COMPUTATIONAL ANALYSIS OF KEY TRANSCRIPTION FACTOR BINDING MODULES IN ALZHEIMER'S DISEASE AND ITS ASSOCIATED GENES

A Major dissertation submitted
In partial fulfilment of the requirement for the degree of

Master of Technology
In
Bioinformatics

Submitted by

Dhiren Pattanayak
(DTU/M.Tech/355)

Delhi Technological University, Delhi, India

Under the supervision of

Dr. Pravir Kumar

Department of Biotechnology
Delhi Technological University
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road
Delhi-110042, INDIA

CERTIFICATE

This is to certify that the M. Tech. dissertation entitled **"Computational Analysis of Key Transcription Factor Binding Modules in Alzheimer's Disease and its Associated Genes"** submitted by Dhiren Pattanayak (2K13/BIO/05) in partial fulfilment of the requirement for the award of the degree of Master of Technology, Delhi Technological University (Formerly Delhi College of Engineering), is an authentic record of the candidate's own work carried out by him under my guidance.

The information and data enclosed in this dissertation is original and has not been submitted elsewhere for honouring of any other degree.

Date: 02nd July, 2015

Dr. Pravir Kumar

Guide, Associate professor

Department of Bio-Technology

Delhi Technological University

Prof. D. Kumar

Head of the Department

Department of Bio-Technology

Delhi Technological University

DECLARATION

I, Dhiren Pattanayak, hereby declare that the work entitled **"Computational Analysis of Key Transcription Factor Binding Modules in Alzheimer's Disease and its Associated Genes"** has been carried out by me under the guidance of Dr. Pravir Kumar, at Delhi Technological University, Delhi.

This dissertation is part of partial fulfilment of requirement for the degree of M.Tech. in Bioinformatics. This is the original work and has not been submitted for any other degree in any other university.

Dhiren Pattanayak

Roll No: 2K13/BIO/05

ACKNOWLEDGEMENT

I am extremely thankful to my mentor, Dr. Pravir Kumar, Associate professor, Department of Biotechnology, Delhi Technological University, Delhi for his exemplary guidance, monitoring and constant encouragement throughout the M. Tech course. I would also like to thank other faculty members of the department for sparing their efforts in completion of my course.

I would like to acknowledge my deep sense of gratitude to Prof. D. Kumar, (Head of Department) Department of Biotechnology, Delhi Technological University, Delhi for giving me an opportunity to study and work in this prestigious Institute.

I am also thankful to my parents, family members and friends whose blessings and support were always with me. At last I thank to God for being a positive source of energy throughout.

Dhiren Pattanayak

Roll No.: 2K13/BIO/05

TABLE OF CONTENTS

List of Figures

List of Tables

List of Abbreviations

AD	Alzheimer's Disease
TF	Transcription Factor
TFBS	Transcription factor binding site
MSA	Multiple Sequence Alignment
APP	Amyloid precursor protein
APOE	Apo Lipoprotein E
BDNF	Brain derived neurotropic factor
PSGL	Problem specific gene list
MRP	Multidrug Resistance Protein
EIF5	Eukaryotic translation factor 5
CLU	Clusterin
CDH1	Cadherin 1
REEP5	Receptor expression enhancing protein
CBP	Calcium Binding Protein

COMPUTATIONAL ANALYSIS OF KEY TRANSCRIPTION FACTOR BINDING MODULES IN ALZHEIMER'S DISEASE AND ITS ASSOCIATED GENES

Dhiren Pattanayak

Delhi Technological University, Delhi, India

ABSTRACT

There is partial understanding of the molecular mechanism underlying Alzheimer's disease (AD) pathogenesis and the genetic risk factors involved in it. Many approaches like genomics, proteomics and functional genomics have been taken in the recent years to identify new factors of AD etiology. Pathway or a particular disease-associated gene might have participated in multiple transcriptional co-regulation networks. The group of such gene can be identified either by literature survey or high throughput molecular techniques like protein interaction mapping or microarrays. The strategy that has been followed in the study includes *in-silico* promoter analysis to define regulatory networks and then locating important co regulated factors without a *priori* knowledge. Using a computational approach it was concluded that different AD related genes have common transcription factor binding site modules. A new bioinformatics workflow was established to identify other unknown co-regulated genes, which may potentially relate to AD. We have constructed some significant TF modules through Genomatix tool, composed of some important transcription factor families: CTCF, SP1F, and EGRF, which are thought to be genetically conserved in between human and mouse or rat APP promoter sequences. We found some new genes like cadherin 1, MRP, REEP5, EIF5 which can potentially regulate AD mechanism. After that to validate our result we have done Multiple Sequence Alignment (MSA) including our predicted genes and established a relation between our predicted genes and key AD causing genes.

Keywords: Alzheimer's disease, MSA, TF, TFBS modules.

CHAPTER1: INTRODUCTION

Alzheimer's disease (AD) is a primary neurodegenerative disorder representing one of the most common causes of dementia among the population above fifty.(Ritchie and Lovestone,2002). The risk of this shocking disease increases and becomes almost twice in incidence with every 5 years of the age after the age of 65. About 30-35 million people are suffering from AD worldwide according to current estimates and increase in life expectancy will lead to increase in the number to three times the current value (Ferri, CP. 2005). AD stands among one of the most expensive diseases in the United States and Europe (Need and Goldstein, 2010). Economical and psychological burden is posed on the caretakers of the patients, thus it is troublesome not only for the patients but also for their caretakers (Masters et al., 1985).

AD is polygenic in nature and it is caused by changes in large number of genes and complex interactions among them, rather than defects in single gene. The contribution of those huge numbers of genes results in a wide range of disease phenotypes ultimately leading to varied symptoms. Thus, it is important to identify specific genes that are responsible and capable of conferring the risk of Alzheimer's disease. Like other neurodegenerative diseases, it is a challenging and cumbersome to elucidate the genetic risk factors and genetic mechanisms involved in AD. Evident suggest probability of regulation of the AD genes at transcriptional level and their correlation to other diseases due to the discovery of the *cis*-regulatory elements in the promoter regions of the genes that are related to the disease.

So our aim was to identify transcription factor binding site modules in the promoters region of AD-linked genes, which may potentially have a say to AD pathogenesis. Therefore, a computational approach, such as promoter analysis was successfully carried out and proved to be an effective procedure for analysing complex diseases.

CHAPTER 2: LITERATURE REVIEW

Ten decades ago, the German psychiatrist Alois Alzheimer described the case of Auguste D., recording an ante-mortem history of weakened memory, problems of speech, paranoia and delusional ideation (Rainulf and Stelzma,1995). An atrophied brain was detected in the post-mortem report neurofibrillary tangles were detected for the first time and were described by Alzheimer after which the disorder came to be known as Alzheimer's disease. AD is the most common cause of dementia in the older population of the world. The cause of the disease continues to remain questionable till date, besides the early discovery of Alois Alzheimer. Currently definitive diagnosis of the disease is possible only after post-mortem. Early diagnosis of the disease can be made possible by identifying a biological marker(s) for AD. The biological marker can also help in observing and measuring the modifications and disproportionateness associated with the disease. Early and accurate diagnosis of the disease can even be supplemented by development of new drug treatments if these biological markers are made available.

Apart from the primary risk factor of AD i:e: aging there are other risk factors such as gross brain shrinkage, low mental achievement in early life, reserve brain capacity and minimal level mentally taxing occupations followed by latter life reduced mental and physical activity. Head injury has also been cited as a risk factor (Mayeux and Stern, 2012). Alzheimer's disease is thought to be associated with hypertension, hypercholesterolemia, cardio vascular disease, atherosclerosis and diabetes, since all these affect the effective supply of blood and are considered to be vascular risk factors. Evidence suggests that diet alteration can counter the risks that predispose a person to dementia. An altered diet can include increased intake of homocysteine-related vitamins (Vitamin B12 and Folate) and antioxidants, such as vitamin C and E; unsaturated fatty acids (Blennow et al., 2006). Sporadic Alzheimer's disease is highly heritable as is suggested by a twin study performed on a large population. Heritability was as high as 79% with the same genetic factors being as influential, irrespective of sex, and non-genetic risk factors like environment (Gatz, M. 2006). The process associated with the occurrence of dementia in both cases of AD i.e. familial and sporadic can be observed in **Figure 2.1** that illustrates the etiology of Alzheimer's disease.

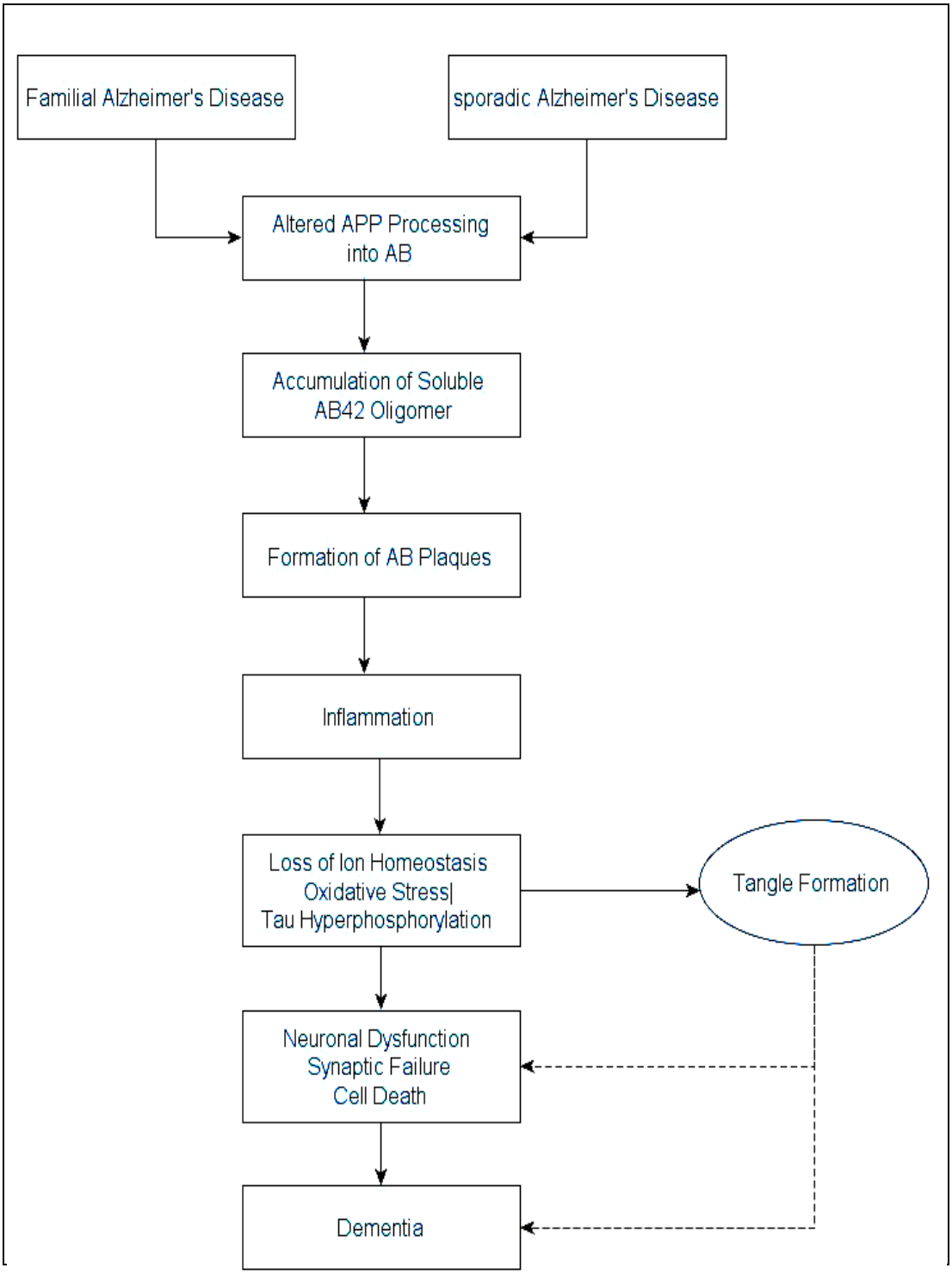

Figure 2.1: Etiology of Alzheimer's disease

Around 50-60% of dementia cases are accounted for by AD in the developed world. Some other causes of dementia include lewy body, vascular, and frontal lobe. Around 44.3 million people around the were estimated to be having dementia in 2013 and this number is expected to be doubled every two decades due to increase in life expectancy**(Figure2.2)**. The estimation shows that population with dementia may rise to161.1 million in 2040 (Ferri, CP. 2005).

