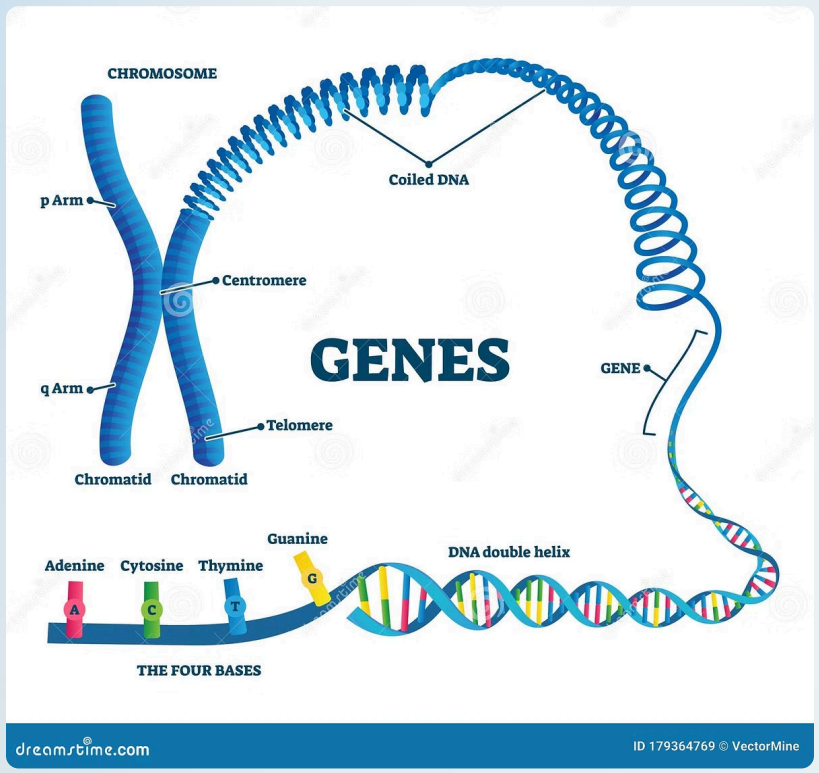


Meta-Analysis of Transcriptome Data Reveals Potential Pathogenic Factors in Alzheimer's Disease

This study aims to identify potential pathogenic factors based on a meta-analysis of transcriptome data from different brain regions of AD patients. Alzheimer's disease (AD) is characterized by the distribution of amyloid plaques and neurofibrillary tangles, which are composed of Aß and hyperphosphorylated tau proteins. Aß is generated from amyloid precursor protein (APP) and exists in two forms: Aß40 and Aß41. Aß monomers have a high binding affinity, leading to the formation of oligomers, fibrils, and plaques. Aß oligomers are recognized as the primary neurotoxic species. Aß directly damages neurons and triggers a series of cellular responses, including alterations in Ca2+ homeostasis and synaptic function. Aß also activates protein kinases to phosphorylate tau proteins, resulting in the formation of neurofibrillary tangles.

by Aryan Sharma

The Amyloid Cascade Hypothesis

The amyloid cascade hypothesis, proposed in 1992, posits that Aß is a causative factor in AD development. This hypothesis is supported by genetic mutations related to Aß formation (APP and PSEN) and clearance (APOEY), which induce early-onset AD. Drug development for AD treatment has been guided by the amyloid cascade hypothesis for the past two decades.

Aß Formation

Mutations in APP and PSEN genes lead to increased Aß production, contributing to AD development.

Aß Clearance

Mutations in APOEY gene impair Aß clearance, leading to its accumulation and AD progression.

Early-Onset AD

Genetic mutations related to Aß formation and clearance cause early-onset AD, supporting the amyloid cascade hypothesis.

Transcriptomic Profiling Approach

The discovery of AD pathogenic factors can be revealed by transcriptomic profiling approaches, such as DNA microarray and RNA sequencing, to identify differentially expressed genes (DEGs) in AD subjects compared to healthy subjects. These approaches provide insights into the molecular changes associated with AD.

1

DNA Microarray

A high-throughput technology that measures the expression levels of thousands of genes simultaneously.

2

RNA Sequencing

A technique that sequences the entire transcriptome, providing a comprehensive view of gene expression.

3

Differentially Expressed Genes (DEGs)

Genes that show significant differences in expression levels between AD and healthy subjects.

Meta-Analysis: Resolving Inconclusive Results

Integrating multiple transcriptomic studies through meta-analysis helps resolve inconclusive results and obtain a generalization. Meta-analysis combines data from different studies to increase statistical power and identify consistent patterns.

1

Data Collection

Transcriptomic datasets from different studies are collected and processed.

