking the nature of memory and the self in dementia**. *Ageing research reviews* 2019:100932.

63. Beaty RE, Schacter DL: **14 Episodic Memory and Cognitive Control: Contributions to Creative Idea Production**. *The Cambridge handbook of the neuroscience of creativity* 2018:249.

64. Baraly KTA, Hot P, Davidson PS, Talmi D: **How emotional arousal enhances episodic memory**. *Learning and memory: A comprehensive reference* 2017:295-324.

65. [http://thebrain.mcgill.ca/flash/capsules/histoire_jaune03.html]

66. Alzheimer A: **Über eigenartige Krankheitsfälle des späteren Alters**. *Zeitschrift für die gesamte Neurologie und Psychiatrie* 1911, **4**(1):356-385.

67. Alzheimer A: **Über eigenartige Krankheitsfälle des späteren Alters: (On certain peculiar diseases of old age**. *History of Psychiatry* 1991, **2**(5):74-101.

68. Grundke-Iqbal I, Iqbal K, Tung Y-C, Quinlan M, Wisniewski HM, Binder LI: **Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology**. *Proceedings of the National Academy of Sciences* 1986, **83**(13):4913-4917.

69. KoSIK KS, Joachim CL, Selkoe DJ: **Microtubule-associated protein tau (tau) is a major antigenic component of paired helical filaments in Alzheimer disease**. *Proceedings of the National Academy of Sciences* 1986, **83**(11):4044-4048.

70. Brion J-P: **Neurofibrillary tangles and Alzheimer's disease**. *European neurology* 1998, **40**(3):130-140.

71. Binder LI, Guillozet-Bongaarts AL, Garcia-Sierra F, Berry RW: **Tau, tangles, and Alzheimer's disease**. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 2005, **1739**(2-3):216-223.

72. Braak H, Braak E: **Diagnostic criteria for neuropathologic assessment of Alzheimer's disease**. *Neurobiology of aging* 1997, **18**(4):S85-S88.

73. Braak H, Braak E: **On areas of transition between entorhinal allocortex and temporal isocortex in the human brain. Normal morphology and lamina-specific pathology in Alzheimer's disease**. *Acta neuropathologica* 1985, **68**(4):325-332.

74. Lindsay J, Laurin D, Verreault R, Hébert R, Helliwell B, Hill GB, McDowell I: **Risk factors for Alzheimer's disease: a prospective analysis from the Canadian Study of Health and Aging**. *American journal of epidemiology* 2002, **156**(5):445-453.

75. Tyas SL, Manfreda J, Strain LA, Montgomery PR: **Risk factors for Alzheimer's disease: a population-based, longitudinal study in Manitoba, Canada**. *International journal of epidemiology* 2001, **30**(3):590-597.

76. Association As: **2015 Alzheimer's disease facts and figures**. *Alzheimer's & Dementia* 2015, **11**(3):332-384.

77. Kawas C, Gray S, Brookmeyer R, Fozard J, Zonderman A: **Age-specific incidence rates of Alzheimer's disease: the Baltimore Longitudinal Study of Aging**. *Neurology* 2000, **54**(11):2072-2077.

78. Koivisto K, Reinikainen K, Hanninen T, Vanhanen M, Helkala E, Mykkanen L, Laakso M, Pyorala K, Riekkinen P: **Prevalence of age-associated memory impairment in a randomly selected population from eastern Finland**. *Neurology* 1995, **45**(4):741-747.

79. Morris JC, Roe CM, Xiong C, Fagan AM, Goate AM, Holtzman DM, Mintun MA: **APOE predicts amyloid-beta but not tau Alzheimer pathology in cognitively normal aging**. *Annals of neurology* 2010, **67**(1):122-131.

80. Schaie KW: **Intellectual development in adulthood: The Seattle longitudinal study**: Cambridge University Press; 1996.

81. Mattson MP, Arumugam TV: **Hallmarks of brain aging: adaptive and pathological modification by metabolic states**. *Cell metabolism* 2018, **27**(6):1176-1199.

82. Sliwinski M, Lipton RB, Buschke H, Stewart W: **The effects of preclinical dementia on estimates of normal cognitive functioning in aging**. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 1996, **51**(4):P217-P225.

83. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack Jr CR, Kaye J, Montine TJ: **Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease**. *Alzheimer's & dementia* 2011, **7**(3):280-292.

84. Wegmann S, Bennett RE, Delorme L, Robbins AB, Hu M, McKenzie D, Kirk MJ, Schiantarelli J, Tunio N, Amaral AC: **Experimental evidence for the age dependence of tau protein spread in the brain**. *Science advances* 2019, **5**(6):eaaw6404.

