Integrated Bioinformatics Analysis to Identify Critical Genes and Potential Drug Candidate Discovery

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE

OF

Mater of Technology

In

Bioinformatics

Submitted by

Diksha Semwal

(2K18/BIO/05)

Delhi Technological University, Delhi, India

Under the guidance of

Prof. Pravir Kumar

Department of Biotechnology

Delhi Technological University

Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India

DECLARATION

I, Diksha Semwal, 2K18/BIO/05 student of M.Tech. Bioinformatics, hereby declare that the Dissertation titled –"**Integrated Bioinformatics Analysis to identify critical genes and potential drug candidate discovery**" which is submitted by me to Department of Biotechnology, Delhi Technological University, Delhi in the partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed for the basis for the award of any degree, Diploma Associate-ship, Fellowship or other similar title or recognition.

Date: 15 August, 2020

DIKSHA SEMWA L
(2K18/BIO/05)
DEPARTMENT OF BIOTECHNOLOGY
DELHI TEHNOLOGICAL UNIVERSITY

CERTIFICATE

This is to certify that the Project Dissertation entitled "**Integrated Bioinformatics Analysis to Identify Critical Genes and Potential Drug Candidate Discovery**" submitted by Diksha Semwal (2K18/BIO/05) in partial fulfillment of the requirement for the award of the degree of Master of Technology from Delhi Technological University, is an authentic record of the candidate's own work carried out by her under my guidance. To best of my knowledge this work has not been submitted in part and full for any Degree or Diploma to this University or elsewhere.

Date: 15 August, 2020

Prof. Pravir Kumar
(SUPERVISOR)
Department of Biotechnology
Delhi Technological University

ACKNOWLEDGEMENT

At the time of submission of my M.Tech. Dissertation, I would first like to thank GOD for giving me patience, strength, capability, and willpower to complete my work. Apart from our efforts, the success of this project depends largely on the encouragement and guidelines of many others. I, therefore, take this opportunity to express my gratitude to the people who have been instrumental in the successful completion of this project.

My initial thank is addressed to my mentor Prof. Pravir Kumar, Professor, Department of Biotechnology, Delhi Technological University, who gave me this opportunity to work in a project under him. It was his enigmatic supervision, constant encouragement and expert guidance, which have enabled me to complete this work. I humbly seize this opportunity to express my gratitude to him.

I would also like to extend my sincere gratitude to Professor Jai Gopal Sharma for providing basic infrastructure and facilities.

I would like to extend my sincere gratitude to Dr. Rashmi Ambasta for her keen observation and continuous advice. Her thoughtful inputs on the project issues have been invaluable for the productive progress of the project. I extend my thanks to technical staff Mr. Jitender Singh and Mr. C.B. Singh, who had been an aid whenever required. Lastly, I wish to extend my thanks to my family and friends who have supported me through the entire process.

DIKSHA SEMWAL

2K18/BIO/05

ABSTRACT

Alzheimer's disease is one of the world's leading causes of dementia. It is one of the most deadly and irreversible neurodegenerative diseases. There is a limited supply of novel drugs introducing in the market because of safety reasons or lower pharmacokinetic properties. Drug repurposing offers the potential to avoid the slow speed of the drug discovery and it's commercialization after clinical trials. Hence, reducing the cost of drug production. The main advantage of adopting the drug repurposing strategy is that the clinical trials have previously been done and safety issues have already been inspected and the drugs have been commercialized and in the market. As this way helps to speed up providing the alternate drugs for AD, this study shows the interaction of the known four cancer drugs (Selinexor, Entrectinib, Lenvatinib, BIIBO21) with two target (FN1 and E1F4A3) involved in AD using in-silico techniques. A computational method was adopted to find the dysregulated genes involved in the disease. The microarray analyses were done on the GSE28146 dataset, which was extracted from the Gene Expression Omnibus (GEO) database, out of all the expressed genes in the array 59 were upregulated and 68 were downregulated. These genes were used to generate a viable gene networks and hub gene network of the top 10 dysregulated genes. Top 4 genes were used as a target for the selected drugs and their interaction was investigated by molecular docking. In this study the best drug was chosen based on the best docking score for both the targets. The current study discloses the potential of Selinexor as a new drug against two targets associated with AD.

Keywords: Alzheimer's disease, microarray analyses, molecular docking, drug repurposing.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Alzheimer's disease (AD) is the most common subtype of dementia which is officially listed as the sixth-leading cause of death worldwide. Recent estimates indicate that AD ranks third behind heart disease and cancer as the major cause of death in the elderly [1]. AD is characterized by deposits of amyloid-beta (Aß) plaques and intracellular neurofibrillary tangles (NFT) in the neocortical and limbic regions of the brain [2]. As a group of unknown primary degenerative and irreversible progressive brain diseases, AD causes neuronal cell apoptosis and brain atrophy and slowly destroys memory, cognitive ability and the ability of the body to perform basic bodily functions such as walking and swallowing, seriously affecting quality of the life [3].

As this happens, the symptoms become more severe over time. There are three stages of disease progression. In the early stages people are sometimes seem experiencing lapses of memory and having problems finding the right words. As the disease progresses, they may become more confused and frequently forgetting the names of people, places, appointments or recent events. Experience frequent mood swings, feel sad or angry abnormally, or due to communication problems. Finding difficulty in carrying out everyday activities- they may get puzzled by checking their change at the shops or become unsure how to work with TV remote. Become more withdrawn, due either to a loss of confidence or due to contamination problems.

As the disease progression is seen, people with Alzheimer's will often need more support from those who care for them. Eventually, they will need help in all their daily affairs and day-to-day activities. While there are some common symptoms of Alzheimer's disease, it is important to note that everyone is unique and might have their own symptomatic features. No two people are likely to experience Alzheimer's disease in the same way as each other. It thus appears that AD is a complex disease for which suitable therapeutic approaches are not currently recognized. Although drug development is improving, the complexity of AD makes therapeutic approaches challenging. To improve AD therapy, a deeper understanding of the molecular mechanisms causing the disease are required. To further

understand the mechanisms of AD pathogenesis, high-throughput gene expression data has been investigated and substantial progress has been made in reconstructing gene regulatory networks.

There are reports linking autophagy with the Alzheimer's disease progression, as the accumulated amyloid-beta is not cleared due to dysfunctional autographic machinery [4].In the current study we analyzed the GSE28146 dataset from Gene expression omnibus and extracted upregulated and downregulated genes. The two genes selected for the study are EIF4A3 and FN1, prior is upregulated and was selected as reports suggest it's involvement in PI3K signaling, activation of this pathway suppresses autophagy function ,thus by inhibiting it the restoration of autophagy function can be achieved [5] while latter one is downregulated and also reported to be regulating PI3K signaling machinery such that high expression of the gene can suppress the neuronal death and apoptosis process thus delaying the disease progression [6]. So these targets can be proposed for targeting the novel drugs in the study to regulate the neuronal death and decay and the molecular docking scores can help in investigating and proposing a potential novel drug candidate for the disease.

CHAPTER 2

REVIEW OF LITERATURE

One of the greatest challenges faced by scientist over the past 50 years is to decode the cognitive and behavioral explanation of dementia and their connection with the responsible brain pathology. These challenges have been increasing ever since as the population is ageing and there has been a substantial increase in age-related dementia in neurodegenerative diseases. Although this disease is prevalent for a thousand years, it was only in the early past century that scientists were able to find out the inside pathology with the development in molecular biology, biostatistics, computational biology over time.

Alzheimer's is recognized by the impairment of cognition and behavior of people over 65 years of age [1]. It is the most common and deadly neurodegenerative disease. It has already affected about 25 million people worldwide and is estimated to affect four times more people by 2050. Thus, researchers are quite keen on deciphering the mechanism behind it and searching for a perfect drug candidate to treat it. The clinical feature for the disease is the accumulation of the amyloid-beta (Aß) extracellularly and formation of neurofibrillary tangles intracellularly due to hyperphosphorylation of tau protein.

2.1 Down the timeline

Alois Alzheimer was born in Germany in a middle-class family. He studied science and graduated in honors in a medical school in berlin. In 1888 he joined Frankfurt mental hospital as an intern.

