LADY DOAK COLLEGE, MADURAI
(*An Autonomous Institution affiliated to Madurai Kamaraj University*)
"College with Potential for Excellence"

'Prediction of Alzheimer's disease using blood gene expression data'

Submitted by

Dr.K.Sujatha & Dr.A.Mahalaskhmi
Assistant Professors
Department of Zoology
Lady Doak College
Madurai

TEAM 2: Ms. Karthika, Mrs.Safna and Mr. Mohammad Basha.

To

EXPERTEZE,
1, Broadway, Cambridge, MA 02142, USA

ACKNOWLEDGEMENT

- First and foremost, we would like to thank God Almighty for giving us the strength and ability to complete our project work successfully.

- We are highly indebted to **Prof. Mohan Venkatraman** for the guidance and constant supervision and for spending precious time and giving valuable suggestion for completion of the project.
- We would like to express our special thanks of gratitude to **Dr. Christianna Singh,** Principal and Secretary, Lady Doak College, who gave us the golden opportunity to do this work.
- Our sincere thanks **to Dr. Priyatharsini Rajendran** Head Dept. Zoology, for giving us a wonderful opportunity to pursue this **Bioinfo-project**.

ABBREVATIONS

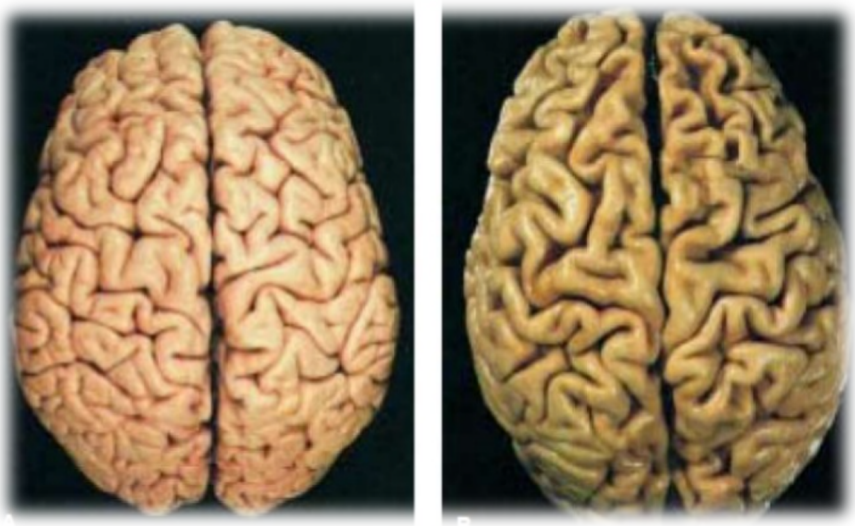
AD : Alzheimer's disease

MCI : Mild Cognitive Impairment

Introduction

"Alzheimer's is a disease for which there is no effective treatment whatsoever. To be clear, there is no pharmaceutical agent, no magic pill that a doctor can prescribe that will have any significant effect on the progressive downhill course of this disease"

David Perlmutter



Normal **AD-Brain**

Short-term memory dysfunction is a key early feature of AD.

As rightly pointed out by David Perlmutter, in the year 2050 (Hebert *et al.,* 2013), 13.8 million individuals in the United States, with 7.0 million being aged 85 years or older will be affected by the common form of dementia i.e. Alzheimer's disease (AD). Also, AD is the fifth leading

cause of death among those age 65 and older (Alzheimer, 2016). Although some drugs showing effectiveness to mitigate the symptoms for a limit time, no treatment can stop the disease. The upcoming reports demonstrated that, the transition from symptom-based to pathophysiology- showed that AD is mainly based on structural brain changes (MRI), molecular neuroimaging changes (positron emission tomography imaging), and alterations in cerebral spinal fluid biomarkers. Although the elucidation of the biological basis of AD has resulted in many advancements, early diagnostic detection of AD and its signal pathway remains challenging.