85. Herrup K: **Reimagining Alzheimer's disease—an age-based hypothesis**. *Journal of Neuroscience* 2010, **30**(50):16755-16762.

86. Crick F: **Central dogma of molecular biology**. *Nature* 1970, **227**(5258):561-563.

87. Crick FH: **On protein synthesis**. In: *Symp Soc Exp Biol: 1958*; 1958: 8.

88. Bird A: **Perceptions of epigenetics**. *Nature* 2007, **447**(7143):396.

89. Nativio R, Donahue G, Berson A, Lan Y, Amlie-Wolf A, Tuzer F, Toledo JB, Gosai SJ, Gregory BD, Torres C: **Dysregulation of the epigenetic landscape of normal aging in Alzheimer's disease**. *Nature neuroscience* 2018, **21**(4):497-505.

90. Koga H, Kaushik S, Cuervo AM: **Protein homeostasis and aging: The importance of exquisite quality control**. *Ageing research reviews* 2011, **10**(2):205-215.

91. Elias MF, D'Agostino RB, Elias PK, Wolf PA: **Neuropsychological test performance, cognitive functioning, blood pressure, and age: the Framingham Heart Study**. *Experimental aging research* 1995, **21**(4):369-391.

92. Chuang Y-F, An Y, Bilgel M, Wong DF, Troncoso JC, O'Brien RJ, Breitner J, Ferruci L, Resnick SM, Thambisetty M: **Midlife adiposity predicts earlier onset of Alzheimer's dementia, neuropathology and presymptomatic cerebral amyloid accumulation**. *Molecular psychiatry* 2016, **21**(7):910-915.

93. Hassing LB, Dahl AK, Thorvaldsson V, Berg S, Gatz M, Pedersen NL, Johansson B: **Overweight in midlife and risk of dementia: a 40-year follow-up study**. *International journal of obesity* 2009, **33**(8):893-898.

94. Kwok C, Loke Y, Hale R, Potter J, Myint P: **Atrial fibrillation and incidence of dementia: a systematic review and meta-analysis**. *Neurology* 2011, **76**(10):914-922.

95. Laitinen M, Ngandu T, Rovio S, Helkala E-L, Uusitalo U, Viitanen M, Nissinen A, Tuomilehto J, Soininen H, Kivipelto M: **Fat intake at midlife and risk of dementia and Alzheimer's disease: a population-based study**. *Dementia and geriatric cognitive disorders* 2006, **22**(1):99-107.

96. Gillette-Guyonnet S, Van Kan GA, Andrieu S, Barberger-Gateau P, Berr C, Bonnefoy M, Dartigues J, De Groot L, Ferry M, Galan P: **IANA task force on nutrition and cognitive decline with aging**. *Journal of Nutrition Health and Aging* 2007, **11**(2):132.

97. Eskelinen MH, Ngandu T, Tuomilehto J, Soininen H, Kivipelto M: **Midlife healthy-diet index and late-life dementia and Alzheimer's disease**. *Dementia and geriatric cognitive disorders extra* 2011, **1**(1):103-112.

98. Wisniewski K, Wisniewski H, Wen G: **Occurrence of Alzheimer's neuropathology and dementia in Down syndrome**. *Annals of Neurology* 1985, **17**(3):278-282.

99. Haass C, Schlossmacher MG, Hung AY, Vigo-Pelfrey C, Mellon A, Ostaszewski BL, Lieberburg I, Koo EH, Schenk D, Teplow DB: **Amyloid β-peptide is produced by cultured cells during normal metabolism**. *Nature* 1992, **359**(6393):322-325.

100. Busciglio J, Gabuzda DH, Matsudaira P, Yankner BA: **Generation of beta-amyloid in the secretory pathway in neuronal and nonneuronal cells**. *Proceedings of the National Academy of Sciences* 1993, **90**(5):2092-2096.

101. Li Y-M, Xu M, Lai M-T, Huang Q, Castro JL, DiMuzio-Mower J, Harrison T, Lellis C, Nadin A, Neduvelil JG: **Photoactivated γ-secretase inhibitors directed to the active site covalently label presenilin 1**. *Nature* 2000, **405**(6787):689-694.

102. De Strooper B, Iwatsubo T, Wolfe MS: **Presenilins and γ-secretase: structure, function, and role in Alzheimer disease**. *Cold Spring Harbor perspectives in medicine* 2012, **2**(1):a006304.

103. Tilley L, Morgan K, Kalsheker N: **Genetic risk factors in Alzheimer's disease**. *Molecular Pathology* 1998, **51**(6):293.