Western Europe, China and the USA are major regions in the world that are most affected by dementia. These countries would contain 55.7% of the total dementia affected population by 2040.

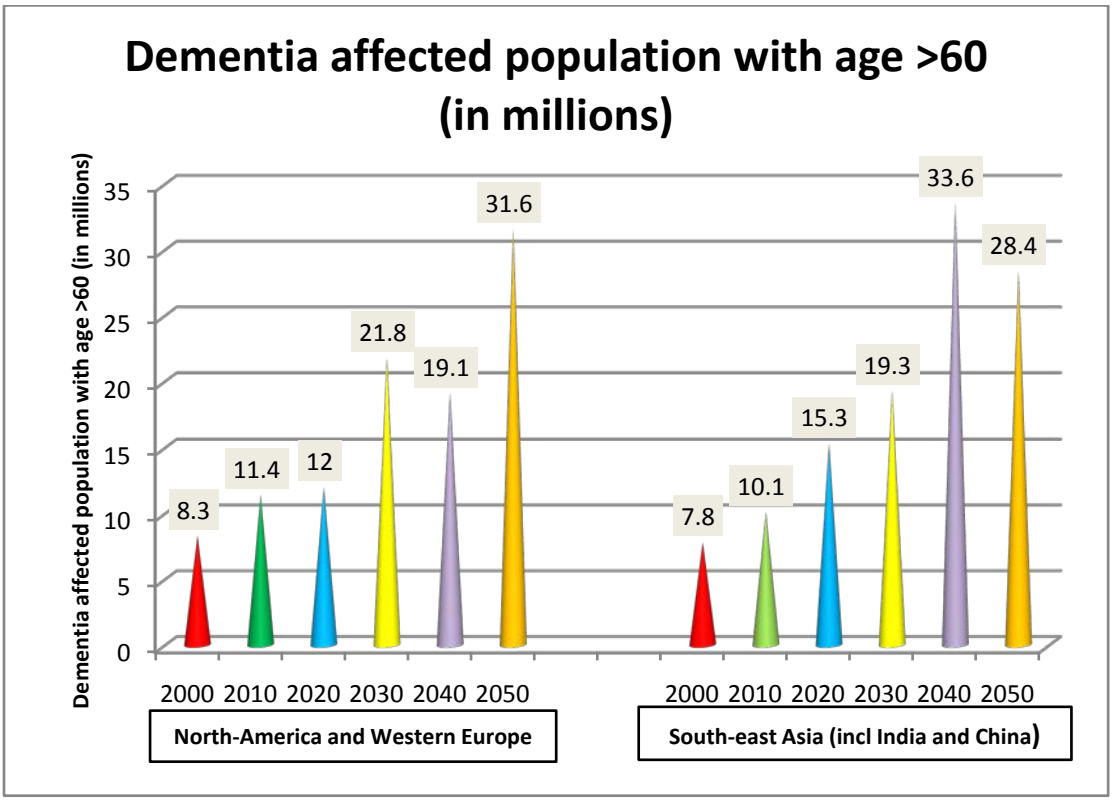

Figure2.2. Estimates of the Population with Dementia over the Years 2000-2050

From figure2.2 it can be observed that the trend-line of dementia affected population increases every decade besides a few exceptions. The global population with dementia was at 24.3 million in 2001 and one case of dementia was detected after every 7 seconds. In north-America and Western Europe the rate of increase in dementia affected population is lower than that of south-east Asian countries. In case of South-east Asia the population affected by dementia is more than double than the affected population two decades earlier. The comparison shows that there will be greater toll of dementia in developing countries in future years than developed countries.

Cognitive deficits and progressive memory impairment are characteristics of clinical manifestation of AD (Billings et al., 2005; Fox et al.,1998). Subtle and poorly recognized memory failure is the initial symptoms of AD (Correa et al., 2015). There is inevitable worsening of memory loss with passage of time. Impaired social judgment, confusion, agitation, language disturbance, withdrawal, impulsivity and irritability are among the other symptoms of dementia. AD pathology is characterized by the presence of extracellular senile plaques composed of amyloid-β (Aβ) peptide and intra neuronal neurofibrillary tangles containing hyper-phosphorylated tau protein (Schenk, SD. 2003). These proteins are considered as inducers and key players of AD by amyloid and the tau hypotheses. Degeneration of cholinergic basal forebrain neurons, cortical atrophy, enlarged ventricles and hippocampal atrophy are the further neuro-pathological features of the disease (Selkoe, DJ. 1994,2001). Developing AD is affected in various ways by the genes. In 1-5% cases autosomal dominant inheritance is observes in familial forms and in ≥95% cases polygenic backgroundis found in sporadic AD. In addition to the genetic component, the risk for developing AD, the age at onset and the course of the disease are influenced by several other factors including socio-demographic, life style, environment and co morbid medical conditions (Papassotiropoulos and Stephan, 2006). Female sex and increasing age pose the risk for developing AD. AD can also be predisposed by low mental ability, poor education, stroke, traumatic brain injury and incidences of depression. However, the genetic factor assumed to be of major importance, since as per twin studies, a significant part of the risk for sporadic AD is genetically determined (Gatz, M. 1997). The amyloid precursor protein (APP) and the presenilin 1 (PSEN1) and 2 (PSEN2) genes are currently known to be implicated in the familial forms of AD (Papassotiropoulos and Stephan, 2006). The understanding of biological mechanisms involved in enhanced Aβ accumulation and senile plaques generation requires the dominant mutations of APP, PSEN1 and PSEN2 to be identified and characterized (Cruchaga et al., 2012; Krüger et al., 2012). The causing factors of all pathological mechanisms including Aβ accumulation remain imprecise in sporadic form unlike familial AD. Pathological processes underlying sporadic AD may be unveiled by multiple heterogeneous susceptibility sets of genes since the genetic model of sporadic AD is complex in nature (Park et al., 2008; Lohmann et al., 2012). However, so far only the apolipoprotein E (APOE) gene has been definitively associated with the risk for AD (Brouwers et al., 2008; Savage et al., 2002). APOE is associated with lipid transport and metabolism and it also plays a major role in the CNS. It is actively involved in neuron development and regeneration and also in some neurodegenerative processes. The polymorphism of the APOE gene results three isoforms of APOE protein (ε2, ε3, ε4) having

different conformation and lipid binding properties (Cedazo-Minguez, C. 2001). The relation found between the risks for developing AD, the age at onset and the number of the inherited ε4 alleles was direct proportionality. The APOE ε4 isoform prefers very low density lipoprotein and it is less effective in cholesterol transport as compared to the other APOE isoforms (Cedazo-Minguez, C. 2001). The enzymatic cleavage of APP protein is controlled by membrane cholesterol. In the presence of the ε4 isoform the poise is tilted to the accumulation of Aβ (Stefani et al., 2009). The predominant model of molecular mechanisms underlying the pathogenesis of AD has been the Amyloid Cascade hypothesis. This model suggests that accumulation and aggregation of the Aβ peptide in the brain is the key factor leading to AD. This model is accepted widely but it lacks in covering all aspects of AD etiology. The key role of other pathogenic mechanisms is also important which includes inflammation, hypoxia, fatty acid dysfunctionality, improper signalling and neuronal dysfunction (Peers et al., 2007; Zhang et al.,2010).Major focus of research in the recent years has been the genetic epidemiology of sporadic AD because its genetic etiology is still less understood and unsolved. The aim of our work was to contribute to this field investigating various genes presumably involved in AD pathogenesis (Gamba et al., 2012). The candidate genes in this study were selected and grouped on the basis of promoter analysis and the pathways that presumably lead to the development of AD: Aβ metabolism, cholesterol metabolism, Neuro-inflammation and neuronal dysfunction (Palop et al., 2012; Martins et al., 2009)