A year later he was joined by a talented Franz Nissl in his research group. He could stain neurons perfectly, a technique very important for studying neuronal biochemistry. He married a woman named Cecilia, and had three children with her. After seven years of their marriage his wife died. And to avoid grief he worked in hospitals and spent most of his time with patients. There he met Aguste Deter, 51- year old female in November 1901 [1].

She exhibited abnormal behavior like amnesia, losing the sense of directions, crying overnight forget to express emotions. He closely observed her symptoms for years. After the patient died he did staining experiments after biopsy and observed plaques and neurofibrillary tangles in the neurons in the cerebral cortex. This part of the brain is responsible for memory, language, and ability to judge and think.

Thus all these aspects were impaired. The plaques were generally observed in 70 and above aged humans. This case was of early-onset. He published his findings and named the disease as Alzheimer's disease.

2.2 Epidemiology

Studies showed that in 2015 around 46.8 million people were affected by Alzheimer's in a report made by the World Health Organization (WHO). The number of new cases is 30% higher than the cases that were presented in the report of 2010. The highest cases were in Asia(49%), Europe (25%)and America(18%) [2][3]. By these reports the researchers have predicted that by 2030 the number of cases will reach to 75.4 million people and by 2050 it will be 131 million cases worldwide.

According to the latest report East Asia and Western Europe have more than 30% of the total cases. According to the surveys done in past it was confirmed that Alzheimer's is proportional to ageing. It was found through statistical studies that 5% of people of 65 years of age were found to be positive with the disease, and 20% of people of 80 years and above were found to be suffering from the disease. Thus, it was concluded that prevalence of the disease doubled every five years [4]. According to the World Alzheimer's Report (2015) [3], the overall cost of care and treatment including social care, medical care and informal care was 818 billion dollars with hike of 31.6% compared to the same study in 2010 report [2]. The cost trend will rise upto 1 trillion dollars in 2030[3].

2.3 Pathophysiology

Even with so much history on Alzheimer's researchers still don't have the exact explanation of the disease and not just that there is no therapeutic treatment to treat the disease. But down the timeline with the advancement of molecular biology and invention of sequencing techniques the scientists were able to find the macroscopic and microscopic markers behind the disease. All this huge biological data of the disease will help them to come close to finding the pathogenesis and finding the treatment [5][6]. After doing brain scans (figure 1) we can see atrophy of hippocampus region and cerebral cortex region as well in old-aged Alzheimer's patients [7][8].

Researchers observed the formation of amyloid-beta and hyperphosphorylated tau protein tangle formation in the neurons which leads to neuronal loss. These are considered as clinical markers for the disease. Recent studies shows that these markers can be a result of genetic defects that are passed through generations as heredity [9], mechanisms containing apolipoprotein E [10] , mechanisms involving the oxidation process [6].

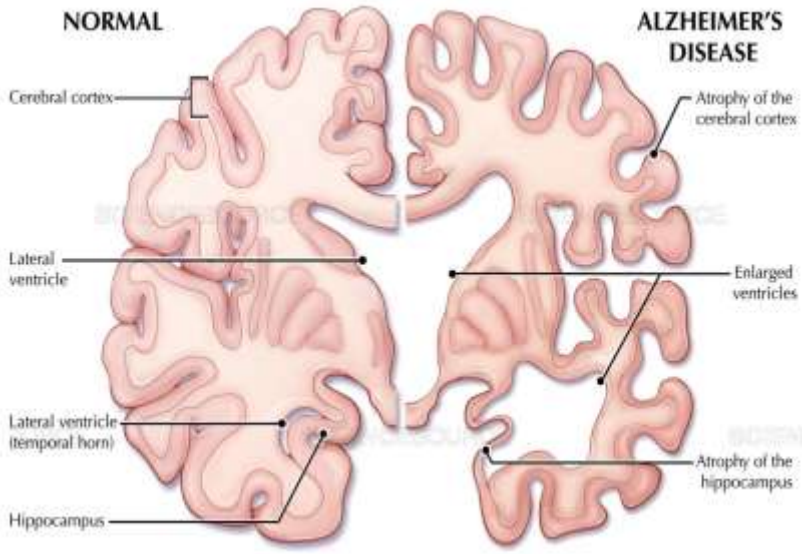

FIGURE 1: Brain scans showing atrophy of hippocampus and cerebral cortex (Science Source Images)

2.3.1 Genetic mutation

A maximum variety of Alzheimer's cases occur due to genetic level abnormalities and are passed on generation to generation. These are rare mutation that occur in the ApoE gene. This gene codes for a protein named apolipoprotein E, which combines with the fats or lipids inside our body to form lipoprotein. The key role of this protein is to package the lipids like cholesterol and help them to travel in the bloodstream easily. It is proved that a mutation in a particular allele of the gene i.e apolipoprotein E4 allele is the primary reason for sporadic Alzheimer's disease [10]. Only 20% of individuals all over the world are ApoE4 carriers, but these account for 65% of the cases of prevalent disease [11]. The mechanism that show the direct connection between AD and ApoE is still unknown but it is shown in few studies that it involve dysregulation in clearance of Aß in such cases [12]. There is also two other alleles present in AD-patients presenilin-1 and presenilin-2 (PSEN-1, PSEN2) and also in people carriers of the disease [13].

These two proteins are homologous and homology studies show that they have 60% similarity. These two proteins are transmembrane proteins and are involved in membrane trafficking, and they have seven transmembrane domain. The proteins are involved in the gamma-secretase activity. Gamma-secretase cleaves the APP protein which was cleaved by either α-secretase or ß-secretase into 42-amino acid long Aß into a 40- amino acid long Aß protein. Now with the mutation leading to dysfunction of γ-secretase, the downstream result of the processing of APP protein leads to the end product which is Aß (42-amino acid long). This is highly prone to aggregation and form plaques in the neurons as compared to the 40-amino acid long Aß.

Another gene involved in the pathogenesis of AD is Amyloid precursor protein (APP), which is metabolically cleaved by two unique pathways- amyloidogenic and non-amyloidogenic involving three separate secretase namely α-secretase, ß-secretase and γ-secretase. Normally it was cleaved by α and γ , while in amyloidogenic processing it is done by ß and γ secretase to produce $Aß_{40}$ and $Aß_{42}$ fragments. Till now 30 different mutation were found in the APP gene and almost every one of them caused AD in individuals.

2.3.2 Amyloid hypothesis and Tau protein

As we know the presence of amyloid-beta and tau tangles is taken as a marker for the disease. These plaque deposits form due to incorrect folding of the Aß protein produced by the dysregulated cleavage of APP protein by the secretase enzymes. APP is a transmembrane protein, present on the cell-like CNS neurons. The cleavage of APP is done by enzymes α-secretase, ß-secretase and γ-secretase [14]. Under normal conditions , APP is cleaved by alpha and gamma secretases, but due to gene defect gamma doesn't work in patients resulting in cleavage by alpha and beta leading to in formation of 42 amino acid long Aß (Figure. 2). This isoform has cytotoxic properties that are linked to neurodegeneration, resulting in the production of free oxyradicals [15].