104. De Strooper B, Saftig P, Craessaerts K, Vanderstichele H, Guhde G, Annaert W, Von Figura K, Van Leuven F: **Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein**. *Nature* 1998, **391**(6665):387-390.

105. Xia D, Watanabe H, Wu B, Lee SH, Li Y, Tsvetkov E, Bolshakov VY, Shen J, Kelleher III RJ: **Presenilin-1 knockin mice reveal loss-of-function mechanism for familial Alzheimer's disease**. *Neuron* 2015, **85**(5):967-981.

106. Saura CA, Choi S-Y, Beglopoulos V, Malkani S, Zhang D, Rao BS, Chattarji S, Kelleher III RJ, Kandel ER, Duff K: **Loss of presenilin function causes impairments of memory and synaptic plasticity followed by age-dependent neurodegeneration**. *Neuron* 2004, **42**(1):23-36.

107. Dermaut B, Kumar-Singh S, Engelborghs S, Theuns J, Rademakers R, Saerens J, Pickut BA, Peeters K, Van Den Broeck M, Vennekens Kl: **A novel presenilin 1 mutation associated with Pick's disease but not β-amyloid plaques**. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 2004, **55**(5):617-626.

108. Ryman DC, Acosta-Baena N, Aisen PS, Bird T, Danek A, Fox NC, Goate A, Frommelt P, Ghetti B, Langbaum JB: **Symptom onset in autosomal dominant Alzheimer disease: a systematic review and meta-analysis**. *Neurology* 2014, **83**(3):253-260.

109. Hirschhorn JN, Daly MJ: **Genome-wide association studies for common diseases and complex traits**. *Nature reviews genetics* 2005, **6**(2):95.

110. Li A, Meyre D: **Challenges in reproducibility of genetic association studies: lessons learned from the obesity field**. *International journal of obesity* 2013, **37**(4):559-567.

111. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, Boland A, Vronskaya M, Van Der Lee SJ, Amlie-Wolf A: **Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing**. *Nature genetics* 2019, **51**(3):414-430.

112. Palmert M, Golde T, Cohen M, Kovacs D, Tanzi R, Gusella J, Usiak M, Younkin L, Younkin SG: **Amyloid protein precursor messenger RNAs: differential expression in Alzheimer's disease**. *Science* 1988, **241**(4869):1080-1084.

113. Lee CD, Daggett A, Gu X, Jiang L-L, Langfelder P, Li X, Wang N, Zhao Y, Park CS, Cooper Y: **Elevated TREM2 gene dosage reprograms microglia responsivity and ameliorates pathological phenotypes in Alzheimer's disease models**. *Neuron* 2018, **97**(5):1032-1048. e1035.

114. Griciuc A, Patel S, Federico AN, Choi SH, Innes BJ, Oram MK, Cereghetti G, McGinty D, Anselmo A, Sadreyev RI: **TREM2 acts downstream of CD33 in modulating microglial pathology in Alzheimer's disease**. *Neuron* 2019, **103**(5):820-835. e827.

115. Ricciarelli R, d'Abramo C, Massone S, Marinari UM, Pronzato MA, Tabaton M: **Microarray analysis in Alzheimer's disease and normal aging**. *IUBMB life* 2004, **56**(6):349-354.

116. Manczak M, Park BS, Jung Y, Reddy PH: **Differential expression of oxidative phosphorylation genes in patients with Alzheimer's disease**. *Neuromolecular medicine* 2004, **5**(2):147-162.

117. Berchtold NC, Cribbs DH, Coleman PD, Rogers J, Head E, Kim R, Beach T, Miller C, Troncoso J, Trojanowski JQ: **Gene expression changes in the course of normal brain aging are sexually dimorphic**. *Proceedings of the National Academy of Sciences* 2008, **105**(40):15605-15610.

118. Liang WS, Reiman EM, Valla J, Dunckley T, Beach TG, Grover A, Niedzielko TL, Schneider LE, Mastroeni D, Caselli R: **Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons**. *Proceedings of the National Academy of Sciences* 2008, **105**(11):4441-4446.

119. Alexander GE, Chen K, Pietrini P, Rapoport SI, Reiman EM: **Longitudinal PET evaluation of cerebral metabolic decline in dementia: a potential outcome measure in Alzheimer's disease treatment studies**. *American Journal of Psychiatry* 2002, **159**(5):738-745.

120. Piert M, Koeppe RA, Giordani B, Berent S, Kuhl DE: **Diminished glucose transport andphosphorylation in Alzheimer's disease determined by dynamic FDG-PET**. *J Nucl Med* 1996, **37**:201-208.

121. Westman G, Blomberg J, Yun Z, Lannfelt L, Ingelsson M, Eriksson B-M: **Decreased HHV-6 IgG in Alzheimer's disease**. *Frontiers in neurology* 2017, **8**:40.

