MAR 21, 2024

OPEN ACCESS

DOI:
dx.doi.org/10.17504/protocols.io.5qpvokxyxl4o/v1

Protocol Citation: Shreya Satyanarayan Bhat, Adarsh Vishal, Spoorthi Anil Bandikatte, Vidya Niranjan 2024. Bioinformatics Protocol for Investigating Novel Biomarkers, Conducting Statistical Analysis, and Exploring Molecular Pathways: A Unified Approach for Alzheimer's Disease. **protocols.io** https://dx.doi.org/10.17504/protocols.io.5qpvokxyxl4o/v1

Bioinformatics Protocol for Investigating Novel Biomarkers, Conducting Statistical Analysis, and Exploring Molecular Pathways: A Unified Approach for Alzheimer's Disease

Shreya Satyanarayan Bhat[1], Adarsh Vishal[2], Spoorthi Anil Bandikatte[2], Vidya Niranjan[1]

[1]R V College of Engineering; [2]RV College of Engineering, Bangalore

Vidya Niranjan
R V College of Engineering

ABSTRACT

This protocol paper presents a comprehensive bioinformatics approach aimed at identifying novel biomarkers associated with Alzheimer's disease (AD) using RNA sequencing (RNA-seq) datasets obtained from the Gene Expression Omnibus (GEO) database. Alzheimer's disease is a complex neurodegenerative disorder characterized by progressive cognitive decline, and identifying biomarkers is crucial for early diagnosis and therapeutic intervention. Leveraging bioinformatics tools and methods, this protocol outlines a step-by-step procedure for analyzing RNA-seq data to uncover potential biomarkers. The protocol includes data collection from the GEO database, quality control assessment, differential expression analysis, functional annotation, pathway enrichment analysis, and molecular pathway identification. Statistical analysis is applied throughout the protocol to ensure the robustness and reliability of the results. By following this protocol, researchers can systematically identify and validate novel biomarkers associated with Alzheimer's disease, ultimately contributing to a better understanding of its underlying molecular mechanisms and facilitating the development of targeted therapeutic interventions.

Protocol status: Working
We use this protocol and it's working

Created: Mar 18, 2024

Last Modified: Mar 21, 2024

PROTOCOL integer ID: 96840

Keywords: Bioinformatic approach, Biomarkers, statistical analysis, Alzheimer's disease, pathways

Identification of RNA-Seq Dataset

1 The RNA-Seq dataset utilized in this study was sourced from the European Nucleotide Archive (ENA) browser, specifically accession number GSE138853. This dataset was selected due to its relevance to Alzheimer's disease research and its availability through the ENA browser.

Dataset	
ALZHEIMERS DISEASE DATASET	NAME
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE138852	LINK

Retrieval of Samples

2 Raw sequencing samples were downloaded in FASTQ format, containing nucleotide sequences and quality scores for each read. This standardized format enables comprehensive bioinformatics analysis, including quality control and differential expression analysis.
Upon selecting the desired dataset, proceed to download the raw sequencing samples associated with it. The samples were retrieved directly from the ENA browser interface, providing access to the original sequencing data generated for the study. Each sample corresponds to a specific biological sample (e.g., tissue sample) that underwent RNA sequencing to generate transcriptomic data. The raw sequencing samples were downloaded in their original format, typically stored as FASTQ files. FASTQ is a standard file format used to represent raw sequencing data, containing both nucleotide sequences and quality scores for each sequenced read.

Dataset

samples retrieval	NAME
https://www.ebi.ac.uk/ena/browser/view/PRJNA577618	LINK

Command

Download samples (linux 20.4)

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/srr/SRR102/012/SRR10278812
```

Quality Check using FASTQC

3 The quality assessment of the 32 samples using the FASTQC tool. FASTQC is a widely used tool for the initial assessment of raw sequencing data quality.

Command

The command fastqc initiates the quality checking process using FASTQC. reads.fastq.gz represents the input raw sequencing data file for each sample. The file is in compressed FASTQ format (.fastq.gz), commonly encountered in sequencing data analysis. (LINUX 20.4)

```
fastqc reads.fastq.gz
```

Conversion of BAM Files to SAM Files Using Samtools

4 This section outlines the comprehensive conversion of BAM files to SAM files utilizing Samtools. SAM files provide a human-readable representation of sequence alignment data, facilitating downstream analysis and interpretation.

Command

This section outlines the comprehensive conversion of BAM files to SAM files utilizing Samtools. SAM files provide a human-readable representation of sequence alignment data, facilitating downstream analysis and interpretation. (linux 20.4)

```
samtools view -h -o output.sam input.bam
```

Command

This command sorts the alignments in the BAM file by their genomic coordinates. (linux 20.4)

```
samtools sort -@ 4 24.bam -o 24_sorted.bam
```

Alignment

5 All 32 RNA-Seq samples were aligned to the human reference genome obtained from the National Center for Biotechnology Information (NCBI). The alignment was performed using HISAT2, a widely used alignment tool known for its efficiency and accuracy in mapping sequencing reads to a reference genome.

Command

The alignment code aligns RNA-Seq samples to the human reference genome using HISAT2. It specifies the index of the reference genome, the input file containing sequencing reads, and the output file for alignment results in SAM format. This process enables mapping of reads to the genome, facilitating downstream analyses of gene expression and regulatory mechanisms. (Linux 20.4)

```
hisat2 -x human_reference_genome_index -U sample.fastq.gz -S sample_aligned.sam
```

Generation of feature Count

6 The generation of feature counts using the FeatureCounts tool. FeatureCounts is a widely used tool for counting the number of reads mapped to genomic features, such as genes, transcripts, or exons, based on the alignment results obtained from aligning RNA-Seq reads to a reference genome. The annotation file used for feature counting was obtained from the National Center for Biotechnology Information (NCBI) Genome database. This file provides genomic annotations, including the locations and characteristics of genes, transcripts, and other genomic features, necessary for accurately assigning reads to features during the counting process.

Command

featureCounts: Initiates the feature counting process using the FeatureCounts tool. -a human_annotation_file.gtf: Specifies the annotation file (in GTF format) obtained from NCBI, containing genomic annotations for the human reference genome. This file guides the assignment of reads to genomic features during counting. -o counts.txt: Specifies the output file name (counts.txt) for storing the feature counts. This file contains the counts of reads assigned to each genomic feature. sample_aligned.sam: Indicates the input SAM file containing alignment results obtained from aligning RNA-Seq reads to the human reference genome.