The following are the risk factors for developing AD

- Increased age (over 65 years age)
- Hypertension (high BP)
- Increased Cholesterol levels
- Coronary artery disease
- Diabetes

Other Risk Factors

- Genetics
- Smoking and alcohol use
- Down syndrome
- Mild Cognitive impairment

Hence, in this mini project work, we performed Gene expression analysis using 'Japanese cohort' to detect the potential blood-based biomarkers upregulation and downregulation for earlier diagnosis of AD. Similarly, pathway analysis results also indicated the changes in the gene expression profile. We strongly believe that, once optimized the results of this work may pave a way for early diagnosis of AD. The following objectives were framed to achieve the goal

Objectives

- Retrieval of gene expression data profiles from NCBI GEO Database.
- Creation of gene expression data of Alzheimer's disease database.
- Gene expression analysis

Methodology Adopted

Identify Suitable Transcriptomics Datasets in Alzheimer's disease

Meta-analysis guidelines as suggested by Ramasamy *et al.* (2008) was followed(Table-1). We performed search on public repositories including ArrayExpress, GEO, Sequence Read Archive (SRA), Stanley Medical Research Institute and DiseaseLand database to identify RNA-Seq or microarray-based datasets on Alzheimer's disease. In total, 30 RNA-Seq or microarray-based datasets on bipolar disorder were included (Table 1).

- GEO: https://www.ncbi.nlm.nih.gov/geo/
- Array Express: https://www.ebi.ac.uk/arrayexpress/
- NACC: https://www.alz.washington.edu/WEB/adc-home.html
- Alzforum: http://www.alzforum.org
- ADNI: http://adni.loni.usc.edu/
- GXA—Gene Expression Database: http://www.ebi.ac.uk/gxa/
- Expression Atlas :https://www.ebi.ac.uk/gxa/experiments?experimentType=differential

Meta-analysis guidelines to be followed for datasets:

Step	Action
Identify suitable microarray studies (Issue 1)	
1	Formulate objectives and a review protocol.
2	Define inclusion-exclusion criteria and suitable keywords.
3	Perform literature search using the keywords on the Web sites listed in Table 2.
4	Search public microarray repositories listed in Table 2.
5	Contact collaborators and experts in the field to help find published and unpublished data.
6	Search the reference section of retrieved studies for other relevant studies.
7	Check the selected study against inclusion-exclusion criteria.
Extract the data from studies (Issue 2)	
8	Scan the literature to identify FLEO data (e.g., CEL, GPR files).
9	If the main text does not contain a link to FLEO data, search the repositories and group/lab's Web pages. If unsuccessful, write to the authors.
10	If multiple publications use overlapping data, identify the most comprehensive one. Combine any training and validation dataset together.
Prepare the individual datasets (Issue 3)	
11	Identify and remove any arrays with poor quality.
12	Preprocess the FLEO data into a GEDM.
13	Check for batch effects among arrays, especially in large studies.
14	Filter out any probes with poor spot quality in the arrays (optional).
15	Aggregate any technical replicates.
16	Check that the processed expression values from multiple platforms are compatible.
Annotate the individual datasets (Issue 4)	
17	Identify either (a) the probe sequence or (b) the most sequence-specific probe annotation information.
18	Either (a) cluster the probe sequences or (b) map the most sequence-specific probe annotation to a gene-level identifier. Use the same mapping build for all datasets.
Resolve the many-to-many relationship between probes and genes (Issue 5)	
19	Discard any probe that does not map to any GeneID.
20	For every GeneID within a study, calculate the study-specific estimate(s).
21	If a probe maps to multiple GeneIDs within a study, "expand" it by replacing it with a new record for each GeneID with the same study-specific estimate(s) or expression profile.
22	For GeneIDs with multiple records within a study, "summarize" them by either selecting one of the records or by aggregating them.
Combine the study-specific estimates (Issue 6)	
23	For every GeneID, identify the studies that provide usable information. Optionally, discard any GeneID that is not found in at least a prespecified number of studies.
24	For every GeneID, combine the study-specific estimates across the studies using a meta-analytic technique. Record the resulting summary statistic(s).
25	Calculate the nominal p-value of the summary statistic(s) for every GeneID and adjust for multiple testing.
Analyze, present, and interpret results (Issue 7)	
26	Examine the sensitivity of results to individual studies with a leave-one-out analysis and by varying the selections made (e.g., type of data available).
27	Present the summary statistics graphically (e.g., forest plot) for genes of interest.
28	Analyze findings using computational tools (e.g., gene set enrichment analysis).
29	If possible, validate using an alternative technology and/or different samples.
30	Consider strength of evidence, limitations, and generalizability of current findings.

