Computational Studies to analyze DNA methylation alterations in Alzheimer's Disease

Enrolment No - **161506**
Name of the student - **Disha Angra**
Name of the supervisor - **Dr. Tiratha Raj Singh**

Submitted in partial fulfilment of the requirement

for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

BIOINFORMATICS

DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS,

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY,

WAKNAGHAT, SOLAN 173234, HIMACHAL PRADESH, INDIA

ACKNOWLEDGEMENT

Firstly, I would like to extend my sincere gratitude to my final year project supervisor Dr. Tiratha Raj Singh for providing me with an opportunity to conduct research work in my area of interest. Also, his continuous guidance, efforts, and invertible suggestions throughout the duration of the project was blessing in disguise.

I give my greatest acknowledgement to my Parents for all round support during my studies. The endless sacrifices from my parents have made me what I am today.

I also thank my PhD scholars Rohit Shukla and Arvind for their encouragement and constant support.

DECLARATION BY THE STUDENT

I hereby declare that the work reported in the B.Tech project report entitled "**Computational Studies to analyze DNA methylation alterations in Alzheimer's disease**" submitted at Jaypee University of Information Technology, Waknaghat, India, is an authentic record of my work carried out under the supervision of Dr. Tiratha Raj Singh. I have not submitted this work elsewhere for any other degree or diploma.

disha

(Signature of the Students)

Disha Angra (161506)

Department of Biotechnology and Bioinformatics,

Jaypee University of Information Technology,

Waknaghat, India

Date: 15/07/2020

CERTIFICATE

This is to certify that the project entitled "**Computational studies to analyse DNA methylation alterations in Alzheimer's Disease**", submitted by Disha Angra is in its partial fulfilment for the award of degree of bachelors of Technology in Bioinformatics to Jaypee University of Information Technology Waknaghat, Solan (H.P.), India is an authentic record of candidate's own work carried out by her under my supervision.

This work has not been submitted partially or fully to any other university or institution in order to achieve any award or other degree.

Dr Tiratha Raj Singh

Associate Professor,

Department of Biotechnology and Bioinformatics,

Jaypee University of Information Technology,

Waknaghat, Dist. Solan (173234), Himachal Pradesh, India

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. ABSTRACT

Alzheimer's disease (AD), a progressive and chronic neurodegenerative disorder that can destroy memory cells and thinking skills. Also the ability to carry out the multiple tasks gets decreased. The disorder affects different areas of cerebral cortex and hippocampus. Accumulations of amyloid-β and neurofibrillary tangles are the major hallmarks of AD. Early onset of AD includes memory loss, difficulty in planning and problem solving, difficulty determining time and place and vision loss

DNA methylation is a mechanism of transferring methyl group to DNA, thereby modifying the functions of genes. Epigenetics contribute to AD and these modifications such as DNA methylation and histone modifications play a role in neurodegenerative chronic diseases such as AD and Parkinson's disease (PD). DNA methylation regulates gene expression analysis by recruiting proteins and inhibiting the binding of transcription factors to DNA. Methylation exists in two forms: DNA hypomethylation which refers to loss of methyl group in the 5-methylcytosine nucleotide and DNA hypermethylation which refers to addition of methyl groups to the DNA molecule. Histone modification is the post translational modification (PTM) to histone which included methylation, phosphorylation and acetylation. The PTM can cause gene expression changes by alteration in chromatin structure and recruiting histone modifiers.

Our study focuses on analysing DNA methylation alterations in AD. We have planned to find out candidate genes associated with DNA methylation mechanism in AD. There were three components in our analysis. An integrated approach was followed which included finding out the candidate CpGs, differential gene expression analysis and functional enrichment analysis. We have planned to compare different datasets to find out similarity between up-regulated and down-regulated genes and determining the level of methylation. This meta analysis will help to identify commonalities between different kinds of experiments related to AD and finally to identify important markers for the regulatory processes involved in AD.

2. INTRODUCTION

Alzheimer's disease is a progressive-unremitting, neurodegenerative disorder that affects wide areas of brain including cerebral cortex area and hippocampus [1]. Alzheimer's disease causes destruction to nerve connections in brain, thus, making it burdensome to do regular things i.e moving around, feeding oneself, swallowing, remembering. Being a disease that destroys memory, it deteriorates mental functioning [1]. Abnormalities are primarily detected in the tissue of frontal and temporal lobes, and then to other areas such as neocortex at different rates for different individuals. In AD, the peptides of amyloid-β (Aβ) and tau protein aggregate, thus, facilitating the formation of plaques and tangles [2]. These plaques and tangles are histopathological hallmarks of AD, respectively. General improvements in lifestyle and medication have increased the life expectancy, substantially in the past few decades. However, the treatment for dementia remains symptomatic, despite the increase in costly trials, with no approval of a single drug or treatment strategy. As per duration, the average duration of AD is 8–10 years, preceded by preclinical stages (clinical symptomatic phases) that extend the duration for over two decades. Being the 6th driving reason for mortality in the United States, there has additionally been generous increment in the quantity of passing from AD by 71%, in this manner, increasing the commitment for understanding the etiologic and pathogenesis of the illness.[3].

Being a part of neurodegenerative diseases, Alzheimer's disease shares many attributes with Parkinson's disease and fronto-temporal dementias. One of the attribute being the complex interactions between environmental and genetic factors in both AD and PD. Prevalence of both the diseases is expected to increase, due to the global population ageing, contributing to the increase in social and economic burden on the society. Now, the question is whether Alzheimer's disease plays an imminent role in normal ageing or does it have a seperate process. The answer lies in the description of the epidemiology, molecular mechanisms that dominate the neurodegenerative processes in AD, as well as the diagnosis, screening and the developing prevention strategies in current management practices [4].

Alzheimer's illness lives among packed up proteinopathies, which is described by aggregation of amyloid β (Aβ) peptides as amyloid plaques and of the protein tau as neurofibrillary tangles [2]. In more straightforward words, AD is demonstrated by the nearness of extracellular amyloid-β (Aβ) plaques just as intracellular neurofibrillary tangles

(NFTs) brought about by protein tau. Age of amyloidogenic peptides by the successive cleavage of the amyloid antecedent protein (APP) by the layer-bound proteins, for example, β-secretase and γ-secretase, aggregates inside the cerebral extracellular space to shape insoluble Aβ plaques [2]. Other than β-and γ-secretase, APP additionally can be separated by α-secretase followed by γ-secretase. This results in creating solvent, non-amyloidogenic Aβ peptides.

Another protein associated in the causation of AD would be microtubule-associated protein tau (MAPT) that forms neurofibrillary tangles [5]. Hence, Aβ and tau pathology prove to be of clinical relevance in diagnosis and so its assessment is being facilitated by increasingly sensitive methods for the validation of therapeutic interventions. Identification of distinct suitable Aβ and tau species to be used as biomarkers in cerebrospinal fluid, blood, urine and saliva serves as a prerequisite for personalized medicine. Thus, determining the Aβ levels in the CSF accurately is essential by help of new technologies, such as Aβ PET imaging, aiding the estimation of prevalence and incidence of AD [4].

Application of methods of analytical epidemiology serves as the only way to understand 'genes versus environs' in Alzheimer's disease aetiology. Molding and the three-dimensional (3D) structure of the genomic material bridges gap between the genes and environment. This aids in improving our understanding on the etiology of AD [4]. Smoking, diabetes mellitus, physical inactivity, mid-life obesity, depression, mid-life hypertension and low educational attainment seem to be potential risk factors for Alzheimer's disease (**Error! Reference source not found.**).

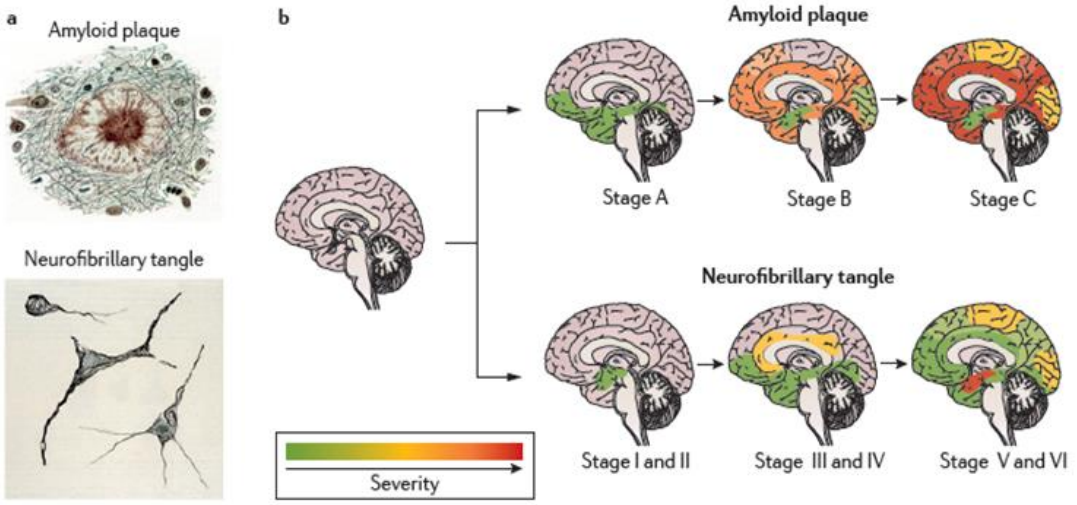


Figure 1 Pathological evolution of Alzheimer's disease.

Progression of Alzheimer's disease by dissemination of Amyloid plaques and neurofibrillary tangles in the brain

AD is characterized based on age, as late-beginning Alzheimer's sickness (LOAD) or beginning stage Alzheimer's ailment (EOAD). Burden happens in people more seasoned than age 65 while EOAD happens in people more youthful than age 65, as a rule at age 40 or 50. EOAD relates with hereditary change occuring in APP, PSEN1 (presenilin-1) or PSEN2 (presenilin-2) [4]. The changes influence the fundamental synergist segment of ϒ-secretase which prompts a patient getting determined to have AD. Both, PSEN1 and PSEN2 qualities code for proteins that make up ϒ-secretase synergist segment, subsequently, any changes could cause extreme outcomes

Mechanism

As mentioned earlier that most neurodegenerative diseases lie in the category of protein-pathies, which are further symbolised by the aggregation of misfolding of proteins. Although majority of the proteins have an inherent property to aggregate, notably when cellular clearance systems experience failure. Especially in the context of ageing, very few of the proteins form fibrillar aggregates. Presently, Aβ, tau and APOE (Apolipoprotein E) serve as three elements having significant evidence for being contributors of Alzheimer's disease. Synaptic loss and selective neuronal death are the neuro-pathological and neuro-chemical hallmarks of Alzheimer's disease, leading to a significant decrease in specific neurotransmitters and increase in the presence of abnormal proteinaceous deposits in neurons popularly known as neurofibrillary tangles(NFTs) in the extracellular space [1]. Besides

being a consequence of proteolytic cleavage, post-translational modifications also affects Aβ neurotoxicity, which is further attributable to specific types of Aβ and are prone to gathering into different assembly states [6]. Likewise Aβ, tau existing as multiple brain isoforms undergoes aggregation to host post-translational modifications, inclusive of acetylation as well as phosphorylation. Cellular and transgenic models exhibit impairment of multiple cellular functions by Aβ and tau, along with the crosstalk between these molecules particularly at synapse (**Error! Reference source not found.**).

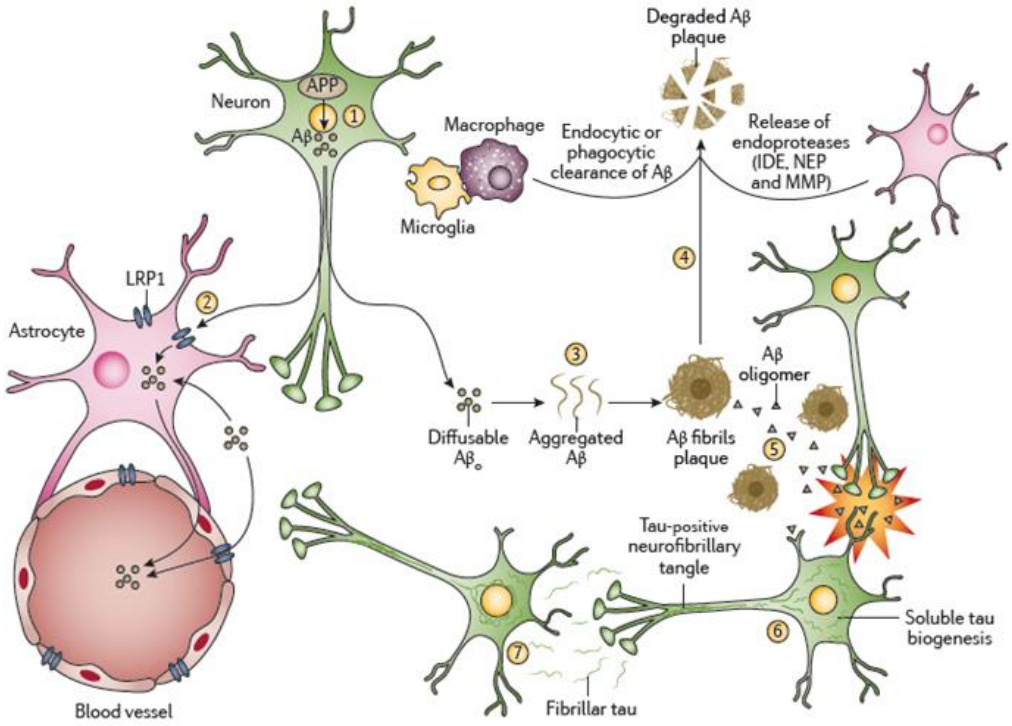


Figure 2 Pathways leading to the formation of plaques and tangles, laying foundation for the amyloid-β theory of Alzheimer's disease.

