ASHWANI KUMAR PhD (BIOINFORMATICS) OCT 2018

COMPUTATIONAL MINING OF GENOMIC AND PROTEOMIC DATA TO GAIN INSIGHT FOR ALZHEIMER'S DISEASE (AD)

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR

DOCTOR OF PHILOSOPHY
IN
BIOINFORMATICS

BY

ASHWANI KUMAR

Enrollment No. 136506

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT, SOLAN-173234, HP, INDIA OCT, 2018

COMPUTATIONAL MINING OF GENOMIC AND PROTEOMIC DATA TO GAIN INSIGHT FOR ALZHEIMER'S DISEASE (AD)

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN

BIOINFORMATICS

BY

ASHWANI KUMAR

Enrollment No. 136506

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT

OCT, 2018

Dedicated to My Family
and Almighty God

DECLARATION

I certify that:

a. The work contained in this thesis is original and has been done by me under the guidance of my supervisor.
b. The work has not been submitted to any other organisation for any degree or diploma.
c. Wherever, I have used materials (data, analysis, figures or text), I have given due credit by citing them in the text of the thesis.

Ashwani Kumar
(Enrollment No. 136506)
Department of Biotechnology and Bioinformatics
Jaypee University of Information Technology, Waknaghat, India

Date:

CERTIFICATE

This is to certify that the thesis entitled, "**Computational Mining of Genomic and Proteomic data to Gain Insight for Alzheimer's Disease (AD)**" which is being submitted by **Ashwani Kumar (Enrollment No. 136506)** in fulfillment for the award of degree of **Doctor of Philosophy** in **Bioinformatics** at **Jaypee University of Information Technology, India** is the record of candidate's own work carried out by her under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Dr. Tiratha Raj Singh
Associate Professor,
Department of Biotechnology and Bioinformatics
Jaypee University of Information Technology, Waknaghat, India
tiratharaj@gmail.com

Date:

ACKNOWLEDGEMENT

This Ph.D has been a lengthy and incredible trip. I am obliged to plenty of people I met on the way, and to those who were with me from the opening. Although this section is little bit long, it is certainly partial.

First and foremost, I repay my deepest gratitude to ***Prof. Vinod Kumar*** *(Vice Chancellor, JUIT);* ***Maj.Gen. Rakesh Bassi*** *(Registrar, JUIT);* ***Prof. S.D Gupta*** *(Dean, Academic and Research, JUIT);* ***Dr. Sudhir Kumar (***HOD, BT & BI***)****; for providing me with the opportunity to pursue my doctorate degree. I would like to sincerely thank* ***Prof. R.S. Chauhan*** *(former Dean and HOD, BT & BI) for his back-up and support thyroughout years and proving to be the prime example of how BT & BI department is held together.*

I would like to express my huge thanks to my supervisor ***Dr. Tiratha Raj Singh*** *for his trust in me, providing me with ideas, inspiration and criticism. I would like to especially thank him for helping me evolve to a skilled computational neurobiologist. His excellent guidance, persistent support in and understanding of complicated issues, his sheer expertise in solving major and minor problems and invaluable advice have made this thesis possible.*

I would like to extend a vote of thanks to the members of my thesis advisory committee; Prof. P.B. Barman, Dr. Harish Changotra, Dr. Raghu M. Yennamalli, I gratefully acknowledge their constant input during my Ph.D. I also want to appreciatete and thanks Dr. Udaybanu M and Vineet from department of Pharmacy for their constant support in fulfilling my Ph.D objectives.

I *would like to thank everyone in the BI Project Lab* ***Ankush***,***Ankita***, ***Kusum***, ***Nupur***, ***Imran and Smriti****. Thanks for all your support and encouragement. I am indebted to be surrounded by such great friends who helped me celebrate and survive all the ups and downs of my research journey. I would especially like to thank* ***Manika Ma'am, Priya Ma'am*** *and acknowledge the assistance of the* ***Somlata Ma'am***, ***teaching***, ***administration*** *and* ***server*** *staff who were always willing to help.*

Acknowledgement

I am deeply indebted to my parents for all their love, support, patience and encouragement throughout my life and entrusting me. I would like to thank my family members ***Manish Kumar, Ashish Kumar,Priyanka Rai, Shweta Singh*** *for supporting me in every way. Huge thanks to my grandfather* ***Late.Shiv Kedar*** *and Grand Mother* ***Late Smt. Ketki Devi.*** *To my little makoons* ***Saattwik ,Srianshi*** *and* ***Reyansh*** *for their smile was enough to light up my days. Their belief and pride in my achievements is something I have no words for. This is for them!*

My wife ***Anupama Anand*** *has been a great support all these years. Her strength, confidence, love and understanding have seen me safely through many of the challenges on my way. Undoubtedly, together with my family, she will be the most relieved person to know that it's finally concluding. She is certainly the recipient of the above and beyond awards because I would not be who I am without her.*

I Would like to thank my friends ***Sanjay****,* ***Ashutosh****,* ***Ankit****,* ***Shalu , Lalit*** *for being with me all the way. It means a lot to know that I can always count on them.*

I appreciate the contribution of countless people I met along the way for making this journey an experience which I will always keep dearly in my heart. I would like to thank College Gaurds and Mess workers for always making me feel home.

Last but not least, I thank the almighty God for bestowing upon me the strength and courage to embark upon this journey of life.

Ashwani Kumar

TABLE OF CONTENTS

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF FIGURES

Figure 4.1: Roadmap of systems biology approach for geneset enhancement and network properties examination of AD genes dataset.

Figure 4.2: **Regression plot between XD-score and fisher q-value**. Here dot represent pathway or process involved in enrichment examination as associated with target geneset of AD. Best linear fit are represented with red line in this plot. Significant threshold for XD-score are marked with horizontal grey line. The top most node in this plot denote AD- pathwayat XD-score= 5.062 and q-value=0.00.

Figure 4.3: Representation of target geneset of AD network environment. Node colors are: gray color for the gene used as target for AD, Blue color represent sandwich node between AD target genes and AD pathway specific genegreen node represent pathway specific genes and orange node represent key genes invole in some pathogenic process of AD. the edge colors are: green edge is used for interacting genes of AD target set with pathway, orange for interactions between the input set and also the growth and gray for others. More than three connected node are used to improve clarity.

Figure 4.4 (A) and 4 (B): Differentiate between properties involved in AD network The focused network are represented through red square. In both cases shortest path length are taken as standard and compare to Node betweenness and Degree respectively.

Figure 5.1(a-h): (a,b) 3D- structure of protein 1B41; (c-h)Active site and Interaction of disease modifying drugs with Acetylcholinesterase (AChE).

Figure 5.2: Protein-Molecule interactions. Normalized stacked bar graph illustration of connection and contacts through the span of mechanical phenomenon (values over 1.0 are attainable as some residue create various contacts of same subtype with atoms).The color bar indexes represent various interaction types, such as green for hydrogen bonds, light violet for hydrophobic interactions, pink for ionic interactions, and blue for water-bridges.

Figure 5.3: Inhibitory Concentration (IC50) value of molecules used in this study. Inhibitory activity of six molecules is represented through IC50 values plotted as dose-response curve. Percent errors are shown in vertical lines. The results shown are of triplicates. Since Begacestat has mM activity shown with y-axis on the right side of the figure, whereas the y- axis on the left side of the

figure is for the remainder of the compounds. Here, except Donepezil which has their activity in nM and remaining molecules has activity in μM.

Figure 5.4: Percent cell viability calculated through MTT-assay. Results are depicted as mean ± SD (n = 3). All the drugs were tested at three concentrations, viz. 1 μM, 10 μM and 25 μM.

Figure 5.5: Effect of drug treatment on HgCl2 induced neurotoxicity calculated through MTT-assay. Results are represented as mean ± SD (n = 3). Statistical significance was determined by one way ANOVA followed by Dunnett's multiple comparison post hoc test at * $p<0.05$, ** $p<0..01$ and *** $p<0.001$ (HgCl2 v/s treatments).

Figure 5.6 (A-D): (A): AO-EtBr (alive-dead) staining. Effect of various treatments on HgCl2 induced neuronal alterations in terms of AO and EtBr fluorescence (A); Total cell fluorescence of AO staining depicted as green fluorescence (B); Total cell fluorescence of EtBr staining depicted as red fluorescence (B, C, D): Ratio of AO:EtBr fluorescence intensity. Results are depicted as mean ± SD (n = 3). Statistical significance was determined by one way ANOVA followed by Dunnett's multiple comparison post hoc test at * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$ (HgCl2 v/s treatments).

Figure 6.1: Overall schema of AD research objectives and their proposed outcomes specifically for early Detection of disease.

LIST OF TABLES

Table 5.1: Properties of molecules used in Autodock 4.2. Important parameters obtained from Pubchem (molecular mass, hydrogen bond donor and acceptor, logarithmic P value and dissociation constant) are listed.

Table 5.2: Summary of the molecular docking studies and the in-vitro inhibition of AChE activity of molecules with AChE.

ABBREVIATIONS

WHO: World Health Organization

CBD: Chronic Brain Disorders

AD: Alzheimer's disease

Aβ: Amyloid Beta

NFT: Neurofibrillary Tangles

ADI: Alzheimer's disease International

ARDSI: Alzheimer related disorder society of India

APOE: Apolipoprotein E

SORL1: Sortilin Related Receptor 1

Ach: Acetylcholine

AChE: Acetylcholinesterase

BBB: Blood Brain Barrier NMDA:

N-methyl-D-Aspartate DMD: Disease

Modifying Drugs BACE: B-Secretase

Cleavage Enzyme

ABCD: Alzheimer's Disease Biomarkers Comprehensive Database

CSS: Cascading Style Sheet

NIH: National Institute Of Health

TF: Transcription factors

HPRD: Human Protein Reference Database

MMSE: Mini Mental State Examination

DT: Decision Tree

GO: Gene Ontology

NGS: Next Generation Sequencing

SPL: Shortest Path Length

NB: Node Betweenness

CC: Clustering Coefficient

EC: Eigenvector Centrality

MD: Molecular Dynamics

GA: Genetic Algorithm

SPC: Simple Point Charge

RMSD: Root Mean Square Deviation

RMSF: Root Mean Square Fluctuation

DMSO: Dimethyl Sulphoxide

DTNB: 5,5’-Dithiobis-(2-Nitrobenzoic acid

AO-EtBr: Acridine Orange- Ethidium bromide

CHAPTER 1

Introduction

ABSTRACT

The effect of Alzheimer disease both on socioeconomic and public health of individual is a great concern. Such disorders affect not only influenced the health sector but also put economic burden on care givers and also affect the life of people affected with AD to live happy and healthy life.

AD is that neurodegenerative disease which put direct burden of about million costs.. Thus, target the management of AD as a proxy for the big issue for approaching the complex challenges that need to be addressed, including scientific and clinical research, technology development, care delivery, public policy, insurance and other financial issues, and psychosocial burdens to caregivers.

Long and long decades were spent and continuously research are going on to understand the undergoing mechanism of AD and develop more effective treatments for this devastating disease. Wide spectreum of intervention are required to reduce or prevent the effect of disease. This could only be gained through joint and collaborative research.

This introduction section will first focus on the clinical and neuropathological characteristics of AD. Subsequently, advances in the understanding of putative molecular mechanisms underlying neurodegenerative events, mostly based on studies of rare familial variants of AD, will be discussed. As mentioned, effective treatment regimes remain conspicuously absent, despite enormous research efforts during the last two decades. Therefore, the final part of this introduction will focus on new research strategies that might help to provide insight in the biological mechanisms that contribute to AD associated neurodegenration.

1.1 INTRODUCTION

1.1.1 Alzheimer's disease(AD) – Multistagial neuronic syndrome

Neurodegenerative disorder (ND) is a wide term for a group of preliminary malfunctioning of neurons with strong characteristics of a progressive damage of working neurons [1]. In contrast, the term dementia is a general term used for the condition of brain affecting coginitive, thinking and memory skill [2]. Dementia affects the human brain in many different ways. It is totally depends on person's way of living and combined effect of environmental factors and states of health. Dementia is divided into 3 types on the basis of their stage of progression

- Latent stage – priliminary year or two
- Middle stage – second to fourth or fifth years
- Severe stage – fifth year and after

AD is the most common form of dementia when compared to other dementia like Lewy body, Frontotemporal dementia and Parkinson Disease (PD) [3]. It involve the damage of neuron and pathological marker are found in form of Neurofibrillary tangle and β-amyloid plaques.

1.1.2 Brief history of AD

AD was first discovered a century ago in year 1906 by Alois Alzheimer, a German neurologist. He was the one who is endorsed for it, the condition received his name. Most of the scientific community at that time does not gave much attention to Alois work toward AD because at that time it was considered as one of rare disease whose harmful effects weer not noticed by medical professionals [4]. Earlier prospective regarding this domain of research changed in 70^{th} century in support of lots of evidences available about that the senile dementia contribute majorily in eldest population [5]. Most and major researches were performed in this field in previous last three decades. [4].

As described earlier that AD is the commonest form of dementia, so that AD draw attention toward itself by their wide spectrum of effects on the society [6]. This disease is now major health concern and is a point of essential discussion.

1.2 DESCRIPTION OF AD

AD is marked by the symptoms appear in the patient suffered from the disease, This symptoms is much different than that of ageing in term of cognitive decline, which is more gradual and associated with less disability. Symptoms appear in this disease are first mild and then become severe [7].

The main physiology lies in this disease is that of neuronal damage that start from hippocampus region of brain and later distribute in different area of brain responsible for different types of activities [8]. Amyloid beta (Aβ), a little by product peptide formed due to malfunctioning of the transmembrane amyloid precursor protein (APP), whose exact function is not known but believed to be involve in neuronic development. Single unit of Aβ are rich in β sheet region transformed into higher tertiary form and later aggregate to form amyloid β-fibrils. This insoluble fibril are deposited outside the neuron and form thick structure called senile plaque or neuritic plaque in extracellular region of blood vessels [9]. Intracellular the neurons are severely damaged by false aggragation or accumulation of tau-protein, which normally help in providing stable structure to microtubule. Activation of tau protein has been done through phosphorylation. Hyperphosphorylated tau accumulate in paired helical filament form and later develop into thick masses inside the nerve cell body famously known as neurofibrillary tangle.

The actual mechanism for formation of plaque and NFT are not known. Senile plaques and NFT like hallmarks promotes the injury and death of essential neurons, which hampered the cogntive function of brain along with change in behavioural activities of AD patients. Abnormal release of glutamate resulting inflammation and neuron death [11]. Neuroinflammation symptom is directly or indirectly linked to AD. Immunological role of AD gene is confirmed by extensive pathological and clinical proof documents available on online literature portal. The molecular changes including increased pro-inflammatory cytokine concentrations within the blood and cerebrospinal fluid [12]. These molecular changes whether promote later stage of disease or not remains to fully understood, however swollowing of brain and associated activities of neuroglia cell for amyloids deposits, has been concerned in dynamic mechanism and pathological processs of AD.

1.3 HAZARDOUS CAUSES FOR AD

(a) Age

Due to better health facilities provided to people they live long and healthier life and as such they live long asa consequence of which number of old aged people increase tremendously and ageing and AD like condition emerges. The Late onset of AD appear after age of 65 years and the rate of occurrence of this disease become high after age of 85 years [13]. It is stated in one sentence as AD grow in exponential fashion after the age of 65.

(b) Role of environment for AD

Environmental factors like diet, activities and way of living have combined effects in occurrence of this disease. This thought was supported by stern. Several diseases like diabetic, smoking, obesity are linked to this AD like condition. A better education and mental quiz along with fiber rich food may decrease the chance of occurrence of disease.

Table 1.1: Influence of different factors and their action of mechanism in possible cause of AD.

Factors	**Directions of AD**	**Possible mechanisms**
Heart disease	upward	Parenchyma destruction, increased Aβ production
Heavy smoking	upward	Cerebrovascular effects, Oxidative stress
stress	upward	Microvascular disease
Type 2 diabeties	upward	Effect on brain vessel, insulin and Aβ act as Competitor.
Obesity	upward	Increased risk of type 2 diabeties inflammations

Head injury	upward	Increased Aβ and APP deposition
Level of education	Downgrade	Provide cognitive reserve
Healthy diet	Downgrade	Antioxidant and Anti-inflammatory
Body Exercise	Downgrade	Activate brain plasticity, promotes brain vascularisation

1.4 EPIDEMIOLOGY AND SOCIAL IMPACT OF AD

After 1998, WHO declared AD as a major health concern in European continents [15]. Statistics of AD shows that numbers of individuals are still increasing, notably among the oldest population age larger than 65 years. The estimated population suffered from this disease is about 44 millions in 2016. According to prior census,now predictable that in future about 70 millions new cases will be registered in 2030 and 120 millions in 2050.

[16]. The total number of new cases of AD in whole world is on average 7.7 million i.e, every four second new cases of AD are reported [17]. On comparison death cases reported by WHO is about 5.3 millions and an Most of the death are due to complication arises as a consequence of AD. In 2016, lumsome 17.5 billion hours of cares are given by family members and unpaid caregiver consist of more than 15 millions to patients sufferer from AD and other dementias, In term of money this service provided by caregiver are about $604 billion..

Cases of AD are now rising in fast growing economy of asian countries like China, India etc [18]. According to statistics figure of Alzheimer related disorder society of India (ARDSI) in 2010 approximately 0.4 million people are diagnosed with AD [19]. Looking at this information, It is greater needed to take urgent steps in this direction. It is assumed that by the end of 2050, 22% of total population aged 60 or over and majority of them will be from Asia, America or Africa [16]. AD is now major cause of death as compared to other harmful diseases as stated by Alzheimers association as indicated in Fig. 1.

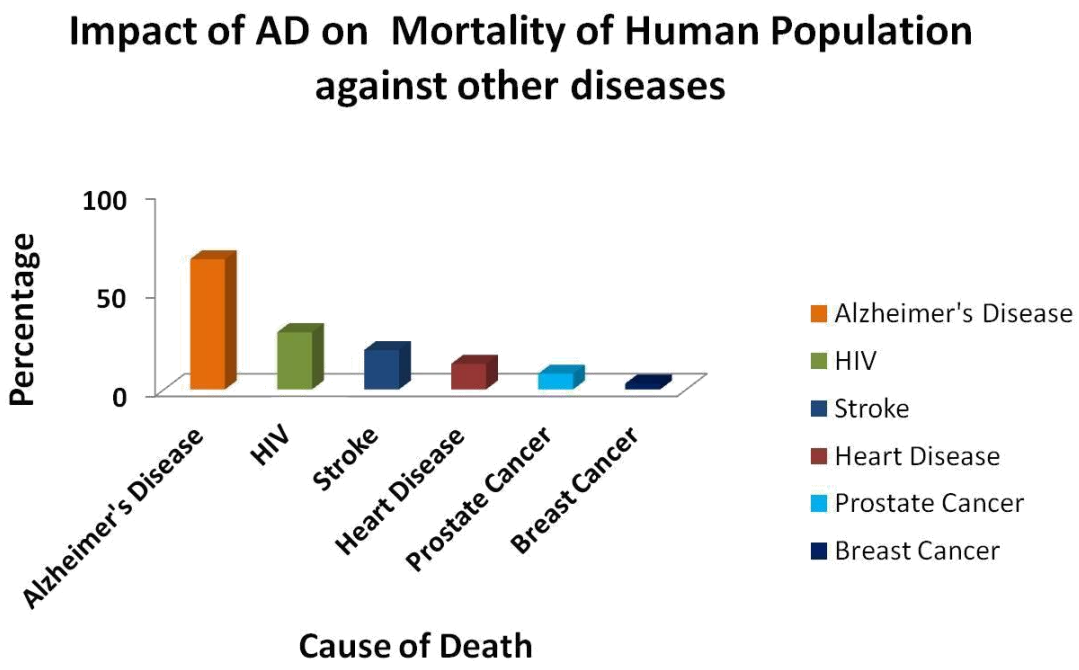


Figure 1.1: Proportion of different factors of death (all ages) between 2010 and 2016

(Source: Alzheimer's Association Report 2016, Alzheimer's disease facts and figures)

1.5 PATHOGENESIS AND SYMPTOMS OF AD

AD is marked by the presence of two signal i.e neurofibrillary tangle (NFT) and amyloid plaque which are formed both intracellularly and extracellularly due to misfolding activity of neuronic proteins.

1.5.1 Amyloid β

Extracellular β-amyloid plaques formation is firstly hypothesised in amyloid cascade hypothesis (Figure 2). Plaques formed outside the surface of neuron by the enzymatic activity of β-amyloid cleaving enzyme (BACE) and as result Aβ-42 type aggregate together to form this structure [20]. The presenilin proteins (PS1 and PS2) plays vital role in formation of β-APP by undesirable cleavage of APP. Familial- AD (FAD) is mainly occur due to specific mutation on PS1 or PS2, as a consequence of which concentration of β-amyloid rises many fold [21]. Allele

of apolipoprotein E type 4 (APOE ε 4) are major culprit on onset of sporadic AD [22]. By the well experimental studies of APOE, PS1 and PS2 describe the molecular mechanism lies in plaque pathology [23]. The amyloid cascade hypothesis is the most popular theory which explained the role of β-amyloid in plaque formation [24]. The theory does not however explain the cause of sporadic AD. Plaques are also common in non-demented individuals [25].

The sortilin related receptor 1 (SORL1), also contribute in APP processing. SORL1 directed APP into recycling process and protect APP from enzymatic activity by BACE and presinilin proteins. Due to less enzymatic activity on APP, lesser amount of Aβ production takes place. However, when the expression of SORL1 is not much expressed than the enzymatic activity of BACE like enzyme increases drastically and higher concentration of Aβ formation takes place. Levels of SORL1 act as biomarker in identification of AD brain and mild cognitive impairment as the concentration of SORL1 found to be reduced have been shown to be reduced in the brains of AD patients and patients suffered from mild cognitive impairment [26]. Another chaperone protein called Clusterin also work as natural biomarker as its expression is directly associated to severity and progressive stage of disease [27].

It is utmost important to identify specific genes which confer the risk of AD. However, there is found always a problem in identification of specific genes as single genes represent low level of risk [22]. GWAS studies overcome this problem by examining whole genome for genes associated with the risk. However, the level of disease pattern varies from person to person.

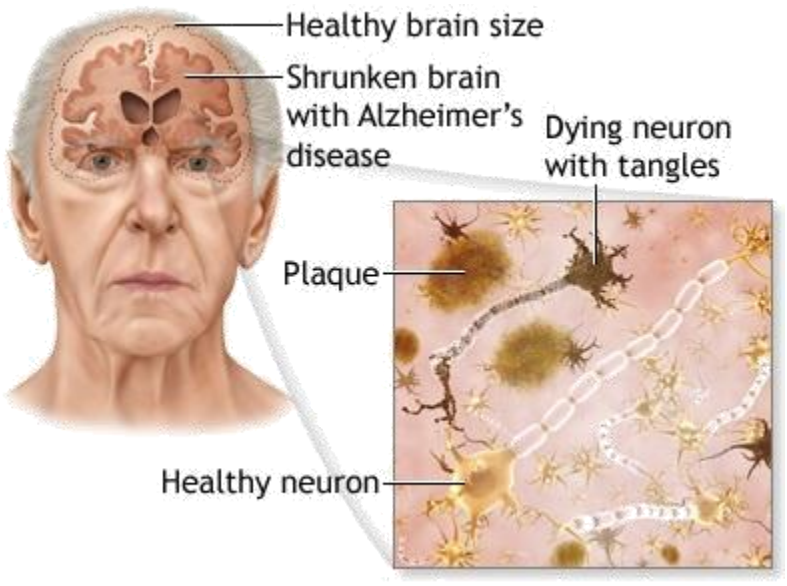


Figure 1.2: Alzheimer's Disease and its major hallmark

(Source: Stem cell Tech)

1.5.2 Tau protein abnormalities

Another positive signal of AD connected changes within the brain is intracellular structure formations referred to as neurofibrillary tangles (NFTs). NFT are made from paired helical filament (PHF). The most important protein molecule of NFTs is that the protein tau, a tubule associated protein (MAP) [28], link with structural protein called microtubule which assist in providing cell stability of neuron, releasing of tau protein from microtubule result in unbound tau protein aggregation [29]. The explanation for the aggregation is explained by tau hypothesis. In physical state, tau proteins are soluble in nature and show dynamicity through phosphorylation and dephosphorylation. Imbalance generated during this dynamic leads to formation of multiplied levels of abnormally hyperphoshorylated tau like P-tau 181, P-tau 199, P-tau 231, P-tau 396, P-tau 404, that successively disturb normal tau and alternative MAPs (MAP1 and MAP2) [30]. Tau aggregate in hyperphosphorylated form and make PHF to establish tangle inside the neuron. Tangle formation take place due to disorganization of microtubules. The mutual impact of tangle formation and disorganisation of microtubules result in compromise in natural and cognitive performance [31]. Amyloid cascade hypothesis depend on high volume of Aβ-protein that influence tau to form NFT [25].

Similar to tau, another protein called α-synuclein help in tubule assembly through binding nature of this protein [32]. When mutation occur in this protein result in lost of

assembly of tubule and tangle formation take place. Tubule is very important for performing role in traditional neural and synaptic connection building, pathology of microtubule also focus in neurodegeneration [33]. The quantity of NFT may be a pathological marker of AD severity.

1.6 Animal models and post mortem studies

Support from transgenic animal model, human tau protein mutation in transgenic mouse model shown tangle formation [34]. Core symptoms associated with AD has biological basis, both plaques and NFT have mutually exclusive behaviour to each other and both grow independent to each other and have different pattern of distribution [35]. On examining the brain during post-mortem shows how the disease progresses from onset to prominent stages with relation to time. The medial temporal cortex, hippocampus, and entorhinal cortex, anterior neural structure are majorly affected whereas other parts are remains unaffected. This distribution or topographically prophetical nature of aggregation is assumed to be a 5 to 6 stage method whereby the primary 3 stages are diagnosis, with symptomatic or clinically identifiable symptoms changing into outstanding from stages 3 head. The symptoms of amnesia are caused due to disturbance in hippocampus region therefore manufacturing early changes in memory, and ultimately the progression to the final stage, wherever the pallium is affected also [36]. Moreover, It has been reported that tangle formation appear before plaques formation. This can be contradictory to amyloid cascade theory which said that tangle formation take place as result of Aβ formation [37]. Support to this statement is found within creation of animal model like "3xTgADAPP","PS1" and "P30IL". The incidence of amyloid deposition precedes tangle formation by abnormalities in tau protein reveal temporal and spatial distribution of taupathy following amyloid deposition [38], both APP and "P301L" models jointly yielded each pairs of pathologies; with amyloid thought-more infective feature of the two, and so a lot of at risk of cause dementia.

1.7 Other famous hypothesis

1.7.1 Oxidative Stress

Oxidative stress is one of major suspect responsible for causing neurodegenerative disease by harming numerous essential bio molecules present in brain region in an uncontrolled pace. One thought that Aβ and NFT essential in initiation of antioxidant defence. Therefore the sign of Aβ deposits and tau hyperphosphorylation could be directly linked to oxidative stress defence mechanism. Animal model study shows that oxidative harm precedes the pathological changes related to AD [39].

1.7.2 Inflammation

Brain regions that are affected by AD shows presence of cytokines and microglia through increased inflammatory cascades [40]. It is not clear whether or not this can be a natural action to control inflammation and associated misregulated immune response. Enzyme COX is a key enzyme of inflammatory phenomenon is targeted by non-steroidal medicinal drug medicines. (NSAIDs), touching COX levels. NSAIDs doesn't cut back the danger or delay the onset of AD [41]. Majorly this process depend on role played by NSAID, although some evidences are there which show inflammation process is independent of inflammation, NSAIDs will cut back Aβ accumulate in animal models of AD. As it is assumed that glia cell of brain activated when AD diagnose at early age in patients and also play role during synaptic disruptions and mild cognitive impairment [42]. The AD medicinal drug bar Trial (ADAPT) looked into the role of NSAIDs in individuals at risk of dementedness, exploitation COX and COX-2 medicine. The trial was associated to heart disease risks [43]. In total, COX enzyme which target NSAID are much better option than that of COX-2 inhibitors [44].

1.7.3 Cholinergic hypothesis

Disturbance in cholinergic pathway lead to discovery of cholinergic hypothesis, it totally based on secretion of AChE that is situated at the basal region of the brain [45]. These neurons arise from basal region to the hippocampus and cerebral cortex that are concerned in each memory disturbance and psychological symptoms. Ach is chopped-off by the protein acetylcholinesterase (AChE). Quantity of acetylcholine is found in trace amount in moderate and severe AD patients, unlike mild patients; inhibition activity of cholinesterase improves transmission capacity of neurons and provides mild relief to AD patients.