122. Itzhaki RF: **Herpes simplex virus type 1 and Alzheimer's disease: increasing evidence for a major role of the virus**. *Frontiers in aging neuroscience* 2014, **6**:202.

123. Readhead B, Haure-Mirande J-V, Funk CC, Richards MA, Shannon P, Haroutunian V, Sano M, Liang WS, Beckmann ND, Price ND: **Multiscale analysis of independent Alzheimer's cohorts finds disruption of molecular, genetic, and clinical networks by human herpesvirus**. *Neuron* 2018, **99**(1):64-82. e67.

124. Caselli RJ, Reiman EM: **Characterizing the preclinical stages of Alzheimer's disease and the prospect of presymptomatic intervention**. *Journal of Alzheimer's Disease* 2013, **33**(s1):S405-S416.

125. Somogyi R, Sniegoski CA: **Modeling the complexity of genetic networks: understanding multigenic and pleiotropic regulation**. *complexity* 1996, **1**(6):45-63.

126. Chen T, He HL, Church GM: **Modeling gene expression with differential equations**. In: *Biocomputing'99.* edn.: World Scientific; 1999: 29-40.

127. Zak DE, Doyle F, Schwaber JS: **Local identifiability: when can genetic networks be identified from microarray data**. In: *Proceedings of the third international conference on systems biology: 2002*; 2002: 236-237.

128. Husmeier D: **Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks**. *Bioinformatics* 2003, **19**(17):2271-2282.

129. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns**. *Proceedings of the National Academy of Sciences* 1998, **95**(25):14863-14868.

130. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: **Gene ontology: tool for the unification of biology**. *Nature genetics* 2000, **25**(1):25-29.

131. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: **Systematic determination of genetic network architecture**. *Nature genetics* 1999, **22**(3):281-285.

132. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: **Systematic determination of genetic network architecture**. *Nature genetics* 1999, **22**(3):281.

133. Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC: **DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists**. *Nucleic acids research* 2007, **35**(suppl_2):W169-W175.

134. Gruca A, Sikora M, Polanski A: **RuleGO: a logical rules-based tool for description of gene groups by means of Gene Ontology**. *Nucleic acids research* 2011, **39**(suppl_2):W293-W301.

135. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E: **PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes**. *Nature genetics* 2003, **34**(3):267-273.

136. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles**. *Proceedings of the National Academy of Sciences* 2005, **102**(43):15545-15550.

137. Rahnenführer J, Domingues FS, Maydt J, Lengauer T: **Calculating the statistical significance of changes in pathway activity from gene expression data**. *Statistical applications in genetics and molecular biology* 2004, **3**(1).

138. Jeanmougin M, De Reynies A, Marisa L, Paccard C, Nuel G, Guedj M: **Should we abandon the t-test in the analysis of gene expression microarray data: a comparison of variance modeling strategies**. *PloS one* 2010, **5**(9).

139. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing**. *Journal of the Royal statistical society: series B (Methodological)* 1995, **57**(1):289-300.

140. Zhang B, Kirov S, Snoddy J: **WebGestalt: an integrated system for exploring gene sets in various biological contexts**. *Nucleic acids research* 2005, **33**(suppl_2):W741-W748.

141. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B: **WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs**. *Nucleic acids research* 2019, **47**(W1):W199-W205.

142. Davis S, Meltzer PS: **GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor**. *Bioinformatics* 2007, **23**(14):1846-1847.

143. Rodrigue KM, Kennedy KM, Park DC: **Beta-amyloid deposition and the aging brain**. *Neuropsychology review* 2009, **19**(4):436.

144. on Aging TNI, on Diagnostic RIWG, Braak H, Coleman P, Dickson D, Duyckaerts C, De La Salpetriere H, Paris F, Gambetti P, of Cleveland UH: **Consensus recommendations for the postmortem diagnosis of Alzheimer's disease**. *Neurobiology of aging* 1997, **18**(4):S1-S2.

145. Wang L, Liu Z-P: **Detecting Diagnostic Biomarkers of Alzheimer's Disease by Integrating Gene Expression Data in Six Brain Regions**. *Frontiers in genetics* 2019, **10**:157-157.

146. Pačínková A, Popovici V: **Cross-platform Data Analysis Reveals a Generic Gene Expression Signature for Microsatellite Instability in Colorectal Cancer**. *BioMed research international* 2019, **2019**.

147. Zhang X-F, Ou-Yang L, Zhao X-M, Yan H: **Differential network analysis from cross-platform gene expression data**. *Scientific reports* 2016, **6**(1):1-12.