Plaque formation and tangles leading to Alzheimer's disease which are the hallmarks of alzheimer's disease.

Various investigations proposed that epigenetic factors additionally have extensive measure of effect on neurodegenerative interminable maladies including AD. Epigenetics, investigation of heritable changes in quality articulation, additionally includes guaranteeing the consistency of the basic DNA succession for example an adjustment in phenotype with no adjustment in the genotype [7]. Epigenetics contains four significant parts, for example, (I) heritability, the capacity of a separating cell to pass its own epigenetic imprints to its little girl cells; (ii) conservation of DNA grouping, in which the arrangement of nucleotides stays

unaltered; (iii) transcriptional guideline, wherein the epigenetic changes affect the translation of qualities in this manner, prompting changes in the phenotype of the cell; and the last one being (iv) steadiness, the protection of epigenetic alterations over time[8]. Some epigenetics factors that may influence AD would be DNA methylation, histone alterations and non-coding RNA.

DNA Methylation

DNA Methylation, being an epigenetic instrument is known to change quality articulation or cell phenotype in a way that would be heritable [9]. The most widely contemplated epigenetic adjustment is DNA methylation and changes of histone proteins. The procedure of DNA methylation incorporates expansion of a methyl bunch at a place that has a cytosine nucleotide that goes before guanines (CpG dinucleotides). All DNA methyltransferase (DNMT) proteins are thought of, among which DNMT1, DNMT3A and DNMT3B are best portrayed to catalyze the exchange of a methyl bunch from S-adenosylmethionine to DNA. An item shaped during this response, i.e S-adenosylhomocysteine, is on the other hand changed over back to S-adenosylmethionine by a chain of responses as it has a noticable impact in the one-carbon digestion cycle (Figure 3) [9]. Essentially, an acceleration in the measure of plasma homocysteine, a middle of the road in this cycle, extrapolates a higher hazard for improvement of dementia just as AD. In this manner, inferring that impact of DNA methylation and changes in one-carbon digestion contributes in AD pathogenesis. Following this, impacts on the capacity of DNA can be considered such to be actuation or suppression of the transcriptional action of a quality. Authoritatively, DNA methylation or adjustment is recognized by a decrease in articulation of a quality, yet some new proof proposes that quality articulation is reliant upon the impact of DNA methylation's setting inside the genome. Albeit, 5-methylcytosine (5mC) is by all accounts the most plenteous altered base in the mammalian genome, more current investigations delineate the distinguishing proof of extra adjusted bases, as N6-methyladenine (N6mA) just as 5-formylcytosine (5fC). Initially, 5fC was seen as a middle in a protein intervened DNA demethylation, can be a steady DNA alteration that influences quality articulation by modifying DNA twofold helix structure [10]. There is an immense variety in the predominance of the change alongside its impact on the quality articulation. In spite of the fact that there has been a great deal of exertion put into these spearheading contemplates,

however the information on these altered bases is as yet fragmented. Need of more research goes about as essential to additionally portray their circulation just as the capacity.

Another kind of change is post translational histone alterations, for example, methylation and acetylation of lysine buildups on histone tails [11]. The histone adjustments assume a vital job in influencing the quality articulation, principally by changing chromatin structure. DNA methylation that depicts a normal level of methylation over the whole genome is named as "worldwide" DNA methylation. Other than this, a normal percent methylation inside a particular quality is named as "quality explicit" DNA methylation [12]. In late DNA methylation concentrates in AD, worldwide DNA methylation is frequently gotten to through counter acting agent based techniques, for example, immunohistochemistry while, the quality explicit DNA methylation was investigated by the use of exhibit based strategies, for example Illumina's Infinium HumanMethylation450 BeadChip exhibit. Other than these strategies, some high-throughput methods that utilization bisulfite transformation are frequently utilized for the estimation of both worldwide and quality explicit DNA methylation [10].

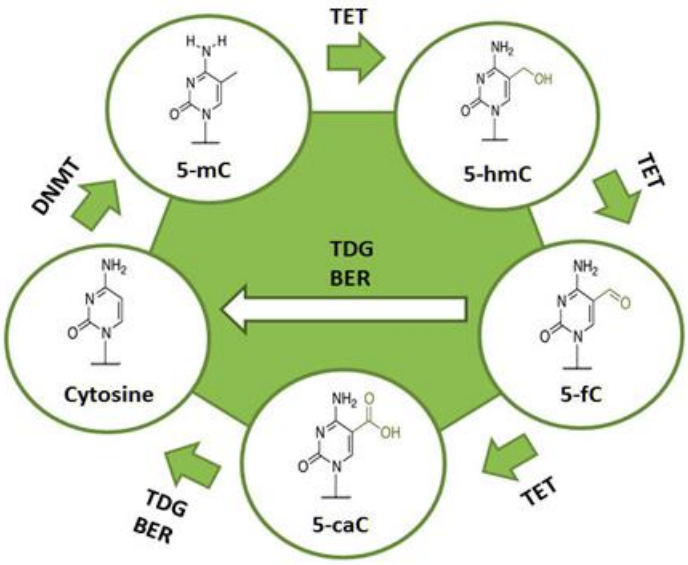


Figure 3 Representation of cytosine's methylation & demethylation processes.

The cycle depicts different modified forms of cytosine, along with the corresponding enzymes formed as intermediates that are responsible for each modifications taking place in DNA methylation.

Results from epidemiological examinations alongside clinical highlights of neurological issue furnishes with an epigenetic commitment to the etiology of separate ailments. Additionally, epigenetic adjustment are normally very much recorded regarding mental health, plastic changes, particularly in the instances of neurodegenerative maladies, for example, AD and PD. Aftereffects of the treatment of AD patients with inhibitors of histone deacetylases (HDAC), being key compound associated with histone acetylation, fills in as the most convincing proof on the job of epigenetics in AD [4]. In any case, the dysregulation of DNA methylation in neurodegenerative infection patients is likewise finely reported. A raised DNA methylation delineates a condition of tedious components in AD patients as appeared in ongoing confirmations.

Various investigations show that epigenetics may add to the pathology of AD, for example, the heterogeneous clinical introduction of AD understanding with fundamentally the same as indistinguishable hereditary foundations [13]. In 2009, one such examination was completed, the investigation included indistinguishable twins that were utilized to separate between the impacts of hereditary qualities and condition. Mastroeni et al. examined DNA methylation in a couple of monozygotic twins dissonant for AD [7]. Further, utilizing immunohistochemical methods for the identification of 5mC, it was discovered that the worldwide DNA methylation inside the unrivaled frontal gyrus and the foremost fleeting neocortex had essentially decreased in the twin managing AD when contrasted and the non-sick, neurologically ordinary twin. Strikingly enough, both of the twins were synthetic architects with practically comparable degrees of instruction, the main separating component would be that the AD twin worked seriously with pesticides yet the non-maniacal twin didn't. Subsequently, the investigation proposes that hereditary and natural variables can add to the advancement of AD, in less complex words i.e, the outcomes shows that other than hereditary transformations, ecological impacts may likewise influence the AD introduction. Despite the fact that, this investigation of indistinguishable twins discorant for AD is significantly liked to particular hereditary just as natural components, is incredibly uncommon. Some normal, investigates that include DNA methylation modifications mind in AD is frequently of random members conflicting for AD. Both, before and since the indistinguishable twin examination

in 2009, worldwide DNA methylation had been broke down in different locales by other for disconnected people conflicting for AD.

Mastroeni et al., as a follow-up to their undefined twin assessment used immunohistochemistry to watch DNA methylation in the entorhinal cortex zone of a model masses suffering with and without AD [8]. Like the finishes drawn from before undefined twin examination, it was assumed that the individuals suffering with AD had an essential decrease in overall DNA methylation when appeared differently in relation to the individuals suffering without AD. DNA methylation changes in AD are territory unequivocal, was exhibited by the immunoreactivity for 5mC which was not extraordinary in AD-influenced regions, for instance, the cerebellum. Disregarding the way that the outcomes of this resulting examination, when was gotten together with the delayed consequences of the indistinct twin assessment in 2009, might lead perusers to expect that the individuals suffering with AD have reduced degrees of DNA methylation in the common cortex when appeared differently in relation to the individuals without AD.

Coppieters et al. in his assessment showed that there was an expansion in the degrees of overall DNA methylation in the brain tissue tests that were taken up from the middle transient gyrus area, when the subjects of AD were diverged from the age-matched, emotionally standard controls. Other than Coppieters study, Lashley et al. in his assessment, saw that there was no meaningful differentiation in overall DNA methylation in the entorhinal cortex area of individuals suffering with and without AD [11]. Another examination drove by Phipps et al., provoked the affirmation of cell-type unequivocal DNA methylation. Immunohistochemistry was used to dismember 5mC and 5hmC in neural and glial cell types in this examination, usually, present in the inferior passing gyrus of human in AD cases similarly as age-facilitated controls. Closures drawn from it, recommended that extranuclear 5mC in neurofilament-stamped pyramidal neurons is vulnerable against the AD pathology, which very reduces in AD circumstances when differentiated and controls. Therefore, Phipps et al. study supports the hypothesis that DNA methylation changes in AD depend on the particular cell type that is analyzed.

Hippocampus

Cerebral decay alongside decay of the hippocampus is one of the different trademark pathologies of AD. Hippocampus is a locale that includes memory arrangement. Subsequently, nearness of decay in the hippocampus district can be especially extreme.

Accordingly, breaking down of the hippocampus as a cerebrum district by analysts delineates its critical effect on AD. Like the investigations referenced before, finishes of studies on DNA methylation inside the fleeting cortex and frontal cortex additionally impacts the finishes of studies that dissect DNA methylation modifications inside the hippocampus district. An examination led by Chouliaras et al. , noticed a sharp diminishing in the degrees of hippocampal DNA methylation in AD [10]. The diminishing was obvious when AD cases were when contrasted with subjective disconnected, age-coordinated controls. Other than for the end drawn, the creators additionally found that DNA methylation in a glial cell was tremendously particular in the CA1 just as CA3 subregions. Additionally, neuronal DNA methylation was further extraordinary just in the CA1 subregion, subsequently, proposing that there might be varieties in the cell type explicit adjustments in DNA methylation relying upon hippocam us.

Quality explicit DNA methylation modifications in AD

A few investigations have been done up till date, breaking down quality explicit DNA methylation in sub areas of the mind tissue of patients enduring with and without AD. Besides, reaching comparative inferences from the examinations that break down worldwide DNA methylation, makes these investigations uncertain. Be that as it may, furnishing with generally solid proof that methylation inside qualities, for example, APP, PSEN1, MAPT might be adjusted in AD makes us one stride nearer to the quantity of basic DNA methylation modifications in qualities that have been seen across examines. Recent investigations of quality explicit DNA methylation contrasts in AD mainly engaged upon qualities related with AD pathology, for example, APP, PSEN1, MAPT and apolipoprotein E (APOE) [6]. In any case, in Iwata et al. study pyrosequencing has been utilized to encourage the investigation of DNA methylation of different CpG destinations of AD-related qualities. The qualities were for the most part present in the second rate transient flap, the predominant parietal projection and the cerebellum in both, AD subjects just as the non-sick control subjects.

In this manner, a lot of distinction was found in DNA methylation profiles of the qualities APP, MAPT [6]. Besides, the creators clarified that these modifications in DNA methylation interprets, causing changes in quality articulation. This gives a potential evidence by which DNA methylation impacts the Alzheimer's malady phenotype in subjects considered.

Gene expression omnibus (GEO) is an international public repository that freely distributes microarray, next-generation sequencing, and other forms of high-throughput functional genomics data submitted by the research community. GEO was used for data collection [14].
For comparison of the datasets, “Bioinformatics and evolutionary genomics” tool was used and Venn diagram indicating common up-regulated and down-regulated genes was obtained.
Further the genes were shortlisted and Metascape tool was used for pathway enrichment analysis [15]. Metascape helps in understanding the common or unique pathways and also provides with giving insights into different protein-protein networks, cluster analysis, customised analysis for the multiple gene list, orthogonal target discoveries [16]. Cluster analysis was also done using Metascape.
The transcription factors with the genes were investigated and the tool used for finding the transcription factor was Transcription Factor Affinity Prediction (TRAP) Web tools developed at Max Planck Institute for molecular genetics [17].

3. MATERIALS AND METHODS

The integrative analysis approach was followed which included three componenets. The first component was finding out the candidate CpGs [18]. CpGs are the regions of DNA where a cytosine nucleotide is followed by a guanine nucleotide in the linear sequence of bases along its 5’-3’ direction. All the results were combined from previous studies and different tissues like peripheral blood, brain were considered for the analysis [19]. Different statstical criterias were considered for the analysis by combining results from multiple studies and this is known as meta analysis (**Error! Reference source not found.**).

Genes associated with those CpGs were considered. The second component of the analysis that was differential gene expression analysis of the affiliated genes [18]. Differential gene expression analysis was done using GEO2R (**Error! Reference source not found.**).

The third component was functional enrichment analysis [18]. This was done using g:Profiler [20]. g:Profiler is a web server for functional enrichment analysis (**Error! Reference source not found.**).