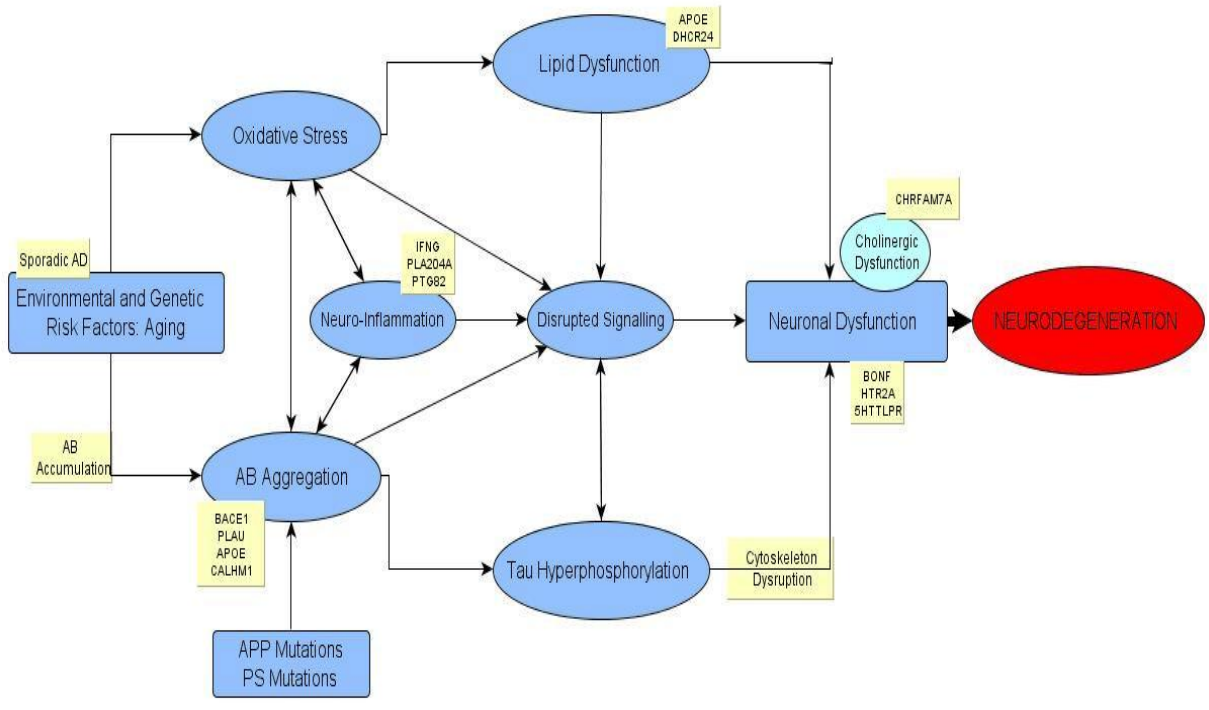

Figure2.3: Process of Neuro-degeneration

2.1 Amyloid-β metabolism and AD

APP is a trans-membrane protein that has characteristics of a cell surface receptor. The principal function of APP is unidentified, but it is supposed to have role at some stage in neuronal development, to be mixed up in synaptic development and repair, signalling and cell-cell linkage (Wesley and Charlene, 2007). APP follows two distinct cleavage pathways competing α- and β-secretases, and both pathways are active in normal metabolism (Selkoe et al., 2001). The α-secretase mediates the predominant cleavage of APP that leads to the generation of non-amyloidogenic products. β-and γ-secretases mediate the cleavage leading to Aβ generation. PSEN1 and PSEN2 are part of the γ-secretase complex. β-secretase is the key and rate-limiting enzyme in Aβ formation cascade. It is encoded by the β-site APP cleaving enzyme-1 (BACE1) situated at locus 11q23.3 (Hendrie et al.,2014; Mielke et al., 2011). Aβ generation may be acted upon by Genetic variations of BACE1 thereby influencing neurodegenerative processes leading to AD. The rs638405 identical polymorphism of BACE1 is a nucleotide change in -8 -exon 5 (C786G) with no alteration at the amino acid level (Val262) (Murphy et al., 2001; Pedersen et al., 2006). But there are reports contradicting the above mentioned role of BACE1 in AD. Some studies have reported that the BACE1 C786G polymorphism influences the risk for AD, especially in those carrying the APOE ε4 allele(Gambhir,S. 2005; Wang et al., 2014). However no such association has been found by other authors. According to the first report on BACE1 C786G polymorphism in the Hungarian population The PLAU gene encoding urokinase-type plasminogen activator (uPA) maps to chromosome 10q22.2, a candidate susceptibility locus in a broad AD linkage region (10q21-24) (Riemenschneider M, 2002; Kojima et al., 2007). The inactive plasminogen is converted to the active plasmin form by the uPA serine protease. uPA serine protease is also capable of degrading Aβ directly. Plasmin promotes α-cleavage of APP, degrades secreted and aggregated Aβ, and thereby blocks Aβ neurotoxicity (Finckh and Mann, 2000; Krut et al., 2013). The frequent single nucleotide polymorphism (SNP) of PLAU (rs2227564, C1788T) is located in exon 6 and results in an amino acid change at codon 141 (Pro141Leu). The association studies investing the Pro141Leu polymorphism in AD have controversial findings. Thus, genotype distribution should be evaluated further. An increasing body of evidence supports the major contribution of the deregulation of calcium homeostasis in accelerating pathological changes in AD, i.e. Aβ accumulation (Bezprozvanny et al., 2013; Jinwal et al., 2010). The calcium homeostasis modulator 1 (CALHM1) at position 10q24.33 has been recently identified as a promising candidate gene for AD. Its location is same as the linkage region where the PLAU

gene is situated. It is localized both in the cell and in the endoplasmic reticulum membranes and it appears to modulate the intracellular calcium levels(Berridge,MJ. 1998). An SNP (rs2986017) at nucleotide 257 (C/T) resulting a non-conservative amino acid substitution at codon 86 (Pro86Leu) has been acknowledged.The Leu allele has been reported to be associated with AD and in vitro demonstrated to result in impaired CALHM1 function leading to decreased calcium permeability and reduced cytoplasmic calcium levels (Betty and Blacks, 1981). The metabolism of APP is also affected by intracellular calcium levels and thus, the level of Aβ also gets affected. Recent association studies failed to support the possibility that CALHM1 polymorphism may represent a susceptibility factor for AD (Nathalie et al., 2012).

2.2 Cholesterol Metabolism and AD

Cholesterol is a significant structural element. It is the cell flexibility modulator of cell membrane. Local de novo synthesis is followed during the production of for the most part of the CNS cholesterol. The role of cholesterol in AD is a contentious topic, but it seems that an optimal quantity of cell cholesterol may be critical for brain homeostasis (Serio,PA. 2008). Aβ generation and cell resistance against Aβ toxicity are significantly affected by the amount of cell cholesterol. Seladin1 (Selective AD Indicator), an enzyme caught up in cholesterol biosynthetic pathway is encoded by the 24-dehydrocholesterol reductase (DHCR24) gene. Seladin1 not only acts as a catalyst for the conversion of desmosterol to cholesterol but also has other relevant biological effects. Seladin1 caspase-3 activity and prevention of p53 degradation, thus conferring resistance against Aβ and oxidative stress induced apoptosis (Serio, PA. 2008). Seladin1 modulates the membrane cholesterol content thereby affecting the Aβ generation. The genetic association between Seladin1 and risk for AD was investigated by Lamsa and co-workers genotyping four SNPs of DHCR24 genes: rs638944 (intron 2 G/T), rs600491 (intron 5 C/T), rs718265 (intron 6 A/G), rs7374 (exon 9 C/T) (Lamsa, KP. 2007). Single SNPs have no significant association with AD risk but in men the risk of AD is increased by the T allele of rs600491. From haplotype analyses it has been revealed that out of two haplotype blocks (block 1: rs638944 and rs600491; block 2: rs718265 and rs7374) only block1 has considerable association with AD risk, being CG the risk haplotype (Lamsa, KP. 2007).

2.3 Neuro-inflammation and AD

An increasing body of evidence supports the major contribution of inflammatory processes in accelerating pathological changes in AD (Lukiw and Bjattacharjee, 2012). The activated microglia driven inflammatory response resulted in an elevated release of various pro-

inflammatory mediators such as cytokines and prostaglandins which may interact at multiple levels with neurodegeneration (Heneka, T. 2007; Liang et al., 2008). An individual's genetic background of inflammatory mediators mediates the degree of inflammatory response. The cytokine interferon-γ (IFNG) has an important role to play in the induction of immune-mediated inflammatory response. In human neuroblastoma cells pre-treatment with IFNG increased the expression of cytosolic phospholipase A 2, group IVA -10 -(PLAG4A) resulting in an elevated release of arachidonic acid and a subsequently elevated level of prostaglandins(Bate et al., 2006). Increase in the level of IFNG is correlated with the T allele of the IFNG T874A polymorphism. The hypothesis that the IFN-γ T874A polymorphism may represent a risk factor for AD has not been supported by other associated studies (Belkhelfa et al.,2014; Rota et al.,2006). Phospholipase A 2 is a superfamily of enzymes which includes the key modulators of cerebral phospholipid metabolism. . The discharge of arachidonic acid from membrane phospholipids is catalysed by the PLA2G4A. A polymorphic location for BanI restriction enzyme in PLA2G4Agene is an A to G base change (Wei, J. 2004).As per our knowledge this is the first report on PLA2G4ABanI polymorphism in AD. Prostaglandin endoperoxide synthase 2 (PTGS2) is a key enzyme in prostaglandin biosynthesis converting arachidonic acid to prostaglandin E2 (PGE2). A functional polymorphism (G-765C) in the promoter region of the PTGS2gene has been identified and significantly subordinate promoter activity has been reported for the C allele (Papafilli A, 2002). Decreased AD risk is associated with the possession of C allele. PLA2G4Aand PTGS2genes are located at the same 1q25region and they are involved in the same pathway of prostaglandin synthesis. Thus, they interact in multiple ways. PLA2G4A expression is elevated by IFNGs, thus it might also play a role in the generation of prostaglandins. Hence, a collective effect with PLA2G4A and PTGS2 can be postulated. APOE helps in mediating brain inflammatory response by playing an isoform-specific role (Zhang et al; 2010; Ryan et al., 2014). Cell culture studies reported association between the presence of APOE ε4 and increased inflammatory response (Jofre et al, 2007). Mouse macrophages transfected with APOE ε4 secrete significantly more pro-inflammatory cytokines as compared to those transfected with APOEε3 (Tsoi, LM. 2007).

2.4 Neuronal Dysfunction and AD

2.4.1 Cholinergic Dysfunction

The association of AD has been observed with a progressive loss of cholinergic neurons and a resultant deficit of acetylcholine, specifically in the parietal and temporal hippocampus and neocortex (Schliebs et al., 2011). Up to 95% loss of the cholinergic innervation to the cortex is caused by the excessive degeneration of the cholinergic neurons in the medial septal nucleus and the basal nucleus of Meynert. There is consequential decrease in the availability of acetylcholine due to reduction in cholineacetyltransferase activity. It is also supposable that cholinergic transmission can be regulated by the altered activity of acetylcholine receptors ascertained by genetic variations. The α7 nicotinic acetylcholine receptors are homopentamer, ligand gated cationic channels. They are commonly articulated in the central nervous system with soaring levels in the regions related to memory functions and caught up in processing of sensory information, such as hippocampus (Weiland, NG. 1992). It has been confirmed that Aβ binds to α7 nAChR with high affinity and hey both are present in senile plaques. It has been demonstrated that Aβ binds to α7 nAChR with high affinity and they both are present in senile plaques (Wang and Smith, 2009). Many neurochemical processes are altered by their interaction, such as acetylcholine release and Ca^{2+} homeostasis. Thus, neuronal physiological functions implicated in memory processes are modulated. Aβ acts as the chronic inhibitor of cholinergic signalling and thus, could contribute to the cognitive defects associated with AD. The α7 nAChR subunit gene (CHRNA7) at region 15q13.1is duplicated from exon 5 to 10 (Robert-Freedman, SL. 2001). The partially duplicated CHRNA7 and four other exons originated from Family with sequence similarity 7A (FAM7A) gene form a hybrid gene (CHRFAM7A). CHRFAM7A is not present on every human chromosome and some individuals lack one (30%) or both (5%) copies (Riley, BW. 2002). A two bp deletion polymorphism at position 497-498 in exon 6 was identified, which is specific to CHRFAM7A and does not occur in CHRNA7 (Leonard, SG. 1998). A frame shift mutation is caused by the two bp deletion leading to the introduction of a stop codon within exon 6 and thus, a truncation is caused in a putative gene product. Reportedly CHRFAM7A is expressed as an mRNA. So, probable regulatory effects should be taken into account. Liou and co-workers failed to find association between CHRFAM7A -2bp deletion polymorphism and AD, investigating a relatively low number of cases in an Asian population (Liou W., 1997).

2.4.2 Serotonergic Dysfunction

History of depression, and in particular with first onset prior to 60 years of age, represents risk for susceptible to AD later in life (Geerlings, MI. 2008). Possible links between AD and depression may be the long term occurrence of inflammatory processes, and the involvement of serotonergic disturbances. The involvement of serotonergic in AD is supported by findings that include cerebrospinal fluid alteration of serotonin and loss of 5HT receptors that of and synthesizing neurons (Geldenhuys et al.,2011; Oliveira et al.,1999). 5HTtransporter (5HTT) is encoded by the SLC6A4 gene which is the 4th member of the solute carrier family6 at locus 17q11.1-q12. The SLC6A4 gene displays a 22 bp tandem repeat polymorphism that is nominated as 5HTT gene-linked polymorphic region (5HTTLPR) (Meltzer et al 1998; Reinikainen et al., 1988). The two alleles differ in 44 base pairs having 14 and 16 repeats and are denoted as long (L) and short (S) alleles. The S allele is less effective. The 5HTTLPR polymorphism concludes dose-dependent 5HT reuptake from the synaptic cleft. Association studies reported mainly negative results investigating the correlation between AD and 5HTTLPR (Seripa, D. 2004; Rodríguez et al., 2012). The gene HTR2A which is located at 13q14–q21, codes for the 5HT receptor type 2A (5HT2A). The rs6313 polymorphism of HTR2A is a silent mutation. It involves a nucleotide change in exon1 (T102C) and it does not modify the serine at 34th position. The polymorphism being located near to the promoter region may have some role in the regulation of gene expression. Compared to the T allele, the activity of C allele significantly decreases. Association studies did not support the association ofHTR2AT102C and 5HTTLPR polymorphisms with AD, when they were separately analysed (Lam, 2001). However, in another study an interaction between the HTR2A T102C and 5HTTLPR polymorphisms have been found. HTR2A C/C and 5HTTLPR L/L genotype carriers resulted in an increased risk for AD (Thome et al., 2001; Guo et al., 2014).