```
featureCounts -a human_annotation_file.gtf -o counts.txt sample_aligned.sam
```

Differential gene expression analysis

7 We performed differential gene expression (DGE) analysis for all 32 samples using the DESeq2 package. DESeq2 is a widely used tool in bioinformatics for identifying genes that are differentially expressed between experimental conditions or sample groups based on RNA-Seq data. It employs a negative binomial distribution model to accommodate variability in sequencing depth and biological variability across samples. DESeq2 conducts normalization, estimation of dispersion, and statistical testing to identify genes that exhibit significant differences in expression levels between conditions. This approach enables researchers to gain insights into the molecular mechanisms underlying biological processes and diseases by identifying genes that are actively involved in specific conditions or experimental treatments.

Command

Dseq2 Commands (windows 11)

```
# Load count data
count_data <- read.csv('count.csv', header = TRUE, row.names = 1)

# Filter low-count genes
count_data <- count_data[rowSums(count_data[, c('SRR15039666', 'SRR15039668',
'SRR22702939','SRR22702940','SRR22702946','SRR22702961','SRR22801652','SRR228016
54')]) > 50, ]

# Define experimental conditions
condition <- factor(c("AD", "AD", ..., "H"))

# Create a data frame with sample IDs and conditions
coldata <- data.frame(row.names = colnames(count_data), condition)

# Remove rows with missing values
count_data_clean <- count_data[complete.cases(count_data), ]

# Create DESeq2 object
dds <- DESeqDataSetFromMatrix(countData = count_data_clean, colData = coldata,
design = ~condition)

# Perform differential expression analysis
dds <- DESeq(dds)

# Transform count data
vsdata <- vst(dds, blind = FALSE)

# Generate PCA plot
plotPCA(vsdata, intgroup = "condition")

# Plot dispersion estimates
plotDispEsts(dds)

# Extract differential expression results
res <- results(dds, contrast = c("condition", "C", "H"))
```

```
# Filter significant results
sigs <- na.omit(res)
sigs <- sigs[sigs$padj < 0.05, ]

# Print significant results
sigs

# Extract expression matrix of significant genes
sig_expr_matrix <- assay(dds)[rownames(sigs), ]

# Log2 transform expression matrix
log_sig_expr_matrix <- log2(sig_expr_matrix + 1)

# Generate heatmap
heatmap <- pheatmap(log_sig_expr_matrix, cluster_rows = TRUE, cluster_cols =
TRUE, color = colorRampPalette(c("blue", "white", "red"))(100))

# Save heatmap as TIFF file
tiff("heatmap.tiff", width = 10, height = 10, units = "in", res = 600)
print(heatmap)
dev.off()

# Plot MA plot
plotMA(res)

# Generate volcano plot
volcano_plot <- ggplot(res, aes(x = log2FoldChange, y = -log10(padj))) +
    geom_point(aes(color = ifelse(padj < 0.05, "Significant", "Not
significant")), size = 2, alpha = 0.8) +
    scale_color_manual(values = c("Significant" = "red", "Not significant" =
"gray")) +
    labs(title = "Volcano Plot", x = "Log2 Fold Change", y = "log10(p-value)") +
    theme_minimal() +
    theme(
        plot.margin = margin(20, 20, 20, 20),
        axis.text.x = element_text(size = 10),
        axis.text.y = element_text(size = 10),
        legend.position = "bottom",
        legend.direction = "vertical",
        legend.box = "vertical",
        legend.margin = margin(t = 20, r = 20, b = 0.5, l = 20, unit = "pt"),
```

```
        legend.margin = margin(t = 20, r = 20, b = 0.5, l = 20, unit = "pt")
    )