GeneID refers to either the sequence cluster or gene-level identifier used in Step 18. See text for further details. Where possible, we have indicated the step number near the relevant text.
doi:10.1371/journal.pmed.0050184.t001

(Adapted from Ramasamy *et al.*, 2008)

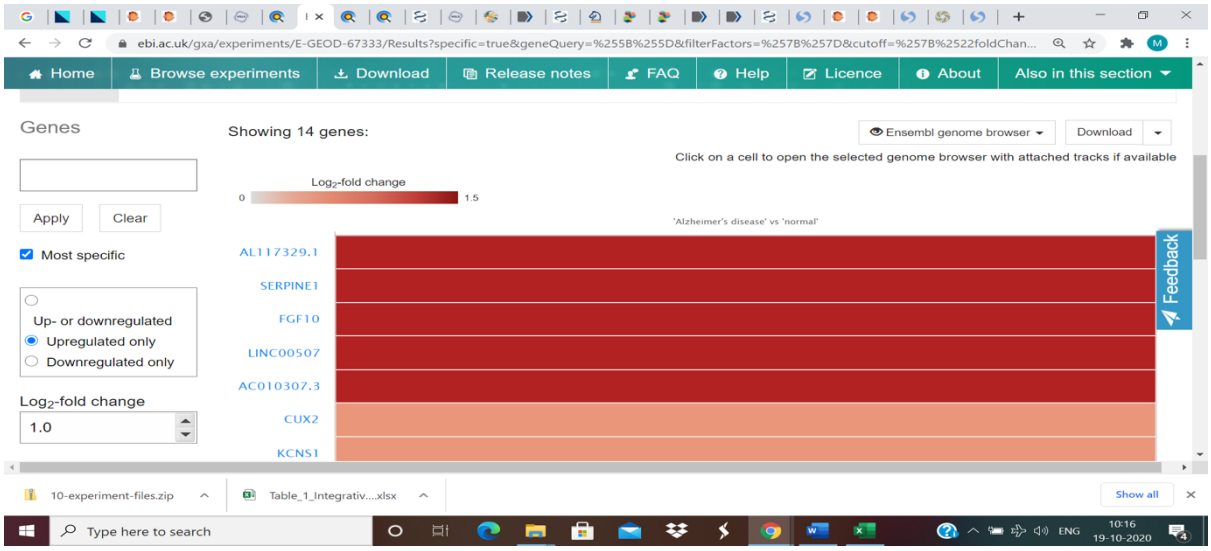
Results and Discussion

Studies are conducted to learn more about plaques, tangles, and other biological features of AD. Advances in brain imaging techniques allow researchers to see the development and spread of **abnormal amyloid and tau proteins in the living brain**, as well as **changes in brain structure and function**. The results of this study focussed on the gene expression profile and changes in the signalling pathways. The entire result will be utilized for the creation of database specific for AD.