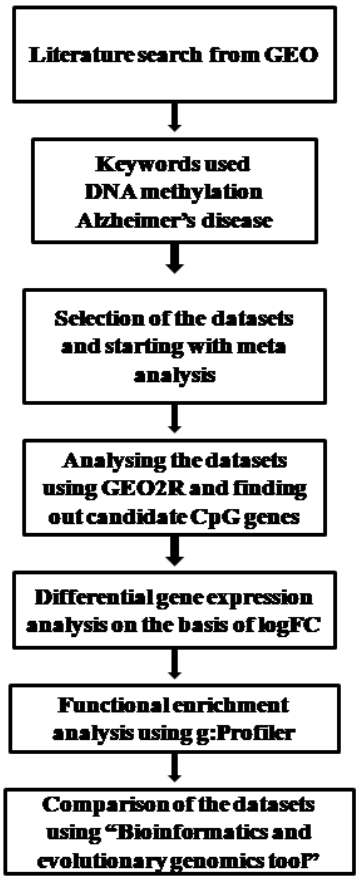


Figure 4 workflow of the analysis.

Workflow showing various steps such as data mining, meta analysis, differential gene expression analysis and functional enrichment analysis

3.1 LITERATURE SEARCH

Data mining comprised of choosing the most relevant datasets in order to proceed with the analysis. "DNA Methylation" and "Alzheimer's Disease" were used as keywords to obtain hits from GEO [14]. For further refining of results, the entry type and study type were specified using series and methylation profiling by array. A total of 10 datasets were found and 2 datasets were chosen for further analysis (**Error! Reference source not found.**).

organism	Type	Sample	Platform	Associated research publication
Homo sapiens	Methylation profiling by array	20	GPL8490	DNA methylation in hippocampus across Alzheimer Braak disease
Homo sapiens	Methylation profiling by array	29	GPL13534	Genome scale profiling of DNA methylation landscape

Figure 5 Datasets chosen for the analysis.

Table representing organism, type, platform, samples and associated research publication

3.2 ANALYSIS OF FIRST DATASET

The first dataset was GSE45775. The illumina DNA methylation beadChip was used to obtain DNA methylation profiles across approximately 27,000 CpGs. There were 20 samples which were divided into 2 groups: Control and Alzheimer's disease. The platform used for GSE45775 was GPL8490. The above dataset was analysed using GEO2R, which allows user to compare two or more groups within the sample-set. Samples were grouped into control and AD accordingly (**Error! Reference source not found.**).

Figure 6 grouping of samples for GSE45775.

Samples were grouped into control and AD and the TOP250 table was obtained

This analysis for GSE45775 dataset resulted into TOP250 table. The TOP250 table contained significant information which included ID, adjusted P-value, P-value, t value, B

value, logFC (logarithmic fold change of gene expression), gene symbol and CpG methylation.

3.3 ANALYSIS OF SECOND DATASET

The second dataset was GSE57360 which had 29 samples. It included DNA methylation analysis of brain samples from patient suffering from Down's syndrome, Dementia of Lewy bodies or Alzheimer's and Parkinson's disease using the Infinium DNA methylation BeadChip platform. The platform used was GPL13534 (PMID: 26784972). The dataset was analysed using GEO2R. Here only Alzheimer's disease and control groups were used for the analysis. There were 5 samples which were grouped into control group and 7 samples were grouped into Alzheimer's disease group (**Error! Reference source not found.**).

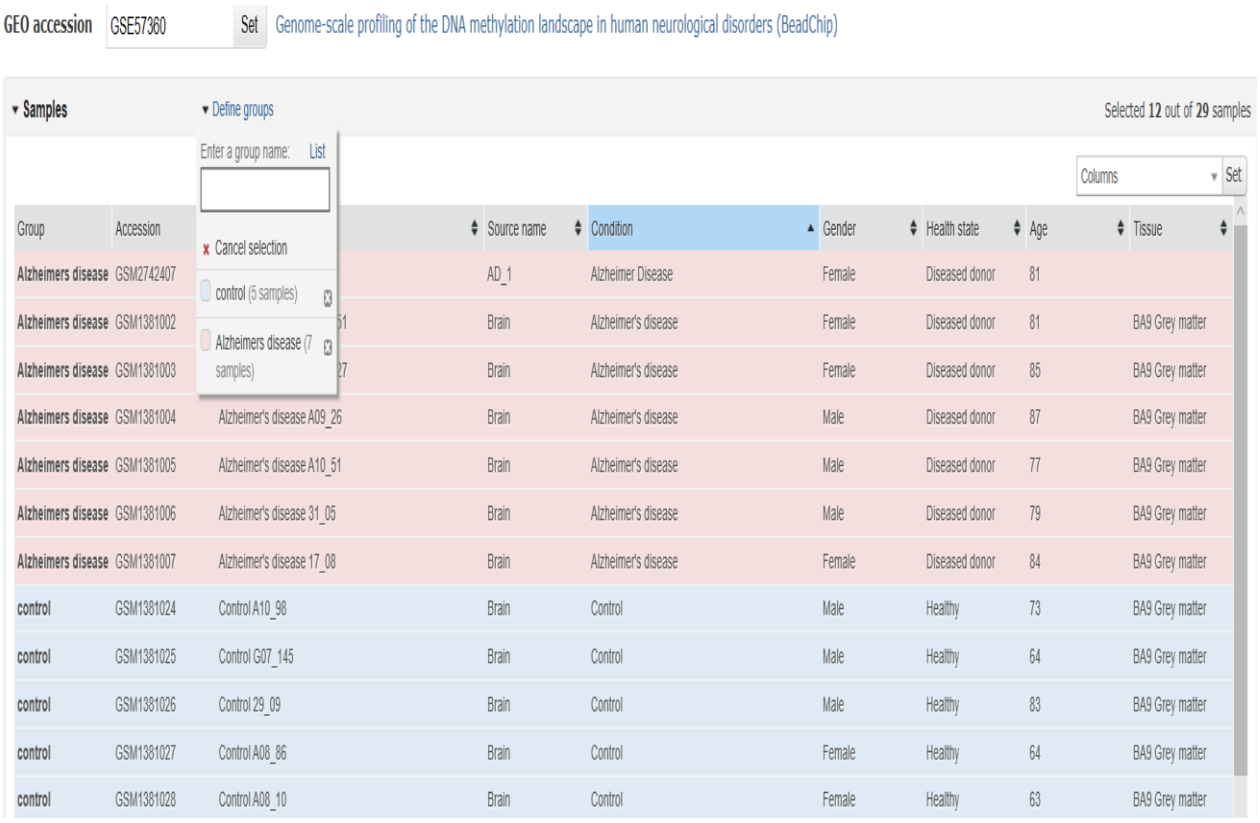


Figure 7 Grouping of samples for GSE57360.

Grouping of samples into control and Alzheimer's disease and the TOP250 table was obtained

In this dataset, the TOP250 table contained ID, logFC, F statistics, methylation, adjusted P-value, P-value, t value, B value and chromosome information, Although, it did not contain gene symbol list. The TOP250 table had the gene ID's which were used to retrieve the gene

symbols. These gene ID's were uploaded on DAVIDGO (Database for Annotation, Visualisation and Integrated Discovery) which provided us with the gene symbols [21]. DAVIDGO provides with comprehensive set of functional and annotated tools for investigators to understand biological meaning behind large gene list.

3.4 CATEGORISATION OF GENES

The gene symbols were available for both the datasets (GSE4775 and GSE57360) [22]. All the genes with P-value <0.01 for both the datasets were considered for the analysis and only those genes that were CpG methylated were taken into consideration [23]. The positive and negative logFC values distinguished up-regulated and down-regulated genes respectively as obtained in the TOP250 table. Further, the genes were categorized as up-regulated and down-regulated based on the logFC values.

3.5 FUNCTIONAL ENRICHMENT ANALYSIS AND VENN DIAGRAM

The up-regulated and down-regulated genes for both the datasets were uploaded on g: Profiler for functional enrichment analysis. It maps genes to known functional sources. g:Profiler provides entire gene ontology which includes cellular components, molecular function and biological processes [24].

The up-regulated and down-regulated genes of the chosen datasets were represented using Venn diagram. Venn diagram was obtained using "Bioinformatics and evolutionary genomics tool". This tool calculates the intersections of list of elements. It will generate a output indicating which elements are in each intersection and are unique to a certain list.

3.6 PATHWAY ENRICHMENT ANALYSIS

After obtaining the Venn diagram, common up-regulated and down-regulated genes were investigated and further these were shortlisted on the basis of a statistical parameter that is the log P-Value. A total of 262 common up-regulated and 92 down-regulated genes were taken into consideration for further pathway enrichment analysis [25].

For pathway enrichment analysis, Metascape was used [26]. Metascape is a tool used for obtaining heatmap, protein-protein interaction networks and enriched ontology clusters. Different parameters were used for obtaining the clusters which includes cluster ID and P-Value [27]. MCODE algorithm was applied to find out protein-protein network and to identify neighbourhood where proteins are densely connected [28].

3.7 CLUSTER ANALYSIS

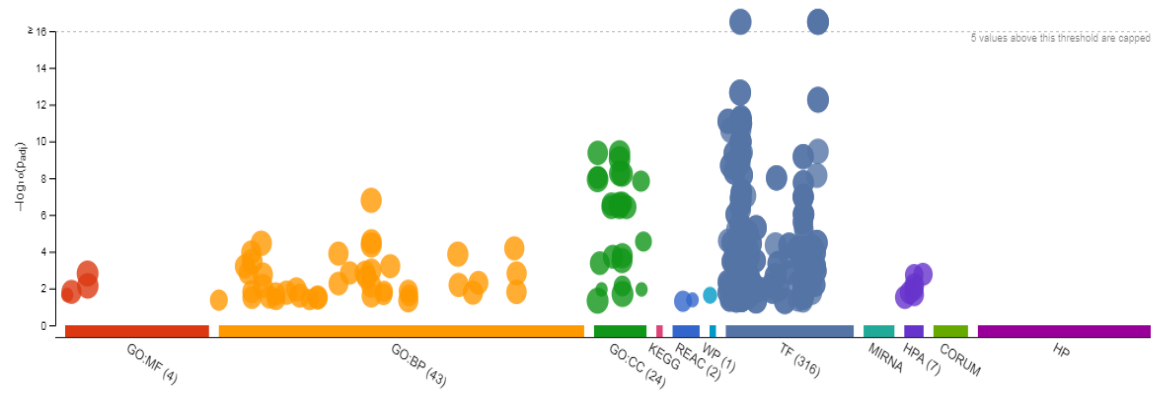
For customised cluster analysis, different parameters were checked which includes pathway and process enrichment where Minimum overlap was of 3, P-Value cut-off was of 0.01 and Minimum enrichment was of 1.5. For protein-protein enrichment, the criteria's used were databases (BIOGRID + In_Web_IM + OmniPath), Minimum network size of 3 and Maximum network size was of 500 [29].

3.8 FINDING TRANSCRIPTION FACTORS

After cluster analysis, transcription factors associated the genes were investigated. For finding out the transcription factors associated with the genes were determined using Transcription Factor Affinity Prediction (TRAP) web tools. These tools were developed at Max Planck Institute for molecular genetics to predict TF affinities. By uploading the up-regulated and down-regulated genes from cluster analysis into the PASTAA database, transcription factors, target genes, association score and P-Value [30].

4. RESULTS AND DISCUSSION

For the dataset GSE45775, out of 20,009 genes, 12,717 genes were up-regulated and 7,292 genes were down-regulated. For the dataset GSE57360, out of 21,270 genes, 10,971 genes were up-regulated and 10,299 genes were down-regulated. The up-regulated and down-regulated genes for both the datasets were uploaded on g:Profiler for functional enrichment analysis (**Error! Reference source not found.**, **Error! Reference source not found.**, **Error! Reference source not found.**, **Error! Reference source not found.**).

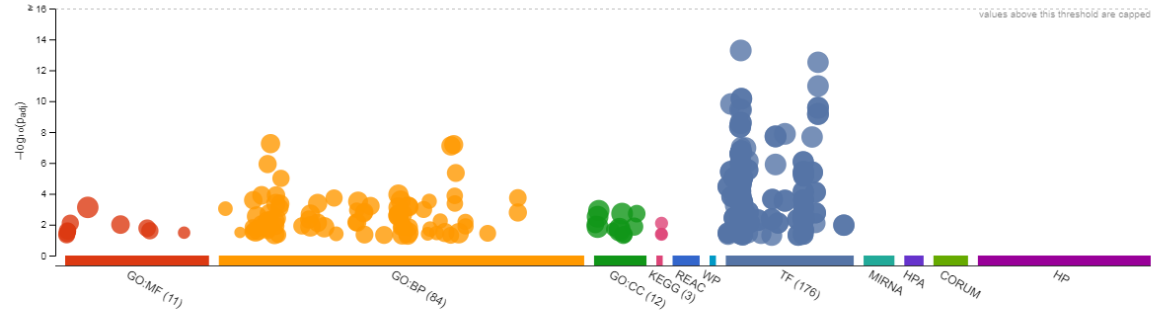


version	e98_eg45_p14_ce5b097
date	11/25/2019, 2:11:16 PM
organism	hsapiens

g:Profiler

Figure 8 Results of g:Profiler for up-regulated genes of the dataset GSE45775

The above represents cellular components, molecular functions, biological processes, KEGG, REAC, WP, TF, MIRNA, HPA and CORUM



version	e98_eg45_p14_ce5b097
date	11/19/2019, 4:07:13 PM
organism	hsapiens

g:Profiler

Figure 9 Results of g:Profiler for down-regulated genes of the dataset GSE45775

The above represents cellular components, molecular functions, biological processes, KEGG, REAC, WP, TF, MIRNA, HPA and CORUM