# Display volcano plot
print(volcano_plot)
```

Gene ontology

8 In this section, we utilized the Gene Ontology (GO) database to perform a functional analysis of the differentially expressed genes identified in the RNA-Seq analysis. GO is a widely used bioinformatics resource that categorizes genes into defined terms based on their molecular functions, biological processes, and cellular components, providing valuable insights into the biological roles of genes. Gene Ontology (GO) helps us understand gene functions. It covers Biological Processes: Genes' roles in cell division, metabolism, development, and immune response. Molecular Functions: Specific activities at the molecular level (e.g., enzyme catalysis, binding). Cellular Components, Where gene products operate within cells (nucleus, mitochondria, etc.). The image shows the Gene Ontology website, where researchers input gene IDs to explore these aspects.

gene ontology website

Analysis Summary: Please report in publication

Analysis Type: PANTHER Overrepresentation Test (Released 20240226)

Annotation Version and Release Date: GO Ontology database DOI: 10.5281/zenodo.10536401 Released 2024-01-17

Analyzed List: upload_1 (Homo sapiens) Change

Reference List: Homo sapiens (all genes in database) Change

Annotation Data Set: GO biological process complete

Test Type: Fisher's Exact Binomial

Correction: Calculate False Discovery Rate Use the Bonferroni correction for multiple testing No correction

Results

	Reference list	upload_1
Uniquely Mapped IDS:	20592 out of 20592	22 out of 22
Unmapped IDs:	0	0
Multiple mapping information:	0	0

Export Table XML with user input ids JSON with user input ids

Displaying only results for FDR P < 0.05, click here to display all results

	Homo sapiens (REF)	upload_1 (Hierarchy NEW!)					
GO biological process complete	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
fever generation	3	2	.00	> 100	+	3.27E-06	1.28E-03
↳response to stimulus	8182	19	8.74	2.17	+	8.93E-06	2.67E-03
↳heat generation	7	2	.01	> 100	+	2.28E-05	5.61E-03
↳multicellular organismal process	6745	18	7.21	2.50	+	3.09E-06	1.24E-03
central nervous system vasculogenesis	5	2	.01	> 100	+	1.09E-05	3.13E-03
↳cell differentiation	3654	14	3.90	3.59	+	2.28E-06	1.02E-03
↳cellular developmental process	3657	14	3.91	3.58	+	2.30E-06	1.00E-03
↳blood vessel morphogenesis	431	5	.46	10.86	+	7.70E-05	1.45E-02
↳anatomical structure morphogenesis	2239	9	2.39	3.76	+	2.77E-04	3.55E-02
↳anatomical structure development	5231	14	5.59	2.51	+	1.72E-04	2.62E-02

The image presents the outcomes of a Gene Ontology (GO) biological process analysis, highlighting processes such as response to stimulus and cell differentiation. Each process is linked to statistical values that signify its significance. This image provides valuable insights into the biological functions influenced by a specific set of query genes, laying the groundwork for understanding their roles within an organism.

Analysis Summary: Please report in publication

Analysis Type: PANTHER Overrepresentation Test (Released 20240226)