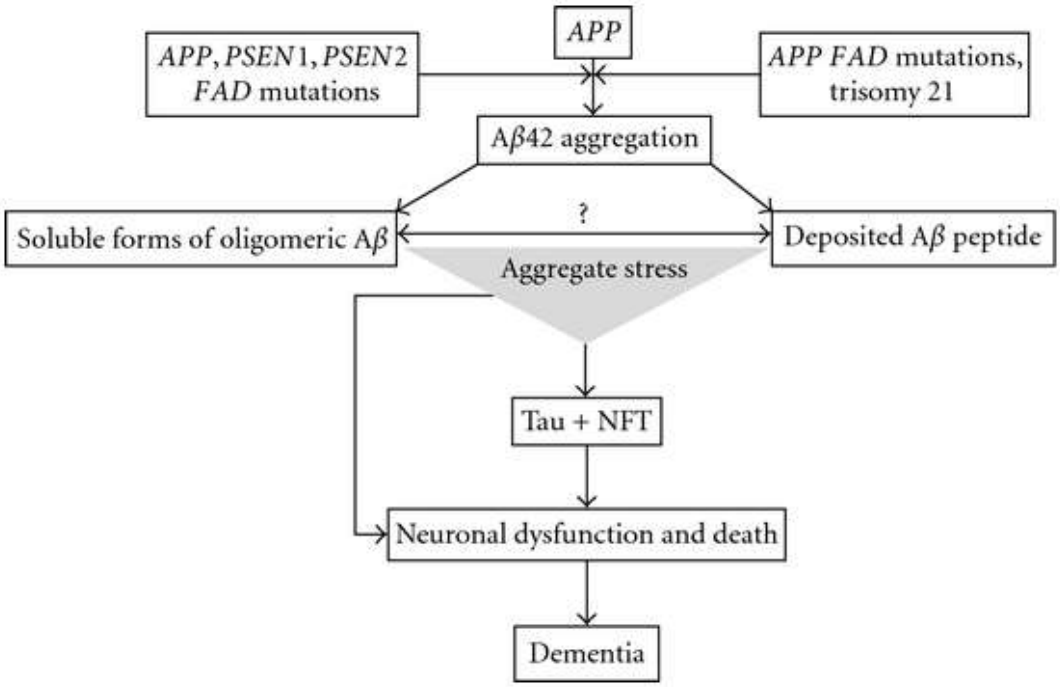

Figure 2: Cascade mechanism for Amyloid in AD [16]

These are toxic and harmful for the neuron cells. As we know,calcium signaling is very important for the transmission of information from one neuron to another through the synaptic cleft. Now, in AD the plaques of amyloid-beta can also function as calcium-permeable channels and result in the high efflux of calcium ions outside the neurons.

2.3.3. Mitochondrial dysfunction and Inflammation process

AD is a disease that is related to the inflammatory process [17]. Several studies have stated that Tau tangle formation also accelerate in time of the acute and severe inflammatory process. These inflammatory process starts around the plaque region of Abeta , induced by cytokines which leads to neurofibrillary tangles.

2.3.4 Oxidative stress

Studies state that oxidative stress is triggered by the accumulation of Abeta and is a very important consequence in the progression and pathogenesis of AD [5]. This stress in cells is the main reason for the inflammation process occurring during progression of the disease. The brain being the most active organ in our body requires a high energy supply, thus has a high amount of mitochondrial activity i.e oxidative phosphorylation going on[18]. This process produces huge amount of oxygen reactive species. Oxidative stress results due to the Overexpression of these species. In AD the protective mechanism involved in clearing out these species are defected due to the accumulation of Abeta plaques, resulting in the building of cytotoxicity and neuronal death.

2.4 Symptoms

AD is the major cause of dementia worldwide. With the progression of the disease the symptoms become severe as the memory decline, and the person loses a sense of movement, lack of language. AD can be categorized into three stages-

a. In the first stage there is the loss of memory is in parts not complete and there is loss of neurons in the brain , hippocampus shrinks by 25% of its original size , this is associated with the decline in the memory gradually as neurons linked with long and short term memory die out.
b. The intermediate stage includes failure in recognizing and talking to people. It is quite a long phase and can count from two to ten years. There is also reduction in acetylcholinesterase level in neurons, which are responsible for long and short term memory.
c. In the final stage, the person fails to perform their normal everyday activities, also has difficulty bin movement and requires constant assistance and care. This phase last up to 2 to 4 years, with the progression for them failing to recognize their friends and family members and eventually leads to the death of the patient.

As the damage has occurred in the limbic system there is no means to recover the lost memory. Due to the lower levels of Acetylcholinesterase which is essential for neurons involved in cognitive, memory processes to communicate with each other, the patients show irritability, sadness, loss of sense of direction and loss of good connection with members to family [19]. Over 95% of patients have issues like anxiety, depression, restlessness, hallucination, sleep disorder, diet disorder.

2.5 Diagnosis

The initial stage diagnosis is important to make sure that the patient has a good and manageable life. Doctors run a few tests such as mental and neuropsychological test to check the cognitive function of the patient. After that, clinical tests like a blood test, neuroimaging, MRI, checking cerebrospinal fluid test to check the presence of plaques and tangles in the neuron .Genetic test can also be done to confirm the stated tests and also in families having history it can be done to check the level of risk a patient has. For the whole cognitive function main clinical test is Mini-Mental State Examination (MMSE). It is the most employed test in the world [20]. This test check the responsiveness of the patient by asking question from the past events, analytical questions and situation questions. This tells a lot about the condition of the individual.

PET (Positron emission tomography) is now being employed in various parts of the world. It operates through a radiation signal to get a 3-D colored image of the body. The patient is injected to a radiotracer, which travels to organs, where it get metabolized and releases positron that is detected by a special camera and a computer to assess tissue and organ functions. So like this we get to see changes at the cellular level and detect the disease way before imaging texts can.

2.6 Treatment

There are six drugs approved by the Food and Drug Administration (FDA) for treating Alzheimer's – Galantamine, Donepezil, Rivastigmine, Memantine, Memantine mixed with Donepezil. These drugs better the condition temporarily by increasing the level of neurotransmitters in the brain region. The magnitude of effectiveness depends on person to person and also for a limited time. In 2004-2012, 244 drugs for Alzheimer's were tested and tested for clinical trials. In the end, only one drug named memantine could pass through the trials and get approved by the FDA [21]. Out of many difficulties scientists get while developing a treatment to disease is, animal models are not that reliable as they don't guarantee that a medication that works on them will work on humans as well. Also the speed of these trials is very slow, taking a lot of time.

Thus, doctors consider a non-pharmacological treatment which includes improving the cognitive functions of the patients and help them to perform there day to day activities-such as practices like listening to favorite music as a way to remember old events, computerized learning as an exercise for the brain. Meta-analyses have shown that aerobic exercise and non-aerobic exercise have shown a positive effect on the condition of patients on the pace of cognitive decline [22].

CHAPTER 3

METHODOLOGY

1. Data selection -

Figure 3 illustrate the flow chart of selecting data. The criteria for selecting were- the title of the study, design of the study, hippocampus based study specially the grey matter area. Treatment based study, blood-based study, and cell-line based study, were excluded.

