Comparative analysis of gene expression in Diabetic vs Non-Diabetic Obese individuals

Durga M*, *Harish* R, *Hindhupriya* K, *Vasanth* T T, *Puniethaa* Prabhu

Department of Biotechnology, K. S. Rangasamy College of Technology, Tiruchengode, Tamil Nadu, India.

Abstract. Diabetes and obesity are closely interconnected epidemics with shared and distinct molecular mechanisms. In this analysis, a comparative gene expression study was carried out among diabetic and non-diabetic obese patients using Gene Expression Omnibus dataset (GSE132831), containing 104 diabetic and 120 non-diabetic obese Patients. Differential Expressed Genes were identified, which results of 509 upregulated genes and 33,885 downregulated genes. Further Gene Ontology terms, such as Biological Process, Cellular Component and Molecular Function were analysed. Visualization of Differential Expressed Genes and pathway enrichment analysis indicated significant associations with metabolic and immune signaling pathways, and fold enrichment Analyses highlighted critical differences in gene activity between the two groups. Protein-Protein Interaction Network is generated, which showed highly connected clusters. These clusters identified the target genes Transmembrane Immune Signaling Adapter Protein (TYROBP) and Receptor for Activated C Kinase (RACK1) with their respective networks. These were considered potential targets for treatment because of the central Positions they occupy in metabolic and immune regulatory pathways associated with diabetic obesity.

Keywords: Gene Expression Omnibus, Diabetes, Obesity, Differential Expressed Genes, Protein-Protein Interaction, Gene Ontology, Clusters

* Corresponding author: durgamurugan433@gmail.com

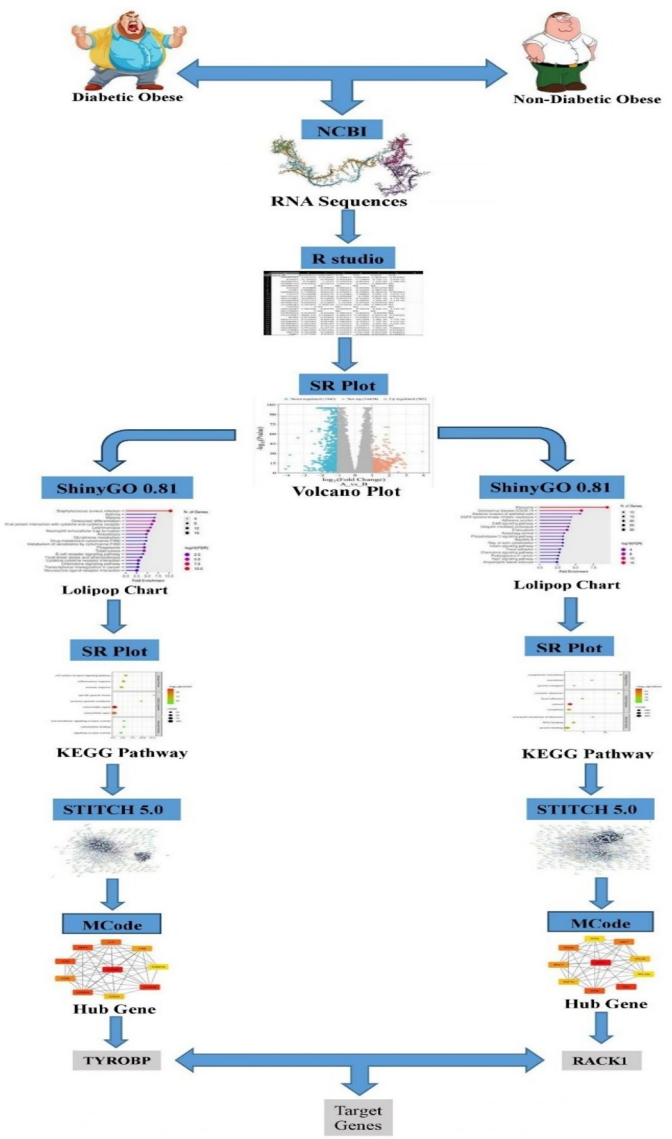

Fig. 1. Graphical abstract – Overall process

1 Introduction

Obesity is a multifactorial condition with a strong association with many metabolic disorders. The most common among these conditions is type 2 diabetes mellitus. The link between obesity and diabetes has been called "diabesity," as the relation between adiposity, insulin resistance, and glucose dysregulation is highly intricate. Diabetic and non-diabetic obesity share common mechanisms, such as chronic inflammation and altered lipid metabolism, but have different molecular and genetic profiles. Understanding these differences is very significant in developing pinpointed interventions and refining therapeutic strategies. RNA sequencing has transformed the field of transcriptomics; it has provided for very high-resolution analyses of expression patterns of genes in most conditions. With the development of this technology, one may identify DEGs and can understand pathways and regulatory networks driving specific phenotypes. The statistical power and reliability of the findings are