Annotation Version and Release Date: GO Ontology database DOI: 10.5281/zenodo.10536401 Released 2024-01-17

Analyzed List: upload_1 (Homo sapiens) Change

Reference List: Homo sapiens (all genes in database) Change

Annotation Data Set: GO molecular function complete

Test Type: Fisher's Exact Binomial

Correction: Calculate False Discovery Rate Use the Bonferroni correction for multiple testing No correction

Results

	Reference list	upload_1
Uniquely Mapped IDS:	20592 out of 20592	22 out of 22
Unmapped IDs:	0	0
Multiple mapping information:	0	0

Export Table XML with user input ids JSON with user input ids

Displaying only results for FDR P < 0.05, click here to display all results

	Homo sapiens (REF)	upload_1 (Hierarchy NEW!)					
GO molecular function complete	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
1-phosphatidylinositol-3-kinase activity	10	2	.01	> 100	+	4.88E-05	2.47E-02
frizzled binding	39	7	.04	> 100	+	8.26E-15	4.19E-11
↳G protein-coupled receptor binding	296	8	.32	25.30	+	4.45E-10	3.76E-07
↳signaling receptor binding	1524	13	1.63	7.98	+	5.03E-10	3.64E-07
co-receptor binding	13	2	.01	> 100	+	8.44E-05	3.89E-02
cytokine activity	236	9	.25	35.69	+	1.28E-12	3.24E-09
↳receptor ligand activity	507	9	.54	16.62	+	1.16E-09	7.34E-07
↳signaling receptor activator activity	514	9	.55	16.39	+	1.31E-09	7.36E-07
↳molecular function activator activity	1137	12	1.21	9.88	+	2.93E-10	2.97E-07

The image reveals the outcomes of a Gene Ontology (GO) molecular function analysis based on query genes. It highlights specific molecular activities associated with statistical significance, providing insights into the functional roles of these genes at the molecular.

Analysis Summary: Please report in publication

Analysis Type: PANTHER Overrepresentation Test (Released 20240226)

Annotation Version and Release Date: GO Ontology database DOI: 10.5281/zenodo.10536401 Released 2024-01-17

Analyzed List: upload_1 (Homo sapiens) Change

Reference List: Homo sapiens (all genes in database) Change

Annotation Data Set: GO cellular component complete

Test Type: Fisher's Exact Binomial

Correction: Calculate False Discovery Rate Use the Bonferroni correction for multiple testing No correction

Results

	Reference list	upload_1
Uniquely Mapped IDS:	20592 out of 20592	22 out of 22
Unmapped IDs:	0	0
Multiple mapping information:	0	0

Export Table XML with user input ids JSON with user input ids

Displaying only results for FDR P < 0.05, click here to display all results

	Homo sapiens (REF)	upload_1 (Hierarchy NEW!)					
GO cellular component complete	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
phosphatidylinositol 3-kinase complex, class IA	9	2	.01	> 100	+	3.90E-05	3.90E-02
↳phosphatidylinositol 3-kinase complex, class I	9	2	.01	> 100	+	3.90E-05	2.60E-02
↳phosphatidylinositol 3-kinase complex	30	3	.03	93.60	+	4.22E-06	8.43E-03
extracellular space	3330	12	3.56	3.37	+	4.09E-05	2.05E-02

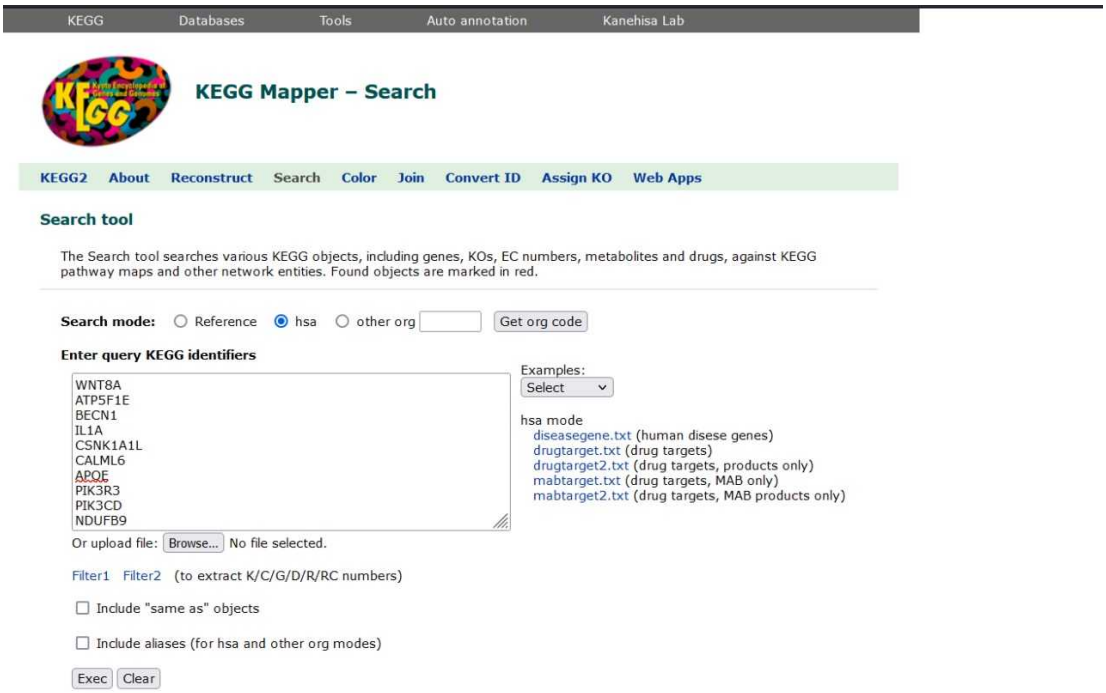
The image reveals the outcomes of a Gene Ontology (GO) cellular component analysis based on query genes. It highlights specific cellular locations associated with statistical significance, providing insights into where these genes operate within cells.