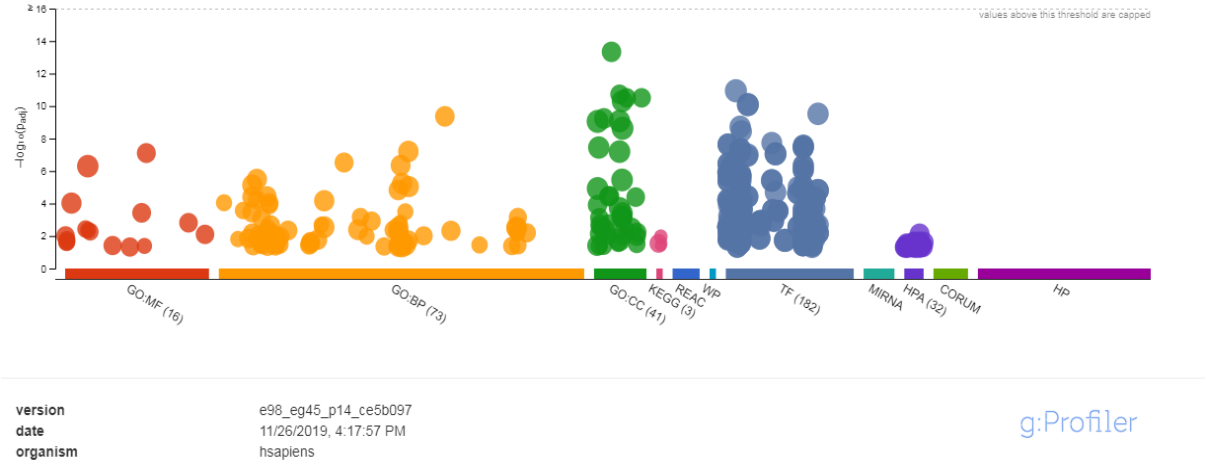


Figure 10 Results of g:Profiler for up-regulated genes of the dataset GSE57360

The above represents cellular components, molecular functions, biological processes, KEGG, REAC, WP, TF, MIRNA, HPA and CORUM

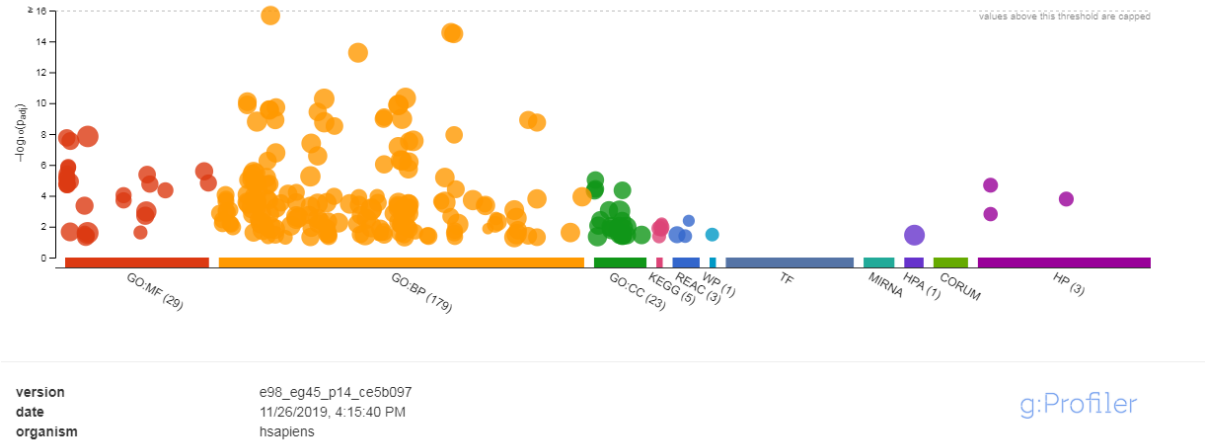


Figure 11 Results of g:Profiler for down-regulated genes of the dataset GSE57360

The above represents cellular components, molecular functions, biological processes, KEGG, REAC, WP, TF, MIRNA, HPA and CORUM

The up-regulated and down-regulated genes for both the datasets were plotted using Bioinformatics and evolutionary genomics tool" (**Error! Reference source not found.**).

The up-regulated and down-regulated genes were sorted on the basis of logFc values for both the datasets. The genes with positive logFC values were up-regulated and with negative

logFC values were down-regulated. There were 262 genes which were common for up-regulated genes of both the datasets (GSE45775 and GSE57360) and 92 genes which were common for down-regulated genes of both the datasets (GSE45775 and GSE57360). The below table represent the top 5 up-regulated and down-regulated genes.

Gene Name	Type	Function
RPL36A	Up-regulated	Anion binding
WAS	Up-regulated	Protein binding
BTK	Up-regulated	Catalytic binding
HSD17B10	Up-regulated	Small molecule binding
APEX	Up-regulated	Carbohydrate derivate binding
TGIF2LY	Down-regulated	Protein binding
PCDH11Y	Down-regulated	DNA binding TF activity
PRKY	Down-regulated	DNA binding TF activity
MAMLD1	Down-regulated	DNA binding TF activity
EIF2S3	Down-regulated	DNA binding TF activity

Table 1 Top 5 up-regulated and down-regulated genes

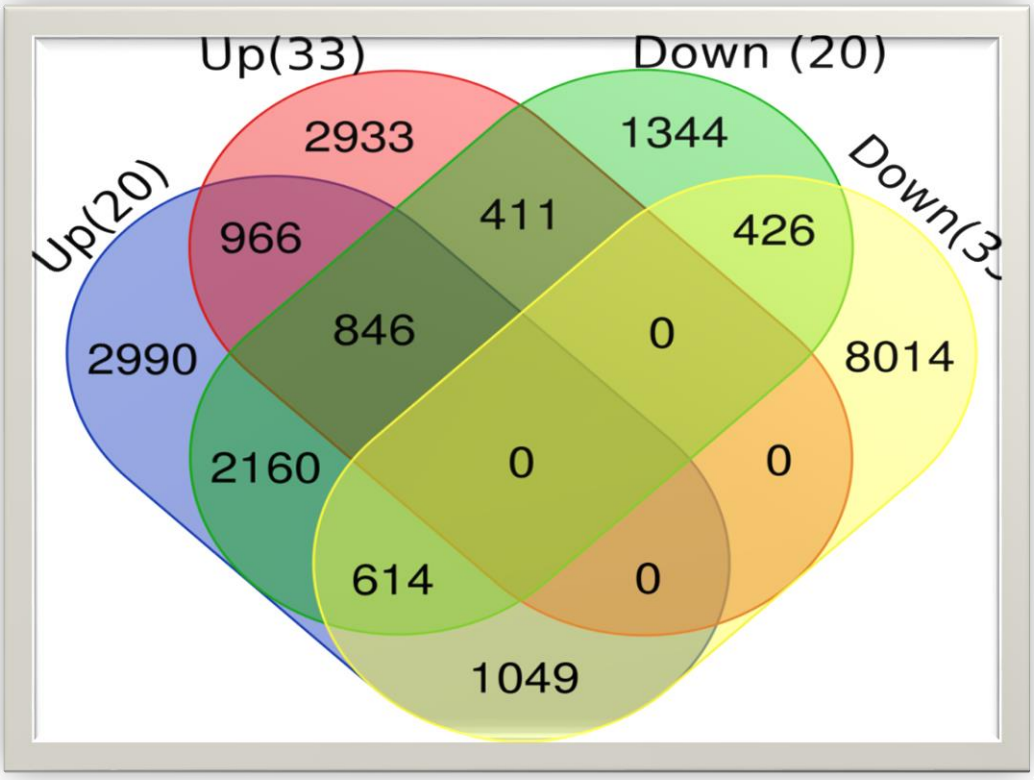


Figure 12 Venn diagram showing all possible overlaps.

All possible overlaps of up-regulated and down-regulated genes of both the datasets (GSE45775 and GSE57360)

The common up-regulated and down-regulated were now narrowed down using a statistical criterion i.e. the log P-value. These up-regulated and down-regulated were uploaded on Metascape. Metascape was used for cluster analysis and provided with enriched ontology clusters, clusters by colour ID, clusters coloured by P-Value, protein-protein interaction network and PPI MCODE network.

Metascape provides with annotation and enrichment results. Annotation results includes Input_ID, Gene_ID, TAX_ID, homologues, gene_Symbol, description, biological_process, kinase_class, protein_function, subcellular_location, Drug and hallmark_gene_sets. Enrichment results include group_ID, category, Term, description, logP and symbols.

Enriched ontology clusters Type 1

Enriched ontology clusters for both up-regulated and down-regulated genes were obtained. Firstly all the statistically enriched terms which includes canonical pathways, KEGG/GO terms, hall_mark gene sets, P-Value, enrichment factors were calculated and used for filtering the results. Significantly essential terms were then hierarchically clubbed into a tree based on κ-statistical similarities among multiple genes. 0.3 κ score was applied to mould the tree into the cluster. A subset of representative terms which include P-value, GO terms, biological terms were considered and converted into a network (Figure 13, Figure 14).

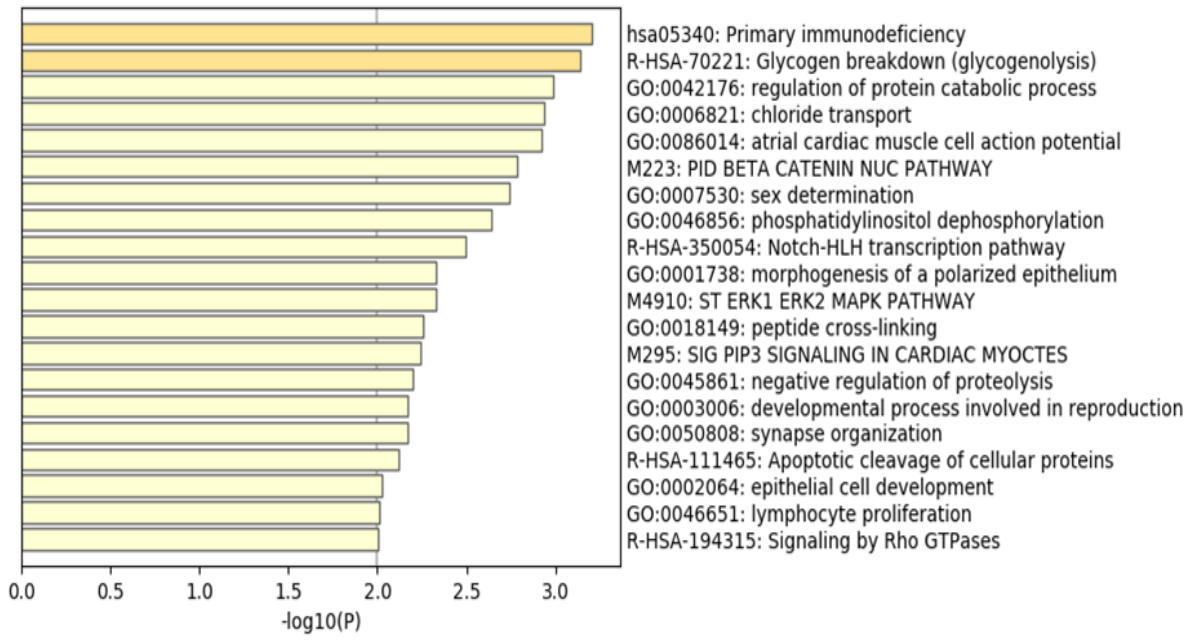


Figure 13 Enriched ontology clusters for common up-regulated genes for both the datasets

Clusters representing collaborative results from different representative terms like P-Value, gene symbols, κ statistics similarities

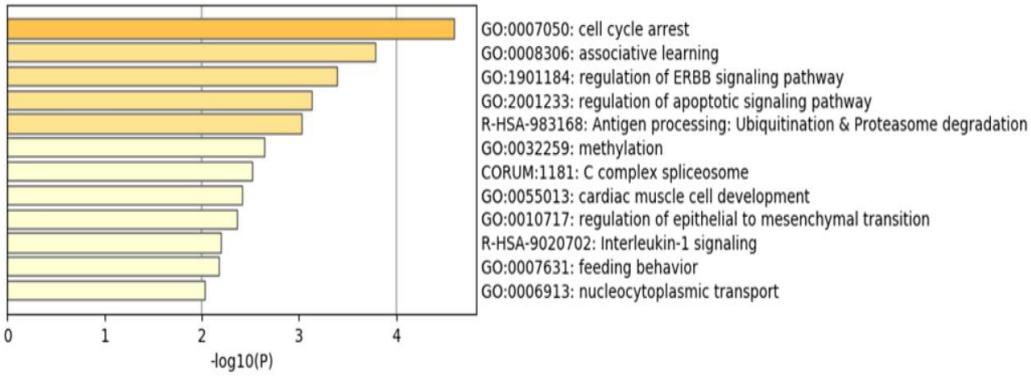


Figure 14 Enriched ontology clusters for common down-regulated genes for both the datasets

Clusters representing collaborative results from different representative terms like P-Value, gene symbols, κ statistics similarities

Enriched ontology clusters Type 2

Enriched ontology clusters were obtained by clustering by colour ID. Each term was represented by circle node and the size is proportional to number of input genes and each colour represents cluster identity. All the node of same colour belonged to the similar cluster and terms with similarity score less than 0.03 were by edge and the thickness of the edge represents the similarity score. The thickness of the edges represents similarity and more the thickness, more the similarity. For common up-regulated and down-regulated genes enriched ontology cluster by colour ID were obtained (Figure 15, Figure 16).