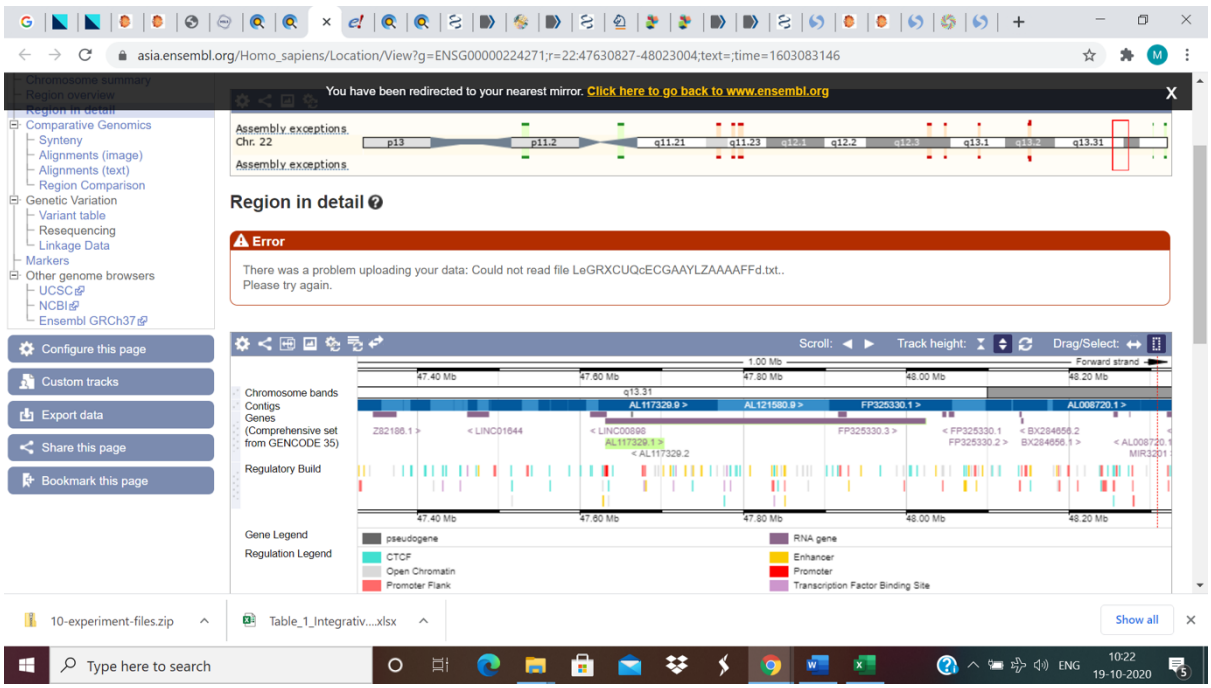
Creation of Alzheimer's specific database is in progress

Upregulated genes:

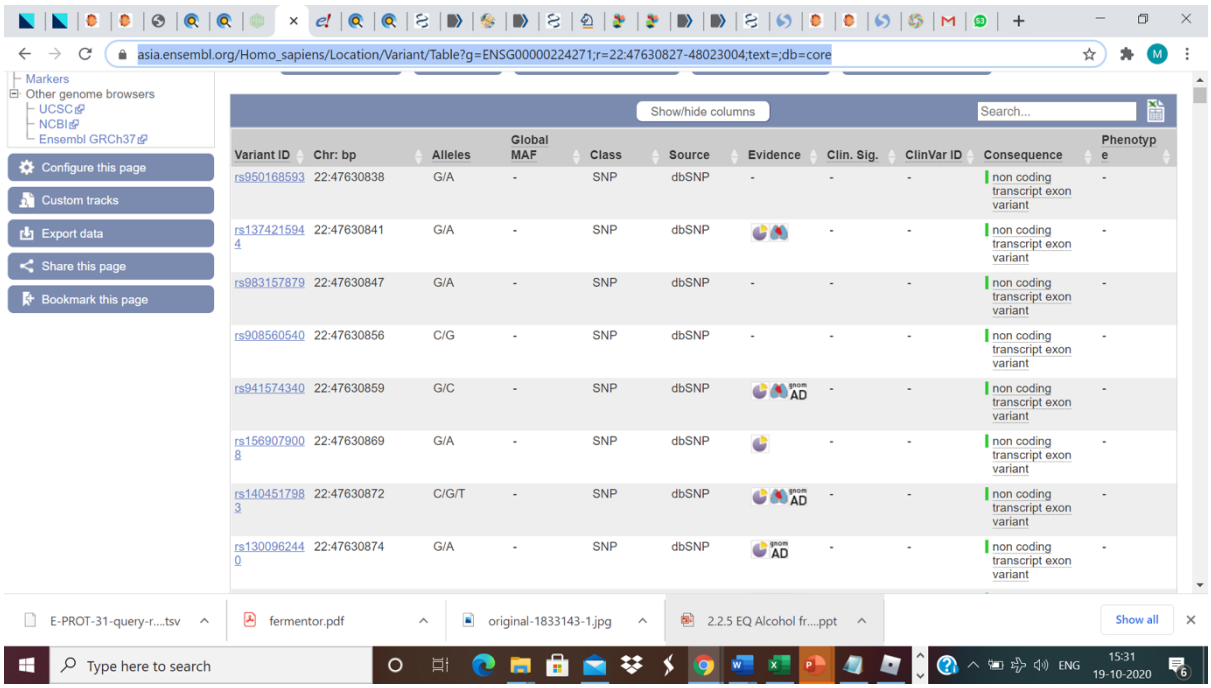
14 genes were upregulated between Alzheimer's disease vs normal