Pathway Analysis using KEGG MAPPER

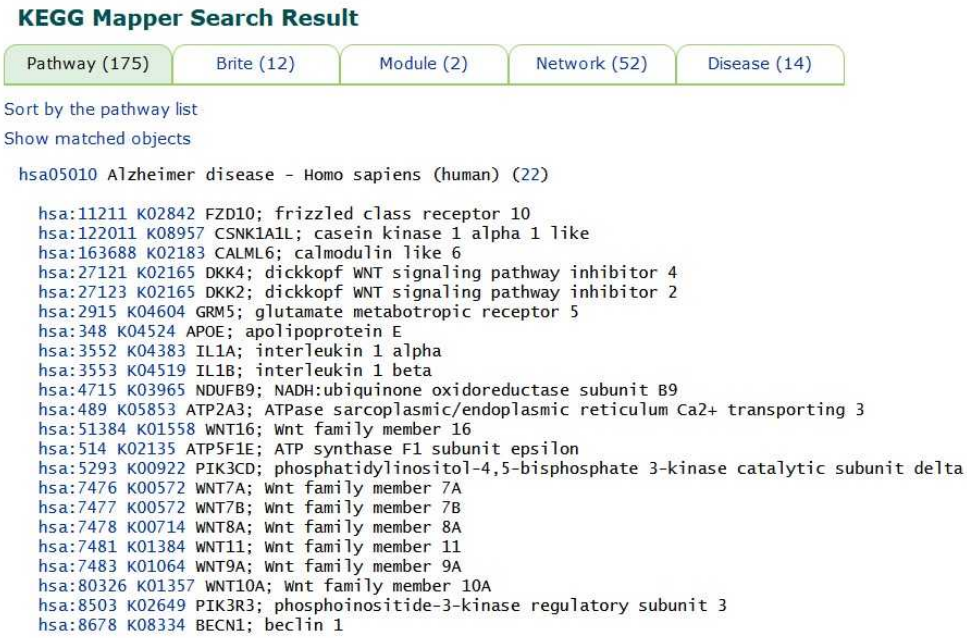
9 Utilization of the KEGG (Kyoto Encyclopedia of Genes and Genomes) database and its pathway analysis tool to investigate the functional significance of the differentially expressed genes identified in our study. KEGG is a comprehensive database that provides information on biological pathways, networks, diseases, and drugs, allowing researchers to elucidate the molecular mechanisms underlying complex biological processes. the list of differentially expressed genes obtained from our analysis as input for the KEGG pathway analysis. KEGG Mapper enables the visualization and interpretation of biological

pathways enriched with differentially expressed genes, providing insights into the biological processes and signaling pathways that are dysregulated under specific experimental conditions.

KEGG MAPPER



The screenshot depicts the KEGG Mapper interface utilized for pathway analysis. The interface provides options for inputting gene IDs, selecting analysis parameters, and visualizing enriched pathways. in the entry query, input a list of gene IDs obtained from differential gene expression analysis. These gene IDs represent the differentially expressed genes identified in the study. select mode hsa (Human Species)

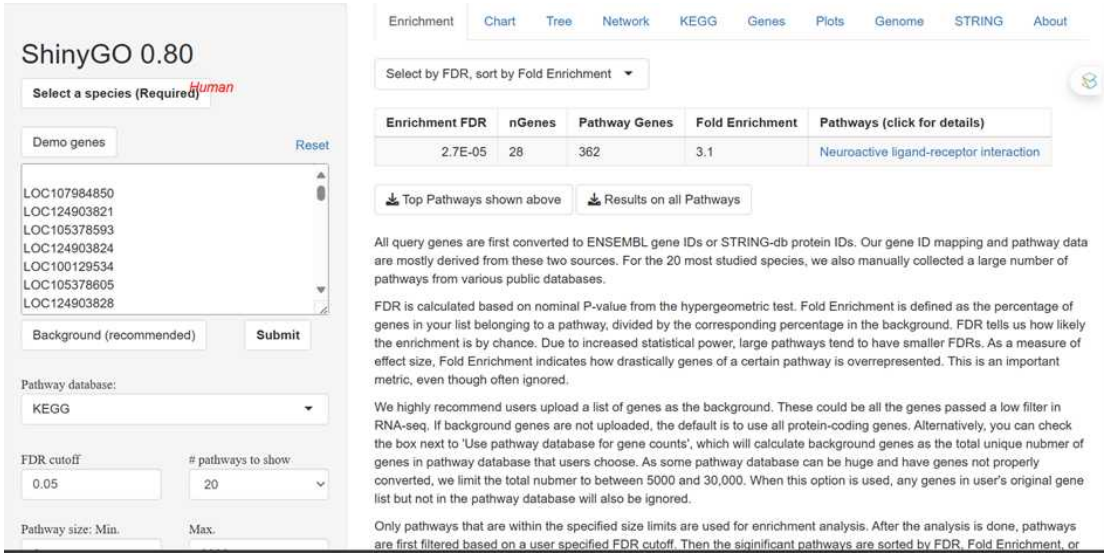


The screenshot illustrates the results obtained from KEGG Mapper analysis, displaying numerous enriched pathways associated with the input gene IDs.The list of enriched pathways identified through KEGG Mapper analysis is presented, showing the pathway names and their corresponding KEGG identifiers (IDs). Each pathway represents a specific biological process, cellular signaling pathway, or disease mechanism that is enriched with the input gene IDs. Key pathways that are significantly enriched with input gene IDs are identified, highlighting their importance in driving the observed changes in gene expression. These key pathways serve as focal points for further investigation into the molecular mechanisms underlying specific biological processes or disease pathways.

Enrichment and Statistical analysis of overexpressing genes

10 We utilized ShinyGO 0.8.0 to perform enrichment analysis of the input gene list obtained from our study. ShinyGO employs statistical algorithms to identify GO terms that are significantly enriched with the input genes compared to what would be expected by chance. This analysis enables the functional annotation of gene sets and helps uncover the biological themes and pathways represented by the input genes. considering factors such as gene count, background gene set, and false discovery rate (FDR) correction.

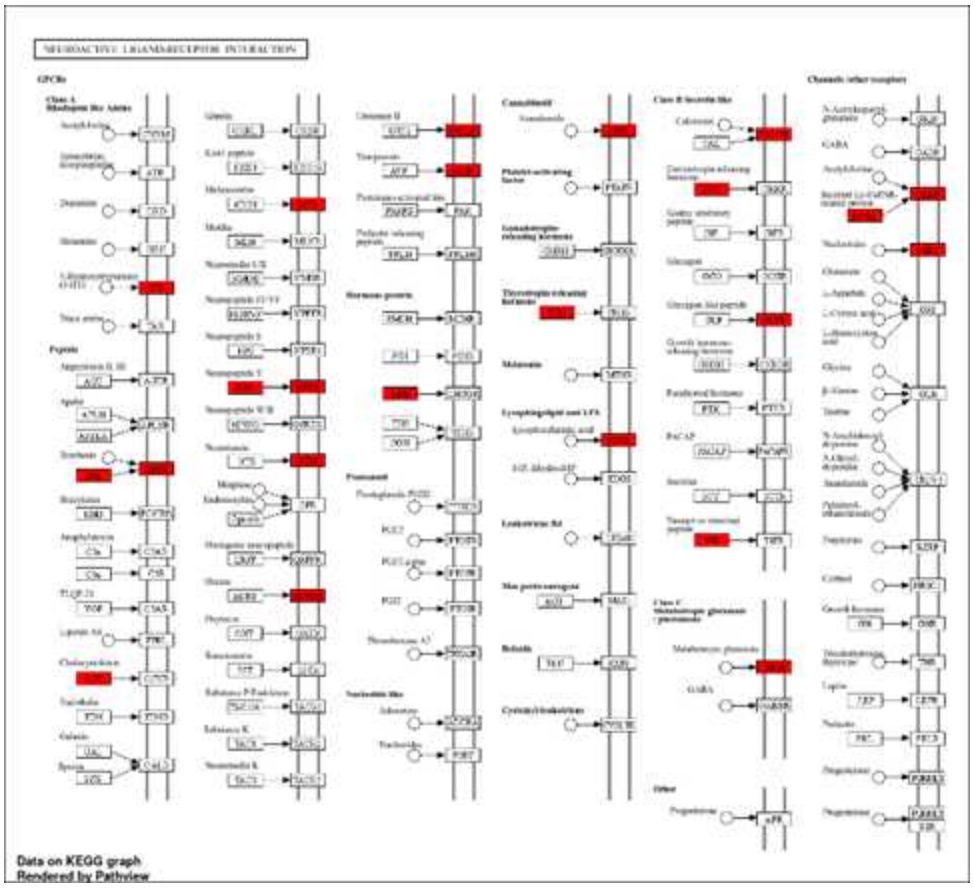
10.1 In the ShinyGO interface, users are provided with a text box or file upload option to input their list of gene identifiers. These gene identifiers can be in various formats, such as Entrez Gene IDs, Ensembl IDs, or gene symbols. Users can either manually enter the gene IDs or upload a file containing the list of gene IDs obtained from their study.