2.4.3 Brain- derived Neurotrophic Factor

The brain-derived neurotrophic factor (BDNF) is a member of the neurotrophin family of growth factors, produced by cortical neurons. In addition to its general role in neurodevelopment, BDNF has essential functions in the adult brain such as promoting the survival and maintaining the structural integrity of neuronal cells (Murer et al., 2001; Laske et al. 2010; Krabbe et al., 2007). The expression of BDNF is activity dependent and it plays a role in modulating synaptic changes associated with memory and learning. The

patients with AD have reduced BDNF levels in the temporal cortex and in the hippocampus as compared to healthy ones. A precursor pro-BDNF is encoded by the BDNF gene (Zhang et al., 2006; Borroni et al., 2009). BDNF protein is formed by the secretion and cleavage of pro-BDNF by extracellular protease. An SNP at nucleotide 196 (G/A) producing a non-conservative amino acid substitution at codon -13 -66 (Val/Met) has been identified (Ventriglia, M. 2002). Even though this SNP is positioned in the 5' pro-BDNF sequence and does not have an effect on the function of the mature BDNF, it has a major impact on the intracellular trafficking and regulated secretion of pro-BDNF (Lukas et al., 2004; Karege et al., 2002). Alterations in hippocampal activity can be resulted through genetic influences on BDNF secretion. Hippocampal function and episodic memory have been found to be associated with BDNF Val66 Met polymorphism. Nevertheless, there are conflicting reports on the correlation between AD and BDNF Val66 Met polymorphism (Akatsu et al., 2006).

2.5 Insight into Regulation of Gene Expression

Gene expression can be regulated in different ways during processing of mRNA, transcription, translation and protein stability. However it is believed that primarily regulation of gene expression occurs at the level of transcription. The eukaryotes have two complementary regulatory elements in their transcriptional machinery: the trans-acting element and the *cis*-acting element (Yaguchi et al., 2014; Tucker et al., 2002). The *cis*-acting elements are DNA sequences in the coding or non-coding regions of the genome (Kucho et al., 2003). The *cis*-acting elements can also be overlaid with epigenetic information. An accessible region in the DNA is created chromatin remodelling and modification in the histones or the DNA sequence. Trans-factors bind to this region to initiate transcription. In some processes trans-factor id restricted from binding to the DNA sequence by creation of an inaccessible chromatin environment. The trans-acting elements are transcription factors or other DNA-binding proteins which recognize and bind to specific sequences in the *cis*-acting elements to initiate, enhance or suppress transcription (Hu et al., 1996; Hyman et al,. 1996). A huge repertoire of precise and unique control patterns are generated either by the regulation of multiple genes by a transcription factor or by binding of the transcription factor to the -regulatory elements at multiple transcription factor binding sites in a complex and combinatorial manner. According to estimates approximately 1800 transcription factors are encoded by the human genome (Citron et al., 2008).

2.5.1 *Cis*-acting Regulatory Elements

There are two distinct elements in the *cis*-regulatory DNA sequences: proximal elements or promoters and distal regulatory regions including silencers or repressors, insulators, enhancers and locus control regions (LCRS) (Coskun et al., 2004; Kel et al., 2003). A coordinated expression pattern of a gene is governed by cooperated actions of these elements. Figure summarizes the actions and mechanisms of all these elements.

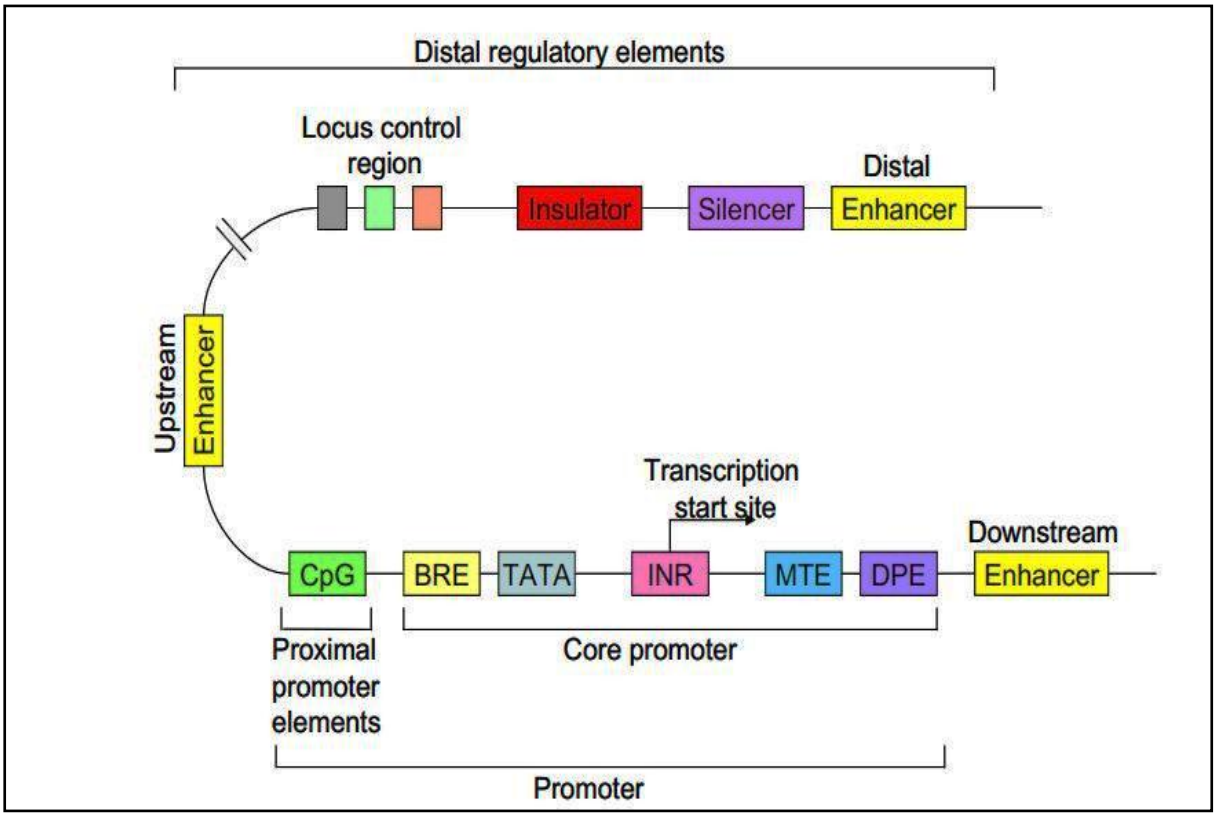

Figure 2.4: Schematic Representation of the *cis*-regulator factors involved in the regulation of Gene Expression

There are core and proximal promoter elements like CpG islands spanning about 1kb around the transcription start site (Dehm et al., 2004; Klingenhoff et al., 1999). The core promoter consists of an initiator element (INR), a TATA box (TATA), a motif ten element (MTE), a downstream promoter element (DPE) and a TFIIB recognition element (BRE). Distal regulatory elements such as silencers, enhancers, LCRs and insulators can be located downstream or upstream or even far away from the transcription start site. Various LCRs, enhancers and silencers act together to repress or activate the promoter activity. The insulators prevent inappropriate regulation by regulatory signals from adjacent genes.

- **Promoters:**

The RNA polymerase II (Pol II) promoter regions consist of the core promoter and the proximal promoter elements. DNA is transcribed to mRNA and snRNA by Pol II promoters (Kadonaga, ST. 2003; Ioshikhes et al; 2005). The location of core promoter is 35bp downstream or upstream of TSS (transcription start site). It serves as the binding site of factors for congregation of the pre initiation complex (PIC).There are a number of elements of the core promoter. TATAAAA, the consensus sequence of the TATA box is located 25-30 bp upstream of the TSS. However, this consensus sequence may vary. Although the TATA box was believed to be a fundamental component of the core promoter, it was revealed that only 32% of the potential human core promoters contain the TATA box (Suzuki, Y. 2001). The initiator element (INR) is located across the transcription start site (denoted as +1) from -3 to +5 having the consensus sequence of PyPy A(+1) N T/A PyPy. The downstream promoter element (DPE) functions in the downstream of TSS, in conjunction with the INR in TATA-less promoters. It is located at +28 to+32 relative to the TSS and possesses the consensus sequence of A/G GA/TC/T G/A/C (S, 2004). DCE was first identified in the human β-globin promoter (Lee, 2005), which is located at +10 to +45 relative to the TSS and acts separately from the DPE. Another element located at +18 to +27 called the motif ten element (MTE) is also relative to the TSS. It functions in a cooperative manner with the INR but independently from the TATA box and the DPE (Lim et al., 2004). The initiation of the recruitment of TFIID (Transcription factor IID) initiation complex to the promoter is done by all the core elements, for gene transcription to occur. TFIIB recognition element (BRE) is another core promoter element that is recognized by TFIIB rather than TFIID. It is located 3-6 bp upstream of the TATA box with the consensus sequence of G/C G/C G/A CGCC.BRE functions as a repressor of basal transcription whose repression is released upon the binding of activators. The existence of the core elements is not entirely universal (Ioshikhes et al., 2005). It is believed that there are still more core elements to be discovered. Higher order structural properties of the DNA sequence are also involved in the recruitment of the PIC.

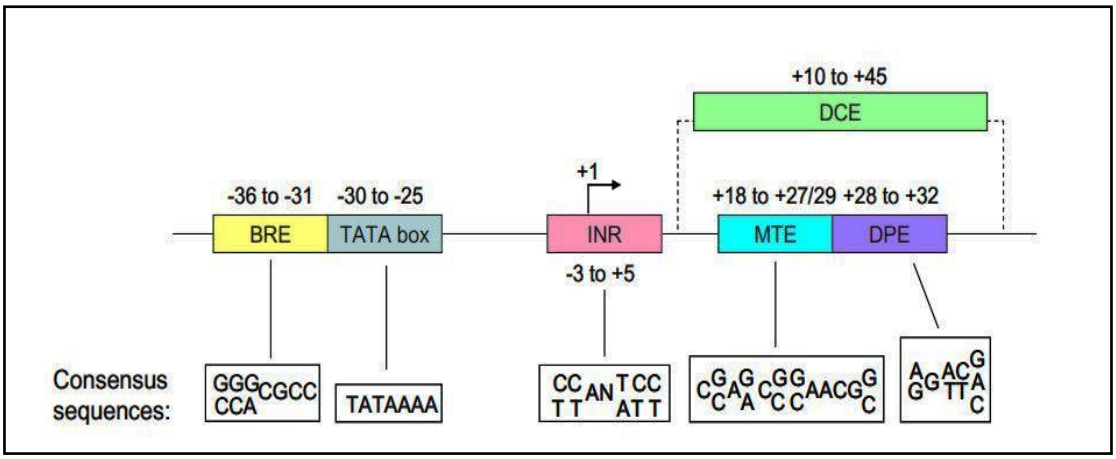

Figure 2.5: The elements of RNA polymerase II core promoter

Figure 2.5 shows the location of INR, TATA box, DPE, MTE, DCE and BRE corresponding to the transcription start site +1. The consensus sequence of each element has been shown in white boxes. The transcription start site (+1) has been indicated by the joint arrow.

The proximal promoter elements are located upstream of the core promoter. These can span up to a few hundred base pair and can be involved in the alteration of the transcription rate. An example, of a proximal element is the CpG island which is 500 base pair to 2 kilo base pairs (kb) in length and is highly GC rich (Kadonaga et al., 2003). Around 60% of human promoters are associated with them. The identification of core elements of CpG islands has not been done but they consist of multiple binding sites for transcription factor SpL. DNA methyl-transferases use CpG dinucleotides as the substrate of methylation but CpG islands are underrepresented since thymine can be formed by the deamination of methylated cytosine. On the other hand, methylation of CpG islands in the proximal promoters does not occur in active genes. Methylation of DNA at CpG islands leads to silencing of transcription and this is implicated in epigenetic imprinting.