2

Meta-Analysis

Statistical methods are used to combine data from multiple studies and identify consistent patterns.

3

Generalization

Meta-analysis provides a more robust and generalizable understanding of the molecular changes associated with AD.

Moradifard et al. (2018) Meta-Analysis

Moradifard et al. (2018) conducted a meta-analysis using the R package Robust Rook Aggy to identify statistically significant DEGs in AD across six microarray datasets. This study provides a comprehensive analysis of gene expression changes in AD.

Study	Database	Platform	Pathogenic Factors
Yan et al. (2019)	GEO	DNA Microarray	BDNF, CACNA1A, CALB, CD44, CDC42, OXT, PDYN, TAC1, TH, VEGFA
Li X. et al. (2018)	GEO	DNA Microarray	NDUFAT, MRPL51, PPL36AL
Rahman et al. (2019)	CMap, GEO	DNA Microarray	AR, CREBBP, E2F1, FOXC1, FOXL1, GATA2, JUN, NFIC, PPARG, RAC1, PPL12, RPL15, RPS11, RPS6, SMAD3, SAF, UBA52, UBC, USF2, YY1
Vargas et al. (2018)	CMap, GEO	DNA Microarray	ATF2, CNOT7, CSRNP2, PARK2, SLC30A9, TSC22D1
Xiang et al. (2018)	Allen Brain Institute, GEO	DNA Microarray	FOS, JUN, MEF2A, MIB2, PCBP1, SMARCA2, SP1, STAT1, TEAD4, ZFHX3, ZINF281
Kawalia et al. (2017)	ArrayExpress, GEO	DNA Microarray	AP2A2, ARAP3, ATP2A3, ATP2B4, HLA-C, HLA-F, ITPR2, RAB11FIP4, STX2

Dataset Selection and Processing

Datasets were collected from GEO and ArrayExpress. Raw DNA microarray and RNA-Seq data reporting the transcriptomes of brain regions from AD and healthy subjects were included. Each dataset was processed through background correction, quantile normalization, and log2-transformation of the averaged expression value of duplicate probes, using the affy, oligo, or limma package. The quality of each array was assessed using the arrayQualityMetrics package. The limma package was used to determine the gene expression ratios between AD and healthy subjects.

Data Collection

Datasets were collected from GEO and ArrayExpress, focusing on transcriptomic data from brain regions of AD and healthy subjects.

Data Processing

Datasets underwent background correction, quantile normalization, and log2-transformation to ensure data quality and consistency.

Quality Assessment

The quality of each array was assessed using the arrayQualityMetrics package to identify and remove outliers.

Meta-Analysis and Biological Enrichment Analysis

The list of DEGs and their upregulated or downregulated states from each study was processed using the meta package. Meta-analysis was conducted under a random-effects model, and their outcomes were logORs with P-values. Genes with logORs above or below 0 were considered upregulated or downregulated, respectively. P-values were adjusted, and those less than 0.05 were regarded statistically significant.

1

DEG List

The list of DEGs and their upregulated or downregulated states from each study was processed.

2

Meta-Analysis

Meta-analysis was conducted under a random-effects model to identify consistent patterns across studies.

3

Biological Enrichment Analysis

The statistically significant DEGs were used to perform a biological enrichment analysis to identify biological pathways among them.

Protein-Protein Interaction Network and Potential Pathogenic Factors

Statistically significant DEGs were used to construct a PPIN (Protein-Protein Interaction Network) based on data from STRING. Only interactions with the highest confidence (0.9) were kept. The PPIN was visualized using Cytoscape. AD seed genes were retrieved from the GWAS Catalog using the keyword "AD" and a significance p-Value cutoff of no more than 1 10–8. The PPIN information and AD seed genes were put into the DIAMOnD algorithm. In each iteration, until the stopping condition was satisfied, the node with the lowest connectivity P-value was treated as the most significantly connected node for output. The nodes within the module that had a high degree (>95th percentile) were treated as potential pathogenic factors.

Protein-Protein Interaction Network (PPIN)

A network that represents the interactions between proteins, providing insights into cellular processes and pathways.