148. Naret O, Chaturvedi N, Bartha I, Hammer C, Fellay J, Study SHC: **Correcting for population stratification reduces false positive and false negative results in joint analyses of host and pathogen genomes**. *Frontiers in genetics* 2018, **9**:266.

149. Piatetsky-Shapiro G, Tamayo P: **Microarray data mining: facing the challenges**. *ACM SIGKDD Explorations Newsletter* 2003, **5**(2):1-5.

150. Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources**. *Nature protocols* 2009, **4**(1):44.

151. Huang DW, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists**. *Nucleic acids research* 2009, **37**(1):1-13.

152. Bolstad BM, Irizarry RA, Åstrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias**. *Bioinformatics* 2003, **19**(2):185-193.

153. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, normalization, and summaries of high density oligonucleotide array probe level data**. *Biostatistics* 2003, **4**(2):249-264.

154. Pan M, Zhang J: **Quantile normalization for combining gene-expression datasets**. *Biotechnology & Biotechnological Equipment* 2018, **32**(3):751-758.

155. Müller C, Schillert A, Röthemeier C, Trégouët D-A, Proust C, Binder H, Pfeiffer N, Beutel M, Lackner KJ, Schnabel RB: **Removing batch effects from longitudinal gene expression-quantile normalization plus combat as best approach for microarray transcriptome data**. *PloS one* 2016, **11**(6).

156. Hu J, He X: **Enhanced quantile normalization of microarray data to reduce loss of information in gene expression profiles**. *Biometrics* 2007, **63**(1):50-59.

157. Hicks SC, Irizarry RA: **Quantro: a data-driven approach to guide the choice of an appropriate normalization method**. *Genome biology* 2015, **16**(1):117.

158. Liang WS, Dunckley T, Beach TG, Grover A, Mastroeni D, Ramsey K, Caselli RJ, Kukull WA, McKeel D, Morris JC: **Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set**. *Physiological genomics* 2008, **33**(2):240-256.

159. Oughtred R, Stark C, Breitkreutz B-J, Rust J, Boucher L, Chang C, Kolas N, O'Donnell L, Leung G, McAdam R: **The BioGRID interaction database: 2019 update**. *Nucleic acids research* 2019, **47**(D1):D529-D541.

160. Lee M-H, Siddoway B, Kaeser GE, Segota I, Rivera R, Romanow WJ, Liu CS, Park C, Kennedy G, Long T: **Somatic APP gene recombination in Alzheimer's disease and normal neurons**. *Nature* 2018, **563**(7733):639-645.

161. Ricciotti E, FitzGerald GA: **Prostaglandins and inflammation**. *Arteriosclerosis, thrombosis, and vascular biology* 2011, **31**(5):986-1000.

162. Ma SL, Tang NLS, Zhang YP, Ji L-d, Tam CWC, Lui VWC, Chiu HFK, Lam LCW: **Association of prostaglandin-endoperoxide synthase 2 (PTGS2) polymorphisms and Alzheimer's disease in Chinese**. *Neurobiology of aging* 2008, **29**(6):856-860.

163. Chen Q, Liang B, Wang Z, Cheng X, Huang Y, Liu Y, Huang Z: **Influence of four polymorphisms in ABCA1 and PTGS2 genes on risk of Alzheimer's disease: a meta-analysis**. *Neurological Sciences* 2016, **37**(8):1209-1220.

164. Michele S, Salluzzo MG, Calogero AE, Raffaele F, Bosco P: **Association study of COX-2 (PTGS2)-765 G/C promoter polymorphism by pyrosequencing in**

Sicilian patients with Alzheimer's disease. *Archives of medical science: AMS* 2014, **10**(6):1235-1238.

165. Etminan M, Gill S, Samii A: **Effect of non-steroidal anti-inflammatory drugs on risk of Alzheimer's disease: systematic review and meta-analysis of observational studies**. *Bmj* 2003, **327**(7407):128.

166. Launer L, Breteler M, Stricker B, Hofman A: **Pharmacologic agents associated with a preventive effect on Alzheimer's disease: a review of the epidemiologic evidence**. *Epidemiologic reviews* 2002.

167. Reines S, Block G, Morris J, Liu G, Nessly M, Lines C, Norman B, Baranak C: **Rofecoxib: no effect on Alzheimer's disease in a 1-year, randomized, blinded, controlled study**. *Neurology* 2004, **62**(1):66-71.

168. Aisen PS: **The potential of anti-inflammatory drugs for the treatment of Alzheimer's disease**. *The Lancet Neurology* 2002, **1**(5):279-284.

169. Rubio-Perez JM, Morillas-Ruiz JM: **A review: inflammatory process in Alzheimer's disease, role of cytokines**. *The Scientific World Journal* 2012, **2012**.