enhanced by the meta-analysis of publicly available RNA-Seq datasets in the NCBI Gene Expression Omnibus (GEO) database. This study will carry out a comparative gene expression profiling between diabetic and non-diabetic obese subjects. High-end bioinformatics tools like David, STRING (Alberto Luiz P. Reyes, 2019), Cytoscape, and MCODE (version 2.0.3) are going to be used for identifying crucial genes, pathways, and protein-protein interactions that might distinguish these phenotypes. The outcome is expected to elucidate the molecular basis of diabesity by pointing out the potential biomarkers and therapeutic targets that may tailor the approaches for treatments. The analysis revealed a gene expression pattern that is uniquely different in the phenotypes, such as TYROBP and RACK1 as central hubs in the PPI networks. These genes, along with their associated clusters, might play pivotal roles in the progression of diabetic obesity. This study integrates advanced bioinformatics tools for data analysis and visualization to provide a detailed molecular understanding of the differences between diabetic and non-diabetic obesity. These findings are not only setting the differences in pathophysiology between these conditions but also establishing grounds for the discovery of new therapeutic targets and biomarkers to deal with the mounting burden of obesity and the related metabolic disorders.

2 Materials and Methods

2.1 Dataset Retrieval

The dataset, via the NCBI Gene Expression Omnibus (GEO) database https://www.ncbi.nlm.nih.gov/gds under accession GSE132831, is an RNA-Seq dataset giving the transcriptomic profiles for enriched enteroendocrine cells (EECs) taken from the jejunum of obese individuals with and without type 2 diabetes mellitus (T2DM) to display differences in gene expression and pathways underlying diabetic and non-diabetic obesity. The dataset includes raw counts or normalized expression data for 224 samples that consist of 104 diabetic obese and 120 non-diabetic obese individuals, important information regarding molecular mechanisms underpinning these conditions. (G, 2023)

2.2 Differential Gene Expression Analysis

The raw count data is analyzed after importing to RStudio and normalization with some tools such as DESeq2 or edgeR for counting library size differences. Lastly, DEG analysis can be done for log2 fold change (Log2FC) and p-value calculations. The results are filtered based on threshold Log2FC, which extracts the upregulated genes with log fold change > 1 and downregulated genes where log fold change <1. Lastly, lists containing the DEGs along with their identifiers and expression value are saved for further further analysis and interpretation.

2.3 Conversion of Gene IDs to Symbols and Functional Annotation

The DAVID tool https://davidbioinformatics.nih.gov/ (Brad T. Sherman, 2022)was used to analyze the DEGs. Lists of both upregulated and downregulated genes were uploaded. For the upregulated DEGs, a total of 509 genes were submitted, and the identifier type was set to Entrez ID. DAVID translated these gene IDs into gene symbols, and the result was two groups: 502 genes were successfully mapped to Homo sapiens, while 7 genes were classified as not mapped or unidentified. Similarly, for the downregulated DEGs, a list of 33,885 genes was uploaded. Of these, 32,326 genes were identified as Homo sapiens, whereas 1,559 genes were classified as unknown. The results for the known genes in both datasets were downloaded for further annotation and analysis.

2.4 Gene Ontology and Kyoto Enrichment analysis

A DAVID-enriched analysis is done using functional enrichment, and the results of the same are visualized in an SRP plot. The analysis is limited to GO terms, broadly categorized into three groups, namely Cellular Component (CC), Molecular Function (MF), and Biological Process (BP). (Babu, 2023) The statistically significant terms were determined based on p-values ≤ 0.05. Among them, the first 10 terms with maximum significance value in each category is CC, selected, including MF, and BP, were selected for further analysis and interpretation. Among the top 10 from each term, the top 3 are taken for Kyoto Enrichment analysis to analyze the enrichment value for each process

2.5 Visualization of DEGs (Volcano Plot and Pathways)

SRP plots https://www.bioinformatics.com.cn/en are used to determine the volcano plots by de-merging the lists obtained from DEG with Log2FC on the X-axis and -log10 (p-values) on the Y-axis. (Doudou Tang, 2023) Significant upregulation and downregulation genes can be obtained through plots. KEGG pathway enrichment analysis of the DEGs is carried out in order to discover those critical pathways that might involve the disease. The relevant pathways are highlighted and visualized in order to describe the molecular mechanisms of those diseases to further investigation and the possible therapeutic targeting.