10.2

Upon performing enrichment analysis using ShinyGO, users receive a comprehensive summary of enriched Gene Ontology (GO) terms associated with the input gene list. The enrichment results provide valuable insights into the functional significance of the input genes by identifying overrepresented biological processes, molecular functions, and cellular components.

ShinyGO presents a list of enriched GO terms, categorized into biological processes, molecular functions, and cellular components. Each GO term is accompanied by statistical metrics indicating its significance level, such as p-values or false discovery rates (FDR).



The figure depicts a pathway network visualization generated from the enrichment analysis results obtained using ShinyGO. The network diagram illustrates the relationships between enriched pathways based on shared genes or functional associations.

Pathway Nodes:

Each node in the pathway network represents an enriched pathway identified through the analysis. Nodes are labeled with pathway names or identifiers, and their size may correspond to the significance level of the pathway enrichment.

Edge Connections:

Edges connecting pathway nodes indicate shared genes or functional relationships between pathways. Thicker edges represent stronger connections, suggesting a higher degree of gene overlap or functional similarity between pathways.

Node Attributes:

Nodes in the pathway network may be annotated with additional information, such as statistical significance scores or enrichment metrics, providing insights into the biological relevance of each pathway.

Cluster Analysis:

The pathway network may undergo cluster analysis to group related pathways into cohesive functional modules. Clusters of pathways with similar biological functions or regulatory mechanisms are visually highlighted within the network.

Interactive Exploration:

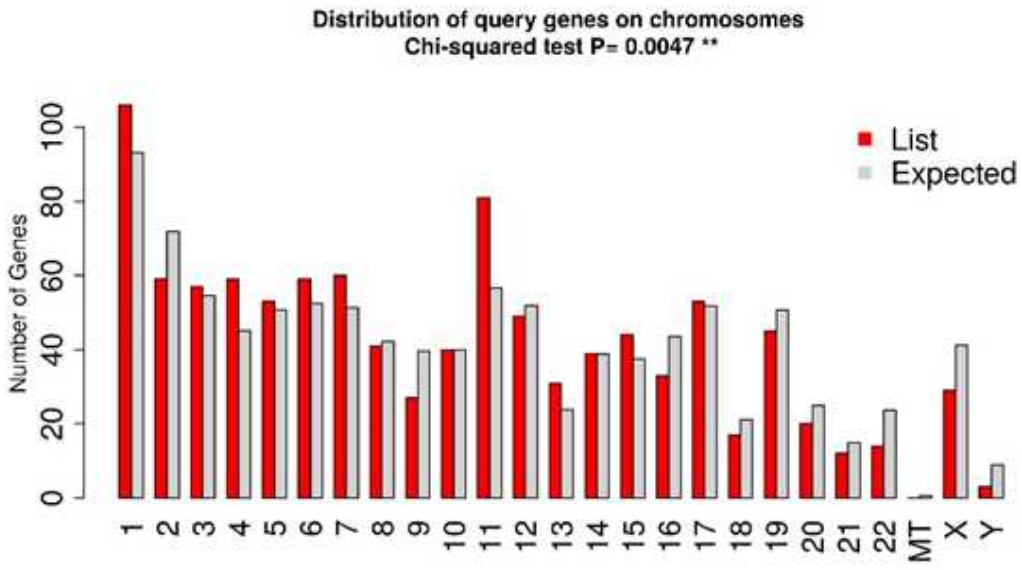
The pathway network visualization may offer interactive features that allow users to explore the network dynamically. Users can interact with nodes and edges to access detailed pathway information, visualize gene overlaps, and navigate between interconnected pathways.

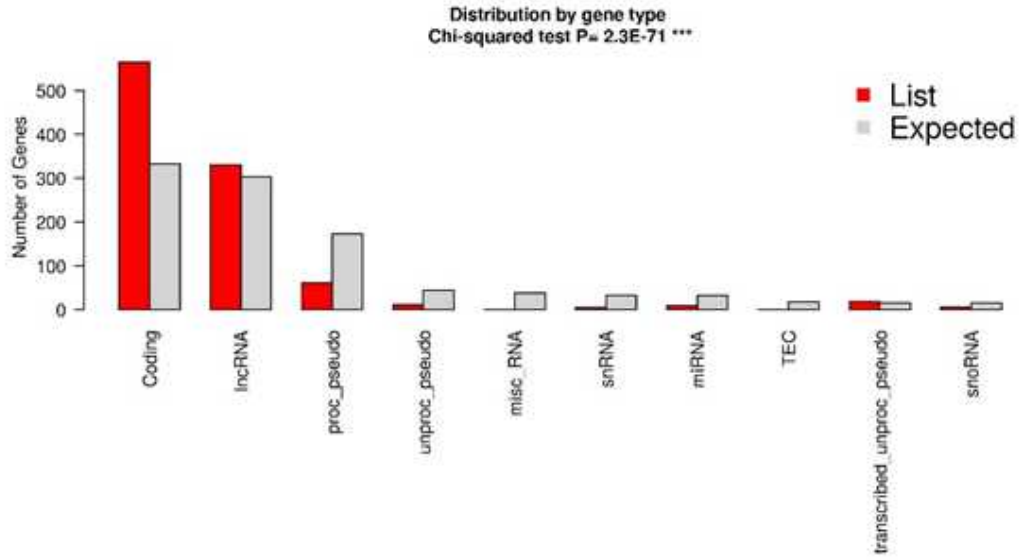
Biological Interpretation:

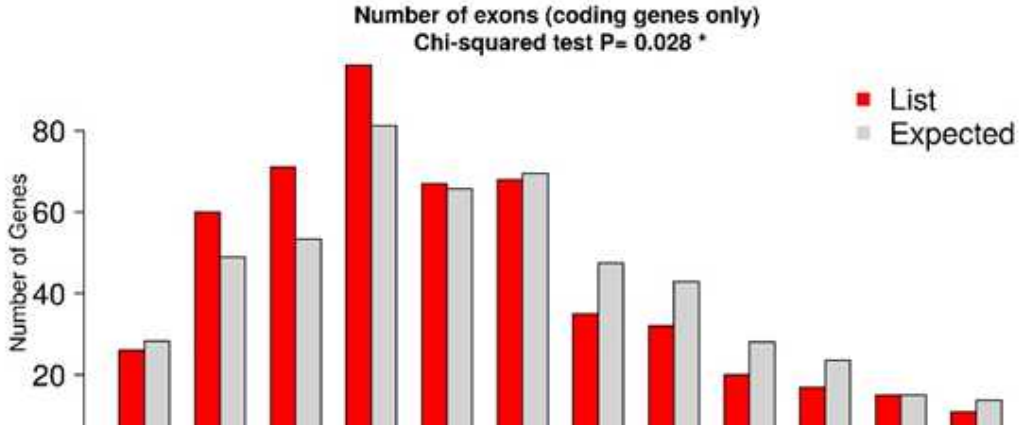
Pathway network analysis facilitates the biological interpretation of enrichment results by contextualizing enriched pathways within broader functional contexts. Users can

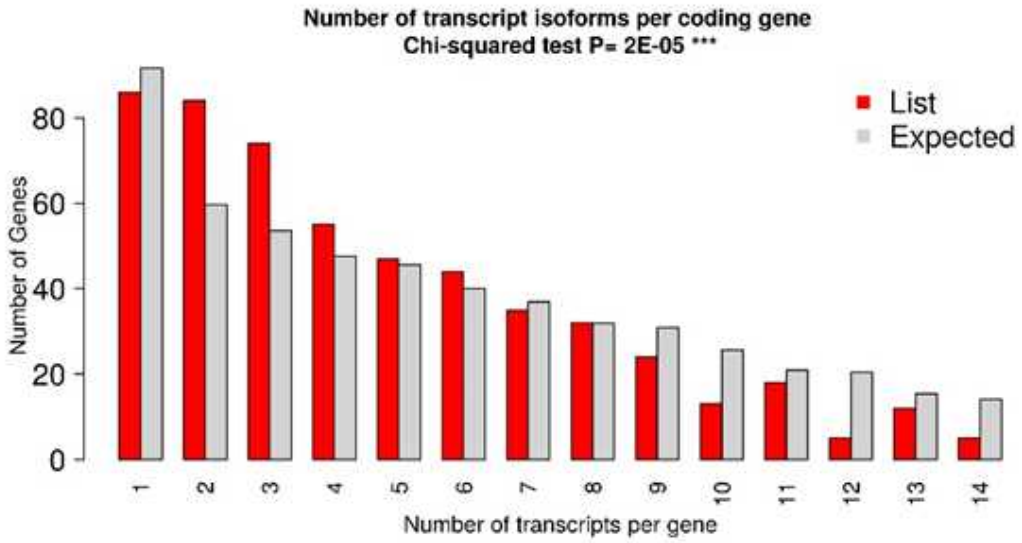
by contextualizing enriched pathways within broader functional contexts. Users can identify key pathway modules, hub pathways, or cross-talk interactions between pathways, providing insights into the underlying biological processes and regulatory mechanisms represented by the input gene list.