Location of the selected genes in the genome



Variant table observed:



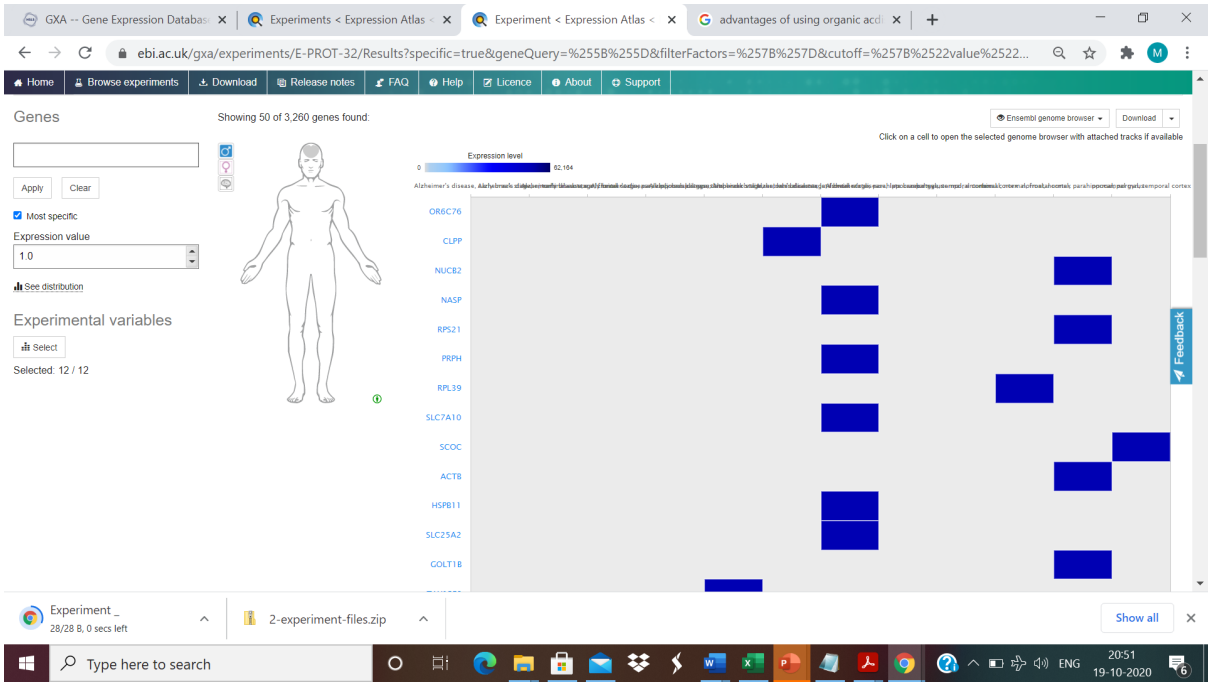
https://asia.ensembl.org/Homo_sapiens/Location/Variant/Table?g=ENSG00000224271;r=22:47630827-48023004;text=;db=core