2.6 Fold Enrichment Analysis

DEGs that were identified in the analysis from RNA-Seq are classified into upregulation and downregulation groups, which are further uploaded to the Shiny application https://bioinformatics.sdstate.edu/go/ for further analysis. Fold enrichment is used in order to determine the degree of GO terms and KEGG pathways. Enriched terms and pathways are visualized, keeping an eye on those concerning the metabolic pathways of study. This method, hence, provides an accurate description of the biological process and pathways that have caused the observed changes in gene expression.

2.7 PPI network and Hub gene analysis

Retrieved the PPI networks between the DEGs from STRING database http://stitch.embl.de/. The STRING database gives interaction scores based on evidence from experiments, co-expression, and text mining and curated databases. The interaction data were exported as TSV file and then imported into Cytoscape version 3.9.1, (Paul Shannon, 2001) which is the software platform for visualizing molecular interaction networks. In Cytoscape, nodes refer to proteins that are gene products, whereas edges represent their associations. Target proteins are represented by nodes in protein-protein interaction (PPI) networks, whereas interactions between proteins are represented by edges. An edge's thickness is directly correlated with the interaction's total score. The number of nodes that are directly connected to a node is known as its degree. A node with a higher degree was considered more significant. After **it** was developed, this network was then loaded into Cytoscape https://cytoscape.org/.so that its structure could be seen and examined. The Cytoscape website can be accessed to get the CytoHubba software https://apps.cytoscape.org/apps/cytohubba. Using CytoHubba, the degree was computed to determine the primary targets. The MCODE plugin was used to determine the densely connected subregions within the network, which refer to functional groups of genes. MCODE ranked clusters based on node connectivity and size besides picking out some of the most important groups of genes. Important clusters were highlighted to include TYROBP and RACK1 to be considered as having the potential role in diabetic

obesity. Visualization in Cytoscape further enabled the evaluation of these clusters in a very detailed manner and simply presents candidate gene networks through graphs. The integrated analysis shall be able to identify key hubs and pathways that direct the interpretation in biology as well as targeting therapies.

3 Result and Discussion

3.1 Differential Gene Expression Analysis

To preliminarily understand the mechanism contributing to the diabetic obesity, 224 patients (104 diabetic obese patients and 120 Non-diabetic obese patients) were selected for subsequent analysis. The Differential expressed genes of diabetic obese VS Non-diabetic obese samples from Rstudio were examined using volcano plot. The volcano plot represents the expressed fold change of genes in obese VS Non-diabetic obese samples were plotted against the degree of statistical significance in differential expression (Fig 2). A total of 35953 DEGs with a threshold criterion of log 2FC (> 1 for upregulated and < 1 for downregulated) and p-value less than 0.05 as the cut-off point were diagnosed. Among them, 509 genes were upregulated (7 remain uncharacterized) and 33885 genes were downregulated (1559 remain uncharacterized).

3.2 Gene Ontology and Kyoto Enrichment analysis of Genes and Genomes

The gene ontology resources showed fold enrichment pathways for up and downregulated genes (Fig 3, Fig 4). The KEGG pathway analysis of DEGs for diabetic vs non-diabetic obese subjects revealed biological pathways. The genes that are overexpressed are highly enriched within inflammatory and immune-related pathways, including cytokine-cytokine receptor interaction, suggesting an enhanced state of inflammation with diabetic obesity. Downregulated genes, however, were linked with disturbed metabolic pathways such as degradation of fatty acids, TCA cycle, and oxidative phosphorylation, pointing towards deranged energy metabolism (Fig 5, Fig 6). These pathways were related to insulin signaling and glucose metabolism, thus indicating that there is a dual role for inflammation and metabolic dysfunction in diabetic obesity.

3.3 Construction and analysis of PPI network and Hub genes

STRING database analysis of DEGs in diabetic compared to non-diabetic obese individuals revealed different PPI networks (Fig 7, Fig 8). The up-regulated genes formed a strongly interconnected network with 347 nodes, 2,106 edges, an average node degree of 12.1, and a clustering coefficient of 0.489, which implies strong immune-related interactions. The down-regulated genes formed a network with 477 nodes, 2,538 edges, an average node degree of