1.7.4 Cholesterol metabolism

The association between cholesterol and AD pathogenesis are now widely accepted [46]. Activity of enzymes involve in APP metabolism are greatly affected by cholesterol. Statins which are cholesterol-inhibiting drug are also played a crucial role in lowering the risk of AD [47]. Another function of Apolipoprotein (APOE) is that it is involved in the transportation of cholesterol, and APOE ε4 allele is an authenticated marker which undoubtly increases AD risk. APOE ε4 is also associated with early onset for AD [22]. APOE ε4 both work as genetic risk for AD and also linked to the production and aggregation of both amyloid and tau. E4 allele of apolipoprotein is involved in malfunctioning of neuronic signal and amyloid plaque formation. Individuals who possess two ε4 alleles are thought 7 times more susceptible in developing AD than those who have the E3 allele. Heavy amount of cholesterol present in body may increase risk of AD in age >= 40 years and FDA approve cholesterol free drugs like statin which reduce the risk of occurrence of AD. But impact of statin on improving cognitive function is not much affected [48]. AD researchers are not convinced by the cause which cause brain cell death and tissue loss in AD brain but plaques and tangles are key suspects. Scientists can see the dreadful effects of AD when they visualize at brain tissue under the microscope (Fig.3).

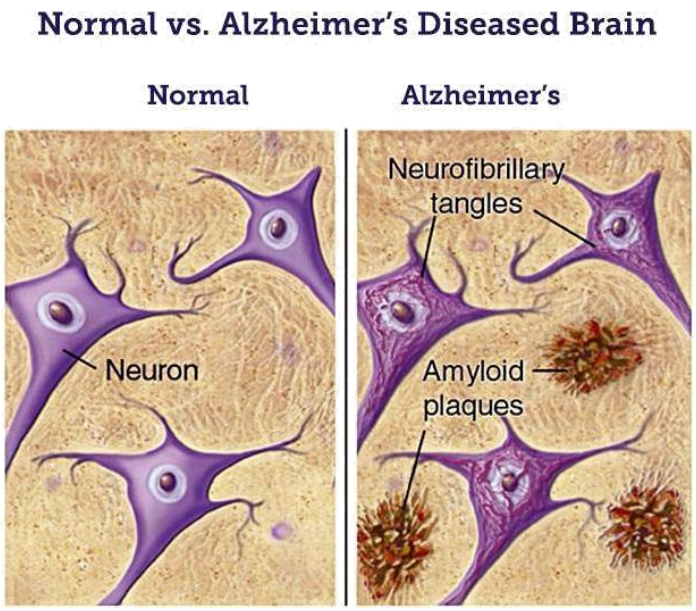


Fig 1.3: Microscopic view of neuron (a) normal condition (b) AD condition from Alzheimer disease International

The actual mechanism for the occurrence of two disease hallmark is still not known. Different mechanism like synaptic failure, oxidative stress, mitochondrial stress may involve indirect association with AD, but is still questionable, Linkage of Aβ with tau is also still not clear [49]. From some AD molecular search it has been found that in tau pathology, once it is started it grow independently without the involvement of Aβ. It also evidence that even after removal of major portion of Aβ deposit around the nerve fibre, The performance of neuron not improve to much extent[50]. These are some of the evidences which are highlighted here but many other hypothesis regarding AD comes from time to time but due to lack of evidences every hypothesis fail to explain actual mechanism of AD. This demonstrates that we have still fragmentary knowledge and need to compile that knowledge to understand the actual cause of AD.

1.8 GENETICS AND RISK FACTORS FOR AD

No one is aware of the actual reason behind AD; however scientists do apprehend that genes are concerned. For several years, researchers have explored for specific genes that are responsible for outcome of commonest dementia. There are some intriguing clues, however additional analysis required to totally perceive the genetic link of disease.

Four chromosomes 1, 14, 19 and 21 on genes are strongly linked to familial AD. The name of genes lies on chromosomes is PS1, PS2, APOE and APP [51].

APOE genes belong to the chromosome responsible for late onset AD that is the most common form of the illness that affects peoples over age 65 years. Several literature available online around the world have shown that when a person possess has one specific allele of the APOE sequence, mentioned as APOE4, it'll increase their chance of obtaining AD at some phase in their lives [52]. But the proof for direct association available is not sufficient. This is confirmed as some people who have APOE4 don't get AD, others have the illness despite the very fact that they don't have APOE4 in their DNA [53]. In alternative words, though the APOE gene clearly influences AD risk, it isn't a true biomarker that someone will have the illness. Scientists have to be compelled to learn lots of relating to the association.

AD sometime appear early in families whose relatives were earlier affected by this disease. By studying the genetic material of those individual the role of gene associated to chromosome 1 and 14 play major role in occurrance of early onset of disease before age of 65 [54]. The chromosome 21 gene is a motivating AD clue attributable to its role in Down syndrome [55]. Individuals with Down syndrome have an additional copy of gene on location 21. As they age, they usually get AD symptoms, although at a younger age than others who get the disease. Their brain cells additionally show identical damage that happens to brains affected with AD [56]. Scientists are still attempting to completely understand the link between these 2 conditions. Genes don't seem to be the sole reason for the disease. Additional analysis can show however DNA, life style, and things within the surroundings play roles in creating individuals additional possible to induce the condition.

1.9 DIAGNOSIS OF AD

The classification and therefore the criteria accustomed diagnose dementia and AD is about move into Diagnostic and Statistical Manual of Mental Disorders (4^{th}, text revision, DSM-IV-TR) (APA) and the International Statistical Classification of Diseases and Health related Problems. The international committee DSM-IV-TR (APA) and ICD-10 define dementia as associated with other impairment in cognitive function of brain [57]. NINCDS-ADRDA criteria classify the identification of symptoms as definite (typical symptoms and microscopic anatomy confirmation), probable (typical symptoms while not microscopic anatomy confirmation) or attainable (atypical symptoms however no different identification a minimum of equally likely).

These classification criteria are considered to be highly reliable as its statistical validation reaching sensitivity and specificity is found to be 81% and 70 % respectively [58].

This criteria work mostly in diagnosing late stage of disease but it has been well accepted in scientific community that disease shows its effect much earlier before the actual diagnosis of disease symptoms [59]. By this way, criteria were fixed to identify mild cognitive impairment (MCI), in this stage sufferer people feel difficulty in performing routine task. But MCI is very unspecific and precede other dementia other than AD [60]. In order to identify AD at early stage of disease there is requirement of efficient biomarkers. In 2011, the NIA-AA workgroup conjointly work on finding the definition and detection of preclinical stages of AD [61], recognizing the importance of biomarkers for the sure shot identification of AD at early stage and causes of other dementia.

Biomarkers can be defined as matter (physiological, biochemical or anatomical) signature used to identify any pathological changes occurred during different stages of diseases [62]. In term of AD 5 potential biomarkers are confirmed till now are as given, decreased concentration of Aβ42 in the cerebrospinal fluid (CSF), increased CSF tau proteins in neuronal cortex regions, decreased fluorodeoxyglucose uptake on PET (FDG-PET), PET Aβ imaging, and structural MRI imaging measures of cerebral atrophy. Evidence available from literature suggests that abnormal biomarker findings ease in finding the disease in years as compared to finding possible factors take decades [63].

1.9.1 CSF as one of preliminary biomarkers

The fluid which found inside the central nervous system are called cerebrospinal Fluid (CSF), therefore when any biochemical changes occur in brain it will be noticed and monitored through CSF, as there is no any hindrance by blood brain barrier (BBB). CSF is extracted from patient brain through lumber puncture procedure. CSF biomarkers comprise components total tau (T-tau), phosphorylated tau (P-tau), and the 42 amino acid isoform of β–amyloid (A β 42) [64].

1.9.2 Level of Tau protein in CSF

Tau present in CSF was used as biomarker from early decade to identify AD. A meta-analysis revealed that over 2200 AD patients had been studied, with 1000 age-matched controls [65]. Consistently CSF-tau level has been significantly elevated (approximately three-fold increase) in AD patients. However, the sensitivity and specificity levels tend to vary. The sensitivity levels range between 40% and 80%, and specificity varies between 65% and 85 % [66]. There is also a link between tau and age related disorders increases in non-demented persons.

Though CSF levels of tau are highly sensitive for AD, they do not possess a high level of specificity against other dementias [67].

1.9.3 Phosphorylated tau protein in CSF

Different form of phosphorylated tau like P-tau 181, P-tau 199, P-tau 231, P-tau 396, and P-tau 404 are found in CSF with varied concentration. P-tau 231 and P-tau 181 were found to be the two robust and authenticated biomarker to facilitate discrimination between AD and other dementias [68]. Sub-groups of AD based on CSF proteins have been highlighted and identified, where following analysis and measurement of the CSF forms of P-tau levels have been shown to cluster indicating different clinical profiles at 88% and 86% sensitivity and specificity respectively [69].

1.9.4 Role of β-Amyloid in CSF

It is now well known that Aβ1-42 is a pathogenic protein, and is found in the CSF. Decrease in concentration of CSF in AD patients due to building of amyloid proteins inside the cerebral cortex. This suggests a connecting-link between extracellular amyloid plaques and CSF Aβ1-42 [50]. In the past decades studies were performed on 900 patients and 500 controls and the findings have been consistent in the levels of Aβ1-42 between the two sets of populations [70]. There was noticed that Aβ1-42 level are much lower in AD population, with a sensitivity range of 78% to 100% and a specificity range of 47% to 80% [71]. The level of Aβ found in CSF is much higher from that of found in plasma [72]. Furthermore plasma Aβ levels are not in sufficient amount to fulfil the criteria as diagnostic measure. Plasma Aβ1-42 levels are found to be similar in both AD-patients and control. A longitudinal study suggested that higher plasma Aβ, a risk issue for developing AD is not thought to be helpful in discrimination between disease and control. Plasma Aβ do not have the sensitivity or specificity compared to CSF-Aβ [73].

1.9.5 Concentration of β-Amyloid antibodies

The level of antibodies present against Aβ in both CSF and blood are lower in AD patients as compare to that of healthy persons. With reference to work from animal-model, active and passive immunization with β- amyloid antibodies resulted in a reduction of plaques development. Active immunization trials on humans with β-amyloid antibodies produce side effects as sub acute meningoencephalitis. However in subset of patients who possess autoimmunity for antibodies of Aβ have shown neuroprotective effect, with improved performance in neuropsychology exam. [74].

1.9.6 Imaging as a biomarker (MRI & PET)

Magnetic Resonance Imaging (MRI) has been used to detect disease state or atrophy in one of the lobe called medial temporal lobe in AD patients. Validity of this physical techniques as biomarkers is measured and reported that it have approx. 85% sensitivity and specificity, and depending on the severity, is found in up to 96% of the AD cases as compared to normal one. The Alzheimer's disease Neuroimaging Initiative (ADNI), one of the central unit of international research as part of a multicentre study are majorly measuring hippocampus size volumes and measuring cortical thickness to differentiate disease and non-disease brains. Available computational technique are used to test MRI scans and differentiate AD and non AD brain with 89% sensitivity, following post mortem confirmation [75]. Neuronal function of brain becomes false in region of temporal parietal and cortical specific area. Positron emission tomography (PET) is also proved to be valid imaging biomarkers utilize live aldohexose metabolism in neuron in differentiating AD brain and healthy control [76].

Although all these are efficient biomarkers but not able to diagnose full pathophysiological process. For example, CSF and Aβ-PET are used to recognise Aβ deposit, but do not provide information on Aβ oligomers, which equally play role in pathophysiology of AD. Neuronal injury in some extent are relate to abnormal tau protein in of brain [77] and the brain atrophy are not specific for AD but also relate to other neuropathy. Furthermore, it may be chance that individual who are marked positive through biomarkers will not necessary that progress to dementia in future [78].

In spite of all these limitation, The development of biomarkers are one of the major step toward AD diagnostic research. Early detection of AD will become easy by applying these biomarkers in-vivo and effective treatment start immediately and it help in reduction of population suffer from this hazardous brain malady [52].

1.10 TREATMENT OF AD

There are no potential drugs available in market for complete prevention of AD, but only drug available in market are symptoms based [79]. Natural compounds are prioritised as compared to chemical compound which improve thinking and memory skill. FDA approve drug for AD are divided into 2 categories, namely the cholinesterase inhibitors Donepezil, Rivastigmine, Galantamine, and the NMDA receptor antagonist Memantine. Study of cholinergic system suggests that deficit of cholinergic results of occurrence of AD [80]. However its efficacy in patient suffers by disease from mild stage to advance stage may be a point of debate. The intake of Memantine base on the hypothesis that neurodegeneration is caused due to toxin radicals, cell damage mediated by excessive Ca^{2+} influx through NMDA receptors. It is not applicable to patients who are in initial stage of disease, another drug called Memantine should be offered as symptomatic cure option for moderate to severe AD. Combined action of these two drugs is proved to be beneficial in controlling fast decline in cognitive function of brain. This treatment is effective just for 6-12 months before the brain function declination start again [81]. AD treatment are divided into both pharmacological and non-pharmacological and it's just mitigate the economic and physical burden on family-members and care-givers, facing both natural and psychological burden of disease and facilitate to perform activities of daily routine till attainable.

Beside treatment of AD on basis of symptoms some drugs are now developed on the basis of disease modifying properties. These drugs follow different strategies in treatment of disease and categorise into 3 phases of clinical trials: (1) Aβ metabolism, (2) hyperphosphorylated tau phenomenon and (3) other remaining strategies. Most drugs which target Aβ production by APP cleavage belongs to phase-2 or phase-3. Beside their role in controlling Aβ production they were primarily developed for other diseases such as diabetes or hypertension, as well as others drugs which are specifically designed for that purpose e.g. to alter the β- or γ-secretase activity, to induce neuroimmune response to Aβ peptides or to stop the formation of Aβ oligomers which are suspected to pose an important neuroprotective toxin.

Despite promising result in preliminary phase of drug development most of the drugs fail in phase III trial stage of drug discovery. Most of the drug is banned in phase III trial of drug development due to adverse effect generated without much improvement in cognitive ability of brain.

This results in a reassessment of preventive methods in AD drug development process. This is supported by updating criteria in diagnosis of disease and majorly focuses on treatment method and techniques for early diagnosis of disease. Implementing these criteria in future trials is predicted to possess a crucial impact on drug development.

1.11 CURRENT RESEARCH

It is critical that AD and connected dementias analysis still accelerate. To confirm that the effort to find better treatments receives the main target it deserves, the Alzheimer's Association funds analysers watching new treatment methods and advocates for a lot of independent funding of AD research [82].

1.11.1 The expectation for perspective drugs

At this time, there are 5 FDA-approved AD medicine that heal the symptoms of disease — temporarily serving to brain activity and thinking problems — with a 6th drug on the market universally [83]. Although these drugs don't operate on cause of disease. In distinction, several of the new medicine in development aim to change the disease development procedure itself, by influencing various changes occurs in brain due to AD. These positive changes provide target for the development of new treatment. Several researchers believe new victory treatment eventually involve a disease modifying drug (DMD) medications aimed toward many targets [84], almost like current progressive treatments for several cancers and AIDS.

1.12 PROBLEM STATEMENT

The problem statement mentioned in this study is to identify the genes and proteins of AD and their connection to the disease. The method and techniques employed in the research, using complex large databases, analyzes the data, and generate the results in a reliable way.

There are two ways of doing the analysis: *in-vitro* analyses and *in-silico* research. The *in-vitro* analysis is capable of diagnosis the disease and additionally suggest a treatment. This needs experimental proofs. The experimental result could generate low quantity of information and therefore the medico, additionally could have some additional information at his/her disposal. With this extra information, medico currently could also be able to analyze the information and turn out results to acquire knowledge regarding the disease and its cure. Because the physicians analyze large information, the results therefore obtained could also be

Inaccurate and insufficient within the method of constructing medical choices. Significantly within the method of drug discovery, the quantity of information to be extracted, stored, manipulated, managed and analyzed is quite large. As a part of these activities, a really great amount of information is to be exchanged and shared. This can be a tough task to find, extract, manipulate, manage and analyze huge volumes of information manually. It's quite apparent that there are ranges of cause concerned in disease.

The in-silico analysis uses associate experimental technique performed on computers or in artificial machine. This approach is capable of extracting, storing, manipulating, managing and analyzing terribly immense amounts of data and manufacture purposeful information, which might be accustomed build purposeful and correct selections. As a part of this, great amount of information might have to be extracted from a range of sources and in numerous formats. The information from varied sources is well-mined using an appropriate acceptable mining technique and also the data be made available for processing.

The current study focuses in finding the genes and proteins responsible for AD as regulatory biomarkers for understanding the pathogenesis of disease. By using various search engines, search and extract data from a wide range of databases available on line for open access. The data extracted is assembled in the desired format by using the computational algorithm for processing, and later perform required computations. The results are represented graphically and provide necessary information for decision making. Using this methodology, the data cannot only be processed but also be interpreted in a variety of formats like pictorial and graphical. The *in-silico* process makes the process more accurate, efficient and reliable.

In this study the data of genes and proteins causing AD and other neurodegenerative diseases are extracted from various online databases assembled and analyzed with different objective in mind by using different approach. From the analysis, the names of the genes/proteins that primarily cause AD are known, inferred and later hold on in web repository.

1.15 OBJECTIVES OF RESEARCH

Keeping in view the gaps existing in efficient biomarkers for AD, the current study was undertaken with the following four objectives:

Objective 1: Build a comprehensive database for AD which includes basic level to advanced level information for genes, proteins, SNPs, miRNA, Animal models, pathways and drugs etc. Along with this AD affected brain image data are also embedded with tissue differentiation according to stage of disease added value to our database.

Objective 2: Machine learning (Decision Tree) approach to unravel the genetic mystery of AD Pathological process using standard diagnostic scoring System. Efficient rule primarily based classification of AD related genes for correct mining of data used for scientific research.

Objective 3: Systems biology approach for gene set enrichment and topological analysis of AD network.

Objective 4: Molecular docking simulation studies of specific chemical group of compounds to find inhibitors for specific target protein Acetylcholinesterase (AChE)**.**

1.16 SCOPE OF THE STUDY

This study focuses on the identifying genes and proteins that play a key role in connection between AD. These genes are used as Biomarker for predicting the complications with which occurrence of one can tell the state of the other one. These genes are identified and extracted from the databases that are obtainable in online repository. Biological aspects are not vividly utilized in this study except to clarify the examples and to grasp the fundamentals of the problem. This study has identified some keystone genes/protein that is responsible for AD.

1.17 LIMITATIONS OF THE STUDY

This dissertation work identifies genes and proteins that are related to the complications but this will not tell that only these genes/proteins related to the diseases and this will relate to only this complication. During the inhibitors identification for Acetylcholinesterase (AChE) this will not tell that only these genes will play an important role in the drug design.

1.18 BRIEFING OF LITERATURE REVIEW

There is an utmost need in identification of AD at early stage and on timely manner. Due to several years of research in this domain clinical profile and associated symptoms are well established. Only demand for effective biomarkers are remaining which correctly diagnose the disease in short time.

Currently available diagnosis techniques take a time of about a month to detect disease rather than hour or day. In term of treatment, clinician only develop drugs which enhance or improve the cognitive ability for short duration by prescribe drugs according to the symptoms and also theses drugs are not recommended to patients in severe stage of disease.

Research shows that AD exist in pre-symptomatic phase so there is higher demand for early diagnosis of this disease and development of new drugs based on complete eradication of this ailment from root. Currently available single or group of biomarkers still promising but are not much sufficient to fulfil their role of diagnosis of disease at preclinical setting, With great hope arise about CSF as biomarkers as it contain all essential biomolecules which assist in identification of disease with great accuracy .

The three key areas of research in AD have yielded important information in the form of data as morphology and molecular understanding of the disease, the pathological events associated with AD, the clinical profiling and the search for biomarkers. This review also focus on discovery of robust biomarker by majorly giving preference to the known or proposed molecular candidates, employing the traditional hypothesis by using scientific approach on small dataset.

With the advancement in technologies researchers can analyse large numbers of data samples in robust manner to interrogate potential biomarkers open new avenue of research direction. In literature discuss that with high throughput technology and big data handling through clusters and heavy servers allow focusing both CSF and braining tissue. It helps in finding both known and unknown biomarkers comprise genome, protein and metabolite level data.

The application of interdisciplinary field of machine learning, systems biology, structural biology and bioinformatics can be applied for discovery of biomarkers for the diseases; this implies the accurate diagnosis of disease state, the ability to sub classify diseases, to elucidate disease aetiology and mechanisms.

1.19 THESIS IN NUTSHELL

The principal title of this work is the identification of important genes and proteins with their role involved in AD and their interaction studies.

In the first chapter a general background is presented, Detail literature about disease, its risk factors, epidemiology, and genetics of disease, also overview the problem statement, scope, significance of the study and limitation of the study.

In the Second, Third, Fourth and Fifth chapters discussed about four different objectives, their methodology, results, conclusions in form of major finding and discussion about their applications in finding potential biomarkers for confirmation of AD.

In the Seventh chapter summarizes the conclusions of the key reports of the study and discusses prospects for further work in these areas.

In summary, the present study has made an attempt to show how applications of different computational and statistical tools help in finding useful knowledge about pathogenetic mechanism of AD. Through structure data of Proteins about different inhibitors enlightened our knowledge about their binding energy and inhibitory potential. This study can useful to generate new biological information. With the exponential growth of sequence and structure data, Bioinformatics can still play a very important role in new biological discovery and in formulating intelligent questions for designing experiments.

REFERENCES

[1] A. Vilalta and G. C. Brown, "Neurophagy - the phagocytosis of live neurons and synapses by glia - contributes to brain development and disease," *FEBS J,* Nov 10 2017.

[2] D. LoGiudice and R. Watson, "Dementia in older people: an update," *Intern Med J,* vol. 44, pp. 1066-73, Nov 2014.

[3] R. J. Castellani, R. K. Rolston, and M. A. Smith, "Alzheimer disease," *Dis Mon,* vol. 56, pp. 484-546, Sep 2010.

[4] H. Hippius and G. Neundorfer, "The discovery of Alzheimer's disease," *Dialogues Clin Neurosci,* vol. 5, pp. 101-8, Mar 2003.

[5] L. Rozzini, B. Vicini Chilovi, E. Bertoletti, M. Conti, I. Delrio, M. Trabucchi*, et al.*, "The importance of Alzheimer disease assessment scale-cognitive part in predicting progress for amnestic mild cognitive impairment to Alzheimer disease," *J Geriatr Psychiatry Neurol,* vol. 21, pp. 261-7, Dec 2008.

[6] E. S. Musiek and S. E. Schindler, "Alzheimer disease: current concepts & future directions," *Mo Med,* vol. 110, pp. 395-400, Sep-Oct 2013.

[7] P. Scheltens, K. Blennow, M. M. Breteler, B. de Strooper, G. B. Frisoni, S. Salloway*, et al.*, "Alzheimer's disease," *Lancet,* vol. 388, pp. 505-17, Jul 30 2016.

[8] R. H. Swerdlow, "Pathogenesis of Alzheimer's disease," *Clin Interv Aging,* vol. 2, pp. 347-59, 2007.

[9] C. L. Masters and D. J. Selkoe, "Biochemistry of amyloid beta-protein and amyloid deposits in Alzheimer disease," *Cold Spring Harb Perspect Med,* vol. 2, p. a006262, Jun 2012.

[10] G. Simic, M. Babic Leko, S. Wray, C. Harrington, I. Delalle, N. Jovanov-Milosevic*, et al.*, "Tau Protein Hyperphosphorylation and Aggregation in Alzheimer's Disease and Other Tauopathies, and Possible Neuroprotective Strategies," *Biomolecules,* vol. 6, p. 6, Jan 06 2016.

[11] C. Pittenger, M. H. Bloch, and K. Williams, "Glutamate abnormalities in obsessive compulsive disorder: neurobiology, pathophysiology, and treatment," *Pharmacol Ther,* vol. 132, pp. 314-32, Dec 2011.

[12] C. Liu, G. Cui, M. Zhu, X. Kang, and H. Guo, "Neuroinflammation in Alzheimer's disease: chemokines produced by astrocytes and chemokine receptors," *Int J Clin Exp Pathol,* vol. 7, pp. 8342-55, 2014.

[13] R. Guerreiro and J. Bras, "The age factor in Alzheimer's disease," *Genome Med,* vol. 7, p. 106, Oct 20 2015.

[14] E. Marcello, F. Gardoni, and M. Di Luca, "Alzheimer's disease and modern lifestyle: what is the role of stress?," *J Neurochem,* vol. 134, pp. 795-8, Sep 2015.

[15] WHO, *mhGAP Mental Health Gap Action Programme Scaling up care for mental, neurological, and substance use disorders*: who press-who, 2008.

[16] A. Alzheimer's, "2016 Alzheimer's disease facts and figures," *Alzheimers Dement,* vol. 12, pp. 459-509, Apr 2016.

[17] K. Y. Chan, W. Wang, J. J. Wu, L. Liu, E. Theodoratou, J. Car*, et al.*, "Epidemiology of Alzheimer's disease and other forms of dementia in China, 1990-2010: a systematic review and analysis," *Lancet,* vol. 381, pp. 2016-23, Jun 08 2013.

[18] Y. T. Wu, C. Brayne, and F. E. Matthews, "Prevalence of dementia in East Asia: a synthetic review of time trends," *Int J Geriatr Psychiatry,* vol. 30, pp. 793-801, Aug 2015.

[19] J. A. shaji kS, Girish N, Srikala Bharath,, "THE DEMENTIA INDIA REPORT 2010 Prevalence, impact, costs and services for dementia," Alzheimer's and Related Disorders Society of India, new delhi2010.

[20] R. J. O'Brien and P. C. Wong, "Amyloid precursor protein processing and Alzheimer's disease," *Annu Rev Neurosci,* vol. 34, pp. 185-204, 2011.

[21] W. Xia, "Role of presenilin in gamma-secretase cleavage of amyloid precursor protein," *Exp Gerontol,* vol. 35, pp. 453-60, Jul 2000.

[22] C. C. Liu, C. C. Liu, T. Kanekiyo, H. Xu, and G. Bu, "Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy," *Nat Rev Neurol,* vol. 9, pp. 106-18, Feb 2013.

[23] M. Kitazawa, R. Medeiros, and F. M. Laferla, "Transgenic mouse models of Alzheimer disease: developing a better model as a tool for therapeutic interventions," *Curr Pharm Des,* vol. 18, pp. 1131-47, 2012.

[24] M. P. M. Nabeela, Mohd Tayyab, Shirin Farheen and Mehdi H Shahi, "Targeting Cell Signalling Pathways: Novel Targets for Alzheimer's Disease," *Int J cell Sci & mol biol,* vol. 1, 2016.

[25] G. P. Morris, I. A. Clark, and B. Vissel, "Inconsistencies and controversies surrounding the amyloid hypothesis of Alzheimer's disease," *Acta Neuropathol Commun,* vol. 2, p. 135, Sep 18 2014.

[26] C. M. Karch and A. M. Goate, "Alzheimer's disease risk genes and mechanisms of disease pathogenesis," *Biol Psychiatry,* vol. 77, pp. 43-51, Jan 01 2015.

[27] M. Thambisetty, Y. An, A. Kinsey, D. Koka, M. Saleem, A. Guntert*, et al.*, "Plasma clusterin concentration is associated with longitudinal brain atrophy in mild cognitive impairment," *Neuroimage,* vol. 59, pp. 212-7, Jan 02 2012.

[28] S. Barghorn, P. Davies, and E. Mandelkow, "Tau paired helical filaments from Alzheimer's disease brain and assembled in vitro are based on beta-structure in the core domain," *Biochemistry,* vol. 43, pp. 1694-703, Feb 17 2004.

[29] E. M. Mandelkow and E. Mandelkow, "Biochemistry and cell biology of tau protein in neurofibrillary degeneration," *Cold Spring Harb Perspect Med,* vol. 2, p. a006247, Jul 2012.

[30] M. Sjogren, P. Davidsson, M. Tullberg, L. Minthon, A. Wallin, C. Wikkelso*, et al.*, "Both total and phosphorylated tau are increased in Alzheimer's disease," *J Neurol Neurosurg Psychiatry,* vol. 70, pp. 624-30, May 2001.

[31] J. Dubey, N. Ratnakaran, and S. P. Koushika, "Neurodegeneration and microtubule dynamics: death by a thousand cuts," *Front Cell Neurosci,* vol. 9, p. 343, 2015.