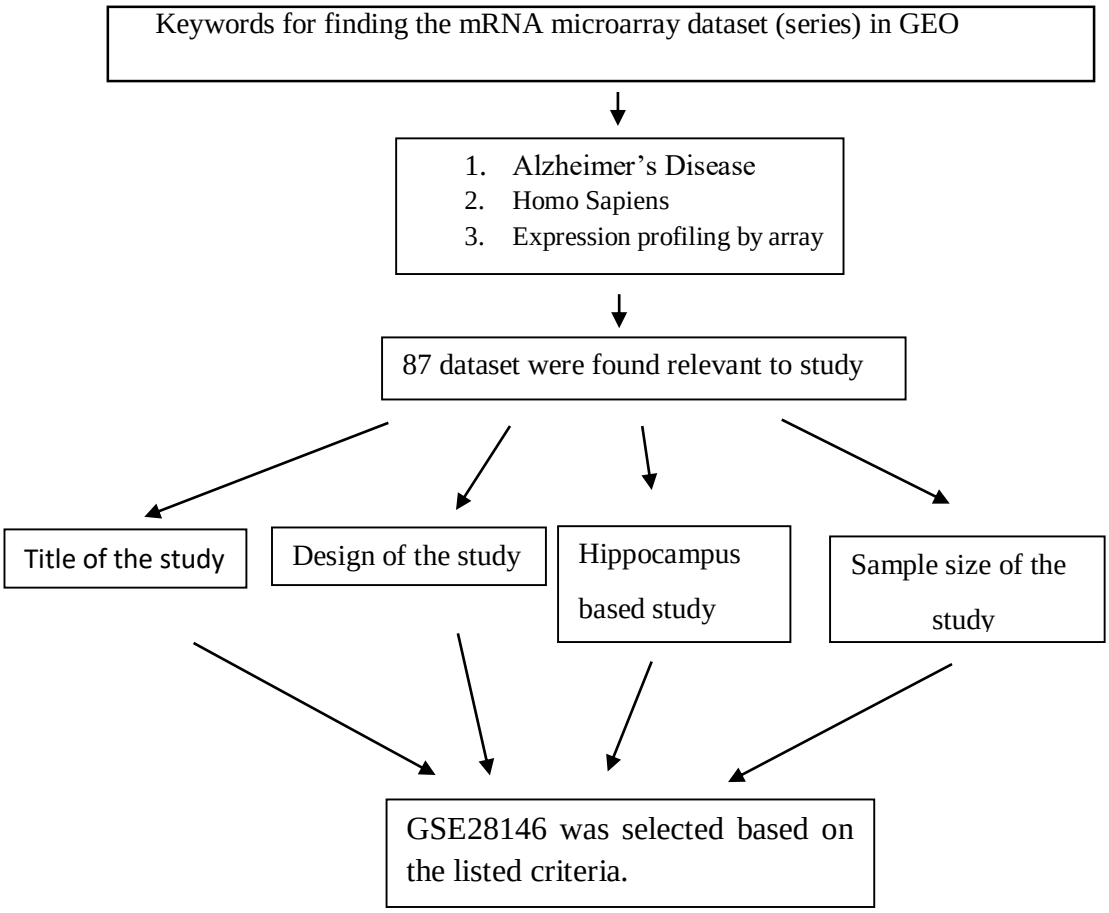

Figure 3: Flow chart of data selection.

2. **Extraction of gene expression data from database source-**

A dataset of hippocampus biopsies (GSE28146) was obtained from Gene expression omnibus (GEO).In it array and sequence-based data are accepted. It also provides us with tools to analyze gene expression profiles. The dataset was a high sample size of 30 produced on an Affymetrix Human Genome U133 Plus 2.0 Array platform. The probe set was annotated and transcription profile was generated using the MAS5 algorithm and annotation data sets [23].

3. **Microarray data processing and screening for DEGs-**

Gene expression omnibus is a public functional genomics data repository. In it array and sequence-based data are accepted. It also provides us with tools to analyze gene expression profiles.

In this study, the expression profile GSE28146 from the GEO database was used to identify DEGs. DEGs were identified by using AltAnalyzer platform (http://www.altanalyze.org/) where a p-value of < 0.05 and $|FC| > 1.25$ was considered as the cut-off. This analysis was performed by using the GEO2R tool where mRNAs that met the statistical criteria of an adjusted P-value (adj. P) <0.01 and $|\log \text{fold change}| > 1.5$ were considered to be DEmRNAs.

4. **Pathway and GO enrichment analysis of DEGs-**

To understand the functional changes of the DEGs, their biological processes (BP), molecular function (MF) and cellular components (CC) were analyzed. By doing this we will be able to decipher the related biological process the DEGs are affecting and their relation to the AD markers i.e. the top differentially expressed genes.

By using FunRich software (http://www.funrich.org/) we can find out pathway enrichment analyzes for all the deferentially expressed genes.

5. PPI network construction-

A PPI network was constructed using the STRING [24] database (https://string-db.org/cgi/input.pl?sessionId=C7InCHtTe3oe&input_page_show_search=on) network gave a diagrammatic description of the degree of connection between all the proteins encoded by the differentially expressed genes.

6. Identification of hub genes using cytoHubba-

The overlapped genes were further enriched in terms of topological factors like Degree, Maximum neighbourhood component, Maximal clique centrality and also on centralities like Closeness, Radiality, Betweenness etc. using CytoHubba plugin from Cytoscape (http://apps.cytoscape.org/apps/cytohubba).

7. Selecting suitable novel drugs for finding drug candidate-

Four FDA approved drugs were selected named BIIBO21, Entrectinib, Lenvatinib and Selinexor. These have proved to work against cancer and as previous studies have stated the genetic mechanism of cancer and Alzheimer's is the same to some extent [25], we can repurpose these drugs for treating Alzheimer's as well.

8. Molecular Docking between hub genes and drug candidate-

The selected drugs were docked on two of the dysregulated target hub gene, FN1 (PDB ID-3M7P) and EIF4AIII (PDB ID-2HXY) using AutoDock Vina on PyRx (https://pyrx.sourceforge.io/downloads), which is an automated docking tool. The drug giving a better binding score can be proposed as a potential drug for the treatment of underlying disease.

9. Analysing Blood Brain Barrier permeability of the drugs-

The drug for AD has to target the degenerating neurons in the brain, but during their course the drug has to pass the Blood-Brain Barrier in order to reach the brain. Thus, the second step after selecting drugs for a neurodegenerative drug is to check the BBB permeability and we investigated it with the PKCSM prediction tool (http://biosig.unimelb.edu.au/pkcsm/prediction).

CHAPTER 4

RESULTS AND DISCUSSION

1. Data download and preprocessing

AD gene expression profiles were download from the GEO database. The mRNA datasets were not normalized and were normalized using the AltAnalyzer tool. The series GSE28146 provides transcription data of 23 different patients from which samples of AD and the normal individual were acquired. DEGs were identified by using the AltAnalyzer platform where a p-value of < 0.05 and $|FC| > 1.5$ was considered as the cut-off.

A total of 45586 genes were identified which included 253 upregulated genes and 99 downregulated genes. The miRNAs that met the statistical conditions of the adjusted P-value (adj. P) <0.01 and $|\log \text{fold change}| > 1.5$ were considered to be DEmRNAs.

2. Functional enrichment analysis of DEG-

To understand the function of the DEGs, their biological processes (BP), molecular function (MF) and cellular components (CC) were analyzed by AltAnalyzer [26] . The DEGs were mainly enriched in protein metabolism (EIF4EBP2; FSTL1; FBXO32; SPINT2), Cell growth and/or maintenance (EIF4EBP2; FSTL1; FBXO32; SPINT2; USP28; PCSK2; RNF125; EIF2S1), Cell communication (GNRH1; PRKCG; ROBO1; EPM2A; ARL2; PLEKHA3; CHRNB1), Neurotransmitter metabolism (ACHE) and Regulation of cell cycle (RAD1) (Table 1,2).

S.No.	Term	P-Value	Genes
	Biological Process		
1	Protein metabolism	0.323635	EIF4EBP2; FSTL1; FBXO32; SPINT2; USP28; PCSK2; RNF125; EIF2S1
2	Cell growth and/or maintenance	0.810738	HOOK3; NF2; FNDC5; CLDN19
3	Neurotransmitter metabolism	0.009797	ACHE
4	Regulation of cell cycle	0.244994	RAD1
5	Regulation of gene expression, epigenetic	0.041759	MBNL1; SUPT3H1
6	Cell communication	0.461811	GNRH1; PRKCG; ROBO1; EPM2A; ARL2; PLEKHA3; CHRNB1; DYRK1B; RAB43; ANAPC1
7	Apoptosis	0.244506	PMAIP1; BNIP1;
8	Cell differentiation	0.117656	GDPD2
9	Transmembrane receptor protein tyrosine kinase signaling pathway	0.008589	NTRK1
10	Signal transduction	0.021977	ESYT1; DEF6; CDH7; CASS4; NMUR2; NRP2; CSNK1G3; PTPN1; EGFR; NXPH3; GDPD2; NTRK1
	Cellular Components		
1	Plasma membrane	0.096754	CD244; ESYT1; DEF6; CDH7; F5; IGSF1; NRP2; SLC46A1; EGFR; GDPD2; NTRK1; PAQR4

2	ER-Golgi intermediate compartment	0.141946	FN1
3	Integral to membrane	0.503764	MFAP3L; CDH7; NMUR2; NRP2; GAD2; NTRK1; RGR; KCNH5; SLC5A12
4	Nucleus	0.842791	RUNX3; DEF6; BNIP1; TYMS; CSNK1G3; PTPN1; EGFR; GATA4; DDX3Y; FN1
5	Actin cytoskeleton	0.491337	PDLIM5
6	Nucleolus	0.80548788	AXIN2; HOOK3; PRKCG; ROBO1; EPM2A; ANKS1B; GSTO1; NF2; FOXL2; NR1I3; HIVEP2; SUFU; POLR2C; NPAT; RAD1
7	Mitochondrion	0.892704	PRKCG; EPM2A; SNRPD3; ANAPC1; EIF2S1; ESYT1; PMAIP1; TYMS; SLC22A4
8	Cytosol	6.944444	EPM2A; SNRPD3; ANAPC1; EIF2S1; ACY1; PTPN1
9	Lysosome	0.409453	SLC46A1; EGFR; LYZ; SERPINB3; GH1; FN1; NPC2; ABCC10; CAT; GSTO1; DPM1; SPINT2
10	Endosome		EGFR; NTRK1; NF2

Table 1: A tabular representation of analysis of gene enrichment of DEG by FunRich.