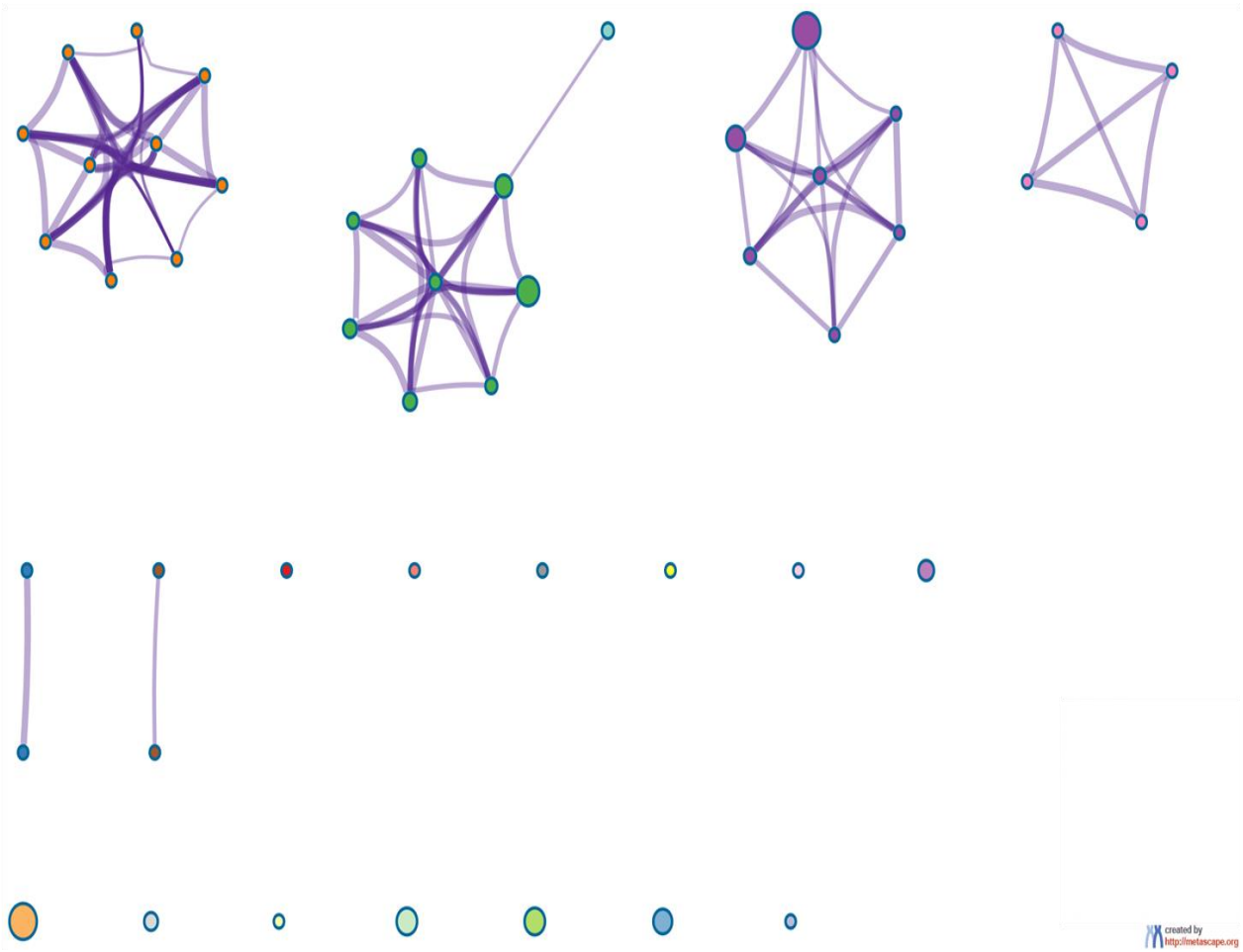


Figure 15 Enriched ontology clusters: coloured by cluster ID for common up-regulated genes.

Clustering genes on the basis of cluster ID and the nodes having same belonged to the similar cluster. Different colours represent different functions like cell cycle regulation, methylation, and regulation of signalling pathway, C-complex spliceosome and cardiac muscle development.

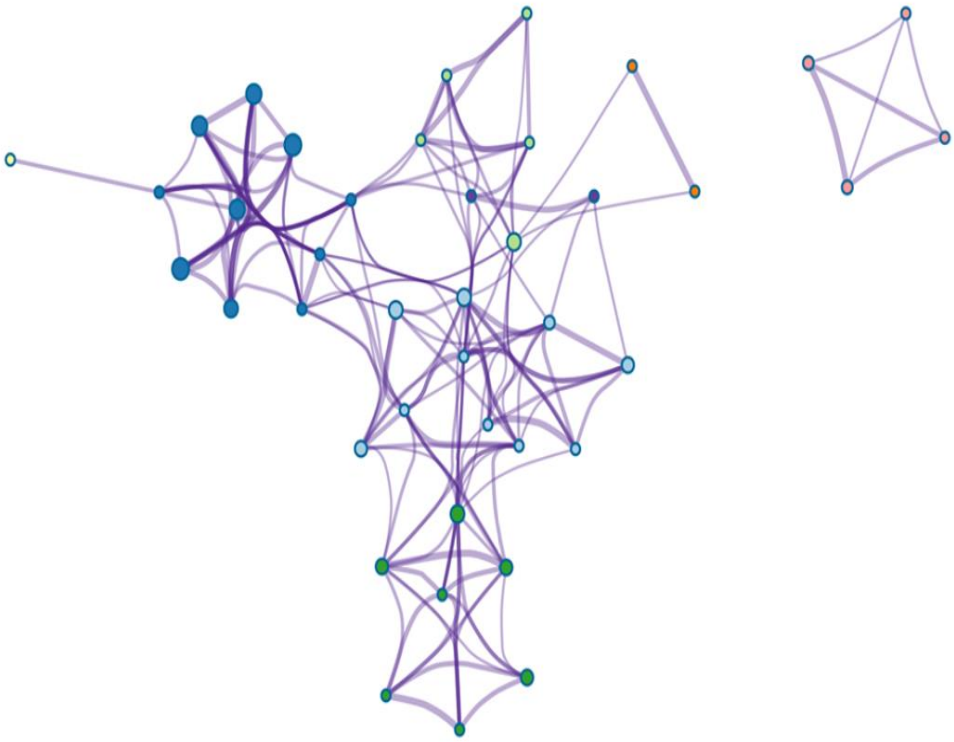


Figure 16 Enriched ontology clusters: coloured by cluster ID for common up-regulated genes.

Clustering genes on the basis of cluster ID and the nodes having same belonged to the similar cluster. Different colours represent different functions like cell cycle regulation, methylation, and regulation of signalling pathway, C-complex spliceosome and cardiac muscle development

Enriched ontology clusters Type 3

Enriched ontology clusters on the basis of P-Value. The enrichment network obtained has nodes coloured on the basis of P-Value and darker the colour, more significant the node. The enriched ontology network coloured by P-Value was obtained for both up-regulated as well as the down-regulated genes (Figure 17, Figure 18).

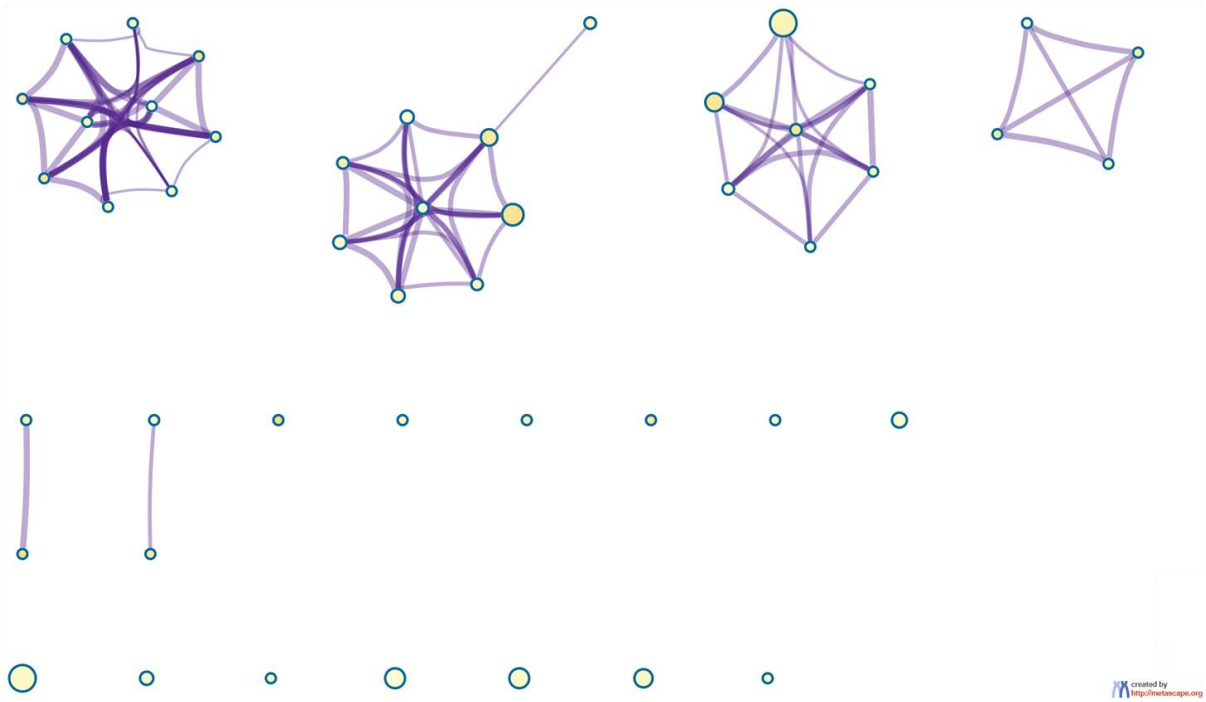


Figure 17 Enriched ontology network coloured by P-Value for up-regulated genes

The network contains the nodes coloured by P-Value and darker nodes represents significant nodes and the above network is for common up-regulated genes.

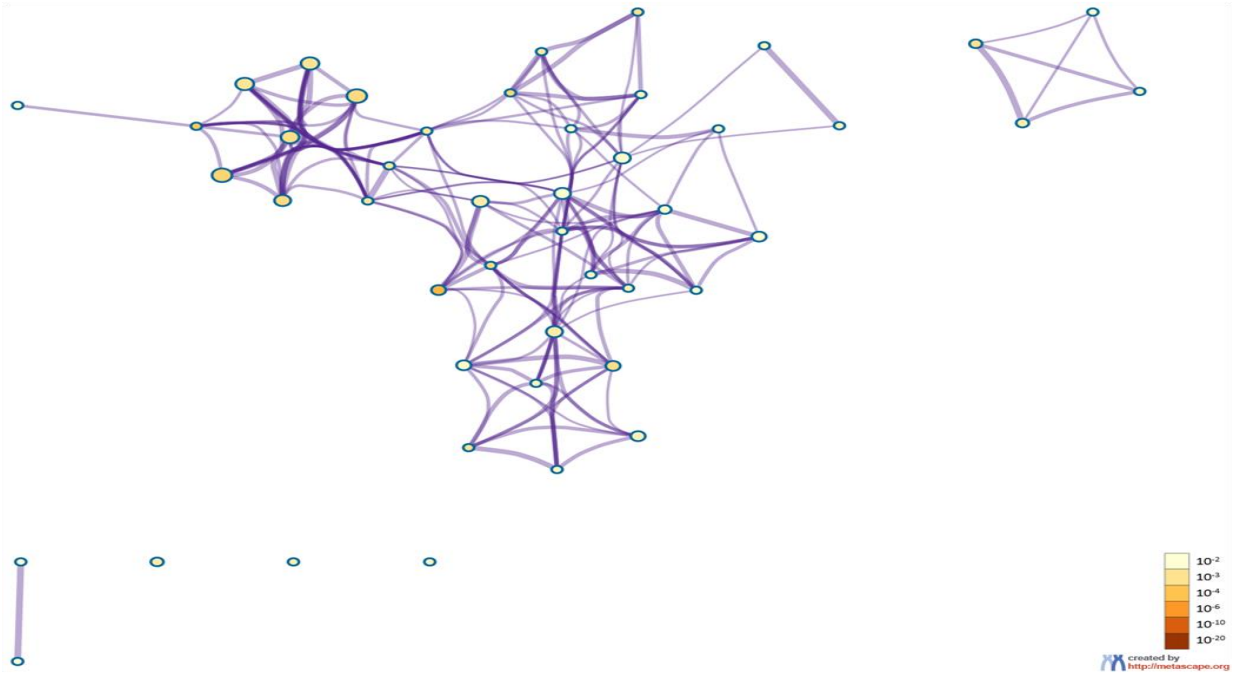


Figure 18 Enriched ontology network by P-Value

The network contains the nodes coloured by P-Value and darker nodes represents significant nodes and the above network is for common up-regulated genes.

PPI network using MCODE

Protein-Protein network (PPI) for common up-regulated and down-regulated genes was obtained. MCODE algorithm was applied to obtain the PPI network. It was used to identify neighbourhoods where the proteins were densely connected (Figure 19, Figure 20).