Downregulated genes

Comparison of downregulated genes between Alzheimer's disease vs. normal

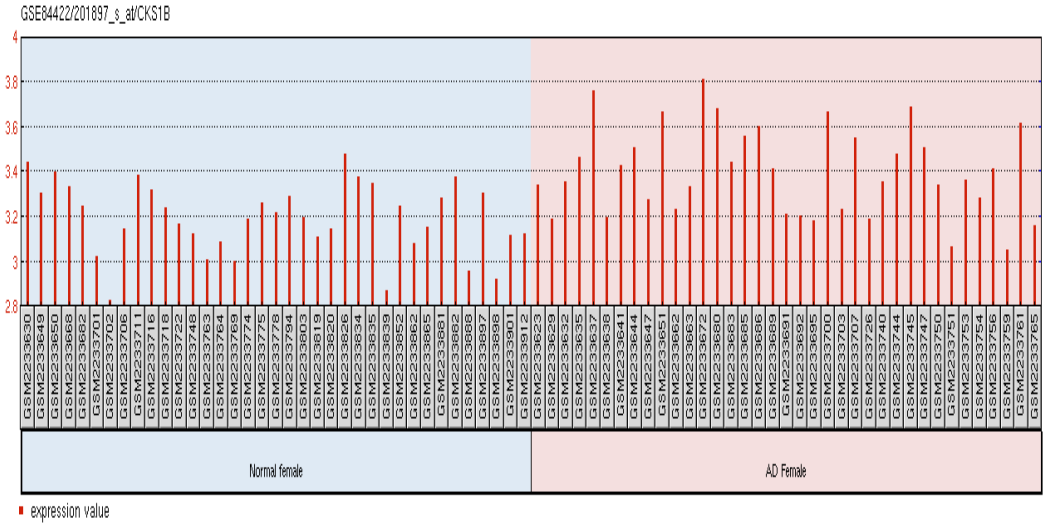
https://www.ebi.ac.uk/gxa/experiments/E-PROT-32/Results?specific=true&geneQuery=%255B%255D&filterFactors=%257B%257D&cutoff=%257B%2522value%2522%253A0.000001%257D

Stage wise comparison of genesets among the various regions of the brain among Alzheimer's disease and normal

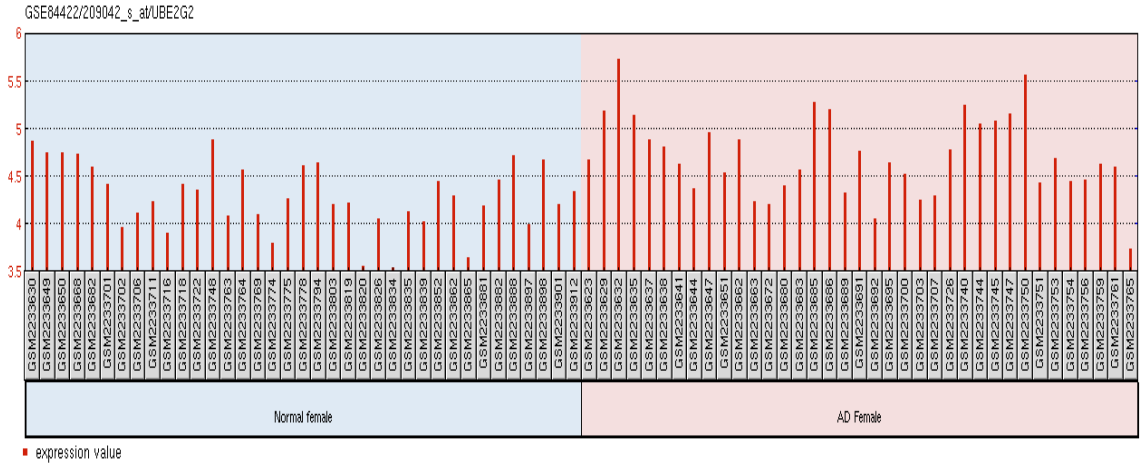


https://www.ebi.ac.uk/gxa/experiments/E-PROT-32/Results?specific=true&geneQuery=%255B%255D&filterFactors=%257B%257D&cutoff=%257B%2522value%2522%253A1%257D