[32] W. S. Kim, K. Kagedal, and G. M. Halliday, "Alpha-synuclein biology in Lewy body diseases," *Alzheimers Res Ther,* vol. 6, p. 73, 2014.

[33] V. Nikoletopoulou, M. E. Papandreou, and N. Tavernarakis, "Autophagy in the physiology and pathology of the central nervous system," *Cell Death Differ,* vol. 22, pp. 398-407, Mar 2015.

[34] T. L. Spires and B. T. Hyman, "Transgenic models of Alzheimer's disease: learning from animals," *NeuroRx,* vol. 2, pp. 423-37, Jul 2005.

[35] T. L. Spires-Jones and B. T. Hyman, "The intersection of amyloid beta and tau at synapses in Alzheimer's disease," *Neuron,* vol. 82, pp. 756-71, May 21 2014.

[36] A. R. Preston and H. Eichenbaum, "Interplay of hippocampus and prefrontal cortex in memory," *Curr Biol,* vol. 23, pp. R764-73, Sep 09 2013.

[37] R. A. Armstrong, "A critical analysis of the 'amyloid cascade hypothesis'," *Folia Neuropathol,* vol. 52, pp. 211-25, 2014.

[38] K. Schmitt, A. Grimm, A. Kazmierczak, J. B. Strosznajder, J. Gotz, and A. Eckert, "Insights into mitochondrial dysfunction: aging, amyloid-beta, and tau-A deleterious trio," *Antioxid Redox Signal,* vol. 16, pp. 1456-66, Jun 15 2012.

[39] X. Wang, W. Wang, L. Li, G. Perry, H. G. Lee, and X. Zhu, "Oxidative stress and mitochondrial dysfunction in Alzheimer's disease," *Biochim Biophys Acta,* vol. 1842, pp. 1240-7, Aug 2014.

[40] W. Y. Wang, M. S. Tan, J. T. Yu, and L. Tan, "Role of pro-inflammatory cytokines released from microglia in Alzheimer's disease," *Ann Transl Med,* vol. 3, p. 136, Jun 2015.

[41] E. Ricciotti and G. A. FitzGerald, "Prostaglandins and inflammation," *Arterioscler Thromb Vasc Biol,* vol. 31, pp. 986-1000, May 2011.

[42] A. Serrano-Pozo, M. P. Frosch, E. Masliah, and B. T. Hyman, "Neuropathological alterations in Alzheimer disease," *Cold Spring Harb Perspect Med,* vol. 1, p. a006189, Sep 2011.

[43] C. J. Hawkey and M. J. Langman, "Non-steroidal anti-inflammatory drugs: overall risks and management. Complementary roles for COX-2 inhibitors and proton pump inhibitors," *Gut,* vol. 52, pp. 600-8, Apr 2003.

[44] A. Zarghi and S. Arfaei, "Selective COX-2 Inhibitors: A Review of Their Structure-Activity Relationships," *Iran J Pharm Res,* vol. 10, pp. 655-83, Fall 2011.

[45] S. H. Barage and K. D. Sonawane, "Amyloid cascade hypothesis: Pathogenesis and therapeutic strategies in Alzheimer's disease," *Neuropeptides,* vol. 52, pp. 1-18, Aug 2015.

[46] B. Allinquant, C. Clamagirand, and M. C. Potier, "Role of cholesterol metabolism in the pathogenesis of Alzheimer's disease," *Curr Opin Clin Nutr Metab Care,* vol. 17, pp. 319-23, Jul 2014.

[47] A. B. Reiss and I. Voloshyna, "Regulation of cerebral cholesterol metabolism in Alzheimer disease," *J Investig Med,* vol. 60, pp. 576-82, Mar 2012.

[48] N. Geifman, R. D. Brinton, R. E. Kennedy, L. S. Schneider, and A. J. Butte, "Evidence for benefit of statins to modify cognitive decline and risk in Alzheimer's disease," *Alzheimers Res Ther,* vol. 9, p. 10, Feb 17 2017.

[49] E. Tonnies and E. Trushina, "Oxidative Stress, Synaptic Dysfunction, and Alzheimer's Disease," *J Alzheimers Dis,* vol. 57, pp. 1105-1121, 2017.

[50] M. P. Murphy and H. LeVine, 3rd, "Alzheimer's disease and the amyloid-beta peptide," *J Alzheimers Dis,* vol. 19, pp. 311-23, 2010.

[51] R. E. T. Deborah Blacker, "The Genetics of Alzheimer Disease:Current Status and Future Prospects," *Neurological Review,* vol. 55, pp. 294-296, 1998.

[52] N. T. Aggarwal, R. C. Shah, and D. A. Bennett, "Alzheimer's disease: Unique markers for diagnosis & new treatment modalities," *Indian J Med Res,* vol. 142, pp. 369-82, Oct 2015.

[53] R. Sherva and L. A. Farrer, "Power and pitfalls of the genome-wide association study approach to identify genes for Alzheimer's disease," *Curr Psychiatry Rep,* vol. 13, pp. 138-46, Apr 2011.

[54] T. D. Bird, "Genetic aspects of Alzheimer disease," *Genet Med,* vol. 10, pp. 231-9, Apr 2008.

[55] K. Gardiner, D. Slavov, L. Bechtel, and M. Davisson, "Annotation of human chromosome 21 for relevance to Down syndrome: gene structure and expression analysis," *Genomics,* vol. 79, pp. 833-43, Jun 2002.

[56] D. Hartley, T. Blumenthal, M. Carrillo, G. DiPaolo, L. Esralew, K. Gardiner*, et al.*, "Down syndrome and Alzheimer's disease: Common pathways, common goals," *Alzheimers Dement,* vol. 11, pp. 700-9, Jun 2015.

[57] B. Reisberg, "Diagnostic criteria in dementia: a comparison of current criteria, research challenges, and implications for DSM-V," *J Geriatr Psychiatry Neurol,* vol. 19, pp. 137-46, Sep 2006.

[58] B. Dubois, H. H. Feldman, C. Jacova, S. T. Dekosky, P. Barberger-Gateau, J. Cummings*, et al.*, "Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria," *Lancet Neurol,* vol. 6, pp. 734-46, Aug 2007.

[59] A. American Diabetes, "Diagnosis and classification of diabetes mellitus," *Diabetes Care,* vol. 32 Suppl 1, pp. S62-7, Jan 2009.

[60] J. G. Goldman, N. T. Aggarwal, and C. D. Schroeder, "Mild cognitive impairment: an update in Parkinson's disease and lessons learned from Alzheimer's disease," *Neurodegener Dis Manag,* vol. 5, pp. 425-43, Oct 2015.

[61] E. D'Angelo, "Abstract VIII Congresso Nazionale Sindem 2013," *Functional Neurology,* vol. 28, p. 63, 2013.

[62] C. R. Jack, Jr., D. S. Knopman, W. J. Jagust, L. M. Shaw, P. S. Aisen, M. W. Weiner*, et al.*, "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade," *Lancet Neurol,* vol. 9, pp. 119-28, Jan 2010.

[63] I. Yakushev, M. J. Muller, H. G. Buchholz, U. Lang, H. Rossmann, H. Hampel*, et al.*, "Stage-dependent agreement between cerebrospinal fluid proteins and FDG-PET findings in Alzheimer's disease," *Curr Alzheimer Res,* vol. 9, pp. 241-7, Feb 2012.

[64] R. Nau, F. Sorgel, and H. Eiffert, "Penetration of drugs through the blood-cerebrospinal fluid/blood-brain barrier for treatment of central nervous system infections," *Clin Microbiol Rev,* vol. 23, pp. 858-83, Oct 2010.

[65] Q. Zhang, R. B. Stafford, Z. Wang, S. E. Arnold, D. A. Wolk, and J. A. Detre, "Microvascular perfusion based on arterial spin labeled perfusion MRI as a measure of vascular risk in Alzheimer's disease," *J Alzheimers Dis,* vol. 32, pp. 677-87, 2012.

[66] T. Muayqil, G. Gronseth, and R. Camicioli, "Evidence-based guideline: diagnostic accuracy of CSF 14-3-3 protein in sporadic Creutzfeldt-Jakob disease: report of the guideline development subcommittee of the American Academy of Neurology," *Neurology,* vol. 79, pp. 1499-506, Oct 02 2012.

[67] A. D. Dekker, J. Fortea, R. Blesa, and P. P. De Deyn, "Cerebrospinal fluid biomarkers for Alzheimer's disease in Down syndrome," *Alzheimers Dement (Amst),* vol. 8, pp. 1-10, 2017.

[68] H. Hampel, K. Blennow, L. M. Shaw, Y. C. Hoessler, H. Zetterberg, and J. Q. Trojanowski, "Total and phosphorylated tau protein as biological markers of Alzheimer's disease," *Exp Gerontol,* vol. 45, pp. 30-40, Jan 2010.

[69] H. Hampel and K. Blennow, "CSF tau and beta-amyloid as biomarkers for mild cognitive impairment," *Dialogues Clin Neurosci,* vol. 6, pp. 379-90, Dec 2004.

[70] S. C. Gupta, S. Patchva, and B. B. Aggarwal, "Therapeutic roles of curcumin: lessons learned from clinical trials," *AAPS J,* vol. 15, pp. 195-218, Jan 2013.

[71] L. D. Maxim, R. Niebo, and M. J. Utell, "Screening tests: a review with examples," *Inhal Toxicol,* vol. 26, pp. 811-28, Nov 2014.

[72] C. D. Aluise, R. A. Sowell, and D. A. Butterfield, "Peptides and proteins in plasma and cerebrospinal fluid as biomarkers for the prediction, diagnosis, and monitoring of therapeutic efficacy of Alzheimer's disease," *Biochim Biophys Acta,* vol. 1782, pp. 549-58, Oct 2008.

[73] C. Humpel, "Identifying and validating biomarkers for Alzheimer's disease," *Trends Biotechnol,* vol. 29, pp. 26-32, Jan 2011.

[74] C. A. Lemere and E. Masliah, "Can Alzheimer disease be prevented by amyloid-beta immunotherapy?," *Nat Rev Neurol,* vol. 6, pp. 108-19, Feb 2010.

[75] G. B. Frisoni, N. C. Fox, C. R. Jack, Jr., P. Scheltens, and P. M. Thompson, "The clinical use of structural MRI in Alzheimer disease," *Nat Rev Neurol,* vol. 6, pp. 67-77, Feb 2010.

[76] S. Shokouhi, D. Claassen, and W. Riddle, "Imaging Brain Metabolism and Pathology in Alzheimer's Disease with Positron Emission Tomography," *J Alzheimers Dis Parkinsonism,* vol. 4, Apr 2014.

[77] A. D. Atul Mallik, Satoshi Minoshima, "Clinical Amyloid Imaging," *Seminar in Nuclear Medicine,* vol. 47, pp. 31-43, 2017.

[78] K. A. Johnson, N. C. Fox, R. A. Sperling, and W. E. Klunk, "Brain imaging in Alzheimer disease," *Cold Spring Harb Perspect Med,* vol. 2, p. a006213, Apr 2012.

[79] D. A. Casey, D. Antimisiaris, and J. O'Brien, "Drugs for Alzheimer's disease: are they effective?," *P T,* vol. 35, pp. 208-11, Apr 2010.

[80] J. B. Standridge, "Pharmacotherapeutic approaches to the treatment of Alzheimer's disease," *Clin Ther,* vol. 26, pp. 615-30, May 2004.

[81] D. Olivares, V. K. Deshpande, Y. Shi, D. K. Lahiri, N. H. Greig, J. T. Rogers*, et al.*, "N-methyl D-aspartate (NMDA) receptor antagonists and memantine treatment for Alzheimer's disease, vascular dementia and Parkinson's disease," *Curr Alzheimer Res,* vol. 9, pp. 746-58, Jul 2012.

[82] N. B. Richard J.Hodges, "Accelerating Medicines Partnership: Alzheimer's Disease (AMP-AD) Knowledge Portal Aids Alzheimer's Drug Discovery through Open Data Sharing," *Expert Opinion on Therapeutic Targets,* vol. 20, 2016.

[83] P. A. Defina, R. S. Moser, M. Glenn, J. D. Lichtenstein, and J. Fellus, "Alzheimer's disease clinical and research update for health care practitioners," *J Aging Res,* vol. 2013, p. 207178, 2013.

[84] D. Borsook, "Neurological diseases and pain," *Brain,* vol. 135, pp. 320-44, Feb 2012.

[85] M. E. Kennedy, A. W. Stamford, X. Chen, K. Cox, J. N. Cumming, M. F. Dockendorf*, et al.*, "The BACE1 inhibitor verubecestat (MK-8931) reduces CNS beta-amyloid in animal models and in Alzheimer's disease patients," *Sci Transl Med,* vol. 8, p. 363ra150, Nov 02 2016.

[86] P. Novak, R. Schmidt, E. Kontsekova, N. Zilka, B. Kovacech, R. Skrabana*, et al.*, "Safety and immunogenicity of the tau vaccine AADvac1 in patients with Alzheimer's disease: a randomised, double-blind, placebo-controlled, phase 1 trial," *Lancet Neurol,* vol. 16, pp. 123-134, Feb 2017.

[87] K. M. Lucin and T. Wyss-Coray, "Immune activation in brain aging and neurodegeneration: too much or too little?," *Neuron,* vol. 64, pp. 110-22, Oct 15 2009.

[88] C. A. Parker, E. A. Rabiner, R. N. Gunn, G. Searle, L. Martarello, R. A. Comley*, et al.*, "Human Kinetic Modeling of the 5HT6 PET Radioligand 11C-GSK215083 and Its Utility for Determining Occupancy at Both 5HT6 and 5HT2A Receptors by SB742457 as a Potential Therapeutic Mechanism of Action in Alzheimer Disease," *J Nucl Med,* vol. 56, pp. 1901-9, Dec 2015.

CHAPTER 2

Development of Comprehensive Biomarkers Database for (AD) - ABCD

ABSTRACT

Alzheimer's disease (AD) is associate in age-related, non-reversible brain disorder. Memory loss, confusion and personality changes are major symptoms noticed. Final consequence of AD leads to a severe loss of mental activities. Due to lack of effective biomarkers, no effective medication available for the complete treatment of AD. There is a need to provide all AD related information to the scientific community. Our resource Alzheimer's Disease Biomarkers Comprehensive Database (ABCD) somewhat accomplished this objective by working as a big repositories of varied data related to AD. The web interface contains information concerning the proteins, genes, transcription factors, SNP's, miRNAs, mitochondrial genes and expressed genes implicated in AD pathogenesis. In addition to this molecular level data, the database has information for animal models, medicinal candidates and pathways involved in AD and some image data for AD patients, ABCD is coupled with some major external resources where user can retrieve additional information about the disease. The database was architected in such a manner that user can extract meaningful information easily. This database is unique in the sense that it is completely dedicated to specific neurological disorder (AD) of a human. Further advance options like AD affected brain image data of affected patients and structural compound level information add values to our database. Features of this database information enable users to extract, analyze and show information associate with a disease in many alternative ways. The database is hosted and accessible at http://www.bioinfoindia.org/abcd.

2.1 INTRODUCTION

Better health facilities within the past few decades have contributed to individual's longer and healthier lives. Advancement in medical technology conjointly increased the variety of individuals with non-communicable disease like dementia [1]. Dementia is a diseased brain condition, typically of severe or progressive nature affects thinking, memory and behavior to perform routine activities. AD is categorized as the most common form of dementia [2]. AD is noticed by an incurable continuous decline in psychological activities. Current ailment of the disease is basically symptom based and relies on three enzyme inhibitors which are donepezil, rivastigmine and galantamine which are affecting acetylcholine based system whereas memantine, affecting the glutamatergic system. Since 2003, no new medicines are passed by US Food and Drugs Administration (FDA) test for the treatment of AD. AD has well enhanced among individuals aged higher than sixty-five years, still there is no permanent cure that stops or reverses the progression of the disease which ultimately worsens the situation, and afterward results in the death of sufferers. Presently, there is no specific marker available which can be considered as medicinal drug target for AD identification and diagnosis that may make sure with a 100 percent certainty regarding AD identification. The challenges in front of scientific community include; understanding of abnormalities in gene regulation, protein-protein interactions, and also the resultant alterations in signal and metabolic pathways result in AD [3].

There is associate imperative would like for substantial advances withinin the research area of biomarkers for assessment of risk, identification of causative factors and disease progression monitoring [4]. Continuing efforts are still going on to achieve success, this includes develops new medicines that might slow or stop the progression of disease. Current studies on AD are underway to spot biomarkers for diagnosing and new medical specialty to hinder disease progression. Acceleration in the biomarker identification, disease prognosis and diagnosis has been observed in various cancers such as colorectal [5] and thyroid cancer [6] with the help of repositories created from the scattered data from public domains and review literature. Therefore, our target was to collect and consolidate this data in a single repository for assessing different AD biomarkers, respective feasibility to be a drug candidates and role in subsequent pathways [7]. With this aim, we designed Alzheimer's Disease Biomarkers Comprehensive Database (ABCD); which is a completely AD dedicated online data resource

for storing and retrieving various components in the form of knowledge to researchers, doctors and caregivers. The amount and quality of data available in ABCD is manually curated and most of the available data are collected from experimentally generated data sources which renders ABCD the foremost comprehensive in itself.

Exponential growth of experimental information with biological studies and also the development of recent tools created it easier for the analysis community to research AD information [8]. The bottleneck of AD study lies in information analysis, as a result of the quality of knowledge analysis depends on multitude of databases, tools and nonuniformity of knowledge concerned within the study [9]. We have a tendency to collate associated data related to AD that is scattered in various web resources for the examination of increasingly massive biological information of AD.

The information available in ABCD is for open access to all and allows user to easily browse the data associated with AD and their related molecular consequences. The platform supports various kind of user input in context of system biology to develop healthcare application for AD. Also, network level understanding of pathways using concrete information can help scientific community to pave a path to resolve AD complexity. Molecular level information retrieval portal can interrogate the information from user-friendly interface. ABCD database management system is shown in Figure1. Its comprehensiveness, standardization, free availability, ease of accessibility and support of various user profiles create ABCD a resource of option to investigate diseases of neurodegeneration.

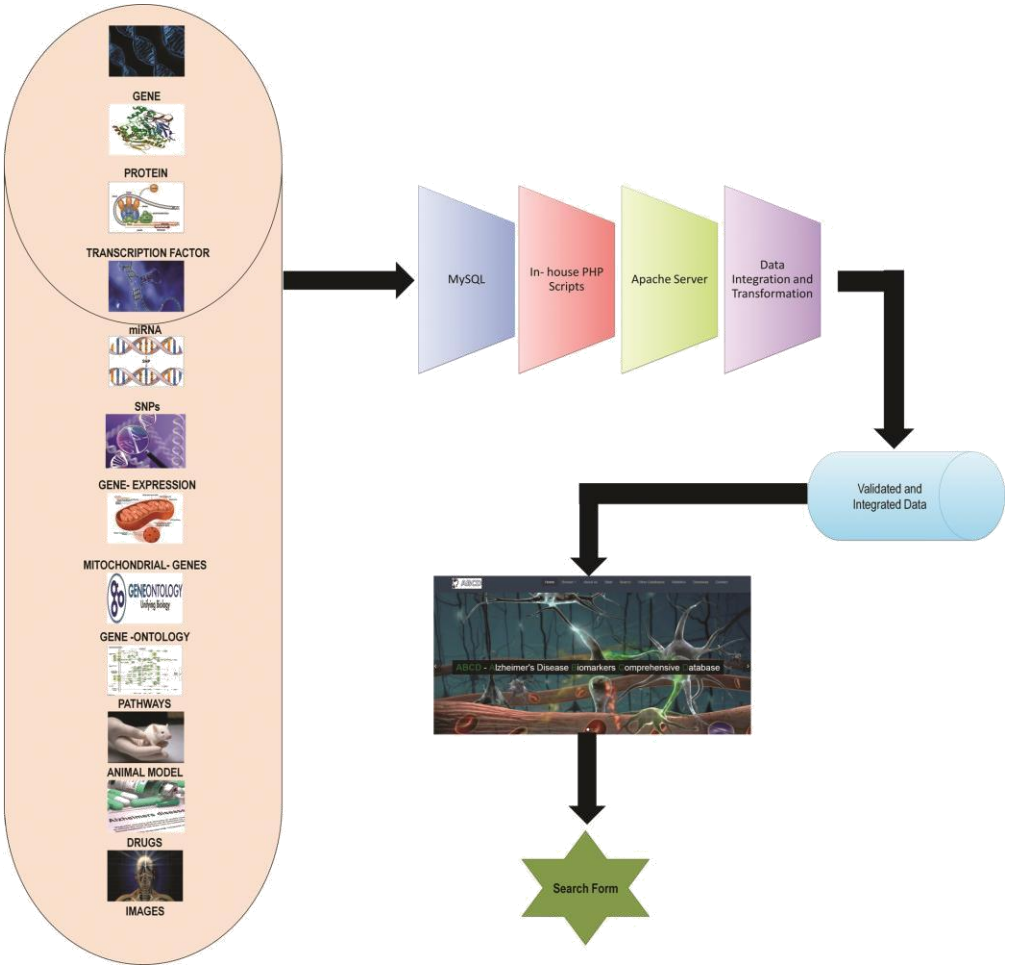


Figure 2.1: Workflow for data storage, processing and mining useful information of Alzheimer's disease from ABCD

2.2 MATERIAL AND METHODS

ABCD is a type of relational database using MySQL for database construction with an online interface that was developed by using HTML, CSS (Cascading Style Sheet) and JavaScript and hosted on Apache http server [10]. Figure 2 shows the manually designed entity–relationship model for ABCD and Figure 3 represent the architecture of ABCD which describes that data extraction from literature and other external resources and their in-between association which allows user to access and retrieve information in easy way without putting much effort.

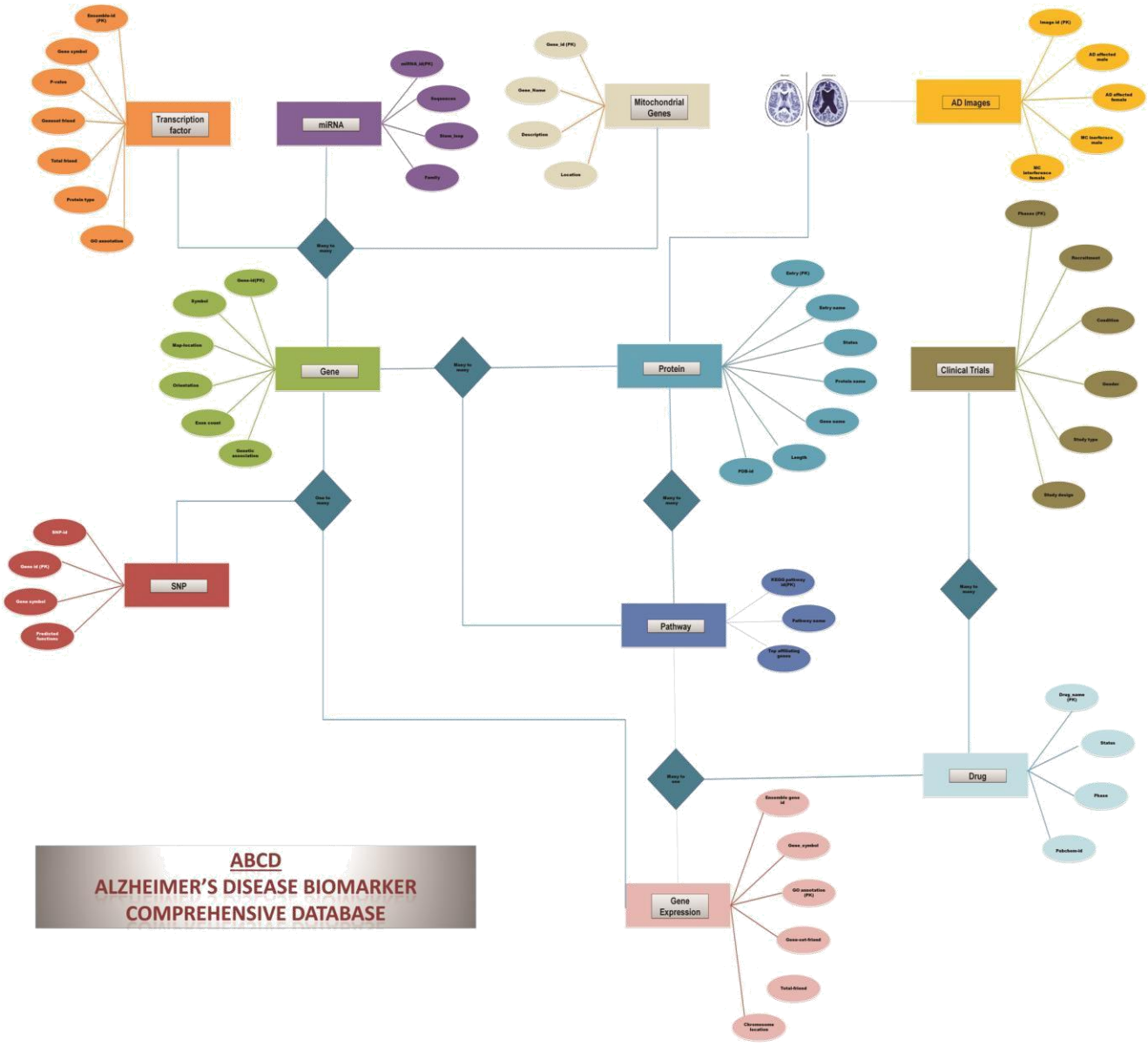


Figure 2.2: Entity-Relationship (E-R) diagram for ABCD

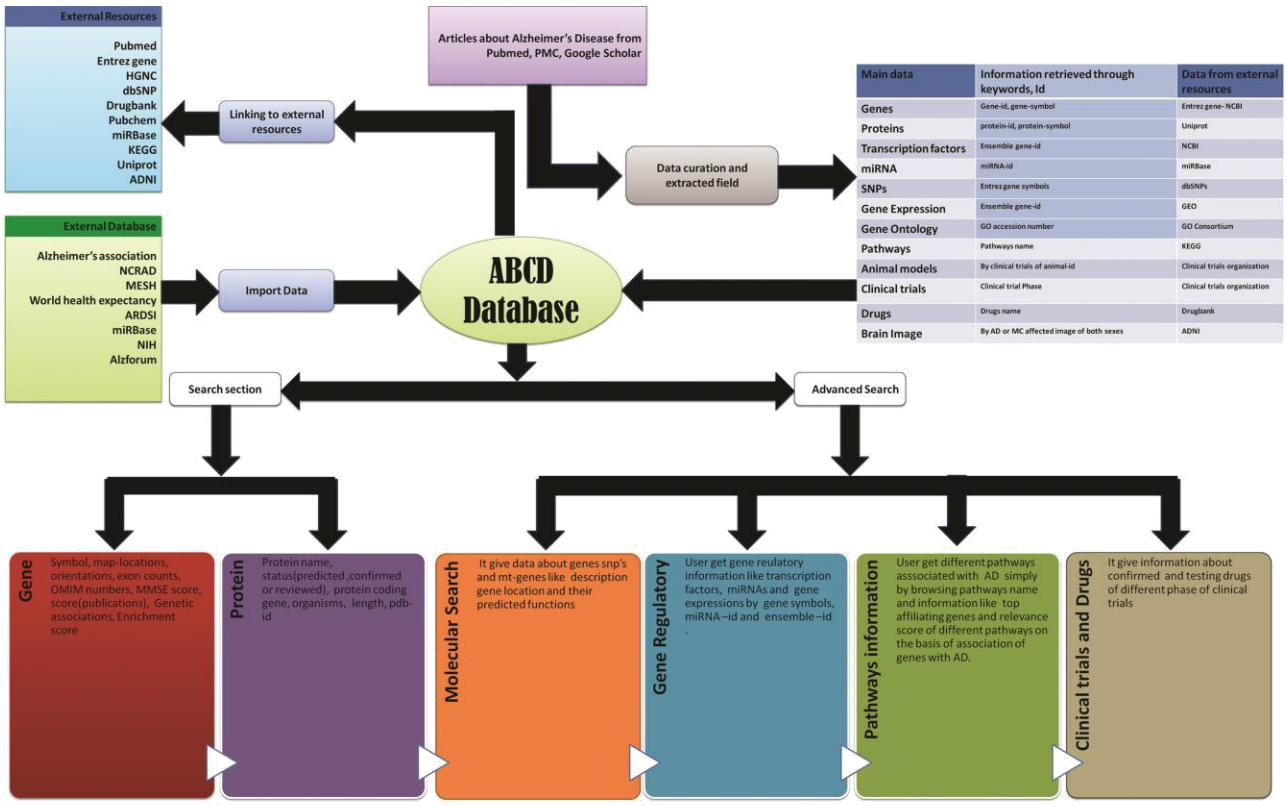


Figure 2.3: ABCD architecture; the main features of the ABCD platform and their integration and user access mode.