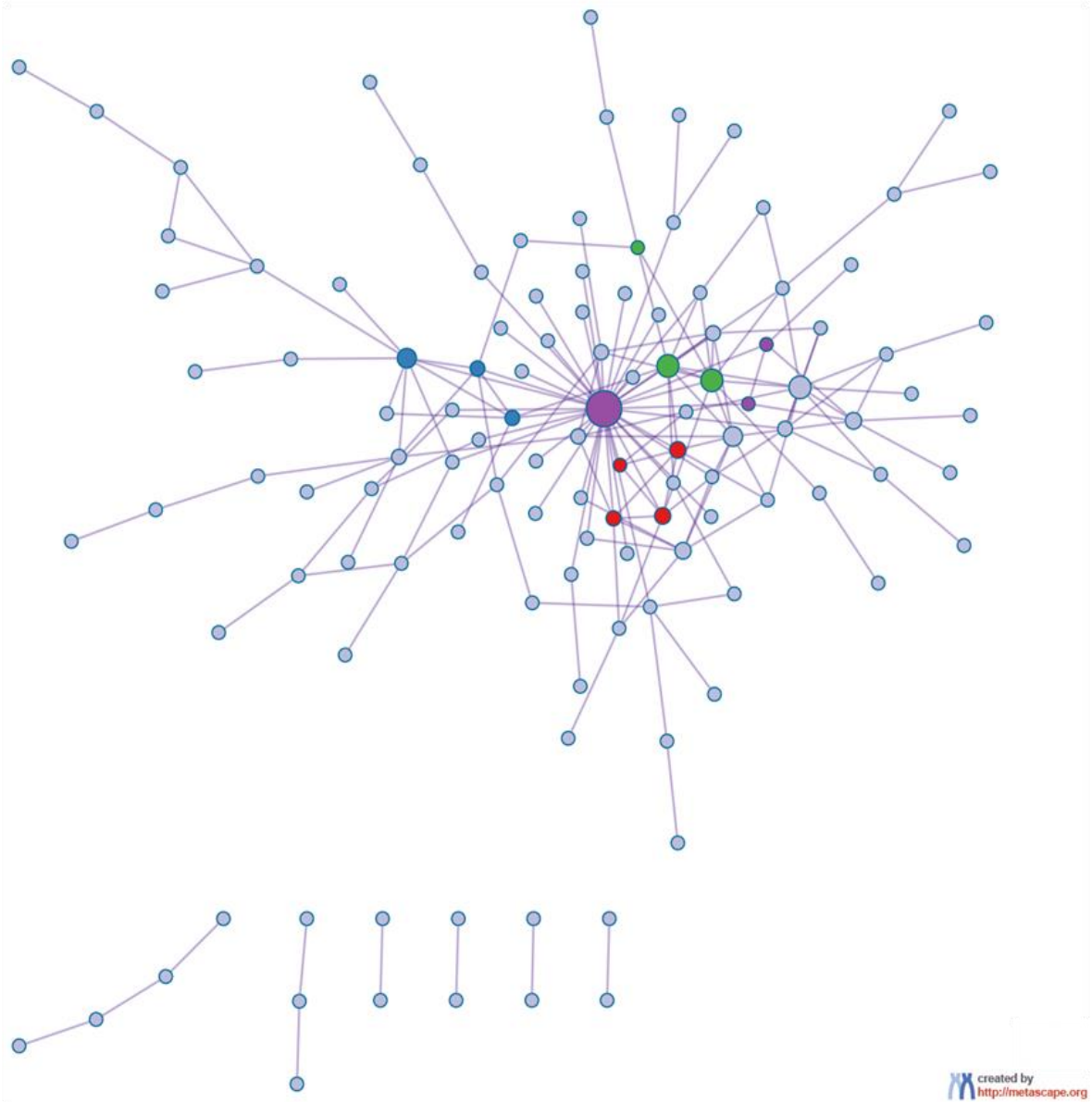


Figure 19 PPI network for common up-regulated genes.

MCODE algorithm was applied to obtain the PPI network for up-regulated genes

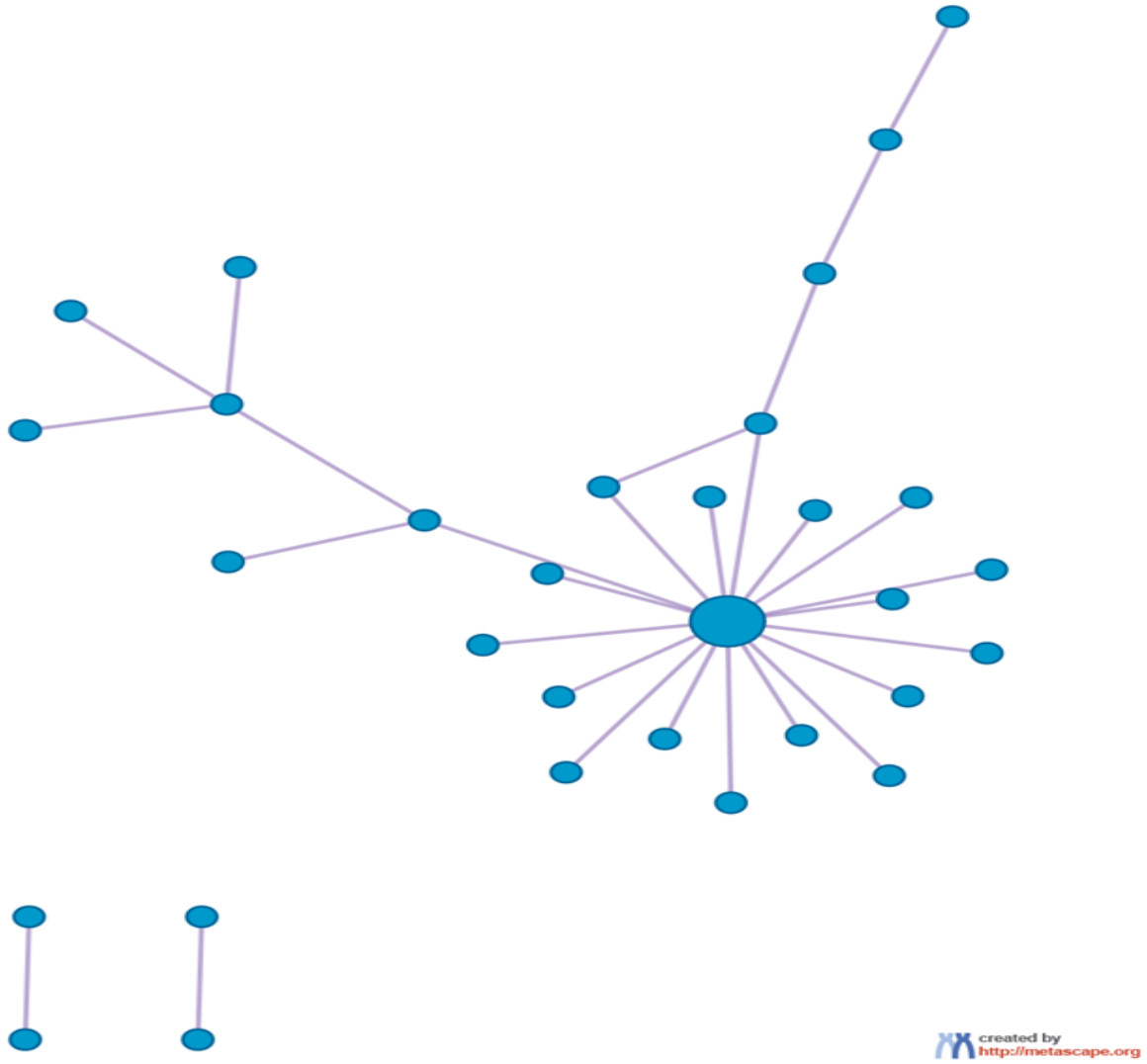


Figure 20 PPI network for down-regulated genes

MCODE algorithm was applied to obtain the PPI network for up-regulated genes

Gene extracted from cluster analysis

Genes which played an important role in DNA methylation were also investigated and presented in the form of a network. The up-regulated genes included IDHG3, STK26, HSD17B10, PDHA1, UBQLN2, H2AW, H2BU1, BTK, FLNA, BCAP31, TAF1 and PGRMC1 (Figure 21). The down-regulated genes included OTC, SLC25A5, CTPS2, PABPC5, ACSL4, IDH3G, YBX1, HNRNPH2, RBMX, PBQ1P and UPF3B. The PPI network follows MCODE algorithm (Figure 22).

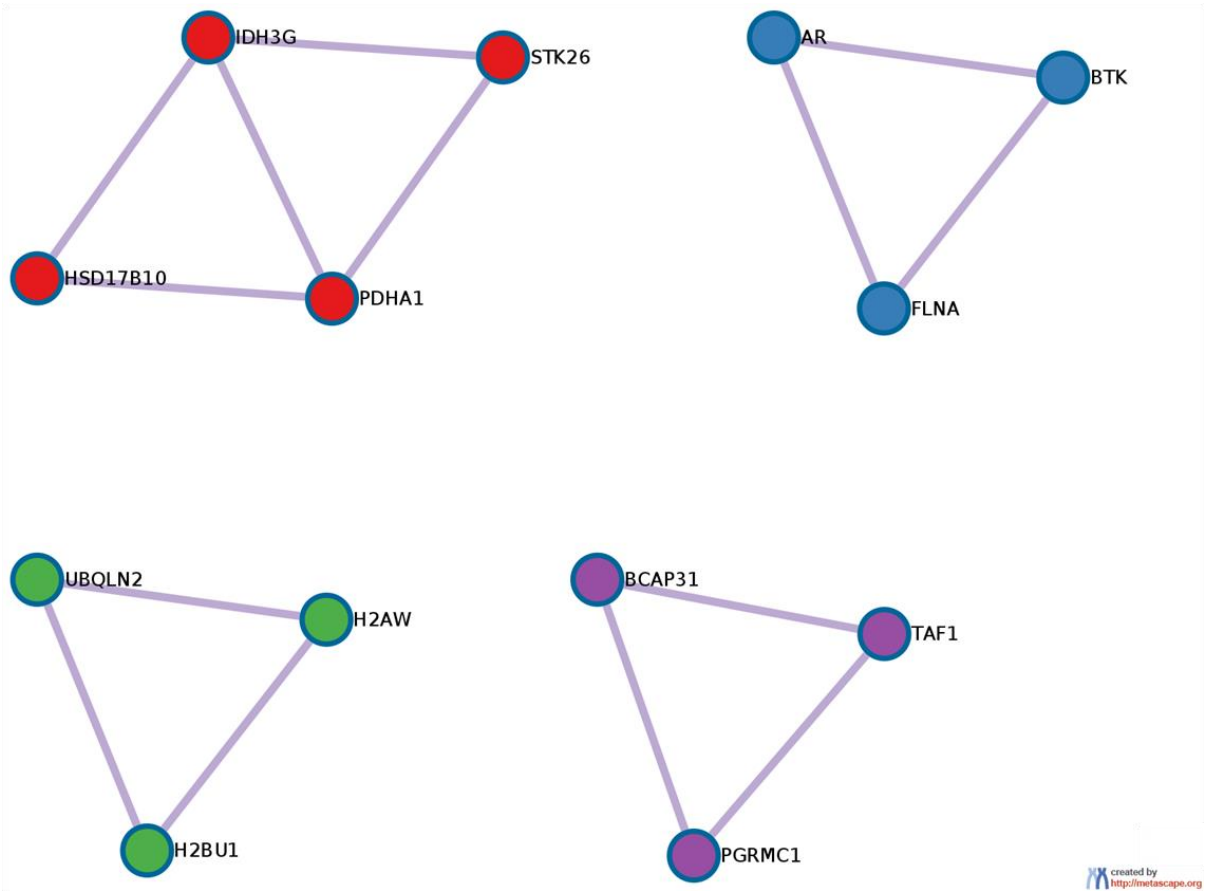


Figure 21 PPI MCODE network for up-regulated genes.

Up-regulated genes responsible for DNA Methylation

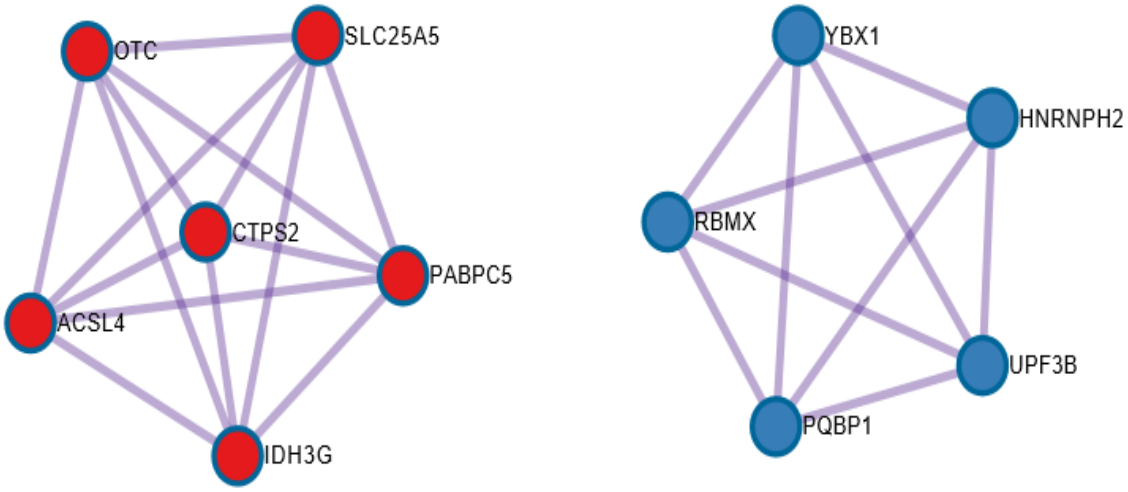


Figure 21 PPI MCODE network for down-regulated genes.

Down-regulated genes responsible for DNA methylation

Customised cluster analysis was performed using Metascape where analysis performed included pathway and process enrichment where Minimum overlap was of 3, P-Value cut-off was of 0.01 and Minimum enrichment was of 1.5. For protein-protein enrichment, the criteria's used were databases (BIOGRID + In_Web_IM + OmniPath), Minimum network size of 3 and Maximum network size was of 500. The results found were similar to that of the metascape analysis. Similar up-regulated and down-regulated genes were found from the analysis.

Transcription factors were calculated for common up-regulated and down-regulated genes. Genes were uploaded on TRAP web tools. Following table contains Gene_symbol, Gene_ID, Transcription_factor and Affinity_Score for up-regulated genes (Table 2).