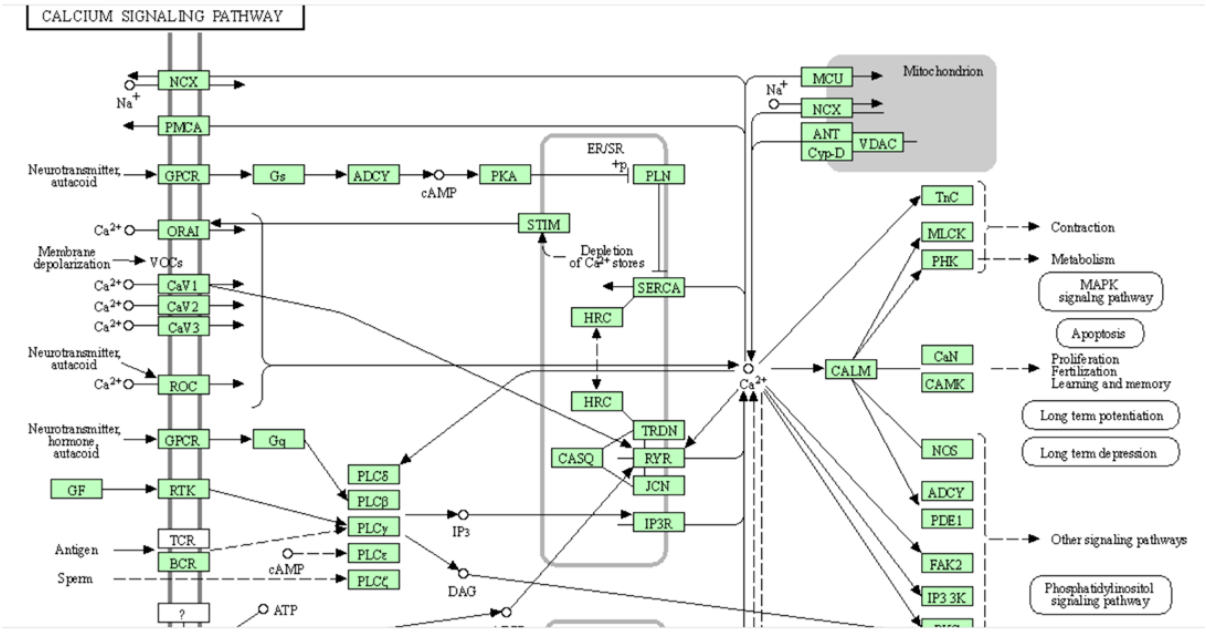
CDC28 protein kinase regulatory protein – showed threefold change