10.3 The chi-squared test assesses the association between categorical variables, such as gene annotations or GO terms, comparing the observed distribution of these characteristics among your genes with the expected distribution based on the entire genome. If the chi-squared test yields a low p-value, typically below 0.05, it indicates a significant difference in the distribution of gene characteristics between your genes and the rest of the genome. Similarly, Student's t-test compares the means of quantitative variables, such as gene expression levels or numerical gene features, between your genes and the broader genome. A low p-value from the t-test suggests a significant difference in the mean values of gene characteristics between your genes and the rest of the genome. Overall, if both tests yield low p-values, it suggests that your genes possess special characteristics, such as enrichment for specific annotations or unique expression patterns, distinguishing them from the general population of genes in the genome and potentially indicating their distinct biological roles.


Distribution of query genes on chromosomes
Chi-squared test P= 0.0047 **
Number of Genes
0
20
40
60
80
100
List
Expected
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
MT
X
Y

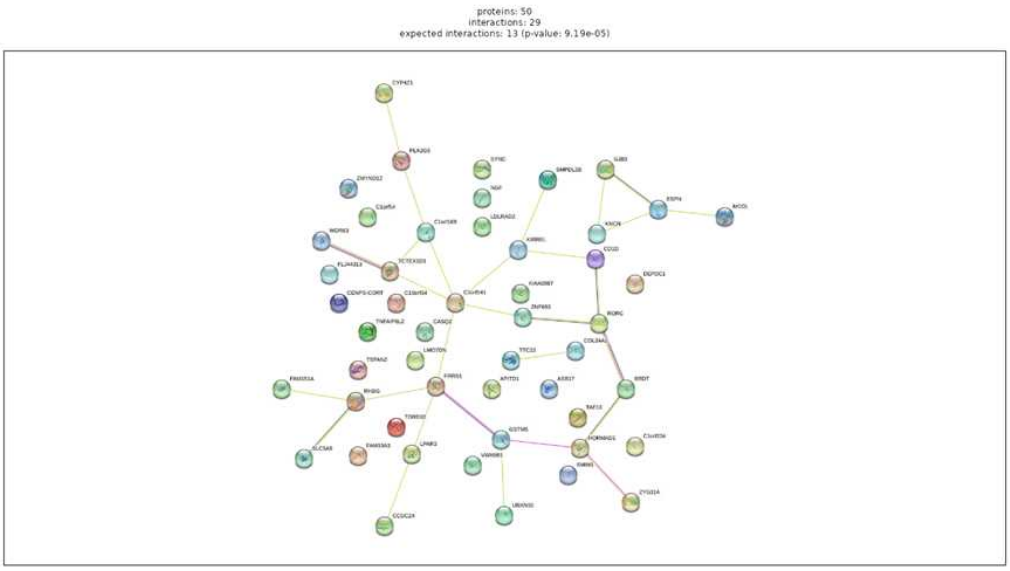

Distribution by gene type
Chi-squared test P= 2.3E-71 ***
Number of Genes
0
100
200
300
400
500
List
Expected
Coding
lncRNA
proc_pseudo
unproc_pseudo
misc_RNA
snRNA
miRNA
TEC
transcribed_unproc_pseudo
snoRNA


Number of exons (coding genes only)
Chi-squared test P= 0.028 *
Number of Genes
20
40
60
80
List
Expected



PPI Network Construction

11 The PPI network construction and hub genes analysis using ShinyGO provide valuable insights into the molecular mechanisms underlying the observed upregulation of genes, helping to elucidate key regulatory pathways and potential therapeutic targets in the studied biological system.The top-upregulated genes are inputted into the platform to construct a PPI network. ShinyGO utilizes existing databases and algorithms to predict protein-protein interactions based on known protein interactions, protein domains, or functional annotations. The resulting PPI network represents the interconnectedness between the proteins encoded by the top-upregulated genes, highlighting potential functional relationships and pathways involved.



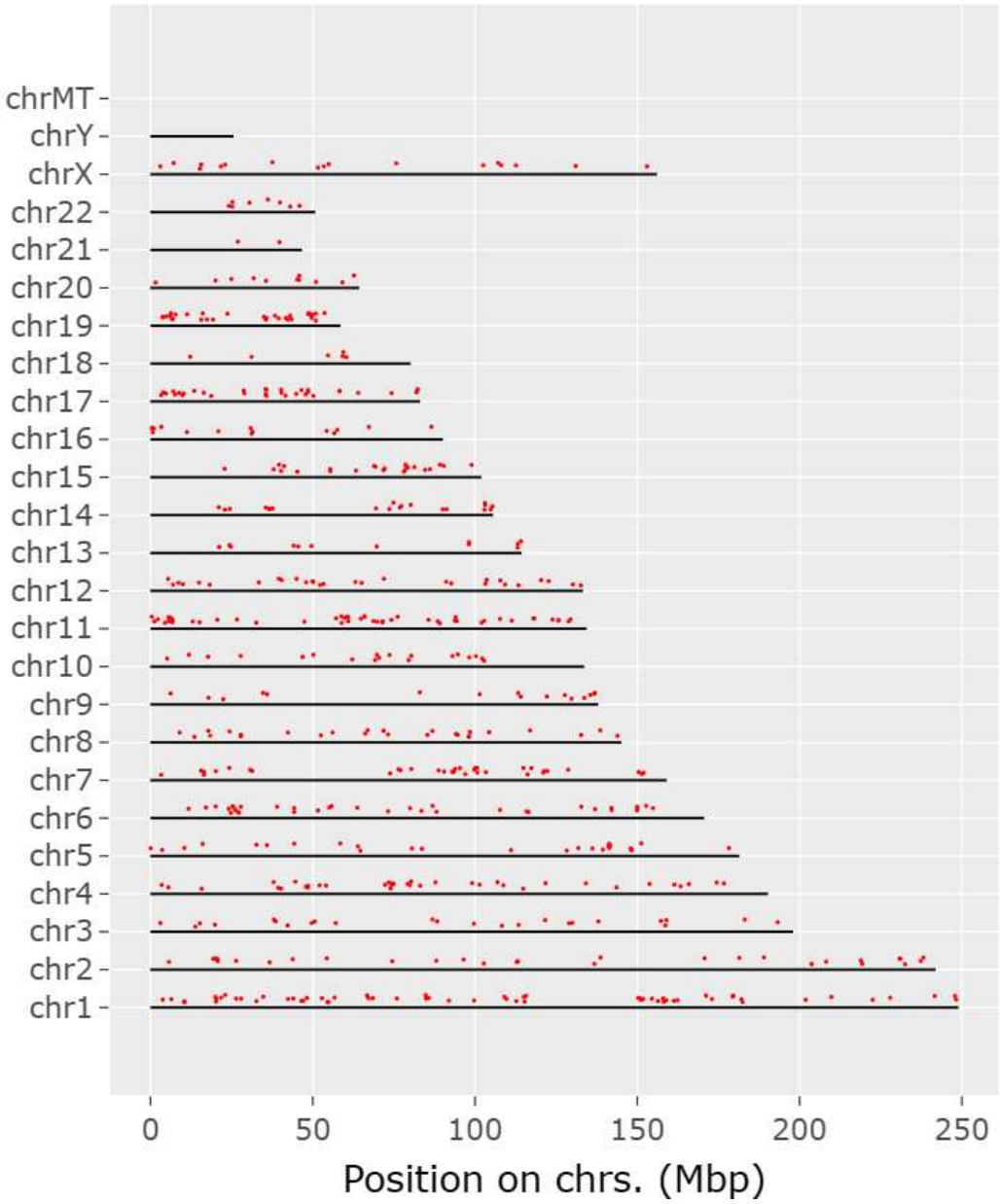
The protein-protein interaction (PPI) network visualized using ShinyGO depicts the interactions among proteins encoded by the top upregulated genes identified from differential expression analysis. Each node represents a protein, with node size indicating its degree of connectivity. Edges between nodes represent predicted or known interactions, with thicker lines indicating stronger interactions. Hub proteins, characterized by high connectivity, are highlighted within the network. Functional annotations or enriched pathways associated with the proteins are overlaid onto the network, providing insights into the biological processes or molecular functions represented.

Chromosomal enrichment

12 The chromosomal positions of genes are visualized using ShinyGO. Genes are represented as red dots on the chromosomes, providing a spatial overview of their distribution across the genome. Purple lines highlight regions where these genes exhibit statistically significant enrichment compared to the background gene density.