Gene_Symbol	Gene_ID	Transcription_Factor	Affinity_Score
BCAP31	ENSG00000185825	SUH_01, ABI4_01, HSF1_Q6, NRF2_01, & MINI19_B	0.4616
FLNA	ENSG00000196924	HAC1_Q2, BZIP911_02, CP2_01, BZIP910_02, & FACBCB_Q2	0.2894
PGRMC1	ENSG00000101856	GCM_Q2, RAV1_01, TCF11MAFG_01, DR4_Q2, & ATMYB84_01	0.2795
PDHA1	ENSG00000131828	DR4_Q2, E47_02, ERR1_Q2, GATA1_05, & VBP_01	0.2352
UBQLN2	ENSG00000188021	ZAP1_01, NFKB_C, AGL3_02, AGL3_03, & HSF_02	0.05030
BTK	ENSG00000010671	MAF_Q6_01, NFE2_01, NRF2_Q4, ZID_01, & GCN4_C	0.005426
TAF1	ENSG00000147133	SRF_01, SRF_C, HAP1_B, USF_Q6, & HBP1A_Q2	0.0002727

Table 2 Transcription factors associated with up-regulated genes. It contains Gene_Symbol, Gene_ID, Transcription_Factor and Affinity_Score.

The following table contains transcription factors for down-regulated genes (Table 3).

Gene_Symbol	Gene_ID	Transcription_Factor	Affinity_Score
ACSL4	ENSG00000068366	ZAP1_Q6, ELK1_01, MIG1_01, CETS1P54_02, & CF2II_02	0.187
CTPS2	ENSG00000047230	FTZ_01, BACH2_01, E2_01, AP4_01, & E2_Q6	0.07876
OTC	ENSG00000036473	HSF_05, HSF_05, ROM_Q2, VJUN_01, & POLY_C	0.002835
SLC25A5	ENSG00000005022	PPARA_01, NERF_Q2, ETS1_B, PPARG_03, & E74A_01	0.001124
RBMX	ENSG00000147274	MYBPH3_01, ABF_C, VMYB_01, MCM1_01, & HAIRY_01	0.001061
IDH3G	ENSG00000067829	HBPA1_Q6_01, OSBZ8_Q6, ATF6_01, FACBALL_Q2, & MTF1_Q4	0.0005767
PABPC5	ENSG00000174740	GAGAFACTOR_Q6, SMAD3_Q6, HSF_05, ABF1_01, & ABF_C	0.0001146

Table 3 Transcription factors associated with up-regulated genes. It contains Gene_Symbol, Gene_ID, Transcription_Factor and Affinity_Score.

5. CONCLUSION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that slowly destroys memory and thinking skills and, eventually, the ability to carry out the simplest tasks. Accumulation of amyloid-β and neurofibrillary tangles are the major hallmarks of AD. DNA methylation is a mechanism of transferring methyl group to DNA, thereby modifying the functions of genes. The integrative analysis approach was followed which included three componenets. The analysis of the datasets were done through GEO. We were able to find out genes associated with CpGs and then on the basis of P-Value and logFC value we were able to sort the up-regulated and down-regulaed genes. For functional enrichment analysis, g:Profiler was used and we were able to retrieve the cellular, molecular and biological function of the up-regulated and down-regulated genes. The up-regulated and down-regulated

genes were represented using venn diagram. Venn diagram was obtained using "Bioinformatics and evolutionary genomics tool". The common up-regulated and down-regulated genes will be further investigated for pathway enrichment analysis. For pathway enrichment analysis, Metascape tool was used. Enriched ontology clusters for common up-regulated and down-regulated genes were obtained. Enriched ontology clusters: coloured by cluster ID and cluster by P-value were also obtained. PPI network using different criteria's for both common up-regulated and down-regulated genes was obtained. Customised cluster analysis, different parameters were checked which includes pathway and process enrichment. After investigating different networks, we were able to conclude 13 up-regulated genes and 11 down-regulated genes responsible for DNA methylation alterations in Alzheimer's disease.

These up-regulated and down-regulated genes were further uploaded on TRAP web tools and transcription factors associated with those genes were found.

6. REFERENCES

[1] C. L. Masters, R. Bateman, K. Blennow, C. C. Rowe, R. A. Sperling, and J. L. Cummings, "Alzheimer's disease," *Nat. Publ. Gr.*, vol. 1, pp. 1–18, 2015.

[2] J. C. Polanco, C. Li, L. G. Bodea, R. Martinez-Marmol, F. A. Meunier, and J. Götz, "Amyloid-β and tau complexity - Towards improved biomarkers and targeted therapies," *Nat. Rev. Neurol.*, vol. 14, no. 1, pp. 22–40, 2018.

[3] P. Hof and J. Morrison, "Quantitative analysis of a vulnerable subset of pyramidal neurons in Alzheimer's disease," *J. Comp. Neurol.*, vol. 301, no. 1, pp. 55–64, 1990.

[4] P. L. De Jager *et al.*, "Alzheimer's disease: Early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci," *Nat. Neurosci.*, vol. 17, no. 9, pp. 1156–1163, 2014.

[5] C. E. Humphries *et al.*, "Integrated whole transcriptome and DNA methylation analysis identifies gene networks specific to late-onset Alzheimer's disease," *J. Alzheimer's Dis.*, vol. 44, no. 3, pp. 977–987, 2015.

[6] A. Iwata *et al.*, "Altered CpG methylation in sporadic alzheimer's disease is associated with APP and MAPT dysregulation," *Hum. Mol. Genet.*, vol. 23, no. 3, pp. 648–656, 2014.

[7] D. Mastroeni, A. Grover, E. Delvaux, C. Whiteside, P. D. Coleman, and J. Rogers, "Epigenetic changes in Alzheimer's disease: Decrements in DNA methylation," *Neurobiol. Aging*, vol. 31, no. 12, pp. 2025–2037, 2010.

[8] D. Mastroeni, A. McKee, A. Grover, J. Rogers, and P. D. Coleman, “Epigenetic differences in cortical neurons from a pair of monozygotic twins discordant for Alzheimer’s disease,” *PLoS One*, vol. 4, no. 8, pp. 1–6, 2009.

[9] K. X. Wen *et al.*, “The role of DNA methylation and histone modifications in neurodegenerative diseases: A systematic review,” *PLoS One*, vol. 11, no. 12, pp. 1–31, 2016.

[10] L. Chouliaras *et al.*, “Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer’s disease patients,” *Neurobiol. Aging*, vol. 34, no. 9, pp. 2091–2099, 2013.

[11] J. S. Rao, V. L. Keleshian, S. Klein, and S. I. Rapoport, “Epigenetic modifications in frontal cortex from Alzheimer’s disease and bipolar disorder patients,” *Transl. Psychiatry*, vol. 2, no. 7, pp. e132-7, 2012.

[12] A. S. Yokoyama, J. C. Rutledge, and V. Medici, “DNA methylation alterations in Alzheimer’s disease,” *Environ. Epigenetics*, vol. 3, no. 2, pp. 1–11, 2017.

[13] A. R. Smith *et al.*, “Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer’s disease,” *Clin. Epigenetics*, vol. 11, no. 1, pp. 1–13, 2019.

[14] R. Edgar, “Gene Expression Omnibus: NCBI gene expression and hybridization array data repository,” *Nucleic Acids Res.*, vol. 30, no. 1, pp. 207–210, 2002.

[15] M. Kori, B. Aydln, S. Unal, K. Y. Arga, and D. Kazan, “Metabolic Biomarkers and Neurodegeneration: A Pathway Enrichment Analysis of Alzheimer’s Disease, Parkinson’s Disease, and Amyotrophic Lateral Sclerosis,” *Omi. A J. Integr. Biol.*, vol. 20, no. 11, pp. 645–661, 2016.

[16] Y. Zhou *et al.*, “Metascape provides a biologist-oriented resource for the analysis of systems-level datasets,” *Nat. Commun.*, vol. 10, no. 1, 2019.

[17] H. G. Roider, T. Manke, S. O’keeffe, M. Vingron, and S. A. Haas, “PASTAA: Identifying transcription factors associated with sets of co-regulated genes,” *Bioinformatics*, vol. 25, no. 4, pp. 435–442, 2009.

[18] Q. Guo *et al.*, “Using Integrative Analysis of DNA Methylation and Gene Expression Data in Multiple Tissue Types to Prioritize Candidate Genes for Drug Development in Obesity,” *Front. Genet.*, vol. 9, no. December, pp. 1–7, 2018.

[19] P. P. Panigrahi and T. R. Singh, “Computational studies on Alzheimer’s disease associated pathways and regulatory patterns using microarray gene expression and network data: Revealed association with aging and other diseases,” *J. Theor. Biol.*, vol. 334, pp. 109–121, 2013.

[20] J. Reimand *et al.*, “Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap,” *Nat. Protoc.*, vol. 14, no. 2, pp. 482–517, 2019.

[21] D. W. Huang *et al.*, “Open Access Clifford Lane ¶ and Richard A Lempicki * The DAVID Gene Functional Classification Tool : a novel biological module-centric algorithm to functionally analyze large gene lists,” vol. 8, no. 9, 2007.

[22] *et al.*, "Genes Selection Comparative Study in Microarray Data Analysis," *Bioinformation*, vol. 9, no. 20, pp. 1019–1022, 2013.

[23] N. Asati, A. Mishra, A. Shukla, and T. R. Singh, "Gene Expression Studies to Identify Significant Genes in AR, MTOR, MAPK Pathways and their Overlapping Regulatory Role in Prostate Cancer," *J. Integr. Bioinform.*, vol. 16, no. 3, pp. 1–12, 2019.

[24] J. Reimand, M. Kull, H. Peterson, J. Hansen, and J. Vilo, "g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments," *Nucleic Acids Res.*, vol. 35, no. Web Server issue, pp. W193–W200, Jul. 2007.

[25] M. Paczkowska *et al.*, "Integrative pathway enrichment analysis of multivariate omics data," *Nat. Commun.*, vol. 11, no. 1, pp. 1–16, 2020.

[26] W. Li *et al.*, "Comprehensive bioinformatics analysis of acquired progesterone resistance in endometrial cancer cell line," *J. Transl. Med.*, vol. 17, no. 1, pp. 1–17, 2019.

[27] J. Zhou, X. Hui, Y. Mao, and L. Fan, "Identification of novel genes associated with a poor prognosis in pancreatic ductal adenocarcinoma via a bioinformatics analysis," *Biosci. Rep.*, vol. 39, no. 8, pp. 1–16, 2019.

[28] K. D. Pruitt, C. W. Hogue, M. Groll, C. Hartmann, R. Huber, and S. Fields, "Mcode," *Nucleic Acids Res.*, vol. 29, no. 1, pp. 137–140, 2001.

[29] A. Chatr-Aryamontri *et al.*, "The BioGRID interaction database: 2015 update," *Nucleic Acids Res.*, vol. 43, no. D1, pp. D470–D478, 2015.

[30] G. F. Liu, G. Y. Ruan, M. M. Huang, L. L. Chen, and P. M. Sun, "Genome-wide DNA copy number profiling and bioinformatics analysis of ovarian cancer reveals key genes and pathways associated with distinct invasive/migratory capabilities," *Aging (Albany. NY).*, vol. 12, no. 1, pp. 178–192, 2020.

7. APPENDIX

Annotations file for common up-regulated genes

Gene Symbol	Description	Protein Function (Protein Atlas)	Subcellular Location (Protein Atlas)
IFITM3	interferon induced transmembrane protein 3	Transporters/Accessory Factors Involved in Transport	
PRSS41	serine protease 41	Enzymes/Peptidases/Serine-type peptidases	
HSD17B10	hydroxysteroid 17-beta dehydrogenase 10	Enzymes/ENZYME proteins/Oxidoreductases	
APEX2	apurinic/apyrimidinic endodeoxyribonuclease 2	Enzymes/ENZYME proteins/Lyases	Nucleoli fibrillar center;Nucleus;Vesicles(Supported)
PGK1	phosphoglycerate kinase 1	Enzymes/ENZYME proteins/Transferases	
MTM1	myotubularin 1	Enzymes/ENZYME proteins/Hydrolases	Plasma membrane(Supported)
CHST7	carbohydrate sulfotransferase 7	Enzymes/ENZYME proteins/Transferases	
HTR2C	5-hydroxytryptamine receptor 2C	G-protein coupled receptors	
TCEAL2	transcription elongation factor A like 2		Cytosol;Nuclear speckles(Approved)
AMMECR1	Alport syndrome, mental retardation		
TSC22D3	TSC22 domain family member 3	Transcription factors/Basic domains	Cytosol;Golgi apparatus;Nuclear speckles(Approved)
RENBP	renin binding protein	Enzymes/ENZYME proteins/Isomerase	
GATA	GATA binding	Transcription	

1	protein 1	factors/Zinc-coordinating DNA-binding domains	
STS	steroid sulfatase	Enzymes/ENZYME proteins/Hydrolases	
FAM9A	family with sequence similarity 9 member A		
VBP1	VHL binding protein 1		Cytosol;Vesicles(Supported)
MOSPD1	motile sperm domain containing 1		Nucleoplasm;Vesicles(Approved)
PNMA3	PNMA family member 3		
SYTL4	synaptotagmin like 4		Cytosol;Microtubule organizing center;Nucleoplasm(Approved)
SSX8P	SSX family member 8, pseudogene		
MTMR1	myotubularin related protein 1	Enzymes/ENZYME proteins/Hydrolases	Microtubule organizing center(Approved)
AR	androgen receptor	Nuclear receptors;Transcription factors/Zinc-coordinating DNA-binding domains	Mitochondria(Approved)
ZCCHC12	zinc finger CCHC-type containing 12		Nuclear bodies(Approved)
CDX4	caudal type homeobox 4	Transcription factors/Helix-turn-helix domains	
MAGEH1	MAGE family member H1		Nucleoli(Approved)
MAGED1	MAGE family member D1		Cytosol(Approved)
GPR101	G protein-coupled receptor 101	G-protein coupled receptors/GPCRs excl olfactory receptors	