2.6 Computational Approaches to Study Gene Regulation

Extensive lists of mammalian DNAs, RNAs and proteins is available due to the completion of many projects on whole-genome sequencing. So, these data can be easily obtained and accessed for computational analysis. However it was soon concluded from evident that it is not possible to study the complexities in the genomes of higher organisms by solely studying the number of parts in the genome, rather more sophisticated approach is

needed to study the interactions and networks of the RNAs, DNAs and proteins. Now the ultimate aim of computational biology is to shift focus towards the analysis of gene groups, their products and their network interactions (e.g.: metabolic and signalling networks). Transcriptional co-regulation network represents the most important level at which network connections emerge (Gerstein et al., 2012; Hardy et al., 2002). Microarray experiments have been used substantially to analyse yeast regulatory networks but it is not authentic to transfer these results directly to human system. Thus, mammalian promoter analysis is an emerging field of research that involves various strategies complementing mammalian networks. A reliable approach is thus to analyse genes or gene products and their interactions by defining one or more subsets by some substantiation to a biological process, condition or disorder. But these genes and gene groups are not precisely defined and have several subgroups that are functionally distinct. Such subgroups can be separated by conventional clustering methods. However there may be transcriptional coupling between the genes within such subgroups that contribute to a specific biological process or pathway, to ensure coordinated availability of proteins. Regulation of transcription is primarily done by the binding of transcription factors to their specific binding sites in the promoter/enhancer of the genes. Thus, co-regulated transcription can be traced on the molecular level by promoter analysis which reveals shared assembly of sets of transcription factor binding sites. We can represent such frameworks or modules that can be used to scan sequence databases for genes showing a similar promoter organization. A notable work has been done in this field where a molecular relationship was established between cardio-vascular disease and AD. Similarly disease associated genes have been linked via promoter analysis.

CHAPTER 3: MATERIALS AND METHODS

Our Methodology is based upon the hypothesis that genes that are co-expressed are also co-regulated. Therefore, since the genes specific to certain diseases and those that are common to all the diseases all resided in the same module, they may be co-regulated. This could be the reason for the clustering of these conditions in epidemiological studies. Furthermore, as there are many transcripts common to these diseases/conditions, it is plausible that similar/common biochemical pathways are active in these seemingly different conditions.

3.1 Terminology

3.1.1 Framework

A framework is the arrangement of two or more transcription factor binding sites (TFBSs) in a defined order and orientation and having a distinct distance range between neighbouring TFBSs.

3.1.2 Model

A model is the computational depiction of framework for detecting the occurrence of frameworks in long DNA sequences with computational aid.

3.2 Strategy

3.2.1 Problem-oriented Gene Selection

Initially a problem specific gene list (PSGL) is identified. The PSGL is correlated either with a disease, metabolic pathway or a signalling pathway. It might also be related to any gene group associated with a biological function.

3.2.2 Orthologous Promoter

Three mammalian species i: e: human, rat and mouse have orthologous promoters. These are collected for all the genes in the PSGL. This step is not mandatory but result accuracy will be more if you go through this step.

3.2.3 Transcription Factor Binding Site Prediction

Analysis of orthologous promoter sets is done to identify frameworks consisting of different transcription binding sites. TBFS framework or module refers to a group of transcription factor binding sites.

3.2.4 Models Construction

There should be at least two to three members in the network. Thus, there is further analysis of the models from orthologous promoter sets of the genes if there is a match between at least two additional promoters of the PSGL.

3.2.5 Extension of Models

In this step there is extension of at least one additional TBFS in the models resulting from model construction stage. This results in models of more than two elements. Models are then optimized in the previous step.

3.2.6 Database search with Final Models

The database of all human promoters is used to make a match list for the models that are defined in previous step. Then these models can be used individually against database consisting of human genes.

3.2.7 Multiple Sequence Alignment and Tree Generation

The list of genes which is obtained from step 6 after database search can be fed to *ClustalW* for multiple sequence alignment. After MSA we can know the relevance of our new predicted genes corresponding to AD based upon the distance parameter in tree.

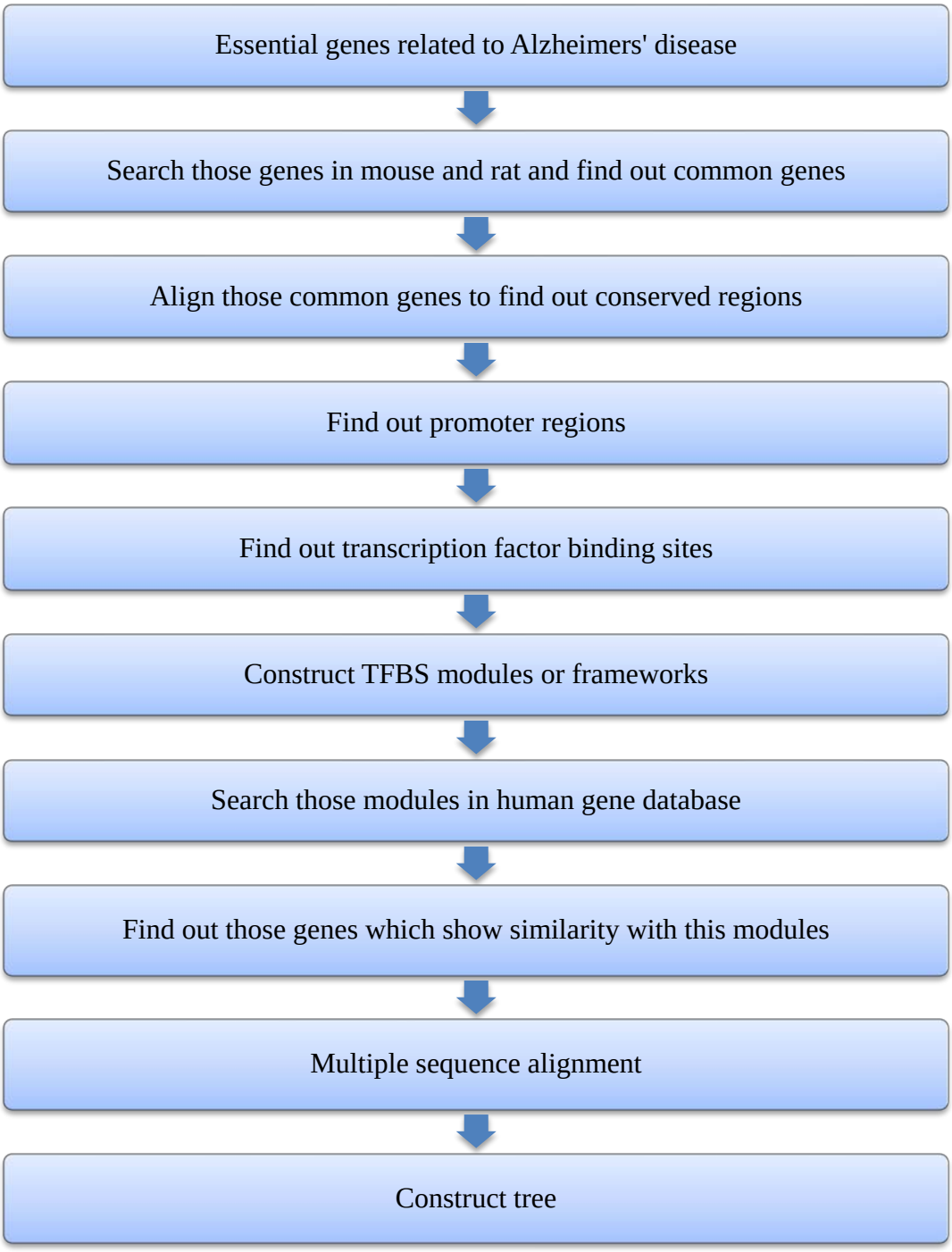

Figure 3.1: Flowchart representing the Overview of Total Process

3.3 Work Process

3.3.1 Literature Analysis Software

The programs *BiblioSpher, LitMiner*and *TM (Genomatix)* were used to collect current data on gene function, gene disease relation and subject-related gene expression from literature. The *LitMiner* has been developed by the GSF group. It is a complete web-

based resource that helps in generating the ranked lists of genes which are thought to be associated with diseases and tissues from summary of scientific publications that can be obtained from *Pub Med*. Around 50 genes were found out that thought to be linked with AD.

3.3.2 Promoter prediction

Prediction of the promoters' location is useful to target regions of interest to study with respect to regulatory interactions. But, it is challenging since the core promoter may be at a distance from the exons and the combination of core elements might differ from promoter to promoter. Analysis of training data sets from recognized promoter sets is a means for identification of functionally defined sequences conserved across promoters. This analysis is the basis of most successful programs. These programs then scan for these conserved signatures in genomic sequence. These include *PromoterInspector* (Scherf, et al., 2000), FirstEF (Davuluri et al, 2001) and Eponine (Down and Hubbard,2002). However, the sensitivity and specificity of these programs is limited for genome-scale analyses because they heavily depend on the data sets of known promoters. Promoters associated with CpG islands are generally well-predicted compared to those which are not (Bajic et al., 2004). Here, We extracted the promoter sequences from human, mouse and rat where available using the 'Comparative Genomics' task of the *ElDoradoTM*database (Genomatix Suite ElDorado TM , release 3.0, Human Genome NCBI build 34, Mouse Genome MGSCv3, Rat Genome NCBI build).

3.3.3 Transcription Factor Binding Site Prediction (TFBS)

TFBSs are conserved sequences. They have a certain degree of degeneracy which is recognized by transcription factors and they bind to it. The binding sites of the most well characterized transcription factors are compiled in online databases such as TRANSFAC, TRRD and COMPEL (Heinemeyer,TW. 1998). Programs such as, MATCH (al, 2003) or online tools such as TESS and TFSEARCH (Akiyama, Y. 1998) use TRANSFAC database for identifying TFBSs in input genomic sequences. One major drawback of these methods there can be a large number of FPs or TNs owing to the data quality initially used to populate the databases. To overcome these drawbacks more intuitive motif discovery approaches were followed in which sets of common sequence motifs located in the upstream of sets of genes that are expected to be co-regulated. The researchers are thus allowed to identify known and novel motifs that may be linked to a

transcription factor. Nucleotide sequence alignment was done by using *DiAlign* task of GEMS launcher so that overall promoter similarity for every orthologous promoter set can be checked.

3.3.4 Model Construction and Optimization

The extension of models was done by using *FastM* task. TFBSs identified by *MatInspector* (Cartharius et al., 2005) analysis were added manually to the existing models.

3.3.5 Database Search with Final Models

Database can be searched with the help of *ModelInspector*(Klingenhoff et al., 1999). A library of predefined models or models defined with *FrameWorker* or *FastM* is used by the *ModelInspector* to scan DNA sequences to track matches for these models. A model consists of various elements like transcription factor binding sites, hairpins or repeats, their sequential order, their orientation and their distance ranges. a proprietary scoring algorithm is used by the *ModelInspector* to include various element types into composite scoring of matches. Thus, IUPAC sequence elements can be successfully combined with different types of structural elements (e.g. hairpins) and weight matrices to assess match quality.

3.3.6 Multiple Sequence Alignment and Tree Generation

Sequence alignment is a standard technique in bioinformatics for visualization of the relationships between residues in an assortment of structurally and evolutionarily related proteins. The residues for each protein in an alignment are displayed on a single line given the amino acid sequences of a set of proteins are to be compared. Gaps are inserted such that equivalent residues appear in the same column. The meaning of equivalence generally depends on the context. The equivalent residue corresponds to having common evolutionary ancestry for the phylogeneticists, it refers to analogous locations on homologous folds in a set of proteins for structural biologists and for molecular biologists it means those residues that play similar functional roles in their equivalent proteins. In all the cases alignment gives an overview of the structural, evolutionary and functional constraints that characterize a protein family in a precise way. In this work, we took around 50 genes including some genes assumed to be related with AD as per our analysis. After that we extracted the amino acid sequence of all those genes in FASTA format from NCBI and conducted MSA by *ClustalW*(Larkin et al., 2007). Then we generate a tree by neighbour joining method which shown us the validation of our analysis.