2.2.1 Large data collection from standard resources for ABCD

The data deposited in ABCD database has been retrieve from literature and standard and authenticated online resources like PubMed, PMC, Google Scholar, MEDLINE, NCBI and other resources [11]. ABCD contains genes, SNPs and miRNAs for providing gene regulatory information to the researcher. For miRNA related data, user get verified information from miRBASE [12]. A custom PHP script helped us to obtain molecular data from NCBI [13]. Information about drugs were obtained from Chembank [14]. ABCD includes manually curated data extracted from the published research articles. The data available in the present version of the ABCD covers 843 publications gave information about 449 genes and 767 miRNAs as well as 404 drugs . ABCD contains manually curated data filter from published research articles. The quantitative data present in current version of database comprises 843 publications contain information about 449 genes and 767 miRNAs as well as 404 drugs [15].

2.2.2 ABCD external links

To enrich the contents of ABCD, association with various external databases was established, which includes (i) World health Expectancy (for statistics of AD and other diseases), medical subject heading (MESH) [16] for medical literatures related to AD; (ii) National cell repository for AD (NCRAD) to retrieve genes that increase risk for AD and dementia funded by National Institute of Ageing (NIA)[17]. (iii) Alzheimer's International (AI) association is the important health organization in AD concern, support and research to have information ragarding causes, risk factors, diagnosing and clinical trials studies like pharmaceutical classes and medical care indications [18]; (vi) Alzforum portal contain repository of biomarkers, literatures, risks, antibodies, animal model, mutation studies and therapeutics [7], National Institute of health (NIH), an international body for human services have quick link to several institutes work under NIH [19]. miRBase has lot of analysis material coupled with miRNAs and endogenous molecules [20]. In addition, Alzheimer's and Related disorders Society of India (ARDSI) link was also connected with ABCD; which is a body to make better life of people who are suffered from dementia and maintain record of the people affected by AD in India [21].

2.3 DATA INCORPORATION

The main modules incorporated in ABCD are; genes enquiry, proteins enquiry, molecular search, gene regulatory related search and advanced Search. These allow the user; (i) to explore the database by searching genes and proteins by their name and id. In advanced search, users can search SNPs regions and mitochondrial gene in molecular search option whereas transcription factors and co-expressed gene under gene regulatory information search. Drug details, pathway catalogs and image information of AD affected brain were provided separately. For each type of enquiry, ABCD help the user by providing sample input to avoid wrong entries. There are 2 levels of filter checking on inputs, a client side and a server side. The input nomenclature relies on the official scientific standard commonly employed by on-line databases. The statistics on the ABCD information are according at the side of a quick documentation on the usage of the information content.

2.3.1 Search and Advanced Search

In search section users obtain their query from ABCD system by genes and proteins. In the advanced search option, users can search ABCD by mitochondrial gene, co-expressed gene, SNPs, Transcription factors (TF), micro RNAs (miRNA), pathways and drug target. The drug target section follow the genes that are targets of medicine utilized in the remedies. Genes are featured by the nomenclature of HGNC [22] and Entrez-gene by NCBI [23]. medicine are searched by their names as reported in Chembank [14]. SNPs are stored in database by rs# number corresponding to the nomenlature in dbSNP [24]. Mitochondrial gene, co-expressed gene, SNPs are also sub-leveled in advanced search option in gene category. Major categories of ABCD listed as below.

Genes: Genes search show the location of genes in chromosome and number of exon count with their genetic association and enrichment score describe about the overrepresentation of gene from geneset and association with disease phenotype [25]. It also reports the co-expressed gene which are expressed in symbiosis, encodes the role of genes.

Proteins: Proteins search provide high quality information about protein physiochemical properties with their structure and function related to AD [26].

Transcription factor: Transcription factors of AD genes are obtained from DAVID gene annotation tools [27].

SNPs: SNPs search show the SNPs out there among the genes and various SNP-ids is coupled to dbSNP. ABCD shows put together the gene name coupled to Entrez-gene card or the microRNAs (miRNAs) access number containing the SNP [28].

miRNA **:** miRNAs are a family of short, 21-22 nucleotides-long non-coding RNAs, constituting concerning 1% of all human genes and therefore the most teeming category of small RNAs in animals. miRNA provide data concerning non-coding genes of AD with chromosomal location and family data concerning several miRNA.

Drugs: Drug search shows the genes a target and connected to DrugBankV4.2 in conjunction with its kind and an overview, if offered. The knowledge are extracted from DrugBank (targets, transporters and enzymes) or computed by the DT-Hybrid algorithmic program [29]. For all different information, this section report details on drug (if any) that in ABCD articles are associated with the searched item. In details, there are gene names coupled to Entrez by NCBI

[23], endogenous molecules involved and additionally the actions that the drug has on the targets. There may also be information regarding the medical aid indications, pharmacodynamics, pharmacologic action and eventually the impact of drug on AD, extracted from the indexed articles.

Clinical trials: It contains information about drugs candidates, their descriptions, clinical testing starting- closing dates, stages of phase trials, etc. [30].

Pathways: Pathways provide the number of pathways involved in AD in which top affiliating genes are involved [31].

Image data: AD affected brain image data are displayed on the basis of patient-id of both male and female [32].

For each specific data out there within the database, users will visualize all details and relations with the genetic components of which might be reported within the results by clicking search options.

2.4 DATA DISTRIBUTION

Statistics section reports the quantity of information by class and Figure 4 shows the ABCD statistics. The quantity regarding the manually curated information is that the following: 499 genes, 259 proteins, 66 SNPs, 404 drugs, 1608 clinical trials information etc. The data which is presented in ABCD are enriched by using external sources like alzforum, NIH, miRBASE etc. Hence, the data collected would be a unified information portal for the AD research community.

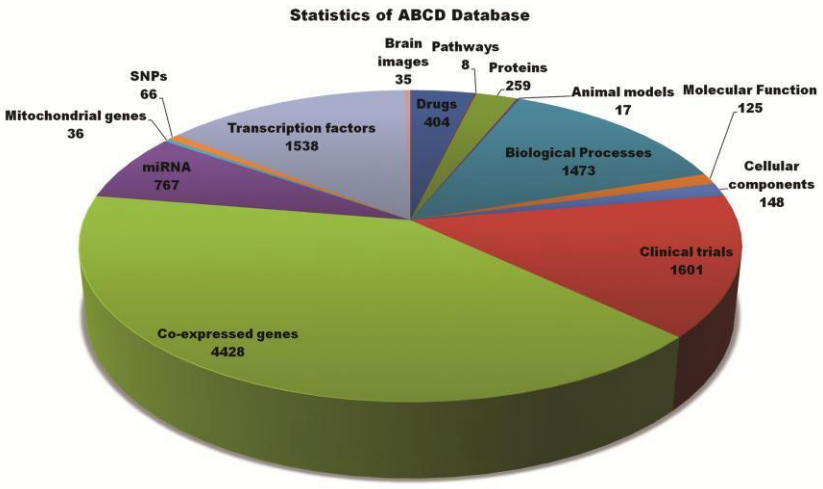


Figure 2.4: Data distribution of ABCD in term of data are represented through pie-chart

2.5 DATA MAINTENANCE

ABCD are unendingly updated, through manual screenings of recent publications on PubMed. Therefore, the manual procedures can extract and appraise genetic and network level information which is able to be incorporated in ABCD. Additionally, researchers will recommend new or missing findings to be inserted within the database by contacting authors by email in our 'Contact us' page. Agile approach has been adopted in designing, so it's simple to update ABCD at any point of time.

2.5.1 Data types distribution

The whole database was dumped along with the SQL scripts in Apache server and available on bioinfoindia.org/abcd. Search results (i.e. by gene, protein, MT-gene, SNPs, transcription factor, co-expressed genes, drug, clinical trials, image and pathways) can be visualized in HTML format through the browse and search section of the menu.

2.6 CONCLUSION

Through manual screening of literature and complimentary information retrieval from online resources, we tend to conclude on the fact that ABCD presents solid and reliable data for AD. Till date, in molecular level most studies are conducted on genes and proteins therefore the limited biomarkers are available for AD. Nevertheless, this study is based on a number of assumptions that are questionable today within the scientific community. To fill this gap; gene expression studies of varied genes are stored in ABCD. It joins bits of missing information scattered publicly archives and associated publications, into an identical, simply accessible and often updated information resource. Our system ABCD will facilitate to offer a comprehensiveness concerning the most genes, proteins, SNPs, medication or miRNAs associated within the pathology. They are extremely necessary for selecting the correct subsets of studies to answer complicated biological queries underlying AD pathology. By querying ABCD, researchers can produce new hypothesis and infer novel knowledge in the form of data. Our results shows that the number of data that are scattered in numerous resources, need in depth manual effort to capture the same. In addition, we tend to report that even with comprehensive manual gather; we weren't able to capture 100% of information to fill for the

fundamental annotation fields. Subsequently, we tend to commit to extend the curation pipeline by adding a lot of options in the form of information for AD particularly; brain image information was also incorporated. Although, microarray studies are the most important contributors to the general public repositories, RNA-Sequence information are quickly growing. We tend to comprehend our database in future so that it will cover all the relevant RNA-Sequence studies, since their massive storage space has contributed to disperse nature of the marker information. The presented database offers great potentials to the scientific world and it is anticipated that ABCD will help the mankind through information dissemination for worldwide monitoring and personalized effective biomarkers search for AD.

REFERENCES

[1] N. Xiao, Q. Long, X. Tang, and S. Tang, "A community-based approach to non-communicable chronic disease management within a context of advancing universal health coverage in China: progress and challenges," *BMC Public Health,* vol. 14 Suppl 2, p. S2, 2014.

[2] D. Aygun and I. L. Gungor, "Why is Alzheimer disease confused with other dementias?," *Turk J Med Sci,* vol. 45, pp. 1010-4, 2015.

[3] S. V. Vasaikar, A. K. Padhi, B. Jayaram, and J. Gomes, "NeuroDNet - an open source platform for constructing and analyzing neurodegenerative disease networks," *BMC Neurosci,* vol. 14, p. 3, Jan 03 2013.

[4] A. P. Privitera, R. Distefano, H. A. Wefer, A. Ferro, A. Pulvirenti, and R. Giugno, "OCDB: a database collecting genes, miRNAs and drugs for obsessive-compulsive disorder," *Database,* vol. 2015, 2015.

[5] A. Shukla, A. Moussa, and T. R. Singh, "DREMECELS: a curated database for base excision and mismatch repair mechanisms associated human malignancies," *PloS one,* vol. 11, p. e0157031, 2016.

[6] A. Bansal and J. Ramana, "TCGDB: A Compendium of Molecular Signatures of Thyroid Cancer and Disorders," *Journal of Cancer Science & Therapy,* vol. 7, pp. 198-201, 2015.

[7] J. Kinoshita and T. Clark, "Alzforum," *Methods in molecular biology (Clifton, NJ),* vol. 401, pp. 365-381, 2007.

[8] J. C. de la Torre, "Is Alzheimer's disease a neurodegenerative or a vascular disorder? Data, dogma, and dialectics," *The Lancet Neurology,* vol. 3, pp. 184-190, 2004.

[9] M. Thangam and R. K. Gopal, "CRCDA—Comprehensive resources for cancer NGS data analysis," *Database,* vol. 2015, 2015.

[10] J. C. Meloni, *Sams teach yourself PHP, MySQL and Apache all in one*: Sams Publishing, 2012.

[11] T. Greenhalgh, "How to read a paper. The Medline database," *BMJ: British Medical Journal,* vol. 315, p. 180, 1997.

[12] A. Kozomara and S. Griffiths-Jones, "miRBase: annotating high confidence microRNAs using deep sequencing data," *Nucleic acids research,* vol. 42, pp. D68-D73, 2013.

[13] D. Maglott, J. Ostell, K. D. Pruitt, and T. Tatusova, "Entrez Gene: gene-centered information at NCBI," *Nucleic Acids Res,* vol. 39, pp. D52-7, Jan 2011.

[14] K. P. Seiler, G. A. George, M. P. Happ, N. E. Bodycombe, H. A. Carrinski, S. Norton*, et al.*, "ChemBank: a small-molecule screening and cheminformatics resource database," *Nucleic acids research,* vol. 36, pp. D351-D359, 2007.

[15] A. P. Ayala, "Polymorphism in drugs investigated by low wavenumber Raman scattering," *Vibrational Spectroscopy,* vol. 45, pp. 112-116, 2007.

[16] C. E. Lipscomb, "Medical subject headings (MeSH)," *Bulletin of the Medical Library Association,* vol. 88, p. 265, 2000.

[17] B. N. Vardarajan, K. M. Faber, T. D. Bird, D. A. Bennett, R. Rosenberg, B. F. Boeve*, et al.*, "Age-specific incidence rates for dementia and alzheimer disease in nia-load/ncrad and efiga families: National institute on aging genetics initiative for late-onset alzheimer disease/national cell repository for alzheimer disease (nia-load/ncrad) and estudio familiar de influencia genetica en alzheimer (efiga)," *JAMA neurology,* vol. 71, pp. 315-323, 2014.

[18] C. R. Jack, M. S. Albert, D. S. Knopman, G. M. McKhann, R. A. Sperling, M. C. Carrillo*, et al.*, "Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's & Dementia,* vol. 7, pp. 257-262, 2011.

[19] K. Strimbu and J. A. Tavel, "What are biomarkers?," *Current Opinion in HIV and AIDS,* vol. 5, p. 463, 2010.

[20] S. Griffiths-Jones, R. J. Grocock, S. Van Dongen, A. Bateman, and A. J. Enright, "miRBase: microRNA sequences, targets and gene nomenclature," *Nucleic acids research,* vol. 34, pp. D140-D144, 2006.

[21] M. Varghese, "MP 6 Dementia: perspectives from developing countries," *J Neurol Neurosurg Psychiatry,* vol. 83, pp. e1-e1, 2012.

[22] E. A. Bruford, M. J. Lush, M. W. Wright, T. P. Sneddon, S. Povey, and E. Birney, "The HGNC Database in 2008: a resource for the human genome," *Nucleic Acids Res,* vol. 36, pp. D445-8, Jan 2008.

[23] D. Maglott, J. Ostell, K. D. Pruitt, and T. Tatusova, "Entrez Gene: gene-centered information at NCBI," *Nucleic acids research,* vol. 39, pp. D52-D57, 2010.

[24] S. T. Sherry, M.-H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski*, et al.*, "dbSNP: the NCBI database of genetic variation," *Nucleic acids research,* vol. 29, pp. 308-311, 2001.

[25] J. M. Hertz, M. Thomassen, H. Storey, and F. Flinter, "Clinical utility gene card for: Alport syndrome–update 2014," *European Journal of Human Genetics,* vol. 23, 2015.

[26] A. Bairoch, R. Apweiler, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro*, et al.*, "The universal protein resource (UniProt)," *Nucleic acids research,* vol. 33, pp. D154-D159, 2005.

[27] G. Dennis, B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane*, et al.*, "DAVID: database for annotation, visualization, and integrated discovery," *Genome biology,* vol. 4, p. R60, 2003.

[28] A. D. Johnson, R. E. Handsaker, S. L. Pulit, M. M. Nizzari, C. J. O'donnell, and P. I. De Bakker, "SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap," *Bioinformatics,* vol. 24, pp. 2938-2939, 2008.

[29] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard*, et al.*, "DrugBank: a comprehensive resource for in silico drug discovery and exploration," *Nucleic acids research,* vol. 34, pp. D668-D672, 2006.

[30] A. R. Jadad, R. A. Moore, D. Carroll, C. Jenkinson, D. J. M. Reynolds, D. J. Gavaghan*, et al.*, "Assessing the quality of reports of randomized clinical trials: is blinding necessary?," *Controlled clinical trials,* vol. 17, pp. 1-12, 1996.

[31] H. Mi, N. Guo, A. Kejariwal, and P. D. Thomas, "PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways," *Nucleic acids research,* vol. 35, pp. D247-D252, 2006.

[32] R. C. Petersen, P. Aisen, L. Beckett, M. Donohue, A. Gamst, D. Harvey*, et al.*, "Alzheimer's disease Neuroimaging Initiative (ADNI) clinical characterization," *Neurology,* vol. 74, pp. 201-209, 2010.

CHAPTER 3

Supervised machine learning approach (Decision Tree) to classify and ontological study of geneset of AD

ABSTRACT

AD is known by frequent memory loss, orientation, thinking ability and language. There's no cure on the market for this disease. Molecular mechanism of this disease remains not clear. Many theories have shown that the abnormal storage of neuronic assisted biomolecules specifically proteins 10 to 20 years prior to symptoms of the disease appear, thus it's extraordinarily necessary to spot changes within the brain before the primary symptoms. Mining useful data and providing scientific inference for the diagnosing and ailment for the malody from the medical series of data is so progressively changing into necessary. We used decision tree to classify 2111 gene dataset from proteinpedia database of Human protein reference database (HPRD). Decision tree model for varied AD genes were generated on the basis of MMSE Scores and alternative very important parameters to identify significant proteins on the basis of expression occur in disease verify the involvement of genes in various AD pathologic process pathways. The robustness of predicted machine learning model for AD geneset is set on basis of information gain with confidence value (0.96), specificity (92 %), sensitivity (98 %) and accuracy (77 %). Along with this classification studies, enrichment analysis of these genes were performed. This study indicates that the quantitativeness of all the genes evaluated in single sets are sufficients to cure AD and can facilitate within the hypothetical prediction of necessary parameters for alternative associative measures.

3.1 INTRODUCTION

Development of various algorithms and data mining techniques helped researchers to resolve bioinformatics problems in efficient way [1]. Advancement of technologies and acceleration within the variety of analysis will increase in progressive manner and huge amount of information in form of data is being generated. So as to investigate and procure results from these data, Machine Learning techniques are essential for obtaining useful insight.

Data mining which is one of popular machine learning approach is applied to datasets for 2 objectives. The prime goal is to classify data and structure using various clustering approaches. Before implementing various data modeling techniques and algorithms, there is a need of preprocessing data for the systems understanding. These preprocessing steps involves data cleaning, aggregation, transformation and dividing it into small subsets to reduce the dimensions[2]. The ultimate process is that the visualization and illustration of main finding. It is to be noted that dimensional reducing tools are proved to be very useful in studies of high dimensional data like NGS information which includes, Genomic information and Transcriptomic information etc.

It has been observed recently that result from the mining of genetic studies provides very interesting outcomes about inheriting properties to contract specific diseases [3]. Alzheimer's disease (AD) is a multistage and inheritable disease. In order to explore available human gene dayasets, we investigated and produced rules for the identification of late-onset AD supported genetic data from the hospital, we have connected decision tree algorithm after the prioritization of some important rule as a desire to broaden the precision rate of classification.

Models for data processing falls in two main categories first one is, prognostic models and second one is expressive model [4]. Prognostic models are helpful in comprehension of class, whose class name isn't perceived yet the utilization of a model construct absolutely in light of recorded measurements which as of now have class names [5]. On other side, clustering model use some patterns to find homologous datasets

[6]. A classification is technique by which user can predict the specific class labels for supplied data or future information. There are various other techniques used in the classification of datasets such as decision Trees, ANN, GA and Naive Bayesian theorem Classification methods.

3.1.1 Decision Tree

Decision Tree is the basic supervised machine learning based techniques which is used for classification purpose for data analysis [7]. Generated rules are clear and simple. Decision tree has root and edges which are identical to the structure of the real tree. Tree structure is mostly used method for data prediction, and specifically used by various researchers in decision making to identify a their goals [8].

The principle motivation behind decision tree is to create rules. It is outstanding technique in among others strategies like Neural Networks, Bayesian Networks and so on. At same time, decision trees are similarly less expensive to remake the answer for a given issue, simpler to translate for producing rules, less demanding to coordinate data with database frameworks and have higher autonomy contrasted with others, they are the extensively utilized techniques for the grouping models. There are numerous different calculations for developing decision trees like C4.5, ID3 and CHAID.

Decision tree performs classification of given datasets and priopritization is done on the basis of root node selection and moving data upto leaf node and this is referred as class [9]. Decision tree usually consist two stages, one is classification of dataset and other is training of dataset.

Decision tree arranges new instances going from the root node and moving it till a leaf node is achieved, which is the label of a class [9]. Decision tree obey two task for classification of data. Beginning stage is that the training phase. At this stage, a training set whose classification names are estimated before is broke down by the order calculations to develop a model. Prepared model is shown as arrangement standards or tree structure. The second stage is classification. In classification step, testing data is utilized to plot the accuracy rate of model. On the off chance that the accuracy rate is fitting, standards can be utilized to classify new data whose class label is obscure [10]. Pictorial representation of decision tree building is shown in figure 3.1.

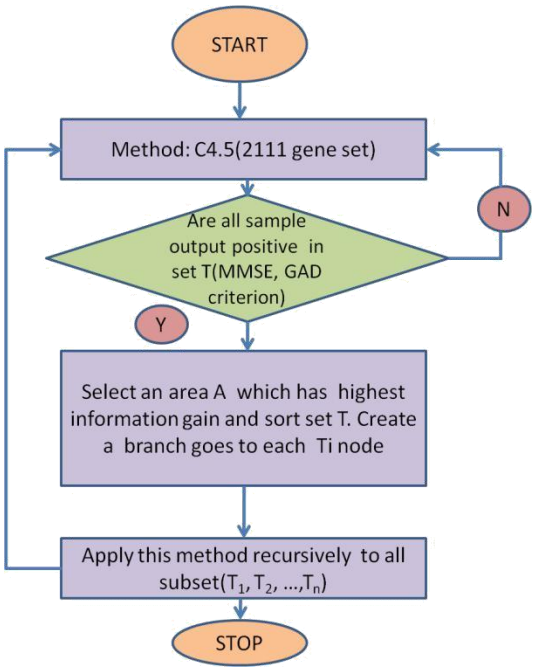


Figure 3.1: Flowchart for Decision Tree building

The decision trees are used in applied field of science as drugs (diagnosis), datastructure, biological sciences for the purpose for arranging different states into categories like high, medium, and low risk groups, and generating rules accordingly [11].

3.2 MATERIAL AND METHODS

3.2.1 Datasets: In this work, we have considered two well recognized databases Human protein reference database and National Centre for Biotechnology Information for data collection. 2111 genes were downloaded in total which are known to be linked with AD. Supplementary table I represents the downloaded gene information which were considered for further analysis in this study. There were different features which were selected on the basis of various literature mining approaches and finally concluded in 14 attributes.[5, 12].

Considered gene set refers the information for genes which are not only or not linked with AD but also linked with other diseases. The attributes for this work, was selected on the basis of various attribute selection methods such as ranker technique. For selected gene set, five different methods were considered to produce significant results. Chi –squired method for qualitative feature selection [13]. Also, to reduce the missing values and effect of biasness and gain, gain ratio was used [14].

Gain (Class, Attribute) = H (Class) - H (Class/Attribute); where H is the information entropy.

3.2.2 C4.5 Algorithm

C4.5 algorithm is invented by Quinlan []15]. Divide and conquer rule and properties of the ID3 algorithm is inherited to given algorithm which helps in construction of predictive models. The prime role of ID3 algorithm is to deal with nominal data and C4.5 techniques were best suited as inherited from ID3 method.

There are various other features in C4.5 algorithm which includes handling of missing values. There are many instances in real world or in simulation preprocessing data which can be handled by given algorithm. Errors of missing values usually dealt with calculating mean or median of the supplied values. Henceforth, by using this C4.5 algorithm missing values in data will a negligible or low rate issue. Also, accuracy of model is optimized and doesn't affect results in reverse way [16].

3.3 ARCHITECTURE AND DESIGNED MODEL

Decision tree model implicated is represented in Figure 3.2. AD genes were collected from various online resources amd then there was a need of preprocessing of data. There were also cases of missing values which were taken care by taking mean or median of the relative values in the downloaded data. Considering downloaded geneset of 2111 genes from various resources, we applied two machine learning tools to optimize and understand the results. One of the tool used is RapidMiner and other one is Weka for various data mining operations. After developing model, there was a need to validate the same, which was done using 10 fold cross validation. Collected gene data consists various parameters which usually referred as descriptors, such as chrom_position, gene names, and association score [17]. Enrichment analysis was subsequently performed by using cluster algorithms. Threshold was kept 0.3 to keep associated genes strongly bound in each cluster, which means anything below this threshold is noise or unknown information. Gene ontology information, which was depicted using biological process, molecular functions and cellular components were considerd for further analysis.

3.4 RESULT AND DISCUSSION

Classification of genedataset was performed separately in RapidMiner and Weka tool applying C4.5 Algorithm and J48 algorithm respectively. This DT classification techniques leads to thproper classification of 950 genes out of 2111 genes. MMSE and Huge navigator were the most important variable as they have low information gain. The classification formula provides

MMSE score cutoff value for a distinct stage of the disease. DT have a tendancy to generate rule every time on iteration by applying different criteria at everytime [18].

In the wake of preprocessing of the data that include cleanup, combination and information reduction, the data mining was finished. All investigations amid this examination include 10-fold cross validationl strategy to check the measures while not pruning the tree. Single dataset of 2111 genes was utilized to fabricate decision tree (DT). The dataset contains information with respect to important attributes identified with genes. To upgrade the execution of classifier , regulated resample classifier was utilized on data to check its execution quality. The quantity of tree was kept consistent at 100 though the measure of choices was kept differed at different points. Validity of built decision tree relies upon precision that was estimated 77% with sensitivity 86% and specificity of the classifier was 81%. On applying the apriori algorithm, classifier produces 10 best standard rule and was also confirmed.

Table 3.1: The relation generate between different classes on the basis of Apriori algorithm and the most appropriate rule obtain from classifier

Minimum support 0.2		
Minimum confidence 0.9		
Number of cycle performed 16		
Best rules found		
a1=false a5=false 24	class=c0 24	conf:(1)
a5=false a8=false 24	class=c0 24	conf:(1)
a5=false a6=false 23	class=c0 23	conf:(1)
a8=false class=c1 22	a5=true 22	conf:(1)
a5=false a7=true 21	class=c0 21	conf:(1)
a5=false a9=false 21	class=c0 21	conf:(1)
a3=false a5=false 20	class=c0 20	conf:(1)
a6=false class=c1 20	a5=true 20	conf:(1)
a2=false a5=false 27	class=c0 26	conf:(0.96)
a4=false a5=false 23	class=c0 22	conf:(0.96)

3.4.1 Initiation of decision tree

C4.5 algorithmic program were used within the RapidMiner statistical tool was utilized to make decision tree. For making the decision tree, those features were decided for classification whose values weren't steady. Also we tend to discretized the data based on recurrence. Totally different color nodes at bottom of the tree reveals the class definition of leaf nodes as appeared in figure 3.2 and summary table 3.2 for decision tree are given below.

Table 3.2: Overview of decision tree classifier (DT-I)

Learning Algorithm	C4.5
Attribute selection criterion	specifies the used method for selecting attributes, we choose gain ratio for this criterion
Minimal size for split	4
Minimal leaf size	1
Minimal gain	0.1
Maximal depth	20
Confidence value	0.25
Number of prepruning	3

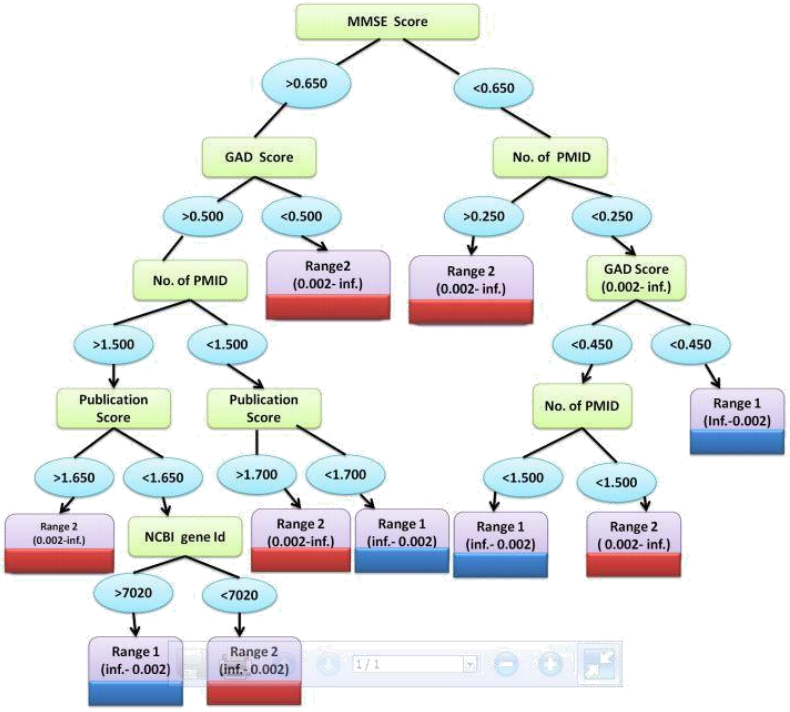


Figure 3.2: DT building for AD. MMSE score are the key point which classification of gene data is done. The cut-off value for generation of DT are obtained from threshold value obtained from C4.5 algorithm, Proper classification required different pubmed id and NCBI gene id.