170. Carpenter S, Carlson T, Dellacasagrande J, Garcia A, Gibbons S, Hertzog P, Lyons A, Lin L-L, Lynch M, Monie T: **TRIL, a functional component of the TLR4 signaling complex, highly expressed in brain**. *The Journal of Immunology* 2009, **183**(6):3989-3995.

171. Zhang X, Feng H, Li Z, Li D, Liu S, Huang H, Li M: **Application of weighted gene co-expression network analysis to identify key modules and hub genes in oral squamous cell carcinoma tumorigenesis**. *OncoTargets and therapy* 2018, **11**:6001.

172. Morgan JI, Curran T: **Stimulus-transcription coupling in neurons: role of cellular immediate-early genes**. *Trends in neurosciences* 1989, **12**(11):459-462.

173. Loebrich S, Nedivi E: **The function of activity-regulated genes in the nervous system**. *Physiological reviews* 2009, **89**(4):1079-1103.

174. Alberini CM: **Transcription factors in long-term memory and synaptic plasticity**. *Physiological reviews* 2009, **89**(1):121-145.

175. Lin Y, Bloodgood BL, Hauser JL, Lapan AD, Koon AC, Kim T-K, Hu LS, Malik AN, Greenberg ME: **Activity-dependent regulation of inhibitory synapse development by Npas4**. *Nature* 2008, **455**(7217):1198-1204.

176. Opsomer R, Contino S, Perrin F, Gualdani R, Tasiaux B, Doyen P, Vergouts M, Vrancx C, Doshina A, Pierrot N: **Amyloid Precursor Protein (APP) controls the expression of the transcriptional activator Neuronal PAS Domain Protein 4 (NPAS4) and synaptic GABA release**. *Eneuro* 2020.

177. Opsomer R, Contino S, Perrin F, Tasiaux B, Doyen P, Vergouts M, Vrancx C, Doshina A, Pierrot N, Octave J-N: **Amyloid Precursor Protein (APP) controls excitatory/inhibitory synaptic inputs by regulating the transcriptional activator Neuronal PAS Domain Protein 4 (NPAS4)**. *bioRxiv* 2018:504340.

178. Drutel G, Kathmann M, Heron A, Schwartz J-C, Arrang J-M: **Cloning and selective expression in brain and kidney of ARNT2 homologous to the Ah receptor nuclear translocator (ARNT)**. *Biochemical and biophysical research communications* 1996, **225**(2):333-339.

179. Ooe N, Saito K, Mikami N, Nakatuka I, Kaneko H: **Identification of a novel basic helix-loop-helix-PAS factor, NXF, reveals a Sim2 competitive, positive regulatory role in dendritic-cytoskeleton modulator drebrin gene expression**. *Molecular and cellular biology* 2004, **24**(2):608-616.

180. Bersten DC, Bruning JB, Peet DJ, Whitelaw ML: **Human variants in the neuronal basic helix-loop-helix/Per-Arnt-Sim (bHLH/PAS) transcription factor complex NPAS4/ARNT2 disrupt function**. *PloS one* 2014, **9**(1):e85768.

181. Zhang S-J, Zou M, Lu L, Lau D, Ditzel DA, Delucinge-Vivier C, Aso Y, Descombes P, Bading H: **Nuclear calcium signaling controls expression of a large gene pool: identification of a gene program for acquired neuroprotection induced by synaptic activity**. *PLoS Genet* 2009, **5**(8):e1000604.

182. Kovalchuk A, Ilnytskyy Y, Rodriguez-Juarez R, Katz A, Sidransky D, Kolb B, Kovalchuk O: **Growth of triple negative and progesterone positive breast cancer causes oxidative stress and down-regulates neuroprotective transcription factor Npas4 and Npas4-regulated genes in hippocampal tissues of tumorgraft mice—an aging connection**. *Frontiers in genetics* 2018, **9**:58.

183. Qiu J, Dunbar DR, Noble J, Cairns C, Carter R, Kelly V, Chapman KE, Seckl JR, Yau JL: **Decreased Npas4 and Arc mRNA Levels in the Hippocampus of Aged Memory-Impaired Wild-Type But Not Memory Preserved 11β-HSD 1 Deficient Mice**. *Journal of neuroendocrinology* 2016, **28**(1).

184. West AE, Greenberg ME: **Neuronal activity–regulated gene transcription in synapse development and cognitive function**. *Cold Spring Harbor perspectives in biology* 2011, **3**(6):a005744.

185. Koo EH, Squazzo SL: **Evidence that production and release of amyloid beta-protein involves the endocytic pathway**. *Journal of Biological Chemistry* 1994, **269**(26):17386-17389.