CHAPTER 4: RESULTS AND DISCUSSION

As we know TFs bind primarily to the upstream region of genes, and each TF has its individual definite binding motif. TFs having analogous binding motif are grouped to form a TF-family. Combinations of TFs in a distinct array, distance range, and in a particular orientation are identified as TFBSs modules. Modules that are common to a set of genes, assumed to act collectively in the same biological context, are able to control the expression of these gene products. On the other hand, the finding regarding the co-regulation in the expression of different genes in a particular biological process may specify functional linkage in this process. This may be potentially used for the prediction of new genes that are linked to a disease. This is the key point of our result tabulated in table 4.1 and 4.2. At first we took four key genes (APP, PSEN1, PSEN2, and APOE) of human, rat, mouse to identify the common TF binding site. After significant analysis, we found that there are approximately 13 modules composed of two or more TFBSs families and the proposed TFBSs families consist of several TFs. The second column specifies the key genes of AD that the TFs of the module bind to, according to the search of the module in all human promoters by *Model Inspector* (Genomatix suit).

TFBS Modules:

Table 4.1:Modules Containing ≥2 TFBS Families

Module/Framework	AD Key Genes-Targets of Module
CTCF-E2FF	BACE1, BACE2, mAPP, hAPP, NCSTN, APH1A
CTCF-EGRE	BACE1, BACE2, mAPP, hAPP, APH1A, PS1, PEN2
CTCF-NRFI	BACE1, BACE2, mAPP, hAPP, PEN-2
CTCF-SP1F	
CTCF-SP1F-E2FF	mAPP, hAPP, NCSTN, APH1A, BACE1
EGRF- E2FF- E2FF	mAPP, hAPP, PEN-2, APH1A, BACE1, BACE2
EGRF- CTCF- E2FF	hAPP, PEN2, APH1A, BACE2, NCSTN
EGRF- CTCF- E2FF	mAPP, hAPP, PEN2, APH1A, BACE1, BACE2
HAND- CTCF-SP1F	BACE2, PS2, mAPP, hAPP
CTCF-SP1F-SP1F	BACE1, BACE2, mAPP, hAPP
CTCF-SP1F-NRF1	BACE1, APH1A, mAPP, hAPP
CTCF-ZBPF-SP1F	BACE2, mAPP, hAPP
CTCF-ZBPF-EGRF	BACE2, PS2, hAPP
CTCF-SP1F- EGRF	BACE2, PS2, hAPP

TFs and its family:

Table 4.2:TF Family Compositions of the Modules

Transcription Factors	Family	Description
CTCF, CTCFL, HMGB1L1	CTCF	Transcriptional Repressor also known as 11-zinc finger protein CCC-TC binding factor. Involved in insulator activity, transcriptional regulation and regulation of chromatin structure
MESP1z, MESP2, LYL1, HAND1, HAND2, NHLH1. NHLH2, TAL1, TAL2, SCXA, SCXB, TWIST1, TWIST2, TCF3, TCF12, TCF15	HAND (Heart- and Neural Crest Derivatives)	Molecular mediators of cardiac development and congenital heart disease
WTI1, ZBTB7A, ZBTB7B, EGR1, EGR2, EGR3, EGR4	EGRF (Epidermal Growth Factor Receptor)	Nerve Growth Factor Induced Protein C and Related factors
TFDP1, TFDP1, TFDP1, E2F1, E2F2, E2F3, E2F4, E2F5, E2F6, E2F7, E2F8	E2FF	Found in higher eukaryotes, Play a major role in G1/S transition
NRF1, NRF2	NRF1	Activates the expression of some key metabolic genes regulating cellular growth and nuclear genes
KLF10, KLF11, KLF16, SP1, Sp2, SP3, SP4, SP5, SP6, SP7, SP8	SP1F	Also known as specificity protein, Used as control protein to compare with when studying estrogen receptor and aryl hydrocarbon receptor

ZNF202, ZNF148, ZKSCAN3, ZNF219, ZNF300, ZNF281	ZBPF (Zinc Binding Protein Factors)	Exert their modulatory effect in the vicinity of any sequence to which the protein domain binds

The first table clearly indicates the proposed modules that are designed by *Genomatix* tool and the target genes of those modules. These modules are constructed by computational algorithm on the basis of permutation and combination. Similarly the second table shows all related transcription factors and their families. Similar transcription factors are grouped to a particular family on the basis of biological relevance.

Genes linked to AD Pathology:

Table 4.3:Genes Linked to AD Pathophysiology

S. No.	Gene Name	Length (No of Amino acids)	Tissue Type
1	APOE	63	Blood
2	APOB	825	Liver
3	A2M	353	Liver
4	ABCB1	1280	Cerebellum
5	AD7C-NTP	375	Neuronal
6	ACHE	546	Brain
7	APP	751	Retinoblastoma
8	BCHE	64	Brain
9	ABCA2	867	Eye
10	BDNF	247	Brain
11	BACE1	501	Brain
12	CCK	115	Pancreas
13	CETP	425	Liver
14	CASP6	293	Lung
15	CLU	449	Brain

16	CHRNA7	321	Brain
17	CHAT	630	Cerebellum
18	CSNK1D	409	Spleen
19	**REEP5**	**189**	Not specified
20	CTNNA3	516	Peripheral Nervous System
21	DCN	359	Liver
22	DBN1	649	Rhabdo-myosarcoma
23	CYP19A1	359	Placenta
24	ESR2	323	Testis
25	FN1	268	Carcinoma
26	GLUL	373	Brain
27	HTR6	440	Fetal Brain System
28	**CDH1**	**647**	Not specified
29	FE65	708	Leimyosarcoma
30	MAPK1	360	Brain
31	GSK3B	74	Fetal Brain System
32	**EIF5**	**431**	Not specified
33	CYCS	105	Carcinoma
34	PLAU	431	Carcinoma
35	**MRP**	**1280**	Plasma membrane
36	MAPT	352	Brain
37	**CLU**	**449**	Not specified
38	NgR	600	Brain
39	UBB	229	Brain
40	IVIg	110	B-Cell
41	PSEN1	463	Melanotic Melanoma
42	**CBP**	**313**	Mouse cell

43	MAPK10	461	Brain
44	STH	128	Cerebellum
45	SNCA	140	Melanotic Melanoma
46	PARP1	250	Melanotic Melanoma
47	PIN1	45	Testis
48	IGFIR	55	Brain
49	GAL	123	Carcinoma
50	ACHE	546	Brain

Table containing the genes related to AD shows the approximately a list of 50 number of genes which are assumed to be associated with Alzheimer's disease pathophysiology. These gene names except those bold ones are extracted through *Litminer* and *BiblioSphere* tool of *Genomatix* suit. Our predicted genes are shown here in bold marks which are obtained by our computational workflow. These genes are Cadherin1, CBP, CLU, REEP5, MRP and EIF5 etc.

Predicted Gene Table:

This table enlists all the predicted genes which are proposed to be associated with Alzheimer's disease pathology as per our analysis.

Table 4.4: Predicted Gene List

GENES	GENEBANK ID	PROTEIN
REEP5	AAH00232.2	Receptor expression protein
EIF5	CAG32993.1	Transcription factor
CDH1	AAI46663.1	Cadherin 1
CLU	AAH19588.1	Clusterin
CBP	AAH01766.1	Calcium binding protein
MRP	NP_000918.2	Multi drug resistance protein

Dendrogram (MSA):

The Dendrogram is depicting the evolutionary relation between around 50 Alzheimer's related genes including our predicted genes through multiple sequence alignment. We have done this with the help of *ClustalW* suit (Larkin et al., 2007). As we know the key gene for Alzheimer's disease is APP which is highlighted here having NCBI id gi41350939, shows a strong evolutionary relationship with genes like Cadherin 1 having NCBI id 4757960 and EIF5 having NCBI id 48145541.We can clearly see it from the highlighted portion of this Dendrogram. Similarly the PSEN1 gene having id gi 15079861 and APOE gene having id gi 705044152 clearly shows strong evolutionary relationship with REEP5 (gi 68087645) and MRP (gi42741659) respectively. There are several other genes which also show evolutionary relationship with APP and can have the potential in AD etiology.

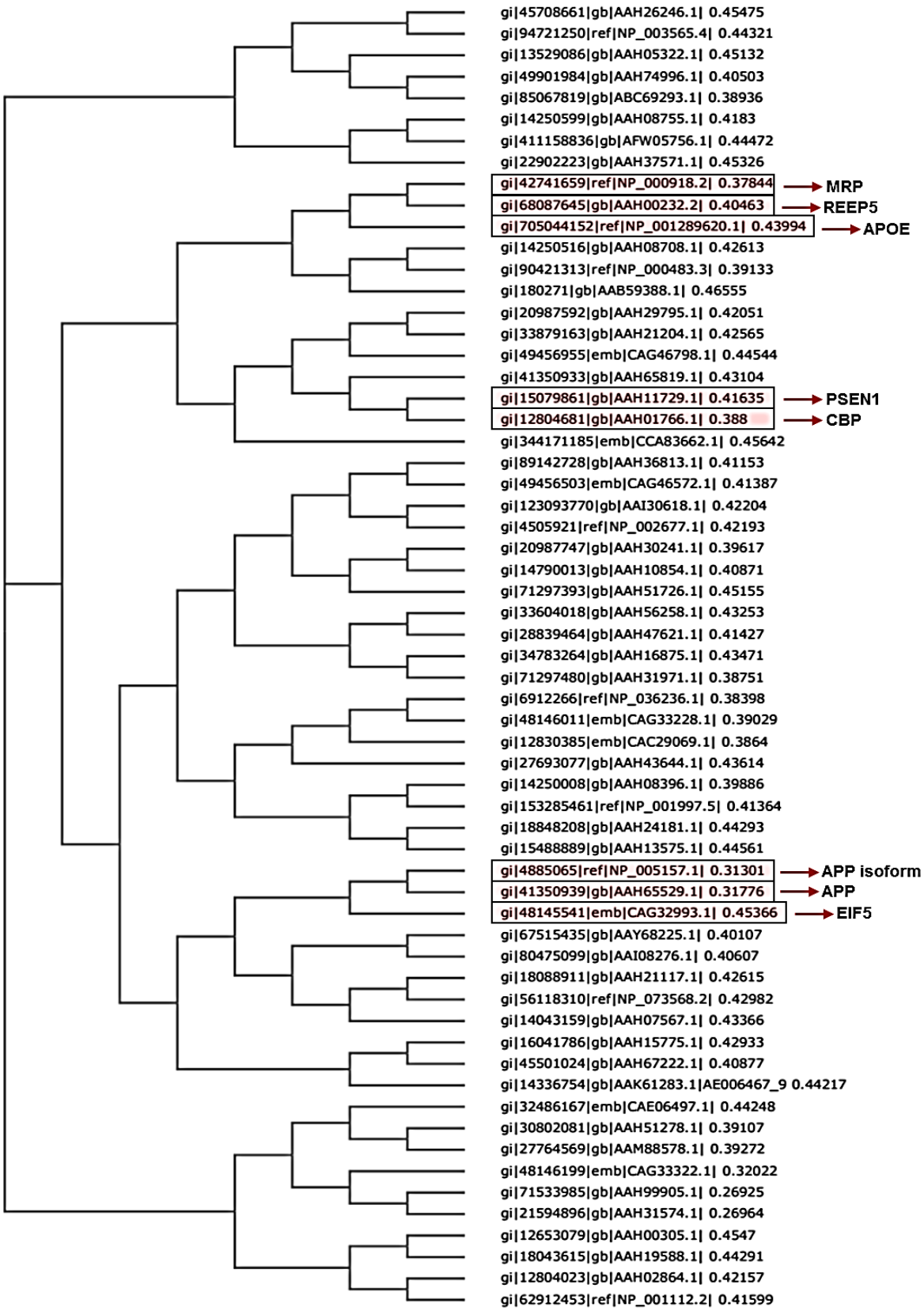

Figure.4: Dendrogram of all set of genes obtained by MSA through *ClustalW*

CHAPTER 5: CONCLUSION AND FUTURE PERSPECTIVE

In this study, we present an intensive computational biology approach to study a complex neurodegenerative disease such as AD. Besides the identification of modules that shed light on higher-order properties of the transcriptome, we identified a module that contained multiple genes known to play prominent roles in AD. We believe that important pathophysiological properties that connect AD and ageing are highlighted in this module. Several *cis* regulatory elements were identified, some of which mapped to the binding sites of known transcription factors involved in neurodegenerative diseases. As described above, the identification of many of the genes which is assumed to be linked to the molecular mechanisms of AD was made possible in the bioinformatics analysis. Specifically, the binding sites of the TF families: CTCF, EGRF, KLFS, SP1F, and ZBPF are proposed for further research study and could present probable candidates for therapeutic treatment of AD and many other neurodegenerative diseases. Previously known regulations of target genes by transcription factors confirm these modules, such as CLU, which is linked to AD. Moreover, several target genes like EIF5 that are possibly involved in AD pathogenesis have not yet been described as AD-related genes. The most likely candidate genes are Cadherin 1, CBP, MRP, CLU, and REEP5, which are derived from our computational study must be analysed further to ensure their relation to AD. Additionally, a modular approach can be practiced in future where organization of genes on the basis of co-expression or co-regulation can prove to be an efficient technique for studying human diseases and analysing results from multiple studies.