3.4.2 Mining association rule from statistical tool

DT gain popularity as they are simple to understand. If the branches of tree will increase, then conditiona of IF-THEN is applied to generate rules from the tree. Rule is designed by paving the path from root to leaf node exploitation of logical condition of AND or IF. Overview of DT-II are given in below table.

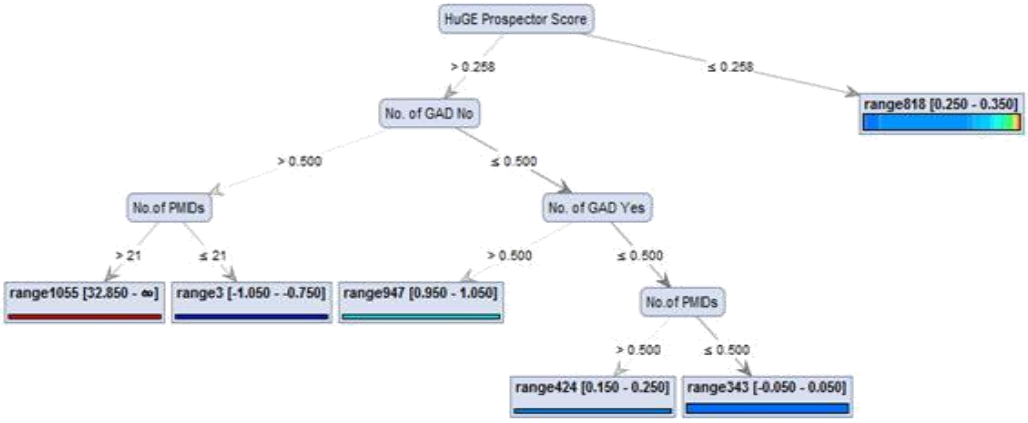


Figure 3.3: DT building on the basis of another important parameter (HUGE) and GAD for classification

Table 3.3: Overview of decision Tree(DT-II)

Applied Algorithm	J48
Attribute selection criterion	Gain ratio
Input	2111 gene dataset
Minimal Gain	0.01
Maximal depth	12
Validation	10-fold cross validation
Minimal split size	5
Minimal leaf size	1
Number of pre-pruning	3

Performance of each decision tree classifier was done through confusion matrix. The first decision tree solely take representative geneset as input take a look at samples accurately at 98.94% using 10- fold cross validation strategy.

Table 3.4: Evaluation table for measuring sensitivity and specificity for DT-I (Without missing values)

	True range	True range	Class precision
Pred. Range	1058	13	98.79%
Pred. Range	2	348	99.43
Class Recall	99.81%	96.40%	

Table 3.5: Evaluation table for measuring sensitivity and specificity for DT-II (With missing values using J48 method

	True range	True range	Class precision
Prediction Range	946	13	75.7%
Pred. Range	2	460	99.43%
Class Recall	75.7%	83.5%	

3.4.3 DT-II built in WEKA tool using J48 algorithm

The built decision tree is evaluated by means of matrix which tell about sensitivity and specificity (3.4; 3.5). The accuracy derived from this DT-I is found to be about 98.94 on applying 10-fold cross validation techniques to check performance of model. From available gene data sample , 10 rules were derived for prediction of disease in DT-I and 6 rules in DT-II. On the idea of normal classification system of MMSE score, performance of the model was improved.

3.4.4 Analysis of data from weka

The geneset contain 499 genes and 14 attributes, now at this time data were classified using J48 algorithmand naïve bayes algorithm. Top feature for selection of attributes were selected on the basis of feature selection techniques. Many techniques are there for feature selection for classification but ranker search method is highly robust among others.Once classification model was built its accuracy are predictively measured and it gave good and better output.

Table 3.6: Attribute Selection method- Attribute ranking

Attribute Evaluator	**Supervised Filtered Attribute evaluator**	
Selection method	**Ranker Search Method**	
Score	**Rank**	**Attribute**
0.255714	6	a5
0.038926	4	a3
0.024319	9	a8
0.009714	3	a2
0.005152	8	a7
0.003551	10	a9
0.003551	2	a1
0.002202	7	a6
0.000531	5	a4
0.000168	1	a0

Robustness of Decision Tree (DT) using C4.5 algorithm in statistical analysis tool Rapidminer

- Model of classification for different AD gene were generated on the basis of root attribute Mini-mental state examination(MMSE) whose information-gain value is high and is key attribute from where classification of geneset were taken place.
- Effectiveness of this generated tree were generated
  by: Confidence value- 0.96
  Specificity- 92%
  Sensitivity- 98%

Accuracy (10-fold cross validation): 77%(Avg.)

On comparing the result, we used same dataset in another popular statistical tool for classification using J48 algorithm analogous to C4.5 Algorithm and applied test on training data of AD genes, the resulted summary result were given below:

- Number of instances: 100 (default)
- Attributes: 14
- Test:2111(Training set)
- Number of Leaves: 10
- Size of the Tree: 5
- Correctly classified instances: 93%
- Incorrectly classified instances: 7%
- Accuracy(10-fold cross validation): 77% (Avg.)

From the above observation we concluded that classification through supervised decision tree methods were popular as it is easy to understandable and perfect for handling missing values. Also it is observable that classification by two different statistical tools gives somewhat similar result and robustness of built classification model were properly maintained.

The exactitude or exactness created from this classification model for gene dataset of 2111 genes is good 77% though applying a C4.5method and J48 algorithm employed totally unique statistical tool that follow branch and bound procedure for classification and characterization. Completely the mistake rate is fantastically low. Specificity and sensitivity are figured to see the nature of planned model and are vital. At long last, from this study it's settled that MMSE and relevance association score are fundamental qualities for grouping of qualities and named them to a chose class. This sort of testing and investigation has been utilized for the decision of genes for expression array, machine-controlled protein information, automatic cancer diagnosing, plant genotype sepration, grouping gene expression profiles and computational model for alteration locale.

3.5 ENRICHMENT ANALYSES

3.5.1 Gene Ontology (GO) Analysis

Enrichment analysis of the gene means giving some meaningful annotation of set of genes. Annotation in GO includes biological process, cellular part and molecular function [19]. Meaningful result for biological process and molecular function are displayed within the barchart as shown in Figs 3.4. and Fig.3.5, respectively.

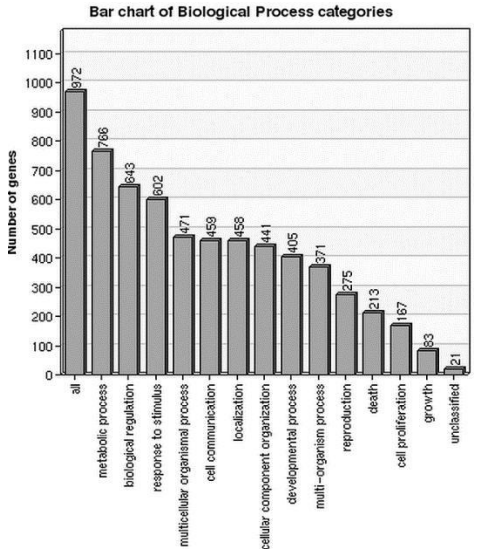


Figure 3.4: Different Biological process involved in GO term with genes

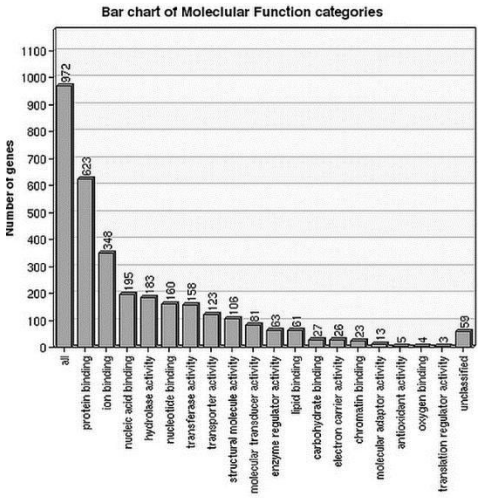


Figure 3.5: Different molecular function associated to GO

Members of independent groups represent one among 2 independent classes. Fisher-q-value take a look at was used to confirm whether each item belong to class is differ by group [20]. To stay away from over-checking copied genes, the Fisher exact test is ascertained relating ensemble gene IDs by which all redundancies in unique Ids were evacuated. All results of chart report were createdwhen going through thresholds (by default, Max. Prob.≤1 and Min. Count ≥ C2) [21]. The edge of EASE score, an adjusted Fisher's exact P value,ranges from 0 to 1 and depicts a statistically significant number of genes from list with pertinence the number inside the population of genes from the list derives. Fisher's exact P value = 0 shows great enhancement. Usually, P-value is equivqlent to or little than 0.05 to be thought of effectively improved inside the individual annotation classes.

3.5.2 Purposeful Classification of Genes

To gain this purpose, hyper geometric numerical procedure which was more popular using Benferroni method at threshold point of 0.05? Out of total sample of 2111 genes 958 genes have foun valid ids, only 74 genes are here found common annotation term profile on the basis of frequency [22]. Lastly, 39 genes come out as output in the criteria of enrichment score by applying certain parameter like kappa threshold of 0.3 and numerous linkage threshold of 0.50 as default setting.

3.6 CONCLUSIONS

For human understanding, it's necessary to get straightforward and logic based classifier, i.e., decision tree fulfills this criteria. However, some classifier have high classification predictive capacity as compare to the DT; however accuracy parameter generated from DT classification model is not much bad by means of C4.5 rule that follow branch and bound algorithm. Absolutely the inaccuracy found to be extremely very little. Specificity and sensitivity are important measure to find the appropriateness of built model. Medical measurements for diagnosis of AD like MMSE and association value are necessary attributes for classification of genes and assign them to a suitable class. This kind of experiment and studies are employed for varied purpose as expression analysis of gene array, comprehensive collection of protein information, malignancies identification, plant genotype discrimination, gene expression profile generation and computational model for modification sites by the means to mine best rule from constructed model. Enhancement investigation changes us with respect to the specific part of genes in term of GO, gene functional annotation, pathway examination, disease affiliation, drug

affiliation and phenotype examination. From gene functional classification, we discovered APOE, PSEN1, GRN, ACE, BCHE, PRNP, IL1A key genes that are emphatically identified with AD whose affiliation score ranges from 526.8 to 19.1. Our built decision tree models and enrichment analysis of target genes can work as a standard for registering biological advancement related with sickness and display their association toward AD conditions and its initial diagnosing.

REFERENCES

[1] K. Raza, "Application of data mining in bioinformatics," *arXiv preprint arXiv:1205.1125,* 2012.

[2] P. Flach, *Machine learning: the art and science of algorithms that make sense of data*: Cambridge University Press, 2012.

[3] M. Richards, "Families, kinship and genetics," *The troubled helix: Social and psychological implications of the new human genetics,* pp. 249-273, 1996.

[4] H. F. Wolcott, *Transforming qualitative data: Description, analysis, and interpretation*: Sage, 1994.

[5] D. G. Altman and P. Royston, "What do we mean by validating a prognostic model?," *Statistics in medicine,* vol. 19, pp. 453-473, 2000.

[6] C. Fraley and A. E. Raftery, "How many clusters? Which clustering method? Answers via model-based cluster analysis," *The computer journal,* vol. 41, pp. 578-588, 1998.

[7] Y. Y. Song and Y. Lu, "Decision tree methods: applications for classification and prediction," *Shanghai Arch Psychiatry,* vol. 27, pp. 130-5, Apr 25 2015.

[8] L. Smith and J. Tansley, "Decision tree analysis," ed: Google Patents, 2003.

[9] R. Duncan, "What is the right organization structure? Decision tree analysis provides the answer," *Organizational Dynamics,* vol. 7, pp. 59-80, 1979.

[10] N. Bhargava, G. Sharma, R. Bhargava, and M. Mathuria, "Decision tree analysis on j48 algorithm for data mining," *Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering,* vol. 3, 2013.

[11] L. Rokach and O. Maimon, *Data mining with decision trees: theory and applications*: World Scientific, 2008.

[12] X. J. Zhou and T. S. Dillion, "A statistical-heuristic feature selection criterion for decision tree induction," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 13, pp. 834-841, 1991.

[13] H. E. Marei, A. Althani, J. Suhonen, M. E. El Zowalaty, M. A. Albanna, C. Cenciarelli*, et al.*, "Common and Rare Genetic Variants Associated With Alzheimer's Disease," *Journal of cellular physiology,* vol. 231, pp. 1432-1437, 2016.

[14] S. L. M. Gutiérrez, M. H. Rivero, N. C. Ramírez, E. Hernández, and G. E. Aranda-Abreu, "Decision trees for the analysis of genes involved in Alzheimer′ s disease pathology," *Journal of theoretical biology,* vol. 357, pp. 21-25, 2014.

[15] D. S. Liu and S. J. Fan, "A modified decision tree algorithm based on genetic algorithm for mobile user classification problem," *ScientificWorldJournal,* vol. 2014, p. 468324, 2014.

[16] S. Ruggieri, "Efficient C4. 5 [classification algorithm]," *IEEE transactions on knowledge and data engineering,* vol. 14, pp. 438-444, 2002.

[17] J. Rogers, L. Kirby, S. Hempelman, D. Berry, P. McGeer, A. Kaszniak*, et al.*, "Clinical trial of indomethacin in Alzheimer's disease," *Neurology,* vol. 43, pp. 1609-1609, 1993.

[18] L. Benuskova and N. Kasabov, "Modeling brain dynamics using computational neurogenetic approach," *Cognitive neurodynamics,* vol. 2, p. 319, 2008.

[19] M. Sehgal and T. R. Singh, "Systems biology approach for mutational and site-specific structural investigation of DNA repair genes for xeroderma pigmentosum," *Gene,* vol. 543, pp. 108-117, 2014.

[20] C.-B. Zhang, P. Zhu, P. Yang, J.-Q. Cai, Z.-L. Wang, Q.-B. Li*, et al.*, "Identification of high risk anaplastic gliomas by a diagnostic and prognostic signature derived from mRNA expression profiling," *Oncotarget,* vol. 6, p. 36643, 2015.

[21] M. Sehgal, R. Gupta, A. Moussa, and T. R. Singh, "An integrative approach for mapping differentially expressed genes and network components using novel parameters to elucidate key regulatory genes in colorectal cancer," *PloS one,* vol. 10, p. e0133901, 2015.

[22] D. Piovesan, M. Giollo, C. Ferrari, and S. C. Tosatto, "Protein function prediction using guilty by association from interaction networks," *Amino acids,* vol. 47, pp. 2583-2592, 2015.

CHAPTER 4

System level investigation by gene set enhancement and network property examination of AD

ABSTRACT

Many complex diseases, like Alzheimer's disease, cancer, and diabetes are multistagial and complex in nature. However, with the advancement in technology and research, the scientific communities are able to decipher these complex diseases as result of not one or two factors but a combination of factors like those from genomics, epigenomic, interactomic and environment. Besides, deeper understanding of the mechanism of action these diseases has led to an understanding that their origin lies in the small variation of biological networks at the cellular level. Alike the observation made for complex disease, cause of AD is majorly driven by genomic susceptibility and changes in both tau networks and central amyloid Precursor protein (APP) processes. Additionally, alterations in any biological network could lead to not only loss of functions and imbalances, but also would help in progression of disease to much severe stages.

In this chapter, we have presented an integrative perspective on AD gene-set using principles of systems biology to explore pathway ranking and conduct multivariate studies based on XD-Score and Fisher-q value. The information from parameters like shortest path length, node betweenness, degree, clustering coefficient, eigenvector centrality for the genes involved in AD have also been discussed. As a result, largely applied statistical scores were found to be significantly high (5.062) at a major threshold (0.74). A graph of linear regression along with enrichment analysis in related genetic pathways was also elucidated. It is clear that gene and pathway enrichment studies help in understanding disease gene regulatory process after considering all essential parameter of experiments. From here, network analysis of AD brain help in finding causal agents (genes, proteins and associated pathways) which are responsible for the fatal state of brain and accordingly help in finding definite biomarkers in order to stop or delay the onset, and even reverse the progression of the disease.

4.1 INTRODUCTION

Many complicated diseases like diabeties, neurodegenerative disease and varied forms of cancer are being discovered as complex in nature involving a mixture of genomic, epigenomic, interactomic and environmental factors [1]. Nowadays increased availability and potency of next generation sequencing (NGS) technologies have delivered within a basic genomic part [2]. However, the major challenge while deciphering these complex diseases is to integrate all parts concerned and elucidate their interactions. This can be solely achieved by holistic approach offered by systems biology [3].

Systems biology integrates information from various biological elements and data into models of the system as an entire. The only real purpose of systems biology is to know the structure and dynamics of a complicated network of protein-protein, gene-protein and gene-gene interaction in associated pathways of network. Systems biology is essentially described as the reason that lies behind the sustainability and viability of complex networks and systems, and the way advanced diseases will arise from altered networks states [4] that are the consequence of complex perturbations, whose dynamics and mechanisms maybe best be studied ab initio experiments, and exploitation of model organisms, with the conclusions later confirmed in humans [5]. In this chapter, we tried to gain insight about the complex network of AD, Which may be a complex disorder of brain with loss of memory, thinking skills and behavior and later death of patients. Understanding network in systems biology is important in finding important pathways and clusters or modules in development of potential theraupetic targets [6]. Comprehensive knowledge of network science help in deep understanding of the elementary principles of the key pathways involved in AD could offer a background required to develop medication.

The goal of the present investigation is to break down this through experiment based network of genes involved in AD from built up logical scientific resources and investigate it for potential target proteins. KEGG, Reactome and Interpro databases were taken as the reference node set for the system. At that point we tend to perform PPI of seed proteins in protein interaction data based on combined score of the neighbourhood and domain fusion for tentatively confirmed outcomes in STRING to see for direct coordination between the reference proteins [7].We examined the system to discover proteins with criticalness based on hereditary affiliation and animal studies. We performed gene-set enhancement analysis for the system in light of XD-score and Fisher q-measure in addition, topological examination.

4.2 MATERIAL AND METHODS

Our main target was to gain deep insight from regulatory mechanism involved in AD pathways and utilize it for pathway ranking. In the present study, we used JEPETTO application of Cytoscape to analyze the gene-set associated with AD. The gene-set related to the disease was a combined filter dataset obtained from Genopedia that may be a a part of the Human Genome Epidemiology (HuGE) encyclopedia and gene analytics. We mined 2705 PubMed publications associated with AD to come up with total number of 1669 human genes enlisted in supplementary table I.

JEPETTO, a Cytoscape plugin, provides two totally different ways to investigate the datasets. The primary methodology finds path that are found to involve a query gene-set in term of interaction network supported XD-score. On the opposite hand, the second methodology is performed to seek out similar set of topological characteristics shared by the pathways. every analysis was performed in independent manner.

Functional annotation of querygeneset is obtained by gene-enrichment and focusing on pathway containing highest score. One would possibly to generate the network of interactions for each pathway of interest and analyze the overlap with the gene-set whereas the protein coding genes may be concatenate as path growth. KEGG, Reactome and GO were some of the databases used as reference database for enrichment and ontological comparison studies. The architectural study of network or topological analysis is performed to check the pattern of interaction between associated AD pathways and input geneset.

We performed a noticeable comparison to envision the distribution of important topological feature across pathways; (Figure 1) illustrate the protocol of enrichment and topological study of gene-sets

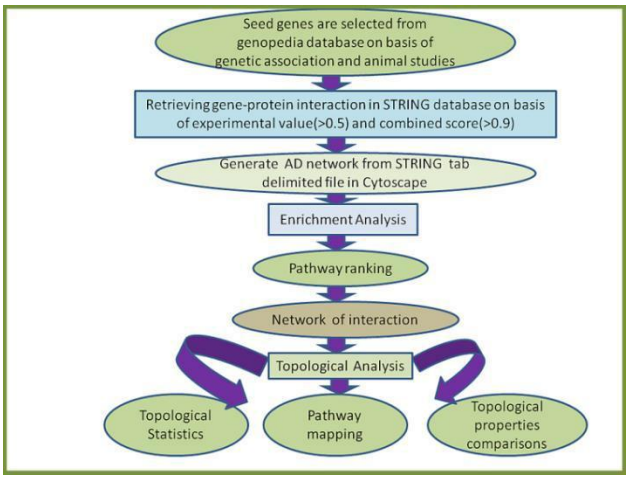


Figure 4.1: Roadmap of systems biology approach for geneset enhancement and network properties examination of AD genes dataset

4.3 RESULT AND DISCUSSION

4.3.1 Pathway enrichment analysis

To perform gene ontology (GO) analysis in term of enrichment studies, firstly the STRING interaction network tool was used for mapping of the target seed gene-set which successfully mapped 49 genes out of 761 and was employed for further analysis. Also, XD-score helped inside the distinguishing proof of the nearly related pathways and cell level process from KEGG (Table1). As XD-score gives a relative normal random walk separate between the entire pathway and in this way to the target genes [8, 9].

Table 4.1: Highly interacting pathways as referenced from KEGG. XD score is used to determine association between pathway and associated genes. Fisher q- value denotes significant overlap of genes. Last column show overlap between genes and pathway.

Pathway or Process	XD-score	q-value	Overlap/Size
Alzheimer's disease	5.06285	0	136/138
Parkinson's disease	4.40451	0	72/99
Oxidative phosphorylation	4.38937	0	67/94
Huntington's disease	2.97584	0	82/149
Cardiac muscle contraction	1.22013	0	21/52
Amyotrophic lateral sclerosis(ALS)	0.98472	0	18/47
Long-term potentiation	0.98265	0	22/63
Long-term depression	0.50646	0.00001	13/57
GnRH signaling pathway	0.4837	0	19/83
Cell death	0.40646	0	19/81
Kreb cycle	0.38646	0.13586	4/26
Dorso-ventral axis formation	0.36201	0.22648	3/20
Prion diseases	0.3382	0.02727	6/35
Vascular smoothmuscle contraction	0.32668	0	17/89

AD signal pathway showed up at the highest position of the pathways ranking chart shown through scatter plot (Figure 2), with a high XD-score of 5.06285, more than three times higher than the significance threshold value of 0.74 determined by the regression plot.

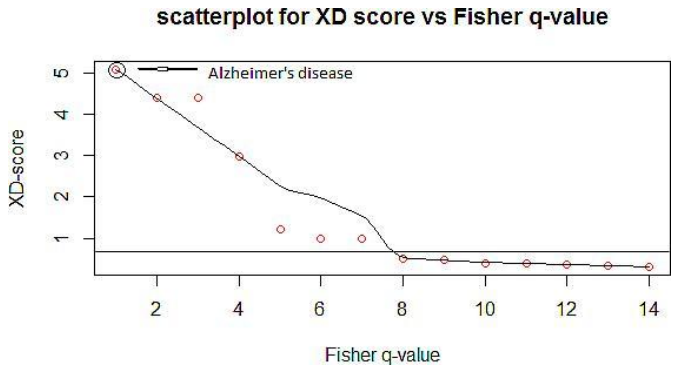


Figure 4.2: **Regression plot between XD-score and fisher q-value**. Here dot represent pathway or process involved in enrichment examination as associated with target geneset of AD. Best linear fit are represented with red line in this plot. Significant threshold for XD-score are marked with horizontal grey line. The top most node in this plot denote AD- pathwayat XD-score= 5.062 and q-value=0.00.

Green nodes found within the right side of the canvas are fixed to the associated pathways and a mixture of target geneset and pathway specific ones, were the essential two genes groups saw inside the network. The labelled orange genes in the middle of the two groups (MMP17, APHIA, APH1B, and TMED10) are the foreseen path expansions. APH1B was at that point present inside the query geneset and was determined yet again by the path expansion algorithm from the investigation of the AD pathway associations [10] appeared in (Figure 3).

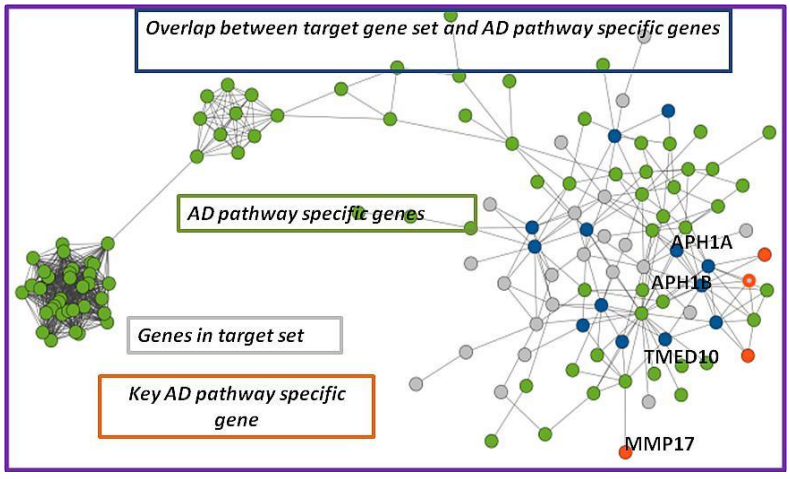


Figure 4.3: Representation of target geneset of AD network environment. Node colors are: gray color for the gene used as target for AD, Blue color represent sandwich node between AD

target genes and AD pathway specific gene, green node represent pathway specific genes and orange node represent key genes invole in some pathogenic process of AD. the edge colors are: green edge is used for interacting genes of AD target set with pathway, orange for interactions between the input set and also the growth and gray for others. More than three connected node are used to improve clarity.

APH1B could be a cofactor that shows an uncommon polymorphism determined in AD population, raising the disease vulnerability throughout connections with the Apolipoprotein E [11]. APH1A compares to aph-1, homolog-A, a gamma-secretase catalyst subunit for cleavage of β-APP and for presenilin protein aggregation. Furthermore, APH1A and preselinins are found to express along in AD pathologic process [12]. TMED10 could be a negative controller of the amyloid β-peptide generation. Restraint of which upgrades the particle accumulation therefore helping in AD movement with related side effects [13]. MMP17 factor was not found to assume any major role inside the Alzheimer's pathology. Notwithstanding, proteins of the MMP17 family are appeared to be equipped for degrading the β-amyloid proteins and it's trusted that location of MMPs inside the cerebrum matter could be a piece of AD mechanism [14]. Accordingly, extra investigations on MMP17 as a disease cofactors for the AD would perhaps demonstrate rewarding.

4.3.2 Topological Analysis

The biggest associated parts (49 nodes) of the enriched network system were utilized as an input data for topological examination. Properties of comparable size estimated random network systems were compared with the system topological properties. A major distinctive feature was found in topological options (Table 2).

Table 4.2: Network properties of enhanced AD network. Vital properties are: shortest path length (SPL), node betweenness (NB), node degree (D), clustering coefficient(CC), eigenvector centrality (EC).

	SPL	NB	D	CC	EC
Data set	3.43	183840	48.11	0.09	0.1
Rand simulation	4.1 ± 0.05	14032 ± 6701	8.63 ± 2.34	0.09 ± 0.03	0.02 ± 0.01
Entire network	4.12 ± 0.94	14669 ± 6893	8.27 ± 16.2	0.11 ± 0.21	0.02 ± 0.04

The features of the topological investigation that are available inside the enriched network were then used to compare and very much characterized biological phenomenon and pathways. Closest topological matches were acquired from KEGG database. Table 3 demonstrates the set of most similar biological phenomenon found. Wnt signal pathway demonstrated the closest topological match (score = 0.03).

Table 4.3: Highly mathched network properties found from KEGG . The distance score related to every match in normalized sum of ranks evaluate from variations between the topological properties. Numbers in parenthesis represent range of nodes overlapped with geneset for AD.