S.No.	Description	P-Value	Genes
1	Cell surface interactions at the vascular wall	0.069753	CD244; FN1;
2	Signal transduction by L1	0.191573	EGFR
3	Hemostasis	0.023529	CD244; IFNA21; F5; PTPN1; GATA4; FN1
4	MicroRNA (miRNA) Biogenesis	0.118815	POLR2C
5	Axon guidance	0.721792	ROBO1

Table 2: A tabular representation of significant KEGG pathways identified for the DEGs in this datasetTerms with p-value < 0.05 were considered as significant.

Meanwhile, significantly enriched pathways highlighted were cell surface interactions at the vascular wall (CD244,FN1), microRNA biogenesis (POLR2C), axon guidance (ROBO1), elongation arrest and recovery (POLR2C), homeostasis (CD244,IFNA21,F5,FN1,GATA4) and neurotransmitter clearance in the synaptic cleft (ACHE), EGFR interacts with phospholipase C-gamma (EGFR) , mitotic telophase cytokines (ANAP1) , G2/M transition (CEP41).

3. Construction of PPI network and its module analysis

STRING database [24] was used to construct a PPI network with a pre-set criteria of displaying only query proteins and a minimum required interaction score of 0.4 as primary parameters.

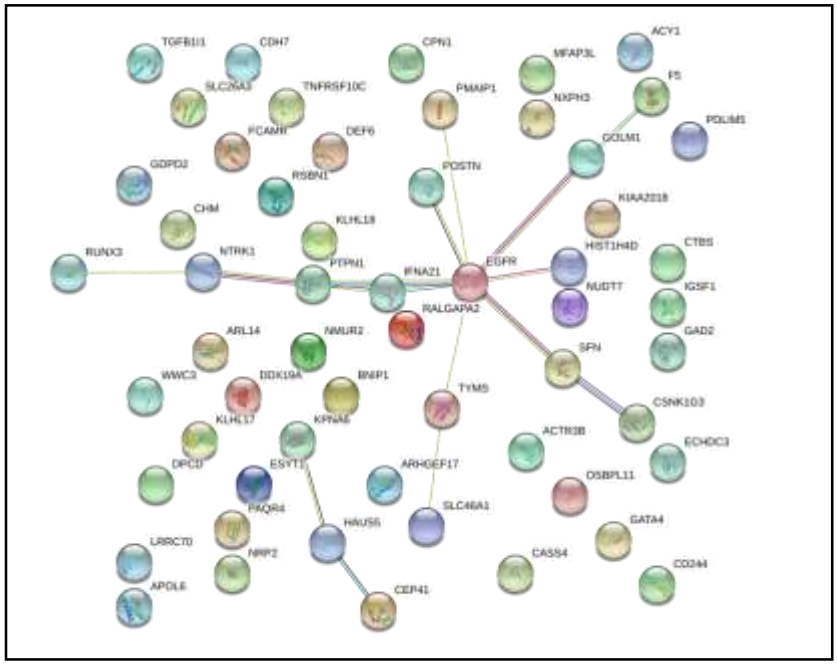

Figure 4: Protein interaction of all the downregulated genes.

The nodes connecting the protein tells about the degree of interaction between them. String database helps in viewing those highly populated nodes.

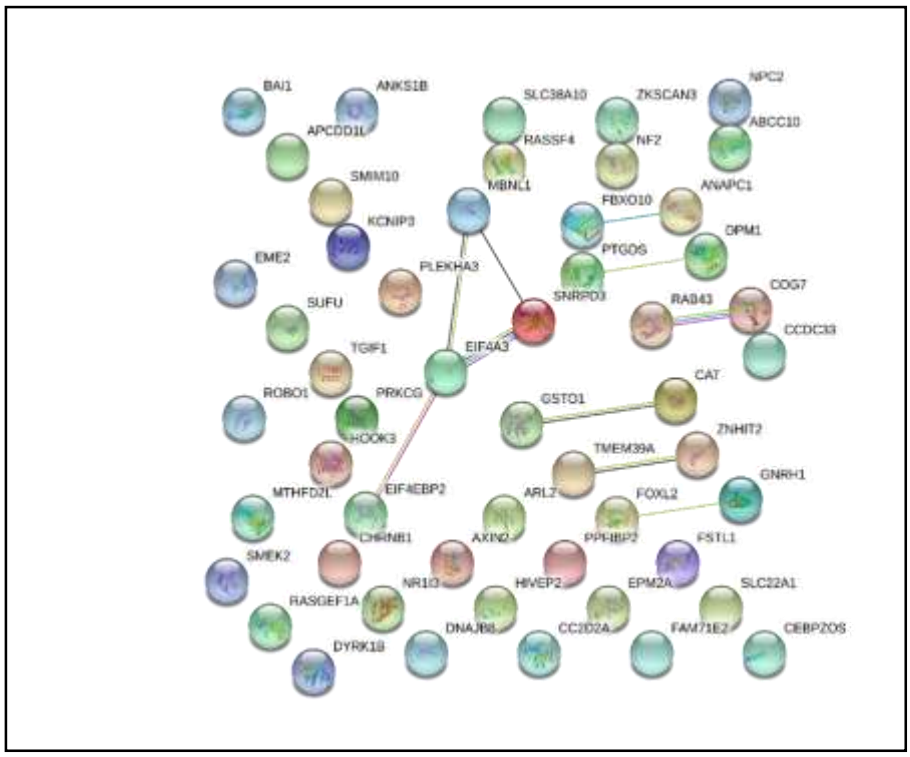

Figure 5: Protein interaction of all the upregulated genes

A network of 220 nodes and 82 edges were included in the DEGs network. We obtained the protein-protein network for both the upregulated and downregulated genes. The obtained network gives a visual representation of the interconnection between the genes.,also revealing the genes which are highly linked to more than two genes like EGFR, FN1, CD44 and PTPN1. These can be taken as potential targets for further studies.

4. Construction of gene network using Cytoscape –

Cytoscape [27] is an online available open-source software used to compile, analyze and visualize complicated networks obtained from complex biological data of gene expression. It performs this by using pre-set criteria of displaying only query genes and the minimum required intersection score of 0.4 as the primary parameter. The nodes are identified based on the centrality scores by looking at parameters like Eccentricity, Closeness, stress, Betweenness, Degree, and Radiality. EGFR has

shown to have 9 nodes. FN1 had 7 nodes.These were the genes with the highest scores.

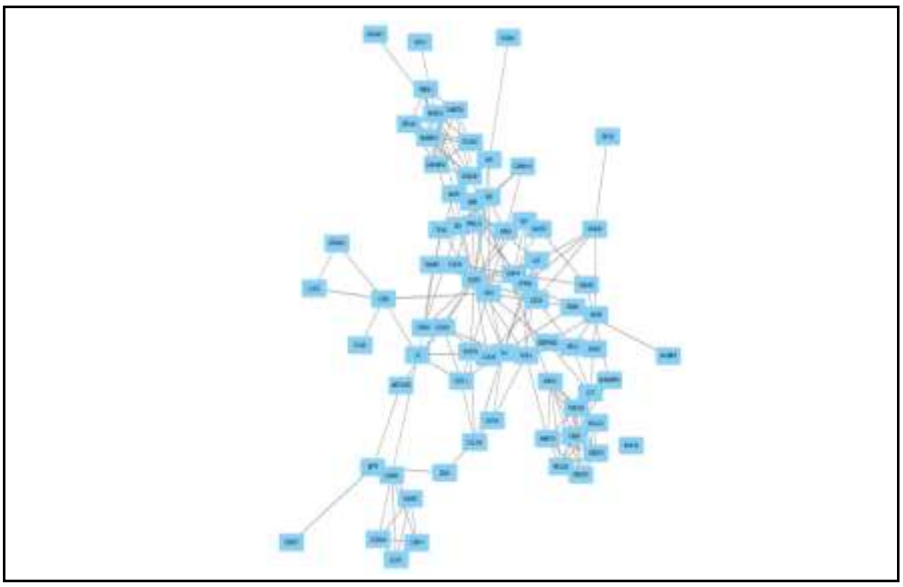

Figure 6: Gene interaction network of DEGs.