186. Thinakaran G, Koo EH: **Amyloid precursor protein trafficking, processing, and function**. *Journal of Biological Chemistry* 2008, **283**(44):29615-29619.

187. Chen X, Wang C, Zhou S, Li X, Wu L: **The impact of EGFR gene polymorphisms on the risk of Alzheimer's disease in a Chinese han population: a case-controlled study**. *Medical science monitor: international medical journal of experimental and clinical research* 2018, **24**:5035.

188. Lazarov O, Marr RA: **Neurogenesis and Alzheimer's disease: at the crossroads**. *Experimental neurology* 2010, **223**(2):267-281.

189. Méndez-López I, Blanco-Luquin I, Sánchez-Ruiz de Gordoa J, Urdánoz-Casado A, Roldán M, Acha B, Echavarri C, Zelaya V, Jericó I, Mendioroz M:

Hippocampal LMNA Gene Expression is Increased in Late-Stage Alzheimer's Disease. *International journal of molecular sciences* 2019, **20**(4):878.

190. Frost B: **Alzheimer's disease: An acquired neurodegenerative laminopathy**. *Nucleus* 2016, **7**(3):275-283.

191. McGaugh JL: **Memory--a century of consolidation**. *Science* 2000, **287**(5451):248-251.

192. Gallo FT, Katche C, Morici JF, Medina JH, Weisstaub NV: **Immediate early genes, memory and psychiatric disorders: focus on c-Fos, Egr1 and Arc**. *Frontiers in behavioral neuroscience* 2018, **12**:79.

193. Sara SJ: **Retrieval and reconsolidation: toward a neurobiology of remembering**. *Learning & Memory* 2000, **7**(2):73-84.

194. von Hertzen LS, Giese KP: **Memory reconsolidation engages only a subset of immediate-early genes induced during consolidation**. *Journal of Neuroscience* 2005, **25**(8):1935-1942.

195. Bozon B, Davis S, Laroche S: **A requirement for the immediate early gene zif268 in reconsolidation of recognition memory after retrieval**. *Neuron* 2003, **40**(4):695-701.

196. Kim S, Kim H, Um JW: **Synapse development organized by neuronal activity-regulated immediate-early genes**. *Experimental & molecular medicine* 2018, **50**(4):1-7.

197. Hayashi Y, Kashiwagi K, Ohta J, Nakajima M, Kawashima T, Yoshikawa K: **Alzheimer amyloid protein precursor enhances proliferation of neural stem cells from fetal rat brain**. *Biochemical and biophysical research communications* 1994, **205**(1):936-943.

198. Wang H, Megill A, He K, Kirkwood A, Lee H-K: **Consequences of inhibiting amyloid precursor protein processing enzymes on synaptic function and plasticity**. *Neural plasticity* 2012, **2012**.

199. Von Rotz RC, Kohli BM, Bosset J, Meier M, Suzuki T, Nitsch RM, Konietzko U: **The APP intracellular domain forms nuclear multiprotein complexes and regulates the transcription of its own precursor**. *Journal of cell science* 2004, **117**(19):4435-4448.

200. Wang M, Wang S, Li Y, Cai G, Cao M, Li L: **Integrated analysis and network pharmacology approaches to explore key genes of Xingnaojing for treatment of Alzheimer's disease**. *Brain and Behavior* 2020:e01610.

201. Kelly J, Moyeed R, Carroll C, Luo S, Li X: **Genetic networks in Parkinson's and Alzheimer's disease**. *Aging (Albany NY)* 2020, **12**(6):5221.

202. Antonell A, Lladó A, Altirriba J, Botta-Orfila T, Balasa M, Fernández M, Ferrer I, Sánchez-Valle R, Molinuevo JL: **A preliminary study of the whole-genome expression profile of sporadic and monogenic early-onset Alzheimer's disease**. *Neurobiology of aging* 2013, **34**(7):1772-1778.

203. Chi L-M, Wang X, Nan G-X: **In silico analyses for molecular genetic mechanism and candidate genes in patients with Alzheimer's disease**. *Acta Neurologica Belgica* 2016, **116**(4):543-547.

204. Satoh J-i, Yamamoto Y, Asahina N, Kitano S, Kino Y: **RNA-Seq data mining: downregulation of NeuroD6 serves as a possible biomarker for alzheimer's disease brains**. *Disease markers* 2014, **2014**.

205. Pettegrew JW, Panchalingam K, Hamilton RL, McClure RJ: **Brain membrane phospholipid alterations in Alzheimer's disease**. *Neurochemical research* 2001, **26**(7):771-782.