Name	Score
Wnt signaling pathway - hsa04310 (123)	0.03
Tight junction - hsa04530 (106)	0.05
Melanogenesis - hsa04916 (80)	0.07
Parkinson's disease - hsa05020 (20)	0.07
Axon guidance - hsa04360 (111)	0.1
Ubiquitin mediated proteolysis - hsa04120 (105)	0.1
Gap junction - hsa04540 (79)	0.11
Long-term depression - hsa04730 (68)	0.11
Maturity onset diabetes of the young - hsa04950 (18)	0.12
ECM-receptor interaction - hsa04512 (73)	0.13
Regulation of actin cytoskeleton - hsa04810 (184)	0.16
GnRH signaling pathway - hsa04912 (82)	0.17
Calcium signaling pathway - hsa04020 (151)	0.17
Amyotrophic lateral sclerosis - hsa05030 (19)	0.17
Natural killer cell mediated cytotoxicity - hsa04650 (117)	0.19
Phosphatidylinositol signaling system - hsa04070 (55)	0.19

Wnt signal acts along with the β-catenin signal during whole process of angiogenesis [15]. Pathway expansion define connection between early onset of AD and MMP17 [16]. MMP17 partly controls the differential development or relapse of veins. The topological similarity of the Parkinson's pathway (score = 0.07) is reflected by the earlier discovered connection of Parkinson's disease. Maturity - Onset diabetes of the youthful (MODY) pathway was a prime match with a score of 0.12. The connection among AD and diabetes had been appeared by increase in amount of the amyloid-β accumulation because of diabetes came about because of AD impacts in Alzheimer transgenic and diabetic mice that was affirmed by recent examinations [17]. Node degree, node betweenness and shortest path length were utilized for visual examination of topological investigation (Figure 4-A,B), demonstrates the objective system in reference to alternative KEGG pathways and processes.

.

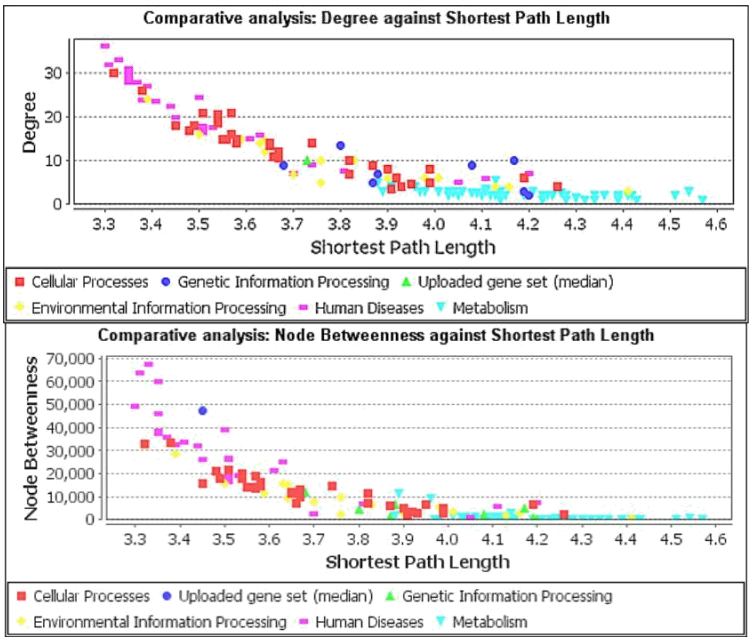


Figure 4.4 (A) and 4 (B): Differentiate between properties involved in AD network the focused network are represented through red square. In both cases shortest path length are taken as standard and compare to Node-betweenness and Degree respectively.

sporadic amyotrophic lateral sclerosis (ALS) shares numerous biological components with AD [18]. Presenilin (PS1 and PS2) joined mutations or the amyloid precursor protein (APP) genes of ALS, displays analogous brain abnormalities kind of similarity to those of sporadic AD. This affiliation is relevant based on topological score (separate score) of 19 genes from AD query genedataset additionally in context of ALS. Components of ECM bind with APP and results in the accumulation of β-amyloid plaques, an ordinary pathology of AD. Moreover, MAPK signal pathway that associated to the brain inflammation in AD is furthermore closer to the target network within the comparative analysis.

4.4 CONCLUSION

We discovered three recognized ailment cofactors (APH1B, APH1A, TMED10) utilizing this Pathway enhancement and topological network system study for candidate genes of an AD. The type of candidate genes like MMP17 and their degrading ability of β-amyloid and pathways like the Wnt signal pathway and Parkinson's disease pathway were moreover affirmed that may be mostly acquainted in pathological processes. Any examinations and exploratory approval are expected to help the theory for firmly associated pathways and known disease cofactors could play out a significant role to uncover the essential job of these gene systems inside the hidden components of the prohibitive procedures of an AD.

REFERENCES

[1] L. Migliore and F. Coppede, "Genetic and environmental factors in cancer and neurodegenerative diseases," *Mutat Res,* vol. 512, pp. 135-53, Dec 2002.

[2] N. Wagle, M. F. Berger, M. J. Davis, B. Blumenstiel, M. DeFelice, P. Pochanard*, et al.*, "High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing," *Cancer discovery,* vol. 2, pp. 82-93, 2012.

[3] G. E. Louridas and K. G. Lourida, "Conceptual Foundations of Systems Biology Explaining Complex Cardiac Diseases," *Healthcare (Basel),* vol. 5, Feb 21 2017.

[4] B. S. Chen and C. C. Wu, "Systems biology as an integrated platform for bioinformatics, systems synthetic biology, and systems metabolic engineering," *Cells,* vol. 2, pp. 635-88, Oct 11 2013.

[5] G. Parra Farré, *Computational identification of genes: ab initio and comparative approaches*: Universitat Pompeu Fabra, 2004.

[6] Y. Wang, H. Liu, Y. Lin, G. Liu, H. Chu, P. Zhao*, et al.*, "Network-Based Approach to Identify Potential Targets and Drugs that Promote Neuroprotection and Neurorepair in Acute Ischemic Stroke," *Scientific reports,* vol. 7, 2017.

[7] D. Szklarczyk, A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguez*, et al.*, "The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored," *Nucleic acids research,* vol. 39, pp. D561-D568, 2010.

[8] K. Lage, E. O. Karlberg, Z. M. Størling, P. I. Olason, A. G. Pedersen, O. Rigina*, et al.*, "A human phenome-interactome network of protein complexes implicated in genetic disorders," *Nature biotechnology,* vol. 25, pp. 309-316, 2007.

[9] B. Gupta and R. B. Mishra, "Protein Network for Associating Genes with Dementia," *International Journal of Computer Applications,* vol. 83, 2013.

[10] E. Y. Chen, C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles*, et al.*, "Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool," *BMC bioinformatics,* vol. 14, p. 128, 2013.

[11] C. Haass and D. J. Selkoe, "Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide," *Nature reviews Molecular cell biology,* vol. 8, pp. 101-112, 2007.

[12] C. M. Tan, E. Y. Chen, R. Dannenfelser, N. R. Clark, and A. Ma'ayan, "Network2Canvas: network visualization on a canvas with enrichment analysis," *Bioinformatics,* vol. 29, pp. 1872-1878, 2013.

[13] W. Yu, M. Clyne, M. J. Khoury, and M. Gwinn, "Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations," *Bioinformatics,* vol. 26, pp. 145-146, 2009.

[14] G. Agapito, P. H. Guzzi, and M. Cannataro, "Visualization of protein interaction networks: problems and solutions," *BMC bioinformatics,* vol. 14, p. S1, 2013.

[15] R. T. Moon, A. D. Kohn, G. V. De Ferrari, and A. Kaykas, "WNT and β-catenin signalling: diseases and therapies," *Nature Reviews Genetics,* vol. 5, pp. 691-701, 2004.

[16] S. Wang, Y. Tong, T.-B. Ng, L. Lao, J. K. W. Lam, K. Y. Zhang*, et al.*, "Network pharmacological identification of active compounds and potential actions of Erxian decoction in alleviating menopause-related symptoms," *Chinese medicine,* vol. 10, p. 19, 2015.

[17] C. Monti, H. Bondi, A. Urbani, M. Fasano, and T. Alberio, "Systems biology analysis of the proteomic alterations induced by MPP+, a Parkinson's disease-related mitochondrial toxin," *Frontiers in cellular neuroscience,* vol. 9, 2015.

[18] J. Sreedharan and R. H. Brown, "Amyotrophic lateral sclerosis: problems and prospects," *Annals of neurology,* vol. 74, pp. 309-316, 2013.

CHAPTER 5

Molecular characterization of specific chemical group of compounds to find inhibitors for specific target protein Acetyl cholinesterase (AChE) both in-vitro and in-vivo

ABSTRACT

Currently approved acetylcholinesterase (AChE) inhibitors for treating Alzheimer's disease (AD) provide only symptomatic relief and are associated with hepatotoxicity and gastrointestinal side effects. An alternative drug discovery method is to develop safer 'disease modifying drugs'. We screened 646 small molecules, already documented drug-like candidates with pharmacological and functional values. Based on *in-silico* AChE interaction study, we predicted quercetin, caffeine, ascorbic acid, gallic acid, begacestat and donepezil to be potential AChE inhibitors and demonstrated quercetin, caffeine, begacestat and donepezil to be efficient AChE inhibitors through *in-vitro* assay. Further, subjecting hippocampal primary neuron cultures to $HgCl_2$ induced toxicity resulted in severe neurodegeneration, which was significantly alleviated by treating cultures with quercetin, caffeine, begacestat and donepezil, however, ascorbic acid and gallic acid proved to be ineffective. In conclusion, quercetin and caffeine can be considered as promising lead molecules and may find a clinical application in AD therapeutics.

5.1 INTRODUCTION

AD is the commonest kind of mental disorder in old aged individuals. Currently, there's no cure for this disease that worsens because it progresses and eventually ends up in mortality [1]. Currently, there are 44 million people who suffer from AD worldwide [2]. It has been reported that the reduced activity of dysfunction in cholinergic neurotransmission is a major factor associated with the development of AD and therefore acetylcholinesterase inhibitors (AChEI), such as Donepezil, Rivastigmine and Galantamine are usually prescribed for AD patients to augment cholinergic signalling. It has also been observed that the effects of AChEI are not universal and possesses high degree of variability among different patients [3]. AChEI are also known to adversely affect patient compliance by causing side effects like nausea, vomiting, diarrhoea, abdominal pain and muscle cramp, reduced appetite, distorted sleep cycle and are contradicted in patients with cardiovascular complications [4]. Therefore, identifying safer molecules that enhances cholinergic activity is a major driving force and need of the hour in drug discovery for AD.

Disease modification in AD refers to the ability of a drug to slow down neurodegeneration, and thus, delaying the onset of dementia and AD. Some disease modifying compounds from the herbal world, some phytochemical as atropine, hyoscyamine, codeine, morphine and ephiderine, are able of crossing the blood-brain barrier (BBB), and therefore, produces central nervous system (CNS) modulating effects [5]. In case of Parkinson's disease (PD) and Multiple sclerosis, disease modifying drugs such as Rosagiline, Selegiline, Coenzyme Q10, Safinamide, and Creatine, have been reported to be effective in slowing the growth process of these diseases [6]. These drugs easily cross the BBB and have been proven to be safe and tolerable. Due to the lack of biomarkers information for AD, it is not possible to prove whether an agent is actually neuroprotective or not [7, 8]. Certain disease modifying drugs have been reported for AD, where they target neuropathological markers, which primarily consist of β-amyloid formation and deposition, formation of neurofibrillary tangles, tau depositions, apoptosis, necrosis, inflammatory stress and oxidative damage [9-14].

Orthogonal approach for the treatment of AD patients, i.e., through dietary effects, has gained popularity in recent past, where several reports demonstrates potential of dietary modulations in delaying onset and progression of AD in both, clinical and laboratory settings. Beneficial effects of dietary modifications has been associated with their potential to modulate neuronal oxidative and inflammatory stress, mitochondrial functioning, metabolism, amyloid-β deposition and tau hyperphosphorylation, besides, decreasing acetylcholine levels in brain [15-17]. In a cohort study of 2,148 elderly subjects without any symptoms of dementia, the intake of Mediterranean diet and physical activity were studied and have been shown to lower the risk of AD [18]. Acetyl-cholinesterase inhibitors (AChEI) are the most common class of drugs used for the management of AD. Rigorous evaluation of AChEIs in several clinical studies have revealed that these drugs are capable of delaying onset and progression of dementia and AD [19-22], however, prevalence of side effects, especially gastrointestinal complication, was severe which limits its clinical benefits [19, 20, 22, 23], besides, having a risk of exaggerating progression dementia and neuropsychiatric disorders after drug withdrawal [24].

Based on these evidences, we screened 646 small molecules, which are consumed abundantly through diet in form of vegetables, fruits and beverages, through in-*silico* docking studies against AChE enzyme and identified quercetin, caffeine, ascorbic acid and gallic acid to possess high potential to modulate enzyme activity. We further conducted *in-vitro* enzyme inhibition and cell line based assays to provide experimental evidence that these molecules can benefit AD. All of these compounds are well known antioxidants and possesses potential to increase the survival rate of neuronic cells and functioning by remove antioxidant agent and stress causing factors, apoptosis, necrosis, and by correcting neuronal insulin signaling, glucose uptake and utilization [25-30], which may further provide added advantage in the Alzheimer's therapeutics. Therefore, present study was aimed to screen out potential molecules having AChE inhibitory activity through computation docking studies and investigate whether or not these molecules actually possesses neuroprotective and AChEI activity through *in-vitro* assays.

5.2 MATERIALS AND METHODS

5.2.1 Molecular Docking

Molecular docking was performed using Autodock version 4.2. The manually handled docking tool like Autodock improvise their functionality by employing GA (Genetic Algorithm)

for global and local search as well to find energy minimized confirmation [31]. The structure of AChE for molecular docking studies was easily taken from Protein Data Bank (pdb id: 1B41) [32]. The protein preparation steps involved removing water molecules, metal ions and cofactors. Partial Kollman charges were applied to the protein after adding the hydrogen ions. The molecules used for molecular docking were Begacestat, Quercetin, Ascorbic acid; Donepezil, Caffeine and Gallic acid were appropriately prepared with molecule properties were given in Table1.

Table 5.1: Properties of molecules used in Autodock 4.2. Important parameters obtained from Pubchem (molecular weight, hydrogen bond donor, hydrogen bond acceptor, logarithmic P value and dissociation constant) are listed.

Molecule Name	**Molecular Weight(g/mol)**	**Number of H-bond donors**	**Number of H-bond acceptors**	**Log P**	**pKa**
Donepezil	379.49	0	4	4.86	3.33
Begacestat	391.73	2	11	3.7	-
Quercetin	302.00	5	7	1.48	7.17
Ascorbic acid	176.00	3	6	-2.15	4.17
Gallic acid	170.00	4	5	0.70	4.4
Caffeine	194.00	0	3	-0.07	10.4

The active site of AChE was identified using the Catalytic Site Atlas database [33]. The group of residues that are listed as active site or catalytic site residues were Gly122, Gly121, Tyr133, Glu202, Ser 203, Ala204, His447, Glu450, and Glu334. A gridbox of dimension 90x90x90 Å was constructed with the α-carbon of Ser203 as its center. Autogrid was performed with the above settings to generate appropriate receptor and molecule maps. All essential parameter for smooth run of Autodock and Dielectric constants were measured for calculation of vander waal and electrostatic potential terms, respectively. Couple simulations study were actioned using genetic algorithm, wherever the initial position, orientation, and torsions of the molecules were

set arbitrarily. The other parameters used for genetic algorithm were number of individual in a population (50), most extreme number of energy assessment (25000000), rate of gene mutation (0.2) and number of GA run (50). Docking calculations were carried out based on the docking free energy and their inhibition constant (K_i) value for the following molecules Donepezil, Begacestat, herbal phytochemical as Ascorbic acid, Caffiene, Gallic acid and Quercetin towards the AChE. Finally, the best confirmation of the docked molecules was picked from the cluster analysis of docked conformations.

5.2.2 Molecular Dynamics simulation

Schrodinger based tool used for MD calculation were Desmond. It works on principal of parallel scalibility and some numerical calculation to gain accuracy in their result. This result help in finding efficient binding pose. Each produced conformer was subjected to a full minimization inside the gas stage with the optimized potential for liquid [36] Simulations all-atom force-field [37]. Individual as well as motion of interacting particle were calculated by medium of MD. The very first coordinates for MD were obtained from Protein complex. The Simple Point Charge (SPC) solvent model was used with each water box measuring orthorhombic dimensions of 10x10x10Å, to make sure a whole solvent coverage of the complex surface creating a box volume of 5,27,647 $Å^3$. To mimic the real behaviour of protein complex MD were performed in presence of water and charges were neutralized through Na+ counter ions.

The whole system was simulated in NPT ensemble with interval of 10 annotations for complete energy in Kcal/mol. The steadiness of the macromolecule determined by using root means sq. deviation (RMSD).

***5.2.3In vitro* Acetylcholinesterase Inhibition Assay**

5.2.3.1 Preparation of stock solution

Stock solutions (0.1 M) of Quercetin, and Begacestat was prepared in DMSO and stock solutions (0.1 M) of Donepezil, Ascorbic acid, Gallic acid and Caffeine was prepared in water and stored at 4ºC until used within 3 days. Working solution of test molecules were prepared by diluting appropriate quantity in phosphate buffer (pH 8) and used fresh.

5.2.4 Acetyl cholinesterase (AChE) Inhibition Assay

Begacestat, Quercetin, Ascorbic acid, Gallic acid and Caffeine were procured from Sigma Aldrich Inc, HiMedia and Loba Chemie. The molecules were tested for AChEI activity

according to the method described with some modifications [38] Donepezil (Sigma Aldrich Inc) is a well-known inhibitor of AChE and was used as a control [39]. AChE inhibition was observed in a 200 µl reaction mixture in a 96-well plate using Ellman's reagent or 5,5'-Dithiobis-(2-Nitrobenzoic acid (DTNB). Reaction mixture consisted of 25 µl AChE [EC 3.1.1.7] solution (25 mU in 0.1 M phosphate buffer; pH 8), 75 µl of 5,5'-dithio-bis(2-nitrobenzoic acid) (DTNB) and (0.01 M in 0.1 M phosphate buffered solution, pH 8, containing 6 mM $NaHCO_3$), 25 µl of test molecules (0.05-0.3 mM) and 50 µl of phosphate buffer (0.1 M; pH 8). The reaction mixture was incubated for 10 minutes at 37ºC. The reaction was started by including 25 µl of acetylthiocholine iodide (0.075 M in water). AChE activity was determined spectrophotometrically by observing the change in absorbance for 4 min at the wavelength of 412 nm. Percent inhibition of AChE activity was calculated. The percentage of DMSO in final concentration of Begacestat were 0.7%, 0.52% and 0.26%; Donepezil were 0.3%, 0.24% and 0.18% and Quercetin were 0.3%, 0.2% and 0.1% respectively.

5.2.5 Preparation of primary neuronal culture

Primary hippocampal neuronal culture was prepared as per our previously described method [29] Entire study was conducted on primary culture developed from E18 embryonic hippocampi. Briefly, E18 embryos were dissected out from the anesthetized rat and their hippocampi were isolated using dissection microscope (Nikon). 6-8 hippocampi were digested for 10 minutes in 2 ml 0.25% trypsin (in PBS) at 37ºC inside humidified incubator. Trypsin was neutralized by adding DMEM to the digested hippocampus, followed by centrifugation at 3000 revolutions per minute for five min. Pellet was resuspended in complete DMEM media supplemented with 100 percent fetal bovine body fluid (having a 100 µg/mL antibiotic drug and a 100 µg/L streptomycin). Cell density was evaluated and 1 × 106 cells/ml was seeded on poly-L-lysine coated plates. Culture plates were incubated at 37ºC within a humidified brooder having 5% carbon dioxide atmosphere for 2 weeks, throughout that growth media was often replaced.

5.2.6 *In-vitro* neurotoxicity assay

Neurotoxicity of the test drugs were evaluated on primary neuronal culture in 96-well plate by using 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide (MTT). For this, growth medium was replaced by serum-free DMEM having completely different concentration of test drugs (Quercetin, Caffeine, vitamin C, acid, Donepezil, and Begacestat). Culture plates were incubated for 24 h at 37ºC in five 5% greenhouse gas atmosphere. Afterward, media was

fastidiously removed and neurons were washed with phosphate cushion saline (PBS; pH 7.4). Cultures were incubated in dark for 3 h with 100 µL DMEM media having 5 mg/mL MTT (37ºC). Cells were lysed by adding DMSO and absorption of purple formazan was recorded at 570 nm using UV-spectrophotometer (Shimadzu 265, Japan). Simultaneously, UV absorption was also recorded for blank (without neurons), and control (without any treatment). We further evaluated the effect of different treatment on mercuric chloride ($HgCl_2$) induced neurotoxicity as per method described by Xu et al. (2012). Briefly, 14-day old cultures were incubated with 10 µM concentration of different treatments for 6 h at 37ºC inside humidified incubator having 5% CO_2. Cells were then exposed to 25 µM $HgCl_2$ and neurons were allowed to grow further for 24 h. Cultures were washed with PBS followed by MTT assay as described above to determine cell viability and percent neuroprotection provided by drug treatments.

5.2.7 Acridine Orange (AO) - Ethidium bromide (EtBr) assay

AO-EtBr assay was employed to determine the effect of different drug treatments on $HgCl_2$ induced neurotoxicity. For this, fully mature cultures were incubated with 10 µM concentration of different drugs for 6 h, after which 25 µM $HgCl_2$ was added to culture media. Neurons were allowed to grow for 24 h at 37ºC. Neuronal cultures were washed thrice with ice-cold PBS, followed by 1 h fixation using 4% paraformaldehyde solution (in PBS).Cells were permeabilized by hatching them with 0.1% triton X-100 for 20 min at ordinary temperature. Neurons were then stained with AO (10 µg/ml) and EtBr (10 µg/ml) in dark, followed by counterstaining with 10 µg/ml DAPI. Corresponding fluorescence images were captured at 200X magnification using fluorescence microscope (Nikon eclipse Ti fluorescence microscope). Results were demonstrated in terms of total fluorescence intensity of AO (green fluorescence) and EtBr (red fluorescence) calculated using image-J software (=Integrated Density − (CA3 area × Mean fluorescence of background readings)) [29]. Further, quantitative relation of AO and EtBr expression was tested to check the impact of assorted treatments on neural integrity throughout HgCl2 evoked neurotoxicity.

5.2.8 Statistical Reading

Statistical reading was performed by GraphPad prism half 6 software. All the information were expressed as mean ± SD and also the statistical reading was evaluated by one-way ANOVA taken after by Dunnett's multiple comparison post hoc test (* $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$).

5.3 Results and Discussion

5.3.1 Molecular Docking

Using AutoDock 4.2 molecular docking simulations, we predicted the varied protein-ligand interactions that play a big role in structure based drug designing. The different scores like binding free energy, inhibition constant, building block energy and static energy values are listed in Table 2. Donepezil, also known as Aricept is an N-benzylpiperidine derivative, specifically inhibits AChE and shows higher binding affinity to AChE. Donepezil when docked with AChE was able to reproduce the earlier reported interactions and had a predicted binding energy of -8.60 Kcal/mol. Previously, the reported docking binding energy of Donepezil to AChE is -5.15 Kcal/mol [40]. The results showed that each one of the molecules showed binding energy range in between -8.87 Kcal/mol to -6.49 Kcal/mol. Also, the electrostatic energy values of the molecules range from -0.29 kcal/mol to -0.03 Kcal/mol.

The tying up calculations of all 5 molecules at the active site of AChE unconcealed that the molecules bind to the active site of protein with lower docking energy compared to Donepezil. Recapturing the reported interactions of Donepezil was used as a standard or a measure to compare the efficacy of binding of other docked molecules, which required changing the parameters of AutoDock. Previous studies of Donepezil against AChE as a control was used to test the efficacy of other molecules as inhibitor of AChE [32, 41, 42]. In one study Donepezil was shown to have unfavourable binding energy and thus postulated that its binding is via weak attractive forces [43]. Nevertheless, other studies have pointed that Donepezil binds through a π-π interaction with Trp84 and a cation-π interaction with Phe330 of AChE [44]. Also, there are water-mediated contacts that are reported as crucial for Donepezil's nanomolar range of binding and specificity.

Observing the best conformation among six docked molecules, Caffeine exhibited higher number of interactions with AChE enzyme and had a binding energy of -8.87 Kcal/mol. Quercetin was ranked second in terms of binding energy of -8.68 Kcal/mol. Autodock make cluster on the

basis of binding energy score. It has been noticed that the predicted binding energies scores from AutoDock are close to the values expected from experimental receptor-ligand binding studies and the predicted inhibitory constant is directly proportional to binding energy [45]. Intermolecular energy has also direct influence on binding free energy. So, from the report it had been clear that each one the herbal phytochemical molecules having promising AChE inhibition activity except begacestat in comparison to Donepezil (standard) with inhibition constant 494.82 nm. Intermolecular energy of all 5 molecules vary between -9.57 to -8.76 kcal/mol that was higher in comparison to the standard -10.39 Kcal/mol. Intermolecular energy is the energy between non-bounded atoms, that is the energy between particles isolated by 3-4 bonds or between particles in various atoms. It was used to predict the strength of interaction between proteins and different molecules. Using Ligplot the protein-ligand interactions were mapped and they are shown in Figure 1c-h, where green dotted line represent H-bond and hydrobhobic interacting residues are depicted in red arcs with diverging spikes. In the case of Begacestat (Figure 1d), Arg296 and Ser293 form hydrogen bonds, whereas Leu289, Val294, Asp74, Trp286, Tyr72, Tyr124, Tyr337, Phe297, Tyr341, Phe295 and Phe338 forms hydrophobic interactions. Similarly Ascorbic acid (Figure 1e) interacts with AChE by making hydrogen bond interactions to Thr238, Arg296, and Trp236; hydrophobic interaction by Val239, Pro235, Phe297, Val300 and Arg247. Caffeine (Figure 1f) forms two hydrogen bonds with Arg296 and Phe295; Gallic acid (Figure 1g), interestingly, does not form any hydrogen bond interactions. Quercetin makes (Figure 1h) hydrogen bond interactions with Trp86, Tyr133, Gly120, Asn87, and Asp74 and hydrophobic interactions with Ile451, Glu202, Val73, Gly448, His447, Gly121, Ser125, Tyr72, Tyr124, Pro88 and Gly126. Comparing these results with Donepezil, the standard used in this study, more hydrogen bond interactions are observed in other molecules, whereas Donepezil (Figure 1c) is observed to make only one hydrogen bond interaction, specifically, with Try72.

Among the active site residues, peripheral anionic site, and the acyl binding site residues (Figure 1a and 1b), the following are observed to make hydrogen bonds with the molecules used in molecular docking study. These are catalytic triad (Ser203, His447, and Glu334); acyl binding site (Trp286, Phe297, Phe295, and Phe338) and peripheral anionic site (PAS) (Gly121, Gly122, and Ala204). These three sites are arranged in a manner that His447 and Glu334 donate electrons to Ser203 for AChE to carry out catalytic action.