5. Identification of Hub Genes-

Hub genes are the genes with high correlation in candidate module. CytoHubba was employed to observe the intersection of five algorithms (degree, MCC, MNC, EPC and Eccentricity) to identify the hub genes present in the network [28]. Based on the given parameters 10 genes were identified out of which 4 were downregulated and 5 were upregulated genes. And based on the biological mechanism associated the genes FN1 and EIF4A3 were selected as the two target for further studies. So by regulating their expression we can control the expression of the other dysregulated genes associated with these genes. Fibronectin (FN) interactions play a very crucial role in synovial related disorders, tumor cell biology and Alzheimer's disease as w well [29]. EIF4A3 is a core protein of an RNA-binding protein Exon junction complex involved in RNA decay-mediated RNA splicing.It binds to a protein RBM8A which in AD patients is found at a high level and cause anxiety and lack of social interaction .

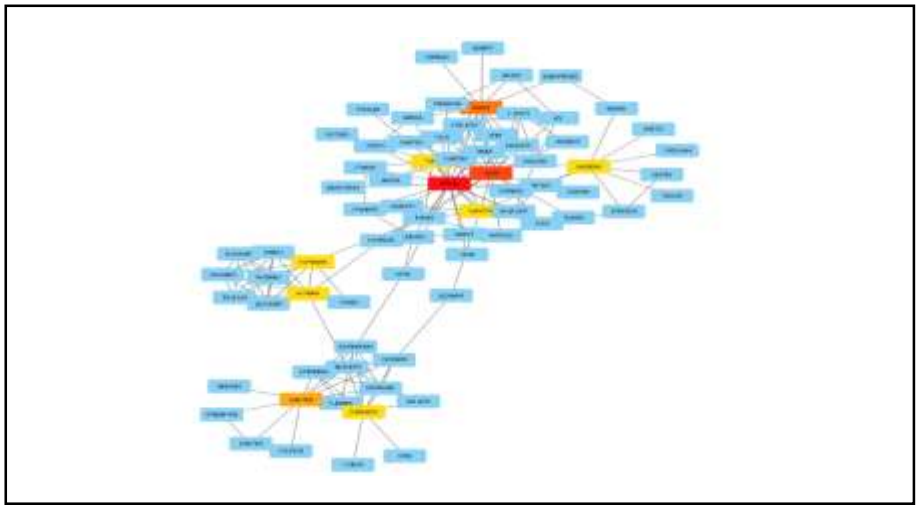

FIGURE 7: PPI network of Hub genes.

RANK	NAME	SCORE
1	EGFR	25
2	FN1	18
3	CD44	12
4	EIF4A3	11
5	MGEA5	9
5	ANAPC1	9
5	CAT	9
5	RNF14	9
5	PTPN1	9
5	POLR2C	9

Table 3: The identified hub genes along with score.

6. Extracting 3-D structure of the ligand-

The 3-D structure of BIIBO21 (investigatory drug) (PubChem ID-16736529), Entrectinib (PubChem ID-25141092), Lenvatinib (PubChem ID-9823820) and Selinexor (PubChem ID-71481097) was obtained from the Protein Data Bank

database. These drugs were selected from the recently approved drugs by the Food and Drugs Administration list [14]. Some studies have revealed cancer and Alzheimer's have very similar genetic mechanisms behind the pathophysiology[15]. Thus drugs working for treating cancer have high chances to work for Alzheimer's as well. These drugs have proved to be successful in treating cancer , and as per the previous studies have revealed cancer and Alzheimer's have very similar genetic mechanism behind the pathophysiology, as both diseases are caused due to disruption of cell growth and cell survival machinery [31].

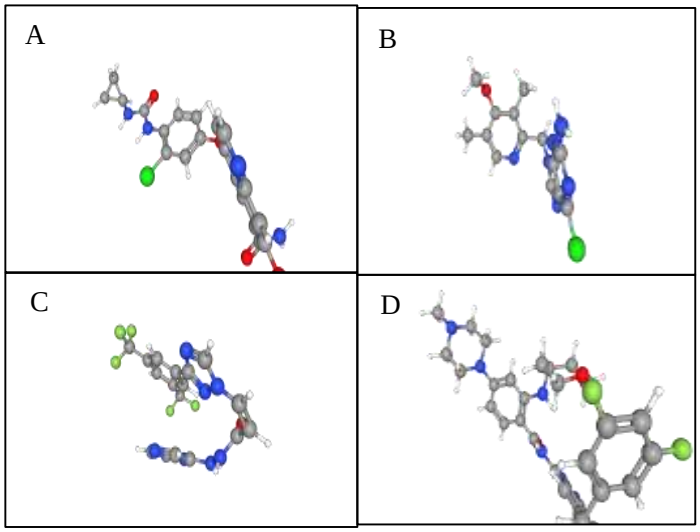

Figure 8: 3-D structure of the ligands; 3(a) 3-D structure of Lenvatinib (b) 3-D structure of BII2021 (c) 3-D structure of Selinexor (d) 3-D structure of Entrectinib.

7. Molecular Docking of the drugs on the target-

Molecular docking is a modelling technique used in bioinformatics that works on two or more small compounds to give a final product along with the binding affinities in terms of scores. Docking was done using AutoDock Vina using PyRx [16]. The binding affinity with the FN1 gene (PDB-2HAZ) for Selinexor was -5.9, BIIB021 was -5.4, Lenvatinib was -5.3 and Entrectinib was -5.7, while the binding affinity for EIF4A3 gene (PDB- 2HXY) for Selinexor was 7.90, BIIB021 was 6.40,

Lenvatinib was -7.80 and Entrectinib was -9.8. Thus, Entrectinib show the highest minimum binding energy for EIF4A3 and the second-highest for FN1. So we can reposition it for the treatment of Alzheimer's.

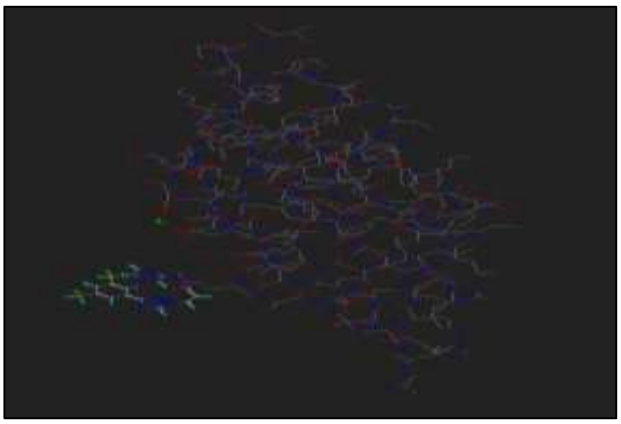

Figure 9: Selinexor docked up on the FN1 receptor.

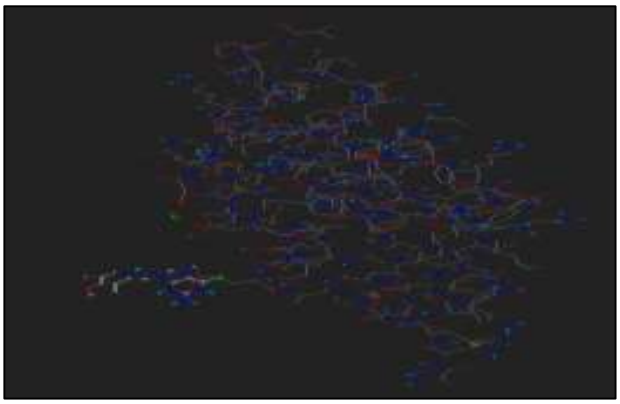

Figure 10: BII021 docked to FN1 receptor.