The genome is scanned using a sliding window approach, with each window subdivided into equal-sized steps for sliding. Within each window, a hypergeometric test is employed to assess whether the genes are significantly overrepresented. Essentially, the genes within each window define a gene set or pathway, and enrichment analysis is conducted to identify statistically enriched regions.

Please note that the chromosomes may only be partially shown, as the last gene's location is utilized to draw the line. This analysis offers insights into the genomic localization and enrichment patterns of the genes, shedding light on potential chromosomal hotspots or regions associated with specific biological processes or pathways.



The figure shows genomic position enrichment analysis results using ShinyGO. Red dots represent genes plotted along chromosomes, while purple lines indicate regions of statistical enrichment compared to background gene density.

EXPECTED OUTCOME OF THE PIPELINE

13 The expected outcome of the study is the comprehensive characterization of top-upregulated genes in Alzheimer's disease, elucidating their roles through gene ontology (GO) analysis, pathway enrichment, and disease intervention pathways. Through rigorous statistical analysis utilizing the tools employed

above, the study aims to identify novel genes associated with Alzheimer's disease pathology. The resulting document will provide detailed functional information on these novel biomarkers, shedding light on their involvement in disease mechanisms and offering insights for potential therapeutic interventions.

Expected result

1. WNT9A - Protein Wnt-9a; It's a Ligand for members of the frizzled family of seven transmembrane receptors. Functions in the canonical Wnt/beta-catenin signaling pathway. It is required for normal timing of IHH expression during embryonic bone development, normal chondrocyte maturation, and for normal bone mineralization during embryonic bone development. It plays a redundant role in maintaining joint integrity.

2. FZD10 - Frizzled-10; Receptor for Wnt proteins. canonical Wnt Functions in the canonical Wnt/beta-catenin signaling pathway. The canonical Wnt/beta-catenin signaling pathway leads to the activation of disheveled proteins, inhibition of GSK-3 kinase, nuclear accumulation of beta-catenin, and activation of Wnt target genes.

3. Probable developmental protein.
It may be a signaling molecule that affects the development of discrete regions of tissues. Is likely to signal over only a few cell diameters.

4. DKK4 - Dickkopf-related protein 4;
DKKs
play an important role in vertebrate development, where they locally inhibit Wnt-regulated processes such as anteroposterior axial patterning, limb development, somitogenesis, and eye formation. In adults, Dkks are implicated in bone formation and bone disease, cancer, and Alzheimer's disease.
5. SEM1 - Putative
protein SEM1, isoform 2; SEM1 26S proteasome complex subunit.