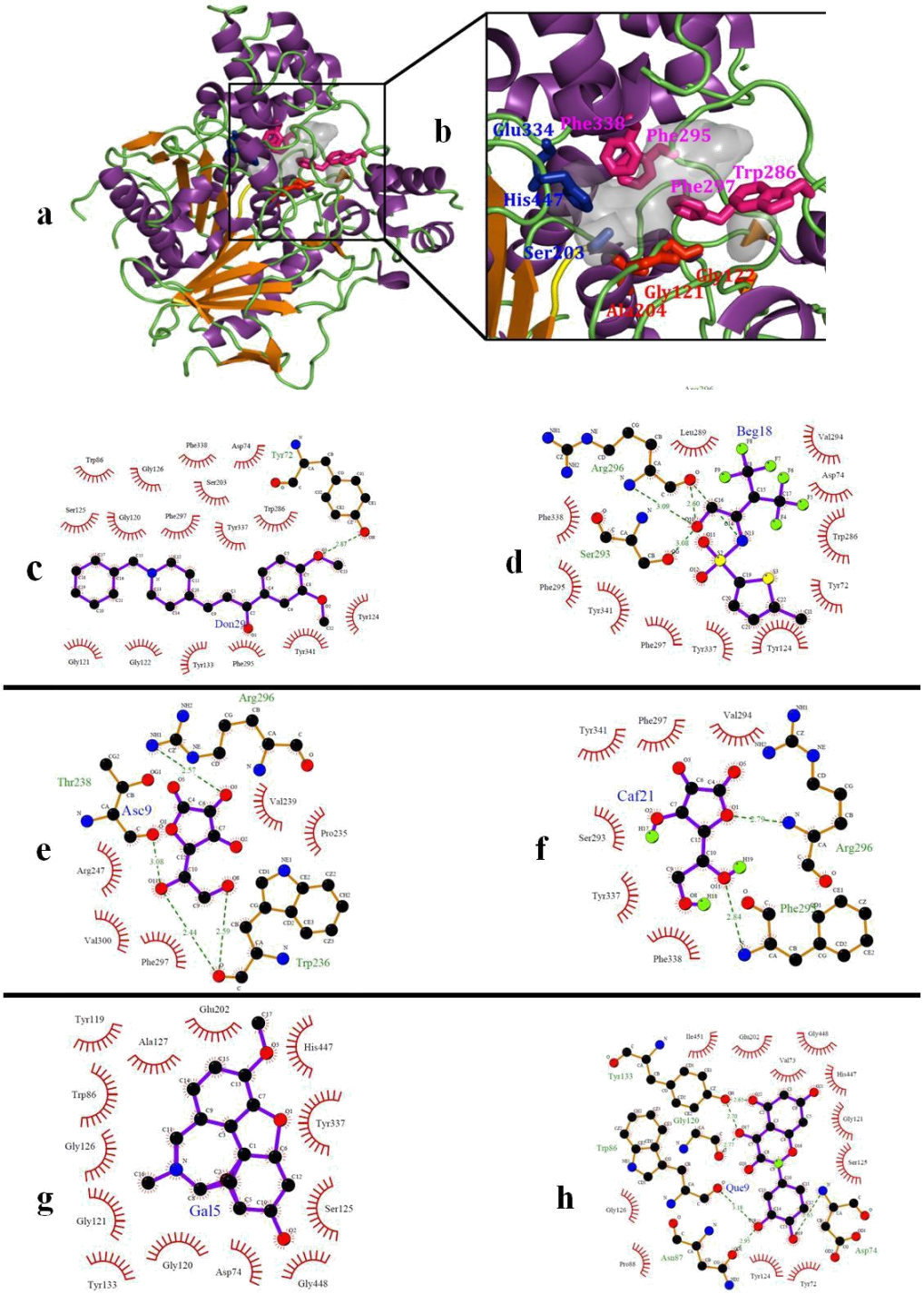


Figure 5.1(a-h): (a,b) 3D- structure of protein 1B41; (c-h)Active site and Interaction of disease modifying drugs with Acetylcholinesterase (AChE).

5.3.2 Molecular dynamics Simulations

MD simulation was mainly used to find dynamicity of all six docked complex and their stability study. Molecular dynamics simulations of the potent docked configurations of the molecules bind with AChE were performed for 20 ns using DESMOND [46]. The time-dependent RMSD values of the complex backbone atoms relative to the corresponding energy minimized structure were calculated. The RMSD plot (Supplementary Figure S1) and the residue fluctuation plot (RMSF shown in Supplementary Figure S2) are a measure to check the stability of the protein-molecule complex. In the case of Donepezil with AChE (Figure S1 (a) and S2 (a)). The equilibration process of MD using default setting of desmond that consists of a series of restrained minimizations and MD simulation relax the whole system and not deviating from the initial coordinates of complex. The whole system was neutralised by adding essential Na+ ion to balance the whole net charge of the system and energy decreased through steepest descent technique using convergence threshold of 1Kcal/mol. The RMSD of backbone atom was stabilized at 20000 ps and the ligand molecule gets stabled at 1.6 Å. The individual residues of protein fluctuate in angstrom range of 1-4 Å. Also, for AChE-Begacestat complex it was observed to achieve equilibrium around 100 ps (Figure S1 (b) and S2 (b)) and the individual residues fluctuated around 1.3Å. In AChE-Ascorbic acid complex, the equilibrium time was also around 100 ps. The structure of AChE-Caffeine complex showed deviation of acceptable value of less than 3Å RMSD with moderately tiny oscillations targeted around it, in whole process of dynamics. Gallic acid and Quercetin complexes also exhibited fluctuations in the beginning of the MD simulations, however the RMSD of the ligands with protein become stable at 3Å and 7Å at 20 ns, respectively (Figure S1). In immersion state, the motions of the projectile structures around the MD average are with respect to ±0.03 Å for every single structure.

Observing the active site residues, peripheral anionic site (PAS) residues, and acyl binding site residues (shown in Figure 1b) in the MD simulation gave some interesting observations. The RMSF captures, for every atom, the fluctuation concerning its average position. This offers insight into the flexibleness of regions of the macromolecule. On RMSF plot, peaks of the wave indicate territories of the protein that change the most amid the simulation. It is more common that the tails (N and C terminal) vary more than some other piece of protein. Specifically by looking at the RMSF plots (Supplementary Figure S2) of the six complex structures, the active

site residue Glu334 and His447 shows less disturbance in almost all structures. Both the residues donate electron to Ser203 for AChE to carry the catalytic action. The total free binding energy and electrostatic interaction demonstrate the solid binding of particle with the receptor [47] .

Little unsettling influence in total energy value demonstrates that the H bonding and optional non-covalent associations exchange among the buildup of the site [48]. We have a tendancy to learned that the most sharing inside the total energy originates from the intermolecular energy from all segments, similar to the protein and in this manner the bound atom, while the qualification of the van der waal energy, bond and desolvation energy is extra negative with connection to the electricity energy, that shows that an ideal design of particle to prevail in the cavity of receptor. High commitment of polar group shows that catalyst inhibition is usually bolstered by the hydrophilic associations. Total energy of Ascorbic acid, Begacestat, Caffeine, Gallic acid and Quercetin were -13.6 Kcal/mol, -13.2 Kcal/mol, -13.0 Kcal/mol, -13.8 Kcal/mol and -13.6 Kcal/mol, respectively. For the standard, Donepezil the total energy was -13.1 Kcal/mol. Thus, while the binding energy and other energy terms are varied, the total energy of the six complexes are similar, this indicates that there are subtle differences between the interactions of each molecule to AChE that gives rise to the specificity and affinity.

Protein–ligand binding are fall into 4 category like H-bonds, hydrophobic, ionic, and water bridges [49]. From the MD results, we plotted these different protein-ligand interactions and analyzed them (Figure 2).

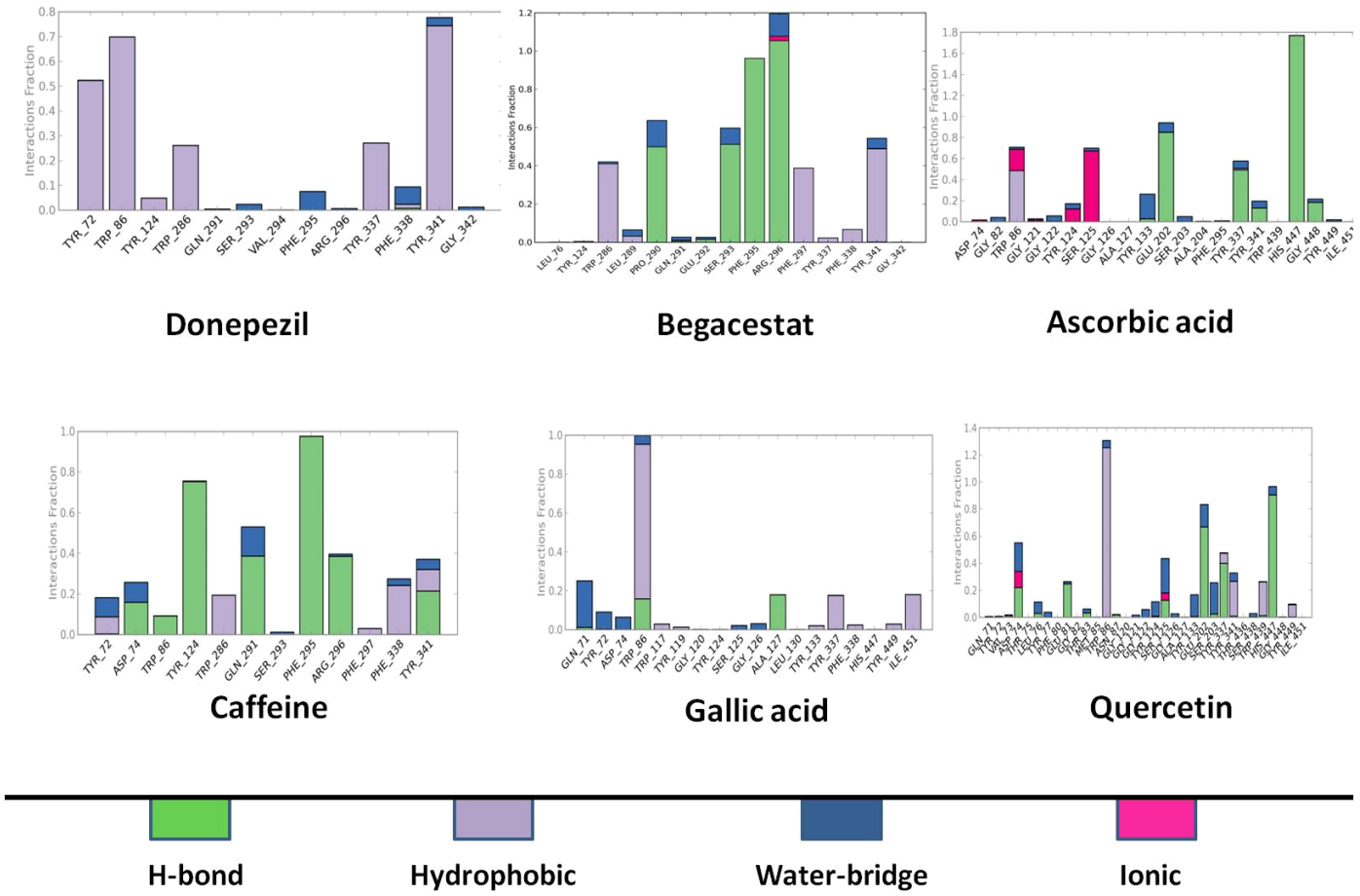


Figure 5.2: **Protein-Molecule interactions.** Normalized stacked bar graph illustration of interactions and contacts over the course of mechanical phenomenon (values over 1.0 are attainable as some residue create multiple contacts of same subtype with molecules).The color bar indexes represent various interaction types, such as green for hydrogen bonds, light violet for hydrophobic interactions, pink for ionic interactions, and blue for water-bridges.

The above given bar charts are normalized over the time duration of the trajectory to indicate if a residue made any interaction and colored based on the type of interaction. When a residue makes more than one type of contacts or multiple interactions of the same type then values over 1.0 are possible and they are shown in Figure 2 accordingly.

The interactions show that water bridges involve as a crucial player in binding of AChE, altogether the models at intervals the situation whereas hydrophobic and chemical element bonding interaction are the most contributors between the molecules and AChE. In Donepezil,

Tyr72, Trp86, Tyr124, Tyr337 and Tyr341 were major hydrophobic contributors whereas Ser293, Phe295 and Phe338 form water bridges with ligand molecule. In the case of Ascorbic acid His447, Gly448, Tyr341, Tyr337, and Glu202 are major H-bond contributor, while Ser125 and Trp86 form hydrophobic interactions. Begacestat form hydrogen bond with residues Arg296, Pro290, and Ser293. Gallic acid form fewer hydrogen bonds and Quercetin's major hydrogen bond contributors are Tyr133, Gly120, Trp86, Asn87, Asp74, His447. Tyr341, Trp439 and Tyr449 form hydrophobic interactions. Trp86 form major contributor (hydrophobic) interaction for ~96% of simulation time with atoms of Gallic acid, Quercetin and Donepezil respectively. Protein-ligand contacts are good indicators of system stability. Interestingly, Begacestat, Quercetin and Ascorbic acid show all four types of interactions, while the rest do not participate in all the 4types (H- bonds, hydrophobic, ionic, and water bridges).

5.3.3 Acetylcholinesterase (AChE) Inhibition Assay

AChE inhibition assay is one of the most common methods used to screen molecules with potential to be implicated in AD therapeutics. The assay is based on the spectrophotometric detection of the formation of yellow colored 5-mercapto-2-nitrobenzoic acid as a result of reaction between Ellamn's reagent or 5,5'-Dithiobis-(2-Nitrobenzoic acid (DTNB) and thiocholine; produced by AChE catalyzed hydrolysis of actylthiocholine in presence or absence of an inhibitor . Dose dependent inhibition of AChE activity was observed for donepezil and herbal molecule and their IC_{50} values are depicted in Table 2.

Table 5.2: Summary of the molecular docking studies and the *in-vitro* inhibition of AChE activity of molecules with AChE

Molecules	**Binding free Energy (Kcal/mol)**	**Inhibition Constant**	**Electrostatic Energy (Kcal/mol)**	**Intermolecular Energy (Kcal/mol)**	**IC_{50}**	**Top Ranked Configuration**
Donepezil	-8.60	494.82nM	-0.14	-10.39	39.46 ± 0.75^{b} nM	10
Begacestat	-6.49	17.54mM	-0.19	-8.87	104.8 ± 16.38^{a} mM	4
Quercetin	-8.68	435.45nM	-0.18	-9.57	250.6 ± 4.65^{b} µM	22
Ascorbic acid	-7.54	2.96mM	-1.03	-9.03	2312 ± 393.6^{b} µM	16
Gallic acid	-8.46	630.93nM	-0.03	-8.76	865.8 ± 2.16^{b} µM	4
Caffeine	-8.87	313.37nM	-0.29	-8.87	239.1 ± 12.95^{b} µM	24

Values are depicted as mean ± SD. Values with different superscript letter on the same row are significantly different from each other ($P > 0.05$).

Donepezil showed very strong inhibition (IC_{50} = 39.46 nM) in the nanomolar range, Caffeine, Quercetin and Gallic Acid showed substantial dose dependent inhibition of AChE (IC_{50} = 239.1, 250.6, and 865.8 µM respectively, shown in Figure 3).

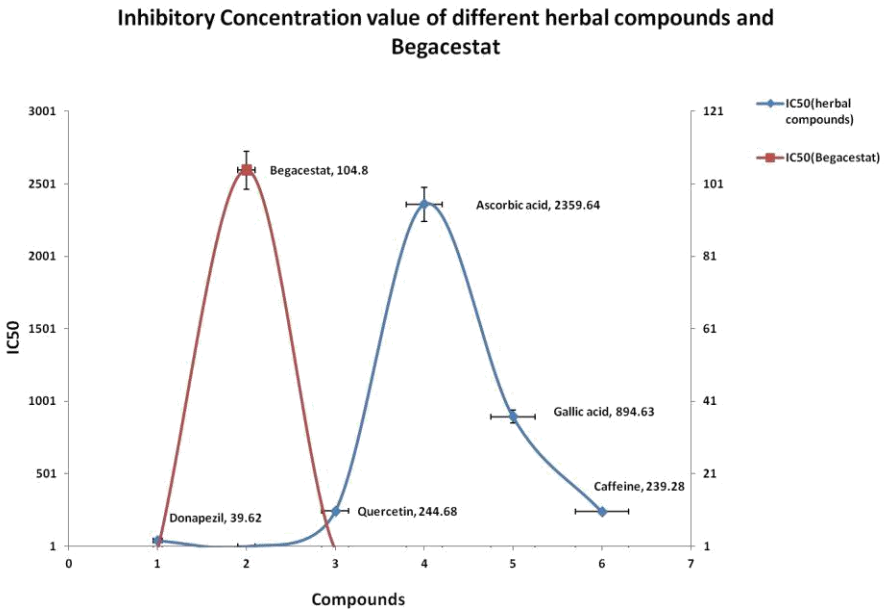


Figure 5.3: **Inhibitory Concentration (IC_{50}) value of molecules used in this study.** Inhibitory activity of six molecules is represented through IC_{50} values plotted as dose-response curve. Percent errors are shown in vertical lines. The results shown are of triplicates. Since Begacestat has μM activity shown with y-axis on the right side of the figure, whereas the y- axis on the left side of the figure is for the remainder of the compounds. Here, except Donepezil which has their activity in nM and remaining molecules has activity in μM.

However, no significant inhibition of AChE activity was observed for Begacestat at the test concentration and results revealed significantly low degree of enzyme inhibition. Further, concentration of DMSO used as a solvent was kept below 0.4%, which had negligible effect on AChE activity, however the previously reported IC_{50} [50] .

These results suggest that quercetin, caffeine and gallic acid could be exploited for AD therapeutics, as AChEI have shown potential to delay AD onset and progression [19-22]. Moreover, antioxidant and anti-inflammatory properties of these molecules are well documented [29, 30], which provides additive advantage in managing neurodegenerative conditions. Therefore, it is worthwhile to evaluate their potential against neurodegenerative condition. Herein, we used $HgCl_2$ induced neurodegeneration on primary hippocampus neuronal culture to mimic AD *in-vitro* and evaluated the neuroprotective potential of these molecules.

The neurotoxicity assay is usually carried out for drugs which may be beneficial, on neuronal cell lines or primary neuronal culture that induces Alzheimer's like neuronal damage through apoptotis, necrosis, and by breaking the cytoskeletal structure [10, 11, 13, 14, 51-53] .

Results of the *in-vitro* cytotoxicity revealed a dose-dependent reduction in percent cell viability for Quercetin, Caffeine, Ascorbic acid, Gallic acid, Donepezil, and Begacestat, when tested at 1, 10 and 20μM concentration (Figure 4).

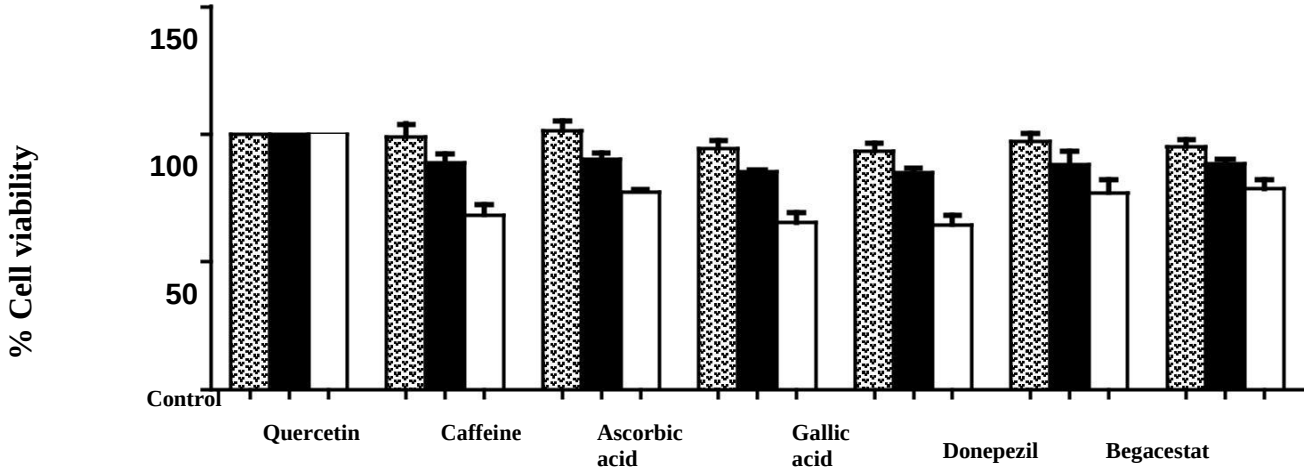


Figure 5.4: Percent cell viability calculated through MTT-assay. Results are depicted as mean ± SD (n = 3). All the drugs were tested at three concentrations, viz. 1 μM, 10 μM and 25 μM.

None of the test molecules showed any significant signs of neurotoxicity at 1μM and 10 μM concentrations.Thus, a 10 μM concentration for each test compound was used for following experiments. Treating the neuronal cells with $HgCl_2$ (1, 25 and 50 μM) resulted in high toxicity and percent cell viability was observed to be below 20%. We further evaluate the effect of test molecules on $HgCl_2$ (25 μM) induced neurotoxicity (Figure 5).

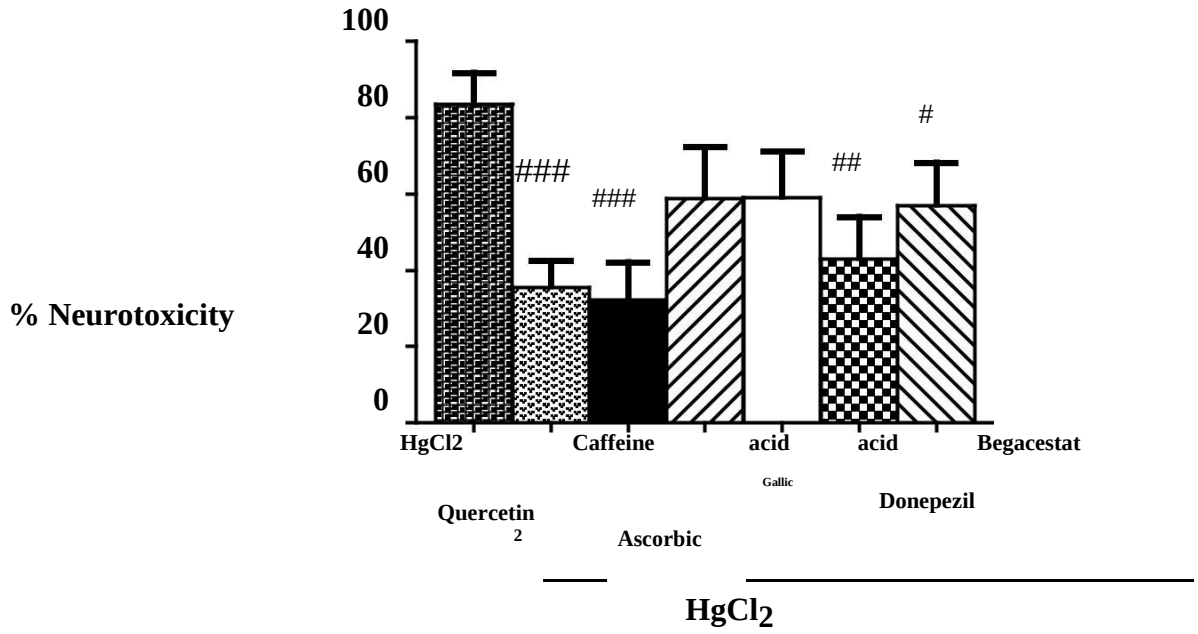


Figure 5.5: Impact of drug treatment on HgCl2 instigated neurotoxicity computed through MTT-measure. Results are shown to as mean± SD (n = 3). Factual criticalness was controlled by one way ANOVA taken after by Dunnet's numerous examination posthoc test at *p<0.05, **p<0.01 and ***p<0.001 (HgCl2 v/s treatments).

These results indicate that Quercetin, Caffeine, Donepezil, and Begacestat possess potential rescue like activity to the neurons from neurodegeneration due to neurotoxins.

5.3.4 Acridine Orange (AO) - Ethidium bromide (EtBr) assay

Fluorescence intensity of AO-EtBr staining were evaluated using image-J software in term of total cell fluorescence (TCF) (Figure 6A). Our results revealed a significant reduction in AO fluorescence intensity in $HgCl_2$ treated neurons, when compared to the control. Primary hippocampal neurons treated with Quercetin, Caffeine and Donepezil showed significantly higher fluorescence intensity of AO stain, indicating higher number of viable neurons. Ascorbic acid, Gallic acid, and Begacestat treatments did not showed any significant difference from $HgCl_2$ treated neurons, suggesting that these drugs are not capable of protecting neurons against $HgCl_2$ induced toxicity (Figure 6B). These findings become further evident from the observed

fluorescence intensity of EtBr staining where $HgCl_2$, Ascorbic acid, and Gallic acid did not showed any significant difference amongst them. Moreover, control cells and neurons treated with Quercetin, Caffeine, Donepezil, and Begacestat revealed significantly lower EtBr fluorescence when compared to $HgCl_2$ treatment (Figure 6C). To get a better insight into the effect of various treatments on $HgCl_2$ induced neurotoxicity, we calculated ratio of AO:EtBr (live:dead) fluorescence intensity, which clearly demonstrated that live to dead ratio was significantly reduced in $HgCl_2$ treated neurons and neurons treated with Ascorbic acid and Gallic acid. Quercetin, Caffeine, Donepezil, and Begacestat showed significantly higher ratio (Figure 6D), suggesting their potential to sustain higher number of viability in treated culture.

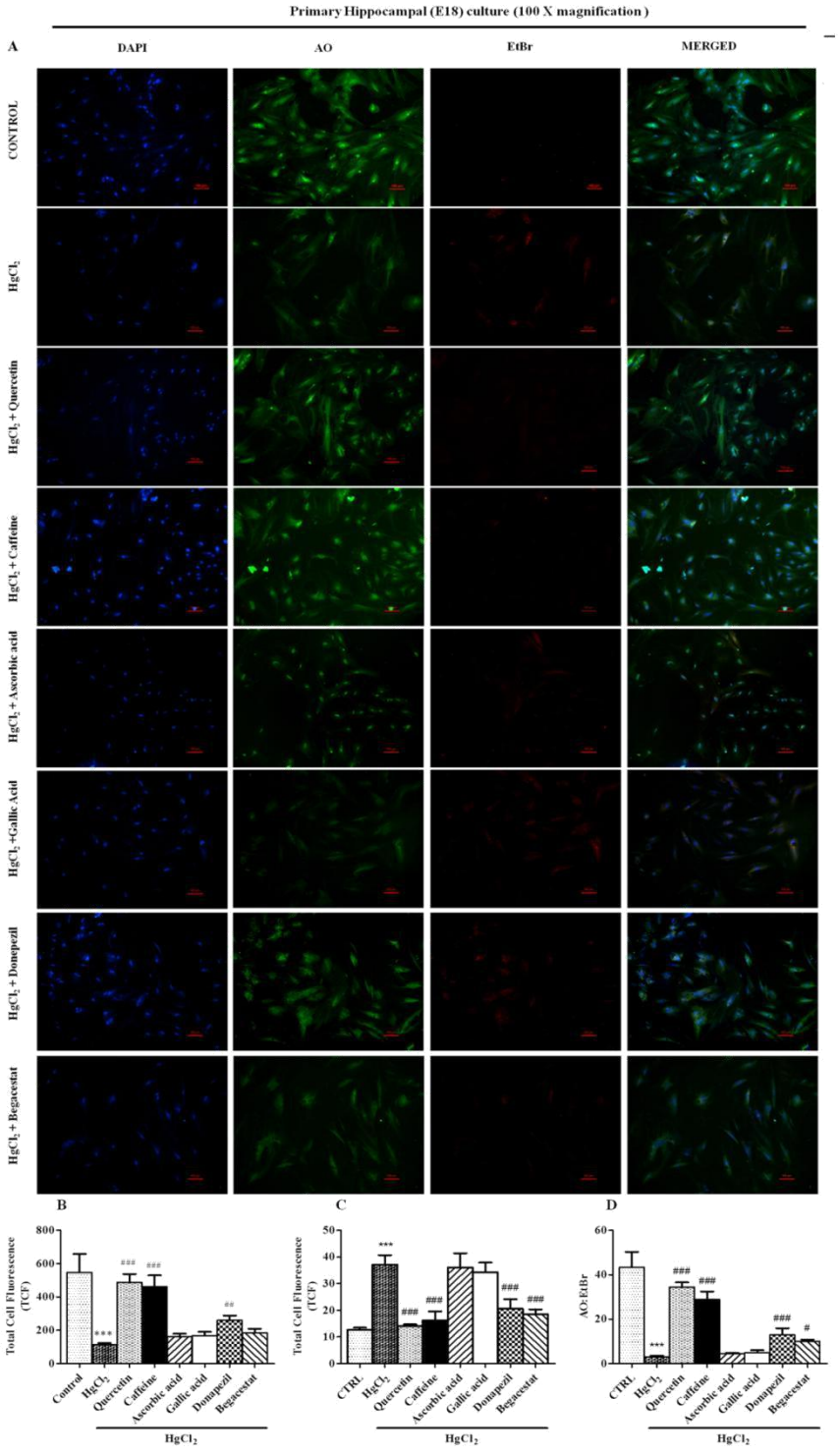


Figure 5.6 (A-D): (A): AO-EtBr (alive-dead) staining. Effect of various treatments on $HgCl_2$ induced neuronal alterations in terms of AO and EtBr fluorescence (A); Total cell fluorescence of AO staining depicted as green fluorescence (B); Total cell fluorescence of EtBr staining

depicted as red fluorescence **(B, C, D):** Ratio of AO:EtBr fluorescence intensity. Results are depicted as mean ± SD (n = 3). Statistical significance was determined by one way ANOVA followed by Dunnett's multiple comparison post hoc test at * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$ ($HgCl_2$ v/s treatments).