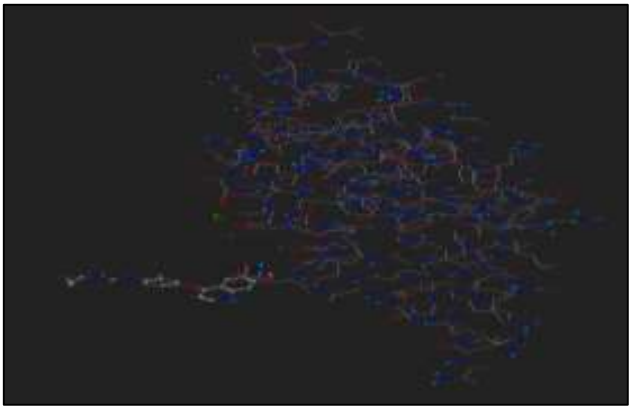

Figure 11: Lenvatinib docked to FN1 receptor.

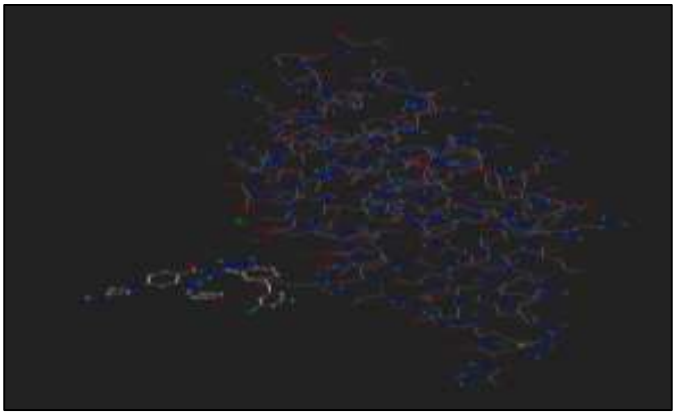

Figure 12: Entrectinib docked to FN1 receptor.

S.NO.	DRUG	BINDING AFFINITY
1	Selinexor	-7.05
2	BII021	-6.8
3	Lenvatinib	-3.23
4	Entrectinib	-3.7

Table 4: Binding affinity of the ligands docked on FN1 gene (kCal/mol).

Figure 13: Selinexor docked to EIF4A3 receptor.

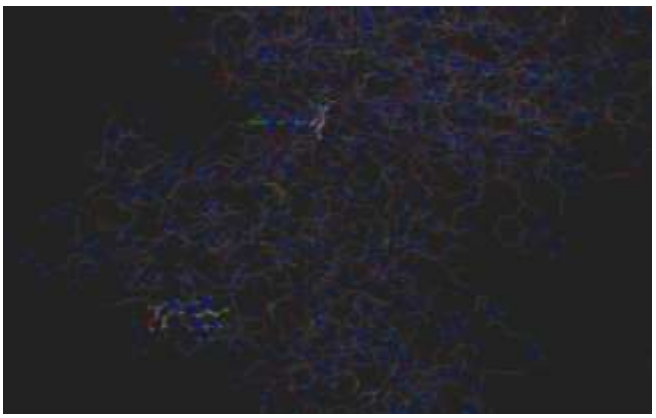

Figure 14: BII021 docked to EIF4A3 receptor.

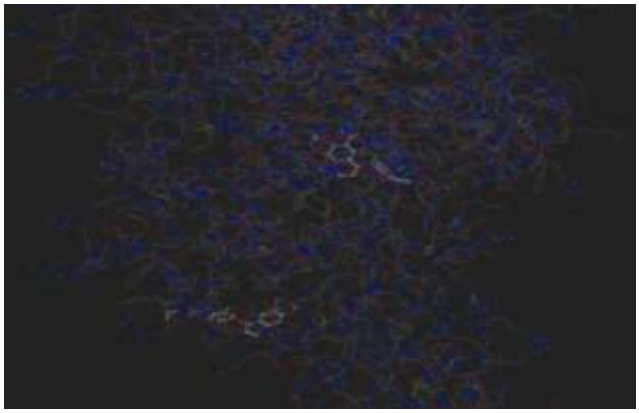

Figure 15: Lenvatinib docked to EIF4A3 receptor.

Figure 16: Entrectinib docked to EIF4A3 receptor.

S.NO.	DRUGS	BINDING AFFINITY
1	Selinexor	-7.40
2	BII021	-7.31
3	Lenvatinib	-3.35
4	Entrectinib	-3.34

Table 5: Binding affinity of the ligands docked on EIF4A3 gene (kCal/mol).

All the different binding affinity score is of the different active site on the receptor in both the cases. AutoDock Vina calculates all the possible binding between the drug and the receptor by calculating the potential active sites. The lower the binding score, the higher is the binding efficiency, which in our case, is the Entrectinib against both the receptor.

8. Analyzing Blood-Brain Barrier-

Investigation of the drugs for crossing the Blood-Brain Barrier is very crucial when we work for neurodegenerative diseases [33]. Thus by using CbLigand prediction tool tested the permeability. For a drug to be permeable it has to have a BBB score 0< All the drugs except Lenvatinib were permeable through the Blood-Brain Barrier.

S.NO.	DRUGS	LOG(BB)
1.	Selinexor	0.203
2.	BII021	0.024
3.	Lenvatinib	-0.024
4.	Entrectinib	0.162

Table 6: Blood-Brain Barrier permeability in terms of BBB score.

9. Comparison with the control set of drugs-

Control drugs that were taken were the drugs already successful in treating Alzheimer's. These were Donepezil (DB00843), Galantamine (DB00674), Rivastigmine (DB00989) and Memantine (DB01043), all these drugs were docked on the two targets using AutoDock Vina in PyRx.

TARGET	Donepezil	Galantamine	Rivastigmine	Memantine
FN1	-5.8	-5.1	-4.2	-4.1
EIF4A3	-6.3	-5.7	-4.6	-4.4

Table 6: Docking score of the control drugs on the targets.

CHAPTER 5

CONCLUSION

In this study, 352 DEGs were identified, including 99 downregulated and 253 upregulated genes. The biological processes associated with each of these genes were analyzed with the help of online software AltAnalyzer. With the help of the protein-protein network obtained from the String database, we were able to find 10 hub genes that were extracted, out of which we selected two drug targets which were previously proved to be involved in the mechanism of the disease. To find potential drugs or ligands for the top two targets, we performed data mining on the official site of the Food and Drug Administration (FDA) for the recently approved drugs for cancer. As it has been stated in a lot of studies that both diseases are inversely associated with each other as the rate of cancer progression is slower in AD patients and the rate of AD development over time is slower if the patient had a cancer history. In cancer, the cell cycle regulation is disrupted with a high rate of cell survival and proliferation whereas AD is associated with high neuronal death either by deposition of amyloid-beta or tau deposits. Thus is are strong chances that in both cases the mechanism of cell survival and death is involved behind the pathogenesis, leading to many genes commonalities between the two. Thus a drug working for the treatment of cancer can be repurposed for the Alzheimer's treatment as well. Thus, the recent four highly successful drugs for cancer which are of BIIBO21 (investigatory drug), Entrectinib, Lenvatinib and Selinexor are docked on the targets identified. The docking score revealed that Entrectinib has the least free binding affinity or the docking score. Thus, it can be termed as a potential new drug candidate for treating Alzheimer's.

REFERENCES

[1] H. D. Yang, D. H. Kim, S. B. Lee, and L. D. Young, "History of Alzheimer's Disease," *Dement. Neurocognitive Disord.*, vol. 15, no. 4, p. 115, 2016, doi: 10.12779/dnd.2016.15.4.115.

[2] A. Wimo and M. Prince, "World Alzheimer Report 2010," *Dementia*, p. 96, 2010, doi: 10.1111/j.0963-7214.2004.00293.x.

[3] martin prince, "World Alzheimer Report," 2015, [Online]. Available: https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf.

[4] K. Inouye, E. S. Pedrazzani, and S. C. I. Pavarini, "Alzheimer's disease influence on the perception of quality of life from the elderly people," *Rev. da Esc. Enferm.*, vol. 44, no. 4, pp. 1093–1099, 2010, doi: 10.1590/S0080-62342010000400034.