REFERENCES

Akatsu, H; Yamagata, HD; Kawamata, J; Miki, T; Kosaka, K. (2006) Variations in the BDNF gene in autopsy-confirmed Alzheimer's disease and dementia with Lewy bodies in Japan. *Dement Geriatr Cogn Disord* , 22(3), 216-222.

Akiyama, Y. (1998) TFSEARCH: searching transcription factor binding sites. *Nucleic Acids Research* , 9, 1-6.

Belkhelfa, M; Rafa, H; Medjeber, O; Arroul-Lammali, A; Behairi, N; Abada-Bendib, M; Makrelouf, M; Belarbi, S; Masmoudi, AN; Tazir, M; Touil-Boukoffa, C. (2014) IFN-γ and TNF-α are involved during Alzheimer disease progression and correlate with nitric oxide production: a study in Algerian patients. *J Interferon Cytokine Res*. 34(11):839-47.

Berridge, MJ. (1998). Neuronal Calcium Signaling. *Neuron, 21*(1), 13-25.

Betty, L; Blacks, LJ. (1981). The Regulation of Endoplasmic Reticulum Calcium Uptake of Adipocytes by Cytoplasmic Calcium. *The Journla of Biological Chemistry* ,33, 322-329.

Billings, LM; Oddo, S; Green, KN; McGaugh, JL; LaFerla, FM. (2005) Intraneuronal Abeta causes the onset of early Alzheimer's disease-related cognitive deficits in transgenic mice. *Neuron.* 3;45(5):675-88.

Blennow, K; De Leon, MJ; Zetterberg H. (2006). Alzheimer's disease.Lancet , 368(9533), 387-403.

Borroni, B; Grassi, M; Archetti, S; Costanzi, C; Bianchi, M; Caimi, L; Caltagirone, C; Di Luca, M; Padovani, A. (2009) BDNF genetic variations increase the risk of Alzheimer's disease-related depression. *J Alzheimers Dis*.18(4):867-75.

Brouwers, N; Sleegers, K; Broeckhoven, C. Molecular genetics of Alzheimer's disease: an update. Ann Med. 2008; 40(8):562-83.

Cartharius, K; Frech, K; Grote, K; Klocke, B; Haltmeier, M; Klingenhoff, A; Frisch, M; Bayerlein, M; Werner, T. (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* 21, 2933-42

Cedazo-Minguez, C. (2001). Apolipoprotein E: a major piece in the Alzheimer's disease puzzle. *J. Cell. Molecular Med* , 11(6), 254-264.

Corrêa, MS; Vedovelli, K; Giacobbo, BL; de Souza, CE; Ferrari, P; Lima Argimon; Walz, JC; Kapczinski, F; Bromberg, E. (2015) Psychophysiological correlates of cognitive deficits in family caregivers of patients with Alzheimer Disease. *Neuroscience.* 12;286:371-82.

Coskun, PE; Beal, MF; Wallace, DC. (2004) Alzheimer's brains harbor somatic mtDNA control-region mutations that suppress mitochondrial transcription and replication. *Proc Natl Acad Sci U S A.* 101(29):10726-31.

Cruchaga, C; Haller, G; Chakraverty, S; Mayo, K; Vallania, FL; Mitra, RD; Faber, K; Williamson, J; Bird, T; Diaz-Arrastia, R; Foroud, TM; Boeve, BF; Graff-Radford, NR; St Jean, P; Lawson, M; Ehm, MG; Mayeux, R; Goate,AM.(2012) NIA-LOAD/NCRAD Family Study Consortium. Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. *PLoS One*.7(2).

Dehm, SM; Hilton, TL; Wang, EH; Bonham, K. (2004) SRC proximal and core promoter elements dictate TAF1 dependence and transcriptional repression by histone deacetylase inhibitors. *Mol Cell Biol*. 24(6):2296-307.

Down, TA; Hubbard, TJ. (2002). Computational detection and location of transcription start sites in mammalian genomic DNA. Genome Res, 12(3), 412-418.

Ferri, CP. (2005). Global prevalence of dementia: A Delphi consensus study. *Lancet* , 366(9503), 2112-2117.

Finckh, U; T. Mann, U. (2000). High prevalence of pathogenic mutations in patients with early-onset dementia detected by sequence analyses of four different genes. *Am. J. Hum. Genet.* ,66(1), 110–117.

Fox, NC; Warrington, EK; Seiffer, AL; Agnew, SK; Rossor, MN. (1998) Presymptomatic cognitive deficits in individuals at risk of familial Alzheimer's disease. A longitudinal prospective study. *Brain*. 121.

Gamba, P; Testa, G; Sottero, B; Gargiulo, S; Poli, G; Leonarduzzi, G. (2012) The link between altered cholesterol metabolism and Alzheimer's disease. *Ann N Y Acad Sci*. 1259:54-64.

Gambhir, S. (2005). Multimodality tumor imaging targeting integrin αvβ3. *Biotechniques* , 1, 6-17.

Gatz, M. (1997). Chemical control of gene expression. *Annual Review* , 48, 89-106.

Geerlings, MI. (2008). History of depression, depressive symptoms, and medial temporal lobe atrophy and the risk of Alzheimer disease. *Neurology* , 12,1258-1264.

Geldenhuys, WJ; Van der Schyf, CJ. (2011) Role of serotonin in Alzheimer's disease: a new therapeutic target? *CNS Drugs*. 25(9):765-81.

Gerstein, MB; Kundaje, A; Hariharan, M; Landt, SG; Yan, K-K; Cheng, C; Mu, XJ; Khurana, E; Rozowsky, J; Alexander, R. (2012). Architecture of the human regulatory network derived from ENCODE data. *Nature* , 489(7414):91-100

Guo, X; Tang, P; Liu, P; Liu, Y; Hou, C; Li, R. (2014) Meta-analysis of the association between two neprilysin gene polymorphisms and Alzheimer's disease. *J Neurol Sci*. 15;346(1-2):6-10.

Hardy, J; Selkoe, DJ. (2002) The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science* 297, 353-356

Heinemeyer, TW. (1998). Databases on transcriptional regulation:TRANSFAC, TRRD and COMPEL. *Nucleic Acids Res* , 362–367.

Hendrie, HC; Murrell, J; Baiyewu, O; Lane, KA; Purnell, C; Ogunniyi, A; Unverzagt, FW; Hall, K; Callahan, CM; Saykin, AJ; Gureje, O; Hake, A; Foroud, T; Gao, S. (2014) APOE ε4 and the risk for Alzheimer disease and cognitive decline in African Americans and Yoruba. *Int Psychogeriatr.* 26(6):977-85.

Hu, YL; Lei, ZM; Rao, CV. (1996) Cis-acting elements and trans-acting proteins in the transcription of chorionic gonadotropin/luteinizing hormone receptor gene in human choriocarcinoma cells and placenta. *Endocrinology*. 137(9).

Hyman, SE; Comb, M; Lin, YS; Pearlberg, J; Green, MR; Goodman, HM. (1998) A common trans-acting factor is involved in transcriptional regulation of neurotransmitter genes by cyclic AMP. *Mol Cell Biol.* (10):4225-33.

Ioshikhes, NI. (2005). Promoter Classifier: software package for promoter database analysis. *Appl. 12, Bioinformatics*

Jinwal, UK; Koren, J; O'Leary, JC; Jones, JR; Abisambra, JF; Dickey, CA. (2010) Hsp70 ATPase modulators as therapeutics for Alzheimer's and other neurodegenerative diseases. Mol Cell Pharmacol 2, 43-46.

Kadonaga, ST. (2003). The RNA Polymerase II Core Promoter. *Annu. Rev. Biochem.* , 34, 449-79.

Karege, F; Perret, G; Bondolfi, G; Schwald, M; Bertschy, G; Aubry, JM. (2002). Decreased serum brain-derived neurotrophic factor levels in major depressed patients. *Psychiatry Res*, 109:143-148.

Kel, AE; Gössling, E; Reuter, I; Cheremushkin, E; Kel-Margoulis, OV; Wingender, E. (2003). MATCH: A tool for searching transcription factor binding sites in DNA sequence. *Nucleic Acids Res* , 31(30), 3576-3579

Klingenhoff, A; Frech, K; Quandt, K; Werner, T. (1999). Functional promoter modules can be detected by formal models independent of overall nucleotide sequence similarity. Bioinformatics 15, 180-186

Klingenhoff, A; Frech, K; Quandt, K; Werner, T. (1999). Functional promoter modules can be detected by formal models independent of overall nucleotide sequence similarity. Bioinformatics 15, 180-186

Kojima, N; Shirao, T. (2007) Synaptic dysfunction and disruption of postsynaptic drebrinactin complex: A study of neurological disorders accompanied by cognitive deficits. *Neurosci Res*, 58.1-5.

Krabbe, K; Nielsen, A; Krogh-Madsen, R; Plomgaard, P; Rasmussen, P; Erikstrup, C; Fischer, C; Lindegaard, B; Petersen, A; Taudorf, S; Secher, N; Pilegaard, H; Bruunsgaard, H; Pedersen, B. Brain-derived neurotrophic factor (BDNF) and type 2 diabetes. *Diabetologia* 2007, 50:431-438

Krüger, J; Moilanen, V; Majamaa, K; Remes, AM. (2012) Molecular genetic analysis of the APP, PSEN1, and PSEN2 genes in Finnish patients with early-onset Alzheimer disease and frontotemporal lobar degeneration. *Alzheimer Dis Assoc Disord*. 26(3):272-6.

Krut, JJ; Zetterberg, H; Blennow, K; Cinque, P; Hagberg, L; Price, RW; Studahl, M; Gissl,M. (2013) Cerebrospinal fluid Alzheimer's biomarker profiles in CNS infections.JNeurol 260, 620-626.

Kucho, K; Yoshioka, S; Taniguchi, F; Ohyama, K; Fukuzawa, H. (2003) Cis-acting elements and DNA-binding proteins involved in CO2-responsive transcriptional activation of Cah1 encoding a periplasmic carbonic anhydrase in Chlamydomonas reinhardtii. *Plant Physiol*. 33(2):783-93.

Lam, S. (2001). Phytophthora genomics consortium. *Phytopathology* , 2(2), 155-158.

Lamsa, KP. (2007). Anti-Hebbian long-term potentiation in the hippocampal feedback inhibitory circuit. *Science* , 15, 1262-1266.

Larkin, MA; Blackshields, G; Brown NP; Chenna, R; McGettigan, PA; McWilliam, H; Valentin, F; Wallace, IM; Wilm, A; Lopez, R; Thompson, JD; Gibson, TJ; Higgins, DG. (2007) *ClustalW and ClustalX version 2.Bioinformatics* 23(21): 2947-2948.

Laske, C; Stransky, E; Leyhe, T; Eschweiler, GW; Wittorf, A; Richartz, E; Bartels, M; Buchkremer, G; Schott K. Stage-dependent BDNF serum concentrations in Alzheimers disease. *J Neural Transm*, 113:1217-1224.