TFE3	transcription factor binding to IGHM enhancer 3	Transcription factors/Basic domains	Nucleoplasm(Supported)Cytosol(Approved)
IL2RG	interleukin 2 receptor subunit gamma	CD markers	Vesicles(Approved)
CLCN4	chloride voltage-gated channel 4	Transporters/Electrochemical Potential-driven transporters	Vesicles(Approved)
BEND2	BEN domain containing 2		
CDKL5	cyclin dependent kinase like 5	Enzymes/{ENZYME proteins/Transferases,Kinases/CMGC Ser/Thr protein kinases}	Nucleoplasm(Enhanced)
CBLL2	Cbl proto-oncogene like 2	Enzymes/ENZYME proteins/Transferases	
BCAP31	B cell receptor associated protein 31	Transporters/Primary Active Transporters	Endoplasmic reticulum(Enhanced)
NXF3	nuclear RNA export factor 3		Nucleoplasm(Supported)
BMP15	bone morphogenetic protein 15		
HMGN5	high mobility group nucleosome binding domain 5		Nucleus(Supported)
SLC25A43	solute carrier family 25 member 43	Transporters/Electrochemical Potential-driven transporters	Nucleus(Approved)
HPRT1	hypoxanthine phosphoribosyltransferase 1	Enzymes/ENZYME proteins/Transferases	Cytosol(Supported)
ARHGEF9	Cdc42 guanine nucleotide exchange factor		Cytosol(Supported)

	2	dases/Metallopeptidases}	
GYG2	glycogenin 2	Enzymes/ENZYME proteins/Transferases	Cytosol;Nucleoli;Nucleus(Approved)
PDHA1	pyruvate dehydrogenase E1 alpha 1 subunit	Citric acid cycle related proteins;Enzymes/ENZYME proteins/Oxidoreductases	Mitochondria(Supported)
FLNA	filamin A	Transporters/Accessory Factors Involved in Transport	Actin filaments;Cytosol;Plasma membrane(Enhanced)
TSR2	TSR2 ribosome maturation factor		Cell Junctions;Cytosol;Nucleoli;Nucleus(Approved)
BTK	Bruton tyrosine kinase	Enzymes/{ENZYME proteins/Transferases,Kinases/Tyr protein kinases}	Vesicles(Supported)
RTL3	retrotransposon Gag like 3		Golgi apparatus;Nucleoplasm;Plasma membrane(Approved)
ZNF157	zinc finger protein 157	Transcription factors/Zinc-coordinating DNA-binding domains	
ARAF	A-Raf proto-oncogene, serine/threonine kinase	Enzymes/{ENZYME proteins/Transferases,Kinases/TKL Ser/Thr protein kinases}	Nucleoli fibrillar center;Nucleus(Approved)
EMD	emerin		Nuclear membrane(Enhanced)Endoplasmic reticulum(Supported)
EFNB1	ephrin B1		Plasma membrane(Supported)
KLHL34	kelch like family member 34		
ELF4	E74 like ETS transcription factor 4	Transcription factors/Helix-turn-helix domains	Nuclear bodies(Supported)Nucleoplasm(Approved)
ANOS1	anosmin 1		
P2RY	pyrimidinergic	G-protein coupled	

4	receptor P2Y4	receptors/{Adenosine and adenine nucleotide receptors,GPCRs excl olfactory receptors}	
PAGE 4	PAGE family member 4		Golgi apparatus;Vesicles(Approved)
PNPLA4	patatin like phospholipase domain containing 4	Enzymes/ENZYME proteins/Hydrolases	
OPHN1	oligophrenin 1		Nucleus;Plasma membrane(Approved)
ESX1	ESX homeobox 1	Transcription factors/Helix-turn-helix domains	Nuclear speckles(Supported)
RTL8C	retrotransposon Gag like 8C		
BGN	biglycan		Endoplasmic reticulum;Golgi apparatus(Approved)
FAM120C	family with sequence similarity 120C		Nuclear speckles(Approved)
NKAPP1	NFKB activating protein pseudogene 1		
ZRSR2	zinc finger CCCH-type, RNA binding motif and serine/arginine rich 2		
MAGEA8	MAGE family member A8		
TCEAL7	transcription elongation factor A like 7		Nucleus(Supported)
MAGEB4	MAGE family member B4		
EIF1A	eukaryotic		Cytosol(Approved)

SRPX	sushi repeat containing protein X-linked		
EBP	EBP cholestenol delta-isomerase	Enzymes/ENZYME proteins/Isomerase	Endoplasmic reticulum(Supported)
MAP7D2	MAP7 domain containing 2		Cytosol;Nucleoplasm(Approved)
STK26	serine/threonine kinase 26	Enzymes/{ENZYME proteins/Transferases,Kinases/STE Ser/Thr protein kinases}	Golgi apparatus(Supported)Centrosome;Cytosol;Nucleus(Approved)
LUZP4	leucine zipper protein 4		
OCRL	OCRL inositol polyphosphate-5-phosphatase	Enzymes/ENZYME proteins/Hydrolases	Cytosol;Microtubules(Approved)
GPC4	glypican 4		Vesicles(Approved)
NSDHL	NAD(P) dependent steroid dehydrogenase-like	Enzymes/ENZYME proteins/Oxidoreductases	Endoplasmic reticulum;Lipid droplets(Enhanced)
CETN2	centrin 2		
RAP2C	RAP2C, member of RAS oncogene family		
WDR44	WD repeat domain 44		Golgi apparatus(Supported)
PIR	pirin	Enzymes/ENZYME proteins/Oxidoreductases	Cytosol(Supported)
SMIM10L2A	small integral membrane protein 10 like 2A		
NR0B1	nuclear receptor subfamily 0 group B member 1	Nuclear receptors;Transcription factors/Zinc-coordinating DNA-binding domains	Nuclear speckles(Enhanced)Microtubule organizing center;Vesicles(Supported)
MSN	moesin		Plasma membrane(Enhanced)

Annotations file for common down-regulated genes

Gene Symbol	Description	Protein Function	Subcellular Location
FBXL19-AS1	FBXL19 antisense RNA 1		
BTBD6	BTB domain containing 6		Nucleoplasm(Approved)
RASSF1	Ras association domain family member 1	RAS pathway related proteins	
ERBB2	erb-b2 receptor tyrosine kinase 2	CD markers;Enzyme	Cytosol;Plasma membrane(Supported)
DAB2IP	DAB2 interacting protein		
SUSD1	sushi domain containing 1		Nucleoplasm(Approved)
TSPYL5	TSPY like 5		Cytosol;Golgi apparatus(Approved)
PLD6	phospholipase D family member 6		
ENTPD3	ectonucleoside triphosphate diphosphohydrolase 3	Enzymes/ENZYME proteins/Hydrolases	Cytosol(Approved)
STRADA	STE20 related adaptor alpha	Enzymes/Kinases/STE Ser/Thr protein kinases	Cytosol;Nucleoplasm(Supported)
OR2L	olfactory	G-protein coupled	

13	receptor family 2 subfamily L member 13	receptors	
SMYD4	SET and MYND domain containing 4		Cytosol;Golgi apparatus;Nucleoplasm;Vesicles(Approved)
PHLDA1	pleckstrin homology like domain family A member 1		Nucleoli(Supported)
GLB1	galactosidase beta 1	Enzymes/ENZYME proteins/Hydrolases	Golgi apparatus;Vesicles(Supported)
NEBL	nebulette		
PLPP2	phospholipid phosphatase 2	Enzymes/ENZYME proteins/Hydrolases	Mitochondria(Approved)
TAMM41	TAM41 mitochondrial translocator assembly	Enzymes/ENZYME proteins/Transferases	Cytosol(Approved)
NAV1	neuron navigator 1		Microtubules(Approved)
CIDEB	cell death inducing DFFA like effector b		
VCY	variable charge Y-linked		
VASN	vasorin		Mitochondria;Nucleoli;Nucleus(Approved)
DHX8	DEAH-box	Enzymes/ENZYME	Nuclear

TDG	thymine DNA glycosylase	Enzymes/ENZYME proteins/Hydrolases	Nucleoplasm(Supported)
RRAGC	Ras related GTP binding C		Nucleus;Vesicles(Supported)
FBXO2	F-box protein 2		Cytosol;Nucleoplasm(Approved)
RPL31	ribosomal protein L31	Ribosomal proteins	
SNRPB	small nuclear ribonucleoprotein polypeptides B and B1		Nucleoplasm(Supported)
NIPSNAP3B	nipsnap homolog 3B		
FCGRT	Fc fragment of IgG receptor and transporter		
UBE3C	ubiquitin protein ligase E3C	Enzymes/ENZYME proteins/Transferases	Mitochondria;Nucleoplasm;Plasma membrane(Approved)
None	None	None	None
GNMT	glycine N-methyltransferase	Enzymes/ENZYME proteins/Transferases	Cytosol(Approved)
TFDP1	transcription factor Dp-1	Transcription factors/Helix-turn-helix domains	Cytosol;Nucleoplasm(Supported)
CRIM1	cysteine rich transmembrane BMP regulator 1		
TRAF	TNF receptor		Nucleoplasm(Approved)

NOL10	nucleolar protein 10		Nucleoli(Supported)
APP	amyloid beta precursor protein	Transporters/Transporter channels and pores	Golgi apparatus;Vesicles(Approved)
LEXM	lymphocyte expansion molecule		
None	None	None	None
None	None	None	None
SMIM14	small integral membrane protein 14		
FAM214B	family with sequence similarity 214 member B		Nucleoli fibrillar center;Nucleus(Approved)
CNTN1	contactin 1		
CBLB	Cbl proto-oncogene B	Enzymes/ENZYME proteins/Transferases	Cytosol(Supported)Nucleoplasm(Approved)
STX2	syntaxin 2		Cytosol;Nucleus(Approved)
CGGBP1	CGG triplet repeat binding protein 1		Nucleus(Enhanced)
C15orf39	chromosome 15 open reading frame 39		Cytosol(Enhanced)
FAM98A	family with sequence similarity 98 member A		Vesicles(Supported)

TMEM126B	transmembrane protein 126B		Mitochondria(Supported)
PNN	pinin, desmosome associated protein	Transporters/Primary Active Transporters	Nuclear speckles(Enhanced)
SIAH2	siah E3 ubiquitin protein ligase 2	Enzymes/ENZYME proteins/Transferases	Nucleoplasm;Vesicles(Supported)
HIKESHI	heat shock protein nuclear import factor hikeshi		
GTF2E2	general transcription factor IIE subunit 2		Nucleoplasm(Enhanced)Cytosol(Supported)
MUC1	mucin 1, cell surface associated	CD markers	
RNF121	ring finger protein 121		Golgi apparatus;Nucleus;Vesicles(Approved)
ALG8	ALG8 alpha-1,3-glucosyltransferase	Enzymes/ENZYME proteins/Transferases	Nucleoplasm(Approved)
RPF2	ribosome production factor 2 homolog		Nucleoli(Supported)Nucleus(Approved)
TFAP2A	transcription factor AP-2 alpha	Transcription factors/Basic domains	Nucleoplasm(Enhanced)

13	receptor family 2 subfamily L member 13	receptors	
SMYD4	SET and MYND domain containing 4		Cytosol;Golgi apparatus;Nucleoplasm;Vesicles(Approved)
PHLDA1	pleckstrin homology like domain family A member 1		Nucleoli(Supported)
GLB1	galactosidase beta 1	Enzymes/ENZYME proteins/Hydrolases	Golgi apparatus;Vesicles(Supported)
NEBL	nebulette		
PLPP2	phospholipid phosphatase 2	Enzymes/ENZYME proteins/Hydrolases	Mitochondria(Approved)
TAMM41	TAM41 mitochondrial translocator assembly	Enzymes/ENZYME proteins/Transferases	Cytosol(Approved)
NAV1	neuron navigator 1		Microtubules(Approved)
CIDEB	cell death inducing DFFA like effector b		
VCY	variable charge Y-linked		
VASN	vasorin		Mitochondria;Nucleoli;Nucleus(Approved)
DHX8	DEAH-box	Enzymes/ENZYME	Nuclear

Novel proteins which were responsible for DNA methylation alterations in Alzheimer's Disease

Initial alias	Transcript ID	Type	Gene symbol
ENSG00000241489	AC244197.3	novel protein	IDS
ENSG00000254536	AL360181.3	novel protein	PAOX
ENSG00000257524	AL157935.2	novel protein	ST6GALNAC6
ENSG00000261832	AC138894.1	novel protein	CLN3
ENSG00000263020	AL662899.2	novel protein	CSNK2B
ENSG00000264668	AC138696.1	novel protein	ZFP41
ENSG00000267335	AC008687.1	novel protein	CGB1
ENSG00000267360	AC012309.1	novel protein	ZNF585B
ENSG00000277726	AL109811.3	novel protein	TARDBP
ENSG00000283761	AC118553.2	novel protein	SLC35A3
ENSG00000284779	AC132217.2	novel protein	IGF2
ENSG00000284969	AL049629.2	novel protein	CD59
ENSG00000285085	AL662884.4	novel protein	PRRT1
ENSG00000285130	AL358113.1	novel protein	FXN
ENSG00000285708	AC097634.4	novel protein	FOXP1
ENSG00000285723	AL034430.2	novel protein	MKKS
ENSG00000288053	AC231657.3	novel protein	WDR45