Ubiquitin conjugating enzyme

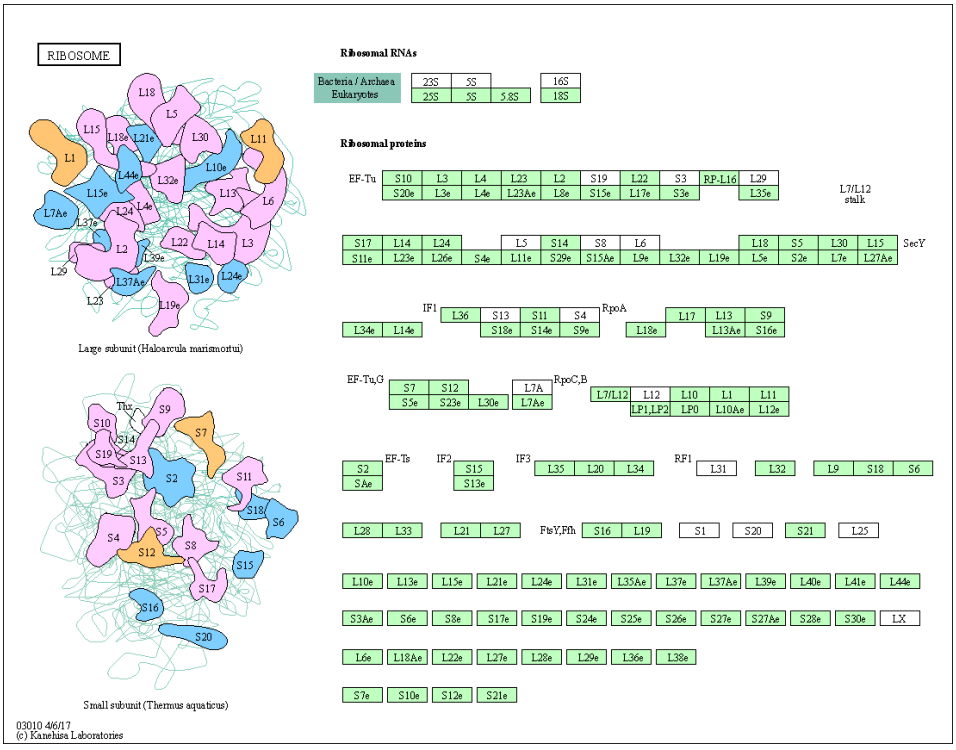


Upregulated KEGG pathways



Down regulated KEGG pathways

Ribosomal pathway



Outcome of the work

Outcome

- The results of this study helpful to identify the **set of genes and related pathways** that may play important role in the development of AD in future.
- Finding **how these genes regulate the downstream DEGs (Differentially expressed genes) is vital for understanding both the pathophysiology of AD** and looking for potential targets for drug therapy.
- It also opens **new avenues for precision medicine- diagnostics** (assessment of risk)
- as well as early treatment (pharmacogenomically informed, personalized, and preventive).
- Other **drug and nutraceutical leads will be identified through** bioinformatic drug repurposing analyses

Suggestions/Recommendations

- Timely diagnosis of AD and early identification of people who are at heightened risk for AD such as those with Mild Cognitive Impairment [MCI]
- As per the recent reports, research work can be focused towards 'herbal therapy' to treat AD disease (Tian *et al*., 2010).

The "Do-It-Yourself" Approach

- Diet control
- Regular exercise
- Stress control/management
- Herbal remedies

References

Alzheimer's, A. (2016). Alzheimer's disease facts and figures. Alzheimers Dement, 2016. 12(4): p. 459-509

Hebert, L. E., Weuve, J., Scherr, P. A. & Evans, D. A. Alzheimer disease in the United States (2010–2050) estimated using the 2010 census (2013). *Neurol.* **80**, 1778–1783.

Tian, J., Shi, J., Zhang, X. *et al.* Herbal therapy: a new pathway for the treatment of Alzheimer's disease. *Alz Res Therapy* **2,** 30 (2010). https://doi.org/10.1186/alzrt54

Kapushesky, M., Emam, I., Holloway, E., Kurnosov, P., Zorin, A., Malone, J., Rustici, G., Williams, E., Parkinson, H., & Brazma, A. (2010). Gene expression atlas at the European bioinformatics institute. *Nucleic acids research*, *38*(Database issue), D690–D698. https://doi.org/10.1093/nar/gkp936.

Rahman, M. R., Islam, T., Shahjaman, M., Rana, M. H. K., Holsinger, R. D., Quinn, J. M., ... & Moni, M. A. Genome-Wide Integrative Analysis Reveals Common Molecular Signatures in Blood and Brain of Alzheimer's. https://doi.org/10.33263/BRIAC112.86868701

Eric M. Blalock, James W. Geddes, Kuey Chu Chen, Nada M. Porter, William R. Markesbery, Philip W. Landfield. Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses Proceedings of the National Academy of Sciences Feb 2004, 101 (7) 2173-2178; DOI: 10.1073/pnas.0308512100.

Moradifard, S., Hoseinbeyki, M., Ganji, S.M. *et al.* Analysis of microRNA and Gene Expression Profiles in Alzheimer's Disease: A Meta-Analysis Approach. *Sci Rep* **8,** 4767 (2018). https://doi.org/10.1038/s41598-018-20959-0

Lavida, R.K.B. and George, I. M. Data-Driven Analysis of Age, Sex, and Tissue Effects on Gene Expression Variability in Alzheimer's Disease. Front. Neurosci., 24 April 2019 | https://doi.org/10.3389/fnins.2019.00392