Leonard, SG. (1998). Further nvestigation of a chromosome 15 locus in schizophrenia analysis of affected sibpairs from the NIMH Genetics analysis of affected sibpairs from the NIMH Genetics Initiative. *American Journal of Medical Genetics American Journal of Medical Genetics* ,62, 308-312.

Liang, WS; Reiman, EM; Valla, J; Dunckley, T; Beach, TG; Grover, A; Niedzielko, TL; Schneider, LE; Mastroeni, D; Caselli, R; Kukull, W; Morris, JC; Hulette, CM; Schmechel, D; Rogers, J; Stephan, DA. (2008). Alzheimers disease is associated with reduced

expression of energy metabolism genes in posterior cingulate neurons. *Proc NatlAcadSci ,USA* .105.4441-4446.

Lohmann, E; Guerreiro,RJ; Erginel-Unaltuna, N; Gurunlian, N; Bilgic, B; Gurvit, H; Hanagasi, HA; Luu, N; Emre, M; Singleton, A. (2012) Identification of PSEN1 and PSEN2 gene mutations and variants in Turkish dementia patients. *Neurobiol Aging*. 33(8):1850.e17-27.

Lukas, P; Beth, A; Venkata, SM; Joseph HC; Bhaskar, SK; Michael, FE; Daniel, RW. (2004). The Brain-Derived Neurotrophic Factor val66met Polymorphism and Variation in Human Cortical Morphology. *The Journal of Neuroscience* , 10099 –10102.

Lukiw, W; Bjattacharjee, S. (2012). Generation of Reactive Oxygen Species (ROS) and Pro-Inflammatory Signaling in Human Brain Cells in Primary Culture. *Alzheimer's Disease & Parkinsonism* , 2(001), 1-6.

Martins, IJ; Berger, T; Sharman, MJ; Verdile, G; Fuller, SJ; Martins, RN. (2009) Cholesterol metabolism and transport in the pathogenesis of Alzheimer's disease. *J Neurochem*. 111(6):1275-308.

Masters, CL; Simms, G; Weinman, NA; Multhaup, G; McDonald, BL; Beyreuther, K. (1985) Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proc NatlAcadSci USA* 82, 4245-4249.

Mayeux, R; Stern, Y. (2012). Epidemiology of Alzheimer Disease. Cold Spring Harb Perspectives in Medicine , 2(8), 1-11.

Meltzer, CC; Smith, G; DeKosky, ST; Pollock, BG; Mathis, CA; Moore, RY; Kupfer, DJ; Reynolds, CF. (1998) Serotonin in aging, late-life depression, and Alzheimer's disease: the emerging role of functional imaging. *Neuropsychopharmacology*. 18(6):407-30.

Mielke, MM; Leoutsakos, JM; Tschanz, JT; Green, RC; Tripodis, Y; Corcoran, CD; Norton, MC; Lyketsos, CG. (2011) Interaction between vascular factors and the APOE ε4 allele in predicting rate of progression in Alzheimer's disease. *J Alzheimers Dis*.26(1):127-34.

Murer, MG; Yan, Q; Raisman, R. (2001). Brain-derived neurotrophic factor in the control human brain, and in Alzheimer's disease and Parkinson's disease. *Prog Neurobiology* ,8(9), 71-124.

Nathalie, C; Schnetz-Boutaud; Joshua, H; Jared, E; Jonthan, L. (2012). Identification and Confirmation of an Exonic Splicing Enhancer Variation in Exon 5 of the Alzheimer Disease Associated PICALM Gene. *Annals of Human Genetics* ,76(6), 448-453.

Need, AC; Goldstein, DB. (2010) Whole genome association studies in complex diseases: where do we stand? *Dialogues Clin Neurosci* 12(1):37-46.

Oliveira, JR; Zatz, M. (1999) The study of genetic polymorphisms related to serotonin in Alzheimer's disease: a new perspective in a heterogenic disorder. *Braz J Med Biol Res*. (4):463-7.

Palop, JJ; Mucke, L. (2010) Amyloid-beta-induced neuronal dysfunction in Alzheimer's disease: from synapses toward neural networks. *Nat Neurosci*. 13(7):812-8.

Papassotiropoulos, A; Stephan, DA. (2006). Common KIBRA alleles are associated with human memory performance. *Science* , 314(5798), 475–477.

Park, HK; Na, DL; Lee, JH; Kim, JW; Ki, CS. (2008) Identification of PSEN1 and APP gene mutations in Korean patients with early-onset Alzheimer's disease. *J Korean Med Sci*. 23(2):213-7

Pedersen, NL; Gatz, M; Mortimer, JA; Renoyolds, CA; Berg, S; Johansson, B.(2006). Potentially modifiable risk factors for dementia in identical twins. *Alzheimer's & Dementia*, 2(2), 110-117

Peers, C; Pearson, HA; Boyle, JP. (2007) Hypoxia and Alzheimer's disease. *Essays Biochem*.43:153-64.

Rainulf, A; Stelzma, HN. (1995). An English Translation of Alzheimer's 1907 Paper. *Clinical Anatomy* , 429-431.

Reinikainen, KJ; Paljärvi, L; Huuskonen, M; Soininen, H; Laakso, M; Riekkinen, PJ. (1988) A post-mortem study of noradrenergic, serotonergic and GABAergic neurons in Alzheimer's disease. *J Neurol Sci*. 84(1):101-16.

Riemenschneider, M. (2002). A polymorphism of the brain-derived neurotrophic factor (BDNF) is associated with Alzheimer's disease in patients lacking the Apolipoprotein E epsilon4 allele. *Molecular Psychiatry* , 782-785.

Riley, BW. (2002). A 3-Mb map of a large segmental duplication over lapping the alpha 7-nicotinic acetylcholine receptor gene (CHRNA7) at human 15q13-q14. *Genomics 79* , 197-209.

Ritchie, K; Lovestone, S. (2002). The dementias. Lancet 360 (9347), 1759-1766.

Robert-Freedman, SL. (2001). Linkage Disequilibrium for Schizophrenia at the Chromosome 15q13-14 Locus of the a7-Nicotinic Acetylcholine Receptor Subunit Gene (CHRNA7). *American Journal of Medical Genetics* , 20-22.

Rodríguez, JJ; Noristani, HN; Verkhratsky, A. (2012) The serotonergic system in ageing and Alzheimer's disease. *Prog Neurobiol*. 99(1):15-41.

Rota, E; Bellone, G; Rocca, P; Bergamasco, B; Emanuelli, G; Ferrero, P. (2006) Increased intrathecal TGF-beta1, but not IL-12, IFN-gamma, and IL-10 levels in Alzheimer's disease patients. *Neurol Sci*. 27(1):33-9.

Ryan, VH; Primiani, CT; Rao, JS; Ahn, K; Rapoport, SI; Blanchard, H. (2014) Coordination of gene expression of arachidonic and docosahexaenoic acid cascade enzymes during human brain development and aging. *PLoS One*. 25; 9(6)

Savage, MJ; Lin, YG; Ciallella, JR; Flood, DJ; Scott, RW. (2002) Activation of c-Jun N-terminal kinase and p38 in an Alzheimers disease model is associated with amyloid deposition. *J Neurosci* 22, 3376-3385.

Schenk, SD. (2003). Alzheimer's disease: molecular understanding predicts amyloid-based therapeutics. *Annual Review Pharmacol. Toxicol.* , 10(1), 545–584.

Scherf, M; Klingenhoff, A; Werner, T; (2000). Highly Specific Localization of Promoter Regions in Large Genomic Sequences by *PromoterInspector*: A Novel Context Analysis Approach. J. Mol. Biol. 297 (3), 599-606

Scherf, U. (2000). A gene expression database for the molecular pharmacology of cancer. *Nature Genetics* , 24(3), 236-244.

Schliebs, R; Arendt, T. (2011) The cholinergic system in aging and neuronal degeneration. *Behav Brain Res*. 10;221(2):555-63.

Selkoe, DJ. (1994) Normal and abnormal biology of the beta amyloid precursor protein.*Annu Rev Neurosci*17, 489-517.

Selkoe, DJ. (2001). Alzheimer's disease results from the cerebral accumulation and cytotoxicity of amyloid β-protein. *Journal of Alzheimer's Disease* , 3(1), 75-80.

Serio, PA. (2008). Neuroprotective effects of the Alzheimer's disease-related gene seladin-1. *J. Mol. Endocrinol. 41* , 251–261.

Seripa, D. (2004). Alzheimer disease risk associated with APOE4 is modified by STH gene polymorphism. *Neurology* , 1631-1633.

Sherman, BT; Huang, DW; Tan, Q; Guo, Y; Bour, S; Liu, D; Stephens, R; Baseler, MW; Lane, HC; Lempicki, RA.(2007) DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. *BM Bioinformatics*, 8(426).

Suzuki, Y. (2001). Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites. *EMBO Rep* , 388-393.

Thome, J; Retz, W; Baader, M; Pesold, B; Hu, M; Cowen, M; Durany, N; Adler, G; Henn, FA; Rösler, M. (2001) Association analysis of HTR6 and HTR2A polymorphisms in sporadic Alzheimer's disease. *J Neural Transm*. 108(10):1175-80.

Tsoi, LM. (2007). Apoprotein E isoform-dependent expression and secretion of pro-inflammatory cytokines TNF-alpha and IL-6 in macrophages. *Arch Biochem Biophys* , 33-40.

Tucker, ML; Whitelaw, CA; Lyssenko, NN; Nath P. (2002) Functional analysis of regulatory elements in the gene promoter for an abscission-specific cellulase from bean and isolation, expression, and binding affinity of three TGA-type basic leucine zipper transcription factors. *Plant Physiol*. 130(3):1487-96.

Ventriglia, M. (2002). Association between the BDNF 196 A/G polymorphism and sporadic Alzheimer's disease. *Mol Psychiatry* , 7(2), 136-138.

Wang, L; Day, J; Roe, CM; Brier, MR; Thomas, JB; Benzinger, TL; Morris, JC; Ances, BM. (2014) The effect of APOE ε4 allele on cholinesterase inhibitors in patients with Alzheimer disease: evaluation of the feasibility of resting state functional connectivity magnetic resonance imaging. *Alzheimer Dis Assoc Disord*. 28(2):122-7.

Wang, X; Smith, MA. (2009). The role of abnormal mitochondrial dynamics in the pathogenesis of Alzheimer's disease. *J Neurochem* , 1(153), 153-159.

Wei, J. (2004). A study of a genetic association between the PTGS2/PLA2G4A locus and schizophrenia. *Prostaglandins Leukotrienes and Essential Fatty* , 70, 410-415.

Weiland, NG. (1992). Estradiol selectively regulates agonist binding sites on the N-methyl-D-aspartate receptor complex in the CA1 region of the hippocampus. *Endocrinology* , 132(2), 662-668.

Wesley, M; Charlene, C. (2007). Where excludability matters: Material versus intellectual property in academic biomedical research. *Research Policy 36* , 1184-1200.

Yaguchi, K; Nishimura-Akiyoshi, S; Kuroki, S; Onodera, T; Itohara, S. (2014) Identification of transcriptional regulatory elements for Ntng1 and Ntng2 genes in mice. *Mol Brain*. 19;7:19.

Zhang, H; Ozbay, F; Lappalainen, J; Kranzler, HR; van Dyck, CH; Charney, DS; Price, LH; Southwick, S; Yang, BZ; Rasmussen, A; Gelernter, J. (2006) Brain derived neurotrophic factor (BDNF) gene variants and Alzheimer's disease, affective disorders, posttraumatic stress disorder, schizophrenia, and substance dependence. *Am J Med Genet B Neuropsychiatr Genet*. 5;141B(4):387-93.

Zhang, X; Le, W. (2010) Pathological role of hypoxia in Alzheimer's disease. *Exp Neurol*. 223(2):299-303.