5.4 CONCLUSION

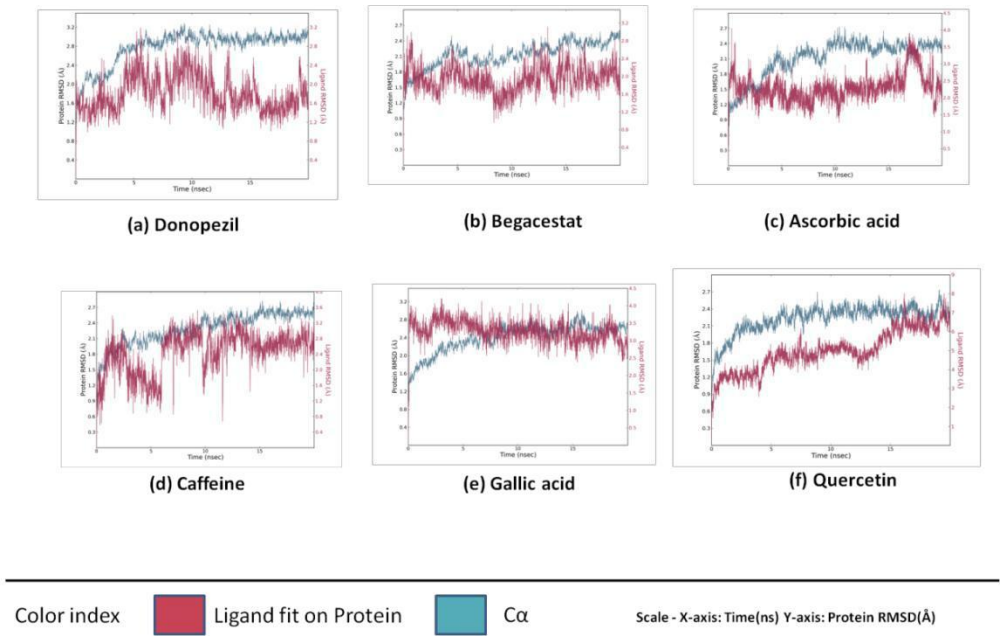
In a first of its kind study, we report that natural compounds can have disease modifying properties and modulate the cholinergic pathway, thereby, delaying the progression of AD. We analyzed the binding mechanisms of 646 small molecules to AChE using complimentary approaches including computational (molecular docking and MD simulations) and predicted quercetin, caffeine, ascorbic acid, gallic acid and begacestat to me most promising one. Results of *in-vitro* AChE inhibition assays confirmed that these molecules could be exploited as a potential therapeutic agent to inhibit AChE activity, especially quercetin and caffeine. Further, we demonstrated none of the molecules possesses neurotoxicity (10 μM) and quercetin and caffeine possesses inherited potential to protect neurons against $HgCl_2$ neurotoxin through their anti-apoptotic and anti-necrotic properties. These findings provide us with a strong evidence that quercetin and caffeine could fine clinical application in the management in AD not only by upregulating cholinergic signaling but also by providing neuroprotection and improving neuronal survival, besides, imparting protection against neuronal oxidative and inflammatory stress.

Supplementary Figures

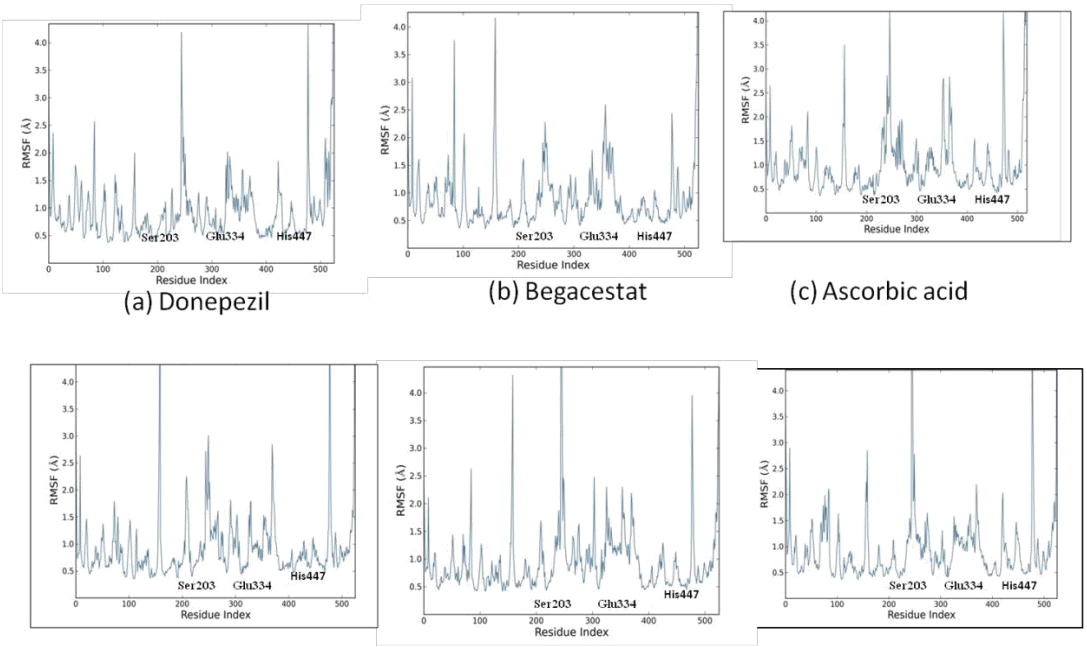
Supplementary Figure S1. Protein-Molecules RMSD. The Root Mean Square Deviation (RMSD) is employed to calculate the average change in displacement of specific range of atoms for a specific frame with relevance to a reference frame. RMSD analysis at 20 ns molecular dynamics simulation runs for Cα atom of AChE protein and different molecules. Cα atoms are represented by blue color arc whereas red color arc represent molecules.

Supplementary Figure S2. Protein RMSF in presence of different molecules. RMSF of 1B41 ranging from 0.5-4.5 angstroms, the active region of the protein shows less fluctuations of Cα

which indicates the stable region. The residue which shows minimum fluctuation were Ser203, Glu334 and His447 (catalytic triad)



Supplementary Figure S1



Supplementary Figure S2

REFERENCES

[1] A. s. Association, "2016 Alzheimer's disease facts and figures," *Alzheimer's & Dementia,* vol. 12, pp. 459-509, 2016.

[2] P. P. Panigrahi and T. R. Singh, "Computational studies on Alzheimer's disease associated pathways and regulatory patterns using microarray gene expression and network data: Revealed association with aging and other diseases," *Journal of theoretical biology,* vol. 334, pp. 109-121, 2013.

[3] M. Bernabei, S. Chiavarini, C. Cremisini, and G. Palleschi, "Anticholinesterase activity measurement by a choline biosensor: application in water analysis," *Biosensors and Bioelectronics,* vol. 8, pp. 265-271, 1993.

[4] H. Kavirajan and L. S. Schneider, "Efficacy and adverse effects of cholinesterase inhibitors and memantine in vascular dementia: a meta-analysis of randomised controlled trials," *The Lancet Neurology,* vol. 6, pp. 782-792, 2007.

[5] M. Iriti, S. Vitalini, G. Fico, and F. Faoro, "Neuroprotective herbs and foods from different traditional medicines and diets," *Molecules,* vol. 15, pp. 3517-3555, 2010.

[6] A. Park and M. Stacy, "Disease-modifying drugs in Parkinson's disease," *Drugs,* vol. 75, pp. 2065-2071, 2015.

[7] D. Galimberti and E. Scarpini, "Disease-modifying treatments for Alzheimer's disease," *Therapeutic advances in neurological disorders,* vol. 4, pp. 203-216, 2011.

[8] E. E. Longbrake, B. J. Parks, and A. H. Cross, "Monoclonal antibodies as disease modifying therapy in multiple sclerosis," *Current neurology and neuroscience reports,* vol. 13, p. 390, 2013.

[9] J. Kalra and A. Khan, "Reducing Aβ load and tau phosphorylation: Emerging perspective for treating Alzheimer's disease," *European journal of pharmacology,* vol. 764, pp. 571-581, 2015.

[10] J. Tong, Y. Wang, and Y. Lu, "In vitro evaluation of inorganic and methyl mercury mediated cytotoxic effect on neural cells derived from different animal species," *Journal of Environmental Sciences,* vol. 41, pp. 138-145, 2016.

[11] E. Van Vliet, S. Morath, C. Eskes, J. Linge, J. Rappsilber, P. Honegger, *et al.*, "A novel in vitro metabolomics approach for neurotoxicity testing, proof of principle for methyl mercury chloride and caffeine," *Neurotoxicology,* vol. 29, pp. 1-12, 2008.

[12] G. Verdile, K. N. Keane, V. F. Cruzat, S. Medic, M. Sabale, J. Rowles, *et al.*, "Inflammation and oxidative stress: the molecular connectivity between insulin resistance, obesity, and Alzheimer's disease," *Mediators of inflammation,* vol. 2015, 2015.

[13] X.-J. Xing, Q. Rui, M. Du, and D.-Y. Wang, "Exposure to lead and mercury in young larvae induces more severe deficits in neuronal survival and synaptic function than in adult nematodes," *Archives of environmental contamination and toxicology,* vol. 56, pp. 732-741, 2009.

[14] F. Xu, S. Farkas, S. Kortbeek, F.-X. Zhang, L. Chen, G. W. Zamponi, *et al.*, "Mercury-induced toxicity of rat cortical neurons is mediated through N-methyl-D-Aspartate receptors," *Molecular brain,* vol. 5, p. 30, 2012.

[15] B. Klimova and K. Kuca, "Multi-nutrient dietary intervention approach to the management of Alzheimer's disease–a mini-review," *Current Alzheimer Research,* vol. 13, pp. 1312-1318, 2016.

[16] J. Mendiola-Precoma, L. Berumen, K. Padilla, and G. Garcia-Alcocer, "Therapies for prevention and treatment of Alzheimer's disease," *BioMed research international,* vol. 2016, 2016.

[17] S. D. Petersson and E. Philippou, "Mediterranean diet, cognitive function, and dementia: a systematic review of the evidence," *Advances in Nutrition: An International Review Journal,* vol. 7, pp. 889-904, 2016.

[18] N. Scarmeas, Y. Stern, M. X. Tang, R. Mayeux, and J. A. Luchsinger, "Mediterranean diet and risk for Alzheimer's disease," *Annals of neurology,* vol. 59, pp. 912-921, 2006.

[19] J. Birks, "Cholinesterase inhibitors for Alzheimer's disease," *Cochrane Database Syst Rev,* p. CD005593, Jan 25 2006.

[20] T. C. Russ and J. Morling, "Cholinesterase inhibitors for mild cognitive impairment," *Cochrane Database of Systematic Reviews,* vol. 9, 2011.

[21] C.-C. Tan, J.-T. Yu, H.-F. Wang, M.-S. Tan, X.-F. Meng, C. Wang, *et al.*, "Efficacy and safety of donepezil, galantamine, rivastigmine, and memantine for the treatment of Alzheimer's disease: a systematic review and meta-analysis," *Journal of Alzheimer's Disease,* vol. 41, pp. 615-631, 2014.

[22] Q.-F. Zhao, L. Tan, H.-F. Wang, T. Jiang, M.-S. Tan, L. Tan, *et al.*, "The prevalence of neuropsychiatric symptoms in Alzheimer's disease: systematic review and meta-analysis," *Journal of affective disorders,* vol. 190, pp. 264-271, 2016.

[23] Y. Li, S. Hai, Y. Zhou, and B. R. Dong, "Cholinesterase inhibitors for rarer dementias associated with neurological conditions," *The Cochrane Library,* 2015.

[24] J. O'Regan, K. L. Lanctôt, G. Mazereeuw, and N. Herrmann, "Cholinesterase inhibitor discontinuation in patients with Alzheimer's disease: a meta-analysis of randomized controlled trials," ed, 2015.

[25] A. Ahmad, S. A Shah, H. Badshah, M. J Kim, T. Ali, G. H Yoon, *et al.*, "Neuroprotection by vitamin C against ethanol-induced neuroinflammation associated neurodegeneration in developing rat brain," *CNS & Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders),* vol. 15, pp. 360-370, 2016.

[26] L. G. Costa, J. M. Garrick, P. J. Roquè, and C. Pellacani, "Mechanisms of neuroprotection by quercetin: counteracting oxidative stress and more," *Oxidative medicine and cellular longevity,* vol. 2016, 2016.

[27] G. Du, Z. Zhao, Y. Chen, Z. Li, Y. Tian, Z. Liu, *et al.*, "Quercetin attenuates neuronal autophagy and apoptosis in rat traumatic brain injury model via activation of PI3K/Akt signaling pathway," *Neurological research,* vol. 38, pp. 1012-1019, 2016.

[28] S. Endesfelder, U. Weichelt, E. Strauß, A. Schlör, M. Sifringer, T. Scheuer, *et al.*, "Neuroprotection by caffeine in hyperoxia-induced neonatal brain injury," *International journal of molecular sciences,* vol. 18, p. 187, 2017.

[29] V. Mehta, A. Parashar, A. Sharma, T. R. Singh, and M. Udayabanu, "Quercetin ameliorates chronic unpredicted stress-mediated memory dysfunction in male Swiss albino mice by attenuating insulin resistance and elevating hippocampal GLUT4 levels independent of insulin receptor expression," *Hormones and behavior,* vol. 89, pp. 13-22, 2017.

[30] V. Mehta, A. Parashar, and M. Udayabanu, "Quercetin prevents chronic unpredictable stress induced behavioral dysfunction in mice by alleviating hippocampal oxidative and inflammatory stress," *Physiology & behavior,* vol. 171, pp. 69-78, 2017.

[31] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, *et al.*, "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function," *Journal of computational chemistry,* vol. 19, pp. 1639-1662, 1998.

[32] G. Kryger, M. Harel, K. Giles, L. Toker, B. Velan, A. Lazar, *et al.*, "Structures of recombinant native and E202Q mutant human acetylcholinesterase complexed with the snake-venom toxin fasciculin-II," *Acta Crystallographica Section D: Biological Crystallography,* vol. 56, pp. 1385-1394, 2000.

[33] C. T. Porter, G. J. Bartlett, and J. M. Thornton, "The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data," *Nucleic acids research,* vol. 32, pp. D129-D133, 2004.

[34] K. J. Bowers, E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, *et al.*, "Scalable algorithms for molecular dynamics simulations on commodity clusters," in *Proceedings of the 2006 ACM/IEEE conference on Supercomputing*, 2006, p. 84.

[35] D. W. Borhani and D. E. Shaw, "The future of molecular dynamics simulations in drug discovery," *Journal of computer-aided molecular design,* vol. 26, pp. 15-26, 2012.

[36] M. J. Robertson, J. Tirado-Rives, and W. L. Jorgensen, "Improved peptide and protein torsional energetics with the OPLS-AA force field," *Journal of chemical theory and computation,* vol. 11, pp. 3499-3509, 2015.

[37] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," *J. Am. Chem. Soc,* vol. 118, pp. 11225-11236, 1996.

[38] A. O. Ademosun, G. Oboh, F. Bello, and P. O. Ayeni, "Antioxidative properties and effect of quercetin and its glycosylated form (Rutin) on acetylcholinesterase and butyrylcholinesterase activities," *Journal of evidence-based complementary & alternative medicine,* vol. 21, pp. NP11-NP17, 2016.

[39] M. Alcolea-Palafoxa, P. Posada-Morenob, I. Ortuño-Sorianob, J. L. Pacheco-del-Cerroc, C. Martínez-Rincónc, D. Rodríguez-Martínezc, *et al.*, "Research Strategies Developed for the Treatment of Alzheimer's Disease. Reversible and Pseudo-Irreversible Inhibitors of Acetylcholinesterase: Structure-Activity Relationships and Drug Design," *Drug Design and Discovery in Alzheimer's Disease,* p. 426, 2015.

[40] S. D., "Insilico Identification of Potential Acetylcholinesterase Inhibitors Fromipomoea Aquatica Forsk for the Treatment of Alzheimers Disease," *International Journal of Research in Pharmaceutical and Biomedical Sciences,* vol. 4, pp. 1002-1010, 2013.

[41] X. Barril, M. Orozco, and F. Luque, "Towards improved acetylcholinesterase inhibitors: a structural and computational approach," *Mini reviews in medicinal chemistry,* vol. 1, pp. 255-266, 2001.

[42] C. Seniya, G. J. Khan, and K. Uchadia, "Identification of potential herbal inhibitor of acetylcholinesterase associated Alzheimer's disorders using molecular docking and molecular dynamics simulation," *Biochem Res Int,* vol. 2014, p. 705451, 2014.

[43] I. Molino, L. Colucci, A. M. Fasanaro, E. Traini, and F. Amenta, "Efficacy of memantine, donepezil, or their association in moderate-severe Alzheimer's disease: a review of clinical trials," *The Scientific World Journal,* vol. 2013, 2013.

[44] M. Bajda, A. Więckowska, M. Hebda, N. Guzior, C. A. Sotriffer, and B. Malawska, "Structure-based search for new inhibitors of cholinesterases," *International journal of molecular sciences,* vol. 14, pp. 5608-5632, 2013.

[45] K. Atkovska, S. A. Samsonov, M. Paszkowski-Rogacz, and M. T. Pisabarro, "Multipose binding in molecular docking," *International journal of molecular sciences,* vol. 15, pp. 2622-2645, 2014.

[46] S. Release, "1: Desmond Molecular Dynamics System, version 3.7," *DE Shaw Research, New York, NY, Maestro-Desmond Interoperability Tools, version,* vol. 3, 2014.

[47] P. U. Lee, H. R. Churchill, M. Daniels, S. C. Jameson, and D. M. Kranz, "Role of 2C T Cell Receptor Residues in the Binding of Self–and Allo–Major Histocompatibility Complexes," *Journal of Experimental Medicine,* vol. 191, pp. 1355-1364, 2000.

[48] R. Jaenicke, "Stability and stabilization of globular proteins in solution," *Journal of Biotechnology,* vol. 79, pp. 193-203, 2000.

[49] U. Yadava, H. Gupta, and M. Roychoudhury, "Stabilization of microtubules by taxane diterpenoids: insight from docking and MD simulations," *Journal of biological physics,* vol. 41, pp. 117-133, 2015.

[50] T. Mohamed, W. Osman, G. Tin, and P. P. Rao, "Selective inhibition of human acetylcholinesterase by xanthine derivatives: in vitro inhibition and molecular modeling investigations," *Bioorganic & medicinal chemistry letters,* vol. 23, pp. 4336-4341, 2013.

[51] A. F. Castoldi, S. Barni, I. Turin, C. Gandini, and L. Manzo, "Early acute necrosis, delayed apoptosis and cytoskeletal breakdown in cultured cerebellar granule neurons exposed to methylmercury," *Journal of neuroscience research,* vol. 59, pp. 775-787, 2000.

[52] T. Stoiber, G. H. Degen, H. M. Bolt, and E. Unger, "Interaction of mercury (II) with the microtubule cytoskeleton in IMR-32 neuroblastoma cells," *Toxicology letters,* vol. 151, pp. 99-104, 2004.

[53] T. Toimela and H. Tähti, "Mitochondrial viability and apoptosis induced by aluminum, mercuric mercury and methylmercury in cell lines of neural origin," *Archives of toxicology,* vol. 78, pp. 565-574, 2004.

CHAPTER 6

Overall Conclusions and Future Prospects

6.1 CONCLUSIONS

We have studied to analyze AD at comprehensive level. The overall goal of this research was to decipher the key biomolecules such as genes and proteins involved in AD at cellular, molecular, structural, functional level and also to find genetic association with disorder for better understanding of the complex molecular mechanism involved in AD. The overall schema and their proposed outcomes to deal with AD using various computational approaches for its system level understanding are given in **Figure 6.1**.

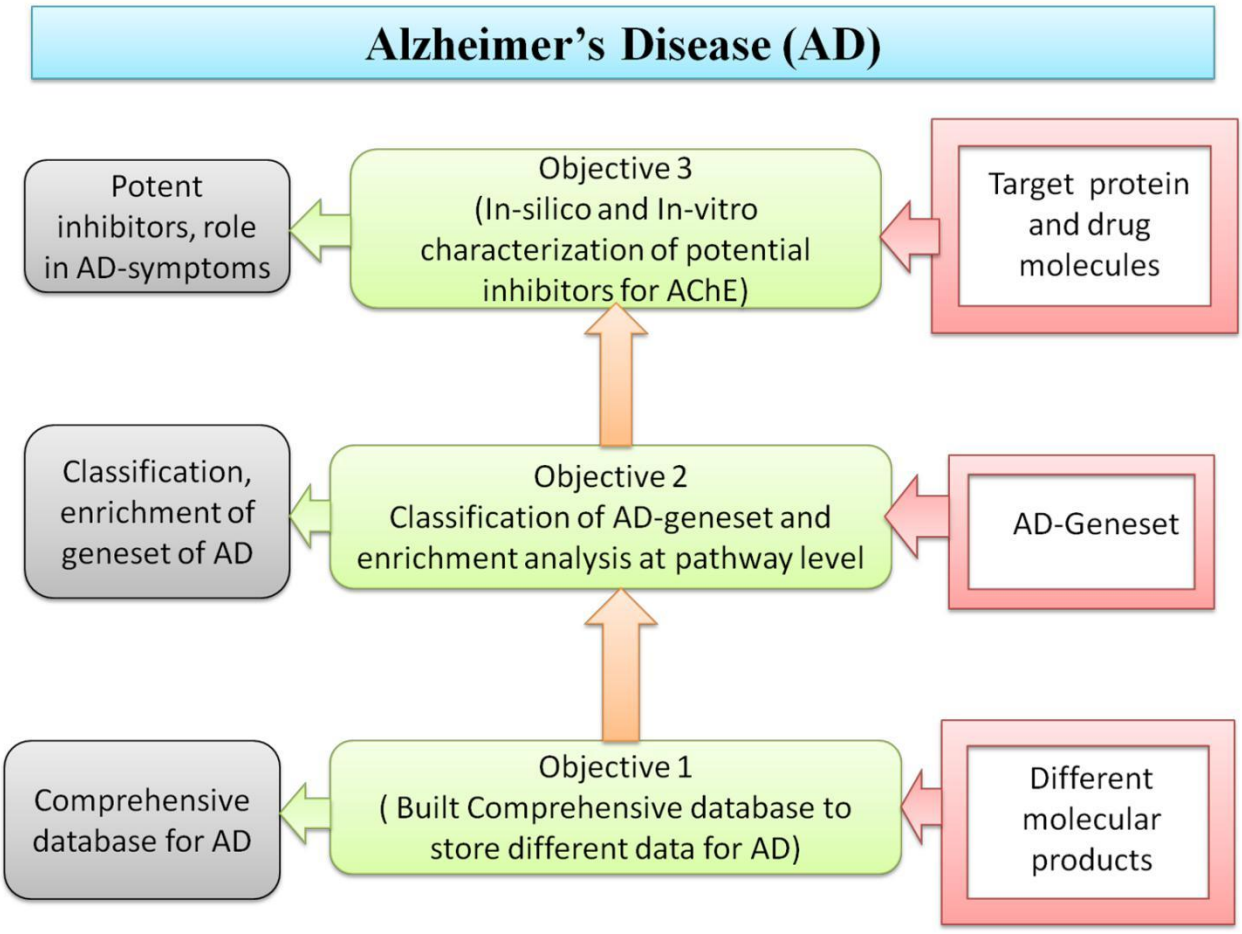


Figure 6.1: Overall schema of AD research objectives and their proposed outcomes specifically for early Detection of disease

Important outcomes of this thesis are being summarized here:-

- Our web interface (ABCD) contains information concerning the proteins, genes, transcription factors, SNP's, miRNAs, mitochondrial genes and expressed genes implicated in AD pathogenesis.

- In addition to this molecular level data, the database has information for animal models, medicinal candidates and pathways involved in AD and some image data for AD patients.

- Our database is coupled with some major external resources where user can retrieve additional information about the disease.

- Their massive storage space has contributed to disperse nature of the biomarker information and help the researcher in finding effective biomarkers search for AD.

- Classification model i.e Decision tree were generated on the basis of MMSE and relevance score help in finding gene expression of various protein involve in AD pathway.

- The resultant decision tree authenticity depends on accuracy, sensitivity and specificity output.

- Gene enrichment analysis found 7 genes, i.e APOE, PSEN1, GRN, ACE, BCHE, PRNP, IL1A are strongly associated to AD on the basis of XD score.

- XD- score and Fisher q-value play very crucial role in deciding relation of gene datasets with associated metabolic pathways of AD.

- Network properties like shortest path length, Node betweenness, Clustering coefficient and eigen vector centrality play important role in describing topological properties..

- Identification of some new genes with their association with different metabolic process could be predicted that AD is directly or indirectly associated with those genes and they may play as role of biomarkers.

- We predicted quercetin, caffeine, ascorbic acid, gallic acid, begacestat and donepezil to be potential AChE inhibitors and demonstrated quercetin, caffeine, begacestat and donepezil to be efficient AChE inhibitors through *in-vitro* assay.

- We report that natural compounds can have disease modifying properties and modulate the cholinergic pathway, thereby, delaying the progression of AD.

- Quercetin and Caffeine possesses inherited potential to protect neurons against $HgCl_2$ neurotoxin through their anti-apoptotic and anti-necrotic properties.

- Quercetin and Caffeine could find clinical application in the management in AD not only by upregulating cholinergic signaling but also by providing neuroprotection and improving neuronal survival, besides, imparting protection against neuronal oxidative and inflammatory stress.

6.2 FUTURE PROSPECTS

- In Future, we comprehend our database by incorporation of High throughput normalized and quality controlled data (NGS), also Image segmented data of MRI for detection and differentiation between AD and MCI in different tissue of brain. We update our database on regular basis to provide state-of-the-art information to the scientific community.

- Beside Decision trees approach of machine learning (ML) techniques, we can apply other techniques of machine learning to develop model for classification of genes and other entities to set standard for early detection of AD.

- Experimental validation are require to support the computational enrichment analysis of genes and for closely connected pathways which could be possibly involve in regulatory mechanism of onset of AD.

- After analyzing the effect of different type of drug compounds on AChE for AD both computationally and experimentally, we can test their effect on different interconnected signaling pathways to AD.

LIST OF PUBLICATIONS

1. **Kumar A** and Singh TR (2016) A new decision tree to solvve the puzzle of Alzheimer's disease pathogenesis through standard diagnosis scoring system. *Interdesciplinary Science: Computational Life Science, 9(1), 1913-2751* **(I.F.: 0.8; SCI and SCOPUS indexed).**

2. **Kumar A** and Singh TR (2017) Analysis for biological network properties of Alzheimer's disease associated genesets by enrichment and topological examination. *International journal of Bioinformatics Research and Applications, 13(3)* **(SCOPUS indexed).**

3. **Kumar A**, Bansal A, Panigrahi P, Singh TR. ABCD: An Open Source Comprehensive Database for Alzheimer's disease. *Neuroinformatics* **(I.F: 2.8; Under Revision).**

4. **Kumar A**, Mehta V, Raj U, Varadwaj PK, Malairaman U, Yennamalli RM, Singh TR. Characterization of Potential Acetylcholinesterase inhibitors: Implications towards delaying the progression of Alzheimer's Disease. *Journal of Computer-Aided Molecular Design* (**I.F: 3.028; under revision).**

CONFERENCE PUBLICATIONS

Ashwani Kumar, Tiratha Raj singh: Systems biology approach for gene set enrichment and topological analysis of Alzheimer's disease pathway. International Conference on Bioinformatics and Computational Systems Biology; BSB-16, IIIT-A; 03/2016.

Ashwani Kumar, Vineet Mehta, Pritish Kumar Varadwaj, Utkarsh Raj, Udayabanu M., Tiratha Raj Singh: Flexible docking and simulation studies of herbal compounds as Acetylcholinesterase inhibitor(AChEI) for Alzheimer's disease. IAN-2015, Panjab University, Chandigarh,Punjab; 10/2015.

Ashwani Kumar, Utkarsh Raj, Pritish Kumar Varadwaj, Tiratha Raj Singh: Insights From Docking And Molecular Dynamic Simulation of Acetyl cholinesterase Inhibitors (AChEI) Structural Model For Possible Therapeutic Of Alzheimer's Disease(AD). National Conference on Bioinformatics Panorama in Agriculture and Health,2015, SHIATS, Allahabad; 10/2015.

Ashwani Kumar, Tiratha Raj Singh: A new decision tree to solve the puzzle of Alzheimer's disease pathogenesis through two standard scoring systems. National network for Mathematical and Computational biology (NNMCB)-2014; IIT-MANDI; 10/2014