AD Seed Genes

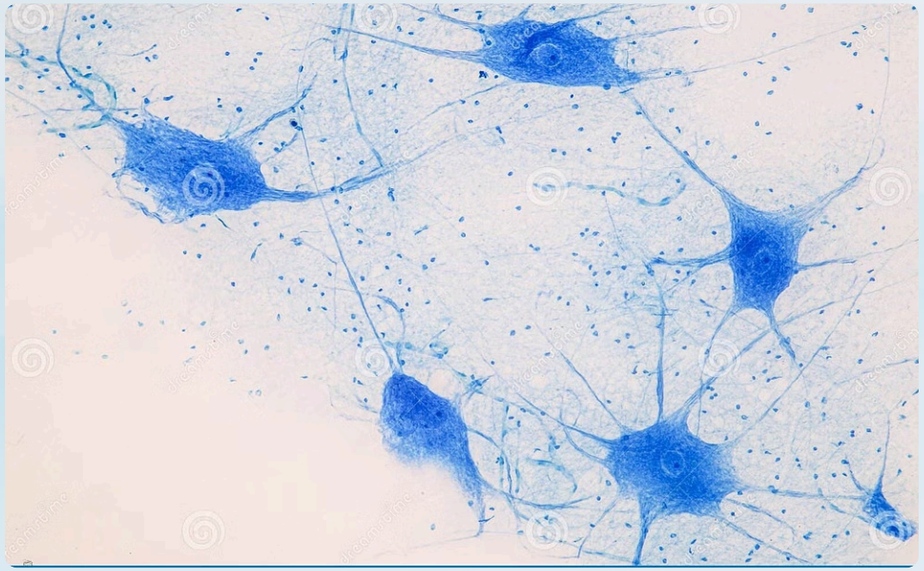
Genes associated with AD identified from the GWAS Catalog, providing a starting point for network analysis.

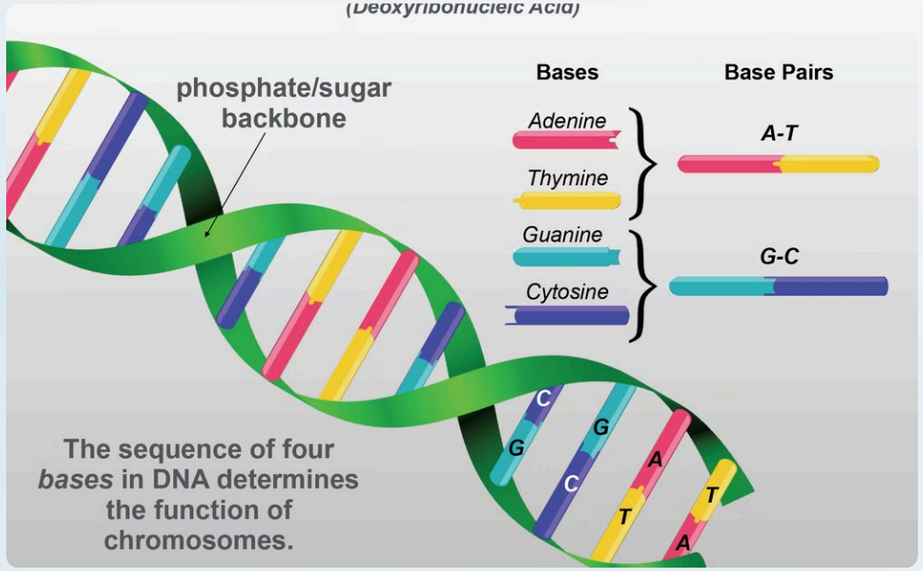
DIAMOnD Algorithm

An algorithm used to identify modules of highly connected nodes within a network, potentially representing pathogenic factors.

Results: Dataset Selection and DEG Identification

From GEO and ArrayExpress, 225 datasets were initially found. 25 datasets (21 for microarray and 4 for RNA-Seq) fulfilled the eligibility requirements after duplicates and datasets that met the exclusion criteria were removed. Among the included datasets, the hippocampus was the most extensively studied brain region. The flow diagram of datasets selection, including identification, screening, eligibility, and inclusion, is shown in the figure.





Dataset Selection

The process of identifying, screening, and selecting relevant datasets for the meta-analysis.

DEG Identification

The process of identifying differentially expressed genes (DEGs) between AD and healthy subjects in each dataset.

Results: Meta-Analysis and Biological Enrichment Analysis

In the meta-analysis of DEGs, 9,298 DEGs were found to be statistically significant, with 4,960 genes downregulated and 4,338 genes upregulated. The most reported downregulated DEGs were DPP6 and FXYD7, which were reported in 16 comparisons; RHOQ was the most reported upregulated DEG across all 15 reported comparisons. A loss of DPP6 or FXYD7 is reported to dysregulate neuronal excitation, while RHOQ is reported to enhance Aβ oligomerization.

1 Downregulated DEGs

Genes that showed decreased expression levels in AD subjects compared to healthy subjects.

2 Upregulated DEGs

Genes that showed increased expression levels in AD subjects compared to healthy subjects.

3 Biological Enrichment Analysis

The process of identifying biological pathways that are enriched with statistically significant DEGs.