[5] J. Liu, Z. Liu, Y. Zhang, and F. Yin, "A novel antagonistic role of natural compound icariin on neurotoxicity of amyloid β peptide," *Indian J. Med. Res.*, vol. 142, no. AUGUST, pp. 190–195, 2015, doi: 10.4103/0971-5916.164254.

[6] A. Forestier, T. Douki, V. De Rosa, D. Béal, and W. Rachidi, "Combination of Aβ secretion and oxidative stress in an Alzheimer-like cell line leads to the over-expression of the nucleotide excision repair proteins DDB2 and XPC," *Int. J. Mol. Sci.*, vol. 16, no. 8, pp. 17422–17444, 2015, doi: 10.3390/ijms160817422.

[7] L. Alves, A. S. A. Correia, R. Miguel, P. Alegria, and P. Bugalho, "Alzheimer's disease: A clinical practice-oriented review," *Front. Neurol.*, vol. APR, no. April, pp. 1–20, 2012, doi: 10.3389/fneur.2012.00063.

[8] J. Mendiola-Precoma, L. C. Berumen, K. Padilla, and G. Garcia-Alcocer, "Therapies for Prevention and Treatment of Alzheimer's Disease," *Biomed Res. Int.*, vol. 2016, no. 2, 2016, doi: 10.1155/2016/2589276.

[9] H. C. Hendrie *et al.*, "Statin use, incident dementia and Alzheimer disease in elderly African Americans," *Ethn. Dis.*, vol. 25, no. 3, pp. 345–354, 2015, doi: 10.18865/ed.25.3.345.

[10] L. Zhu *et al.*, "Phospholipid dysregulation contributes to apoe4-associated cognitive deficits in Alzheimer's disease pathogenesis," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 38, pp. 11965–11970, 2015, doi: 10.1073/pnas.1510011112.

[11] L. Kunz *et al.*, "Reduced grid-cell-like representations in adults at genetic risk for Alzheimer's disease," *Science (80-.).*, vol. 350, no. 6259, pp. 430–433, 2015, doi: 10.1126/science.aac8128.

[12] J. Li *et al.*, "Differential regulation of amyloid-β endocytic trafficking and lysosomal

degradation by apolipoprotein E isoforms," *J. Biol. Chem.*, vol. 287, no. 53, pp. 44593–44601, 2012, doi: 10.1074/jbc.M112.420224.

[13] S. Benhalla, B. El Moutawakil, N. El Kadmiri, and S. Nadifi, "The genetics of Alzheimer's disease," *NPG Neurol. - Psychiatr. - Geriatr.*, vol. 19, no. 110, pp. 83–90, 2019, doi: 10.1016/j.npg.2018.11.006.

[14] M. Cebecauer, M. Hof, and M. Amaro, "Impact of GM1 on Membrane-Mediated Aggregation/Oligomerization of β-Amyloid: Unifying View," *Biophys. J.*, vol. 113, no. 6, pp. 1194–1199, 2017, doi: 10.1016/j.bpj.2017.03.009.

[15] K. Ebrahimi, A. Majdi, B. Baghaiee, S. H. Hosseini, and S. Sadigh-Eteghad, "Physical activity and beta-amyloid pathology in Alzheimer's disease: A sound mind in a sound body," *EXCLI J.*, vol. 16, pp. 959–972, 2017, doi: 10.17179/excli2017-475.

[16] C. Reitz, "Alzheimer's disease and the amyloid cascade hypothesis: A critical review," *Int. J. Alzheimers. Dis.*, vol. 2012, 2012, doi: 10.1155/2012/369808.

[17] P. D. Thorlakur Jonsson, Ph.D., Hreinn Stefansson, Ph.D., Stacy Steinberg, Ph.D., Ingileif Jonsdottir, Ph.D., Palmi V. Jonsson, M.D., Jon Snaedal, M.D., Sigurbjorn Bjornsson, M.D., Johanna Huttenlocher, B.S., Allan I. Levey, M.D., Ph.D., James J. Lah, M.D., Ph. and From, "Variant of TREM2 associated with the risk of AD," *N. Engl. J. Med.*, vol. 368, no. 2, pp. 107–116, 2013, doi: 10.1056/NEJMoa1211103.Variant.

[18] K. J. Barnham, C. L. Masters, and A. I. Bush, "Neurodegenerative diseases and oxidatives stress," *Nat. Rev. Drug Discov.*, vol. 3, no. 3, pp. 205–214, 2004, doi: 10.1038/nrd1330.

[19] N. Butters, J. Hughes, R. Mohs, A. Heyman, and K. Welsh, "Detection of Abnormal Memory Decline in Mild Cases of Alzheimer's Disease using Cerad Neuropsychological Measures," *Arch. Neurol.*, vol. 48, no. 3, pp. 278–281, 1991, doi: 10.1001/archneur.1991.00530150046016.

[20] S. Gluhm, J. Goldstein, K. Loc, A. Colt, C. Van Liew, and J. Corey-Bloom, "Cognitive performance on the mini-mental state examination and the montreal cognitive assessment across the healthy adult lifespan," *Cogn. Behav. Neurol.*, vol. 26, no. 1, pp. 1–5, 2013, doi: 10.1097/WNN.0b013e31828b7d26.

[21] J. L. Cummings, T. Morstorf, and K. Zhong, "Cummings, Jeffrey L_Alzheimer's_drug development candidates failures_2014," pp. 1–7, 2014.

[22] N. Farina, J. Rusted, and N. Tabet, "The effect of exercise interventions on cognitive outcome in Alzheimer's disease: A systematic review," *Int. Psychogeriatrics*, vol. 26, no. 1, pp. 9–18, 2014, doi: 10.1017/S1041610213001385.

[23] E. M. Blalock, H. M. Buechel, J. Popovic, J. W. Geddes, and P. W. Landfield, "Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease," *J. Chem. Neuroanat.*, vol. 42, no. 2, pp. 118–126, 2011, doi: 10.1016/j.jchemneu.2011.06.007.

[24] D. Szklarczyk *et al.*, "STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D607–D613, 2019, doi: 10.1093/nar/gky1131.

[25] O. Shafi, “Inverse relationship between Alzheimer’s disease and cancer, and other factors contributing to Alzheimer’s disease: A systematic review,” *BMC Neurol.*, vol. 16, no. 1, pp. 1–17, 2016, doi: 10.1186/s12883-016-0765-2.

[26] X. Yan, “2018_Pnas_Si_Spe,” *Proc. Natl. Acad. Sci.*, vol. 7, p. 2017, 2017, doi: 10.1073/pnas.

[27] D. Otasek, J. H. Morris, J. Bouças, A. R. Pico, and B. Demchak, “Cytoscape Automation : empowering workflow-based network analysis,” pp. 1–15, 2019.

[28] S. Qi, “Multiple Bioinformatics Analyses of Integrated Gene Expression Profiling Data and Verification of Hub Genes Associated with Diabetic Retinopathy,” pp. 1–12, 2020, doi: 10.12659/MSM.923146.

[29] K. C. H. Dhanani, W. J. Samson, and A. L. Edkins, “Fibronectin is a stress responsive gene regulated by HSF1 in response to geldanamycin,” *Sci. Rep.*, vol. 7, no. 1, pp. 1–13, 2017, doi: 10.1038/s41598-017-18061-y.

[30] E. Lilly, “Hematology / Oncology (Cancer) Approvals & Safety Noti cations,” pp. 1–24, 2020.

[31] M. I. Behrens, C. Lendon, and C. M. Roe, “A Common Biological Mechanism in Cancer and Alzheimer ’ s Disease ?,” pp. 196–204, 2009.

[32] R. Herowati and G. Pamudji, “Molecular Docking Studies of Chemical Constituents o f Tinospora cordifolia on Glycogen Phosphorylase,” *Procedia Chem.*, vol. 13, pp. 63–68, 2014, doi: 10.1016/j.proche.2014.12.007.

[33] R. K. Upadhyay, “Drug delivery systems, CNS protection, and the blood brain barrier,” *Biomed Res. Int.*, vol. 2014, 2014, doi: 10.1155/2014/869269.