206. Jope RS, Song L, Powers RE: **Cholinergic activation of phosphoinositide signaling is impaired in Alzheimer's disease brain**. *Neurobiology of aging* 1997, **18**(1):111-120.

207. Elvekrog MM, Walter P: **Dynamics of co-translational protein targeting**. *Current opinion in chemical biology* 2015, **29**:79-86.

208. Tao Y, Han Y, Yu L, Wang Q, Leng SX, Zhang H: **The predicted key molecules, functions, and pathways that bridge mild cognitive impairment (MCI) and Alzheimer's disease (AD)**. *Frontiers in Neurology* 2020, **11**:233.

209. Lefterov I, Wolfe CM, Fitz NF, Nam KN, Letronne F, Biedrzycki RJ, Kofler J, Han X, Wang J, Schug J: **APOE2 orchestrated differences in transcriptomic and lipidomic profiles of postmortem AD brain**. *Alzheimer's research & therapy* 2019, **11**(1):113.

210. Mookherjee D, Majumder P, Mukherjee R, Chatterjee D, Kaul Z, Das S, Sougrat R, Chakrabarti S, Chakrabarti O: **Cytosolic aggregates in presence of non-translocated proteins perturb endoplasmic reticulum structure and dynamics**. *Traffic* 2019, **20**(12):943-960.

211. Patel SB, Bellini M: **The assembly of a spliceosomal small nuclear ribonucleoprotein particle**. *Nucleic acids research* 2008, **36**(20):6482-6493.

212. Bai B, Hales CM, Chen P-C, Gozal Y, Dammer EB, Fritz JJ, Wang X, Xia Q, Duong DM, Street C: **U1 small nuclear ribonucleoprotein complex and RNA splicing alterations in Alzheimer's disease**. *Proceedings of the National Academy of Sciences* 2013, **110**(41):16562-16567.

213. Nuzzo D, Inguglia L, Walters J, Picone P, Di Carlo M: **A shotgun proteomics approach reveals a new toxic role for Alzheimer's disease Aβ peptide: spliceosome impairment**. *Journal of proteome research* 2017, **16**(4):1526-1541.

214. Bai B: **U1 snRNP alteration and neuronal cell cycle reentry in Alzheimer disease**. *Frontiers in Aging Neuroscience* 2018, **10**:75.

215. McShea A, Lee H-g, Petersen RB, Casadesus G, Vincent I, Linford NJ, Funk J-O, Shapiro RA, Smith MA: **Neuronal cell cycle re-entry mediates Alzheimer disease-type changes**. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 2007, **1772**(4):467-472.

216. Hsieh Y-C, Guo C, Yalamanchili HK, Abreha M, Al-Ouran R, Li Y, Dammer EB, Lah JJ, Levey AI, Bennett DA: **Tau-Mediated Disruption of the Spliceosome Triggers Cryptic RNA Splicing and Neurodegeneration in Alzheimer's Disease**. *Cell reports* 2019, **29**(2):301-316. e310.

217. Wu Y, Yao Q, Jiang G-X, Wang G, Cheng Q: **Identification of distinct blood-based biomarkers in early stage of Parkinson's disease**. *Neurological Sciences* 2019:1-9.

218. Kakati T, Kashyap H, Bhattacharyya DK: **THD-module extractor: an application for CEN module extraction and interesting gene identification for Alzheimer's disease**. *Scientific Reports* 2016, **6**(1):1-11.

219. Salker MS, Nautiyal J, Steel JH, Webster Z, Šućurović S, Nicou M, Singh Y, Lucas ES, Murakami K, Chan Y-W: **Disordered IL-33/ST2 activation in decidualizing stromal cells prolongs uterine receptivity in women with recurrent pregnancy loss**. *PloS one* 2012, **7**(12):e52252.

220. Barichello T, Fagundes GD, Generoso JS, Moreira AP, Costa CS, Zanatta JR, Simões LR, Petronilho F, Dal-Pizzol F, Vilela MC: **Brain–blood barrier breakdown and pro-inflammatory mediators in neonate rats submitted meningitis by Streptococcus pneumoniae**. *Brain research* 2012, **1471**:162-168.

221. Zyla J, Marczyk M, Weiner J, Polanska J: **Ranking metrics in gene set enrichment analysis: do they matter?** *BMC bioinformatics* 2017, **18**(1):256.

222. Roder J, Linstid B, Oliveira C: **Improving the power of gene set enrichment analyses**. *BMC bioinformatics* 2019, **20**(1):257.

223. Tarca AL, Bhatti G, Romero R: **A comparison of gene set analysis methods in terms of sensitivity, prioritization and specificity**. *PloS one* 2013, **8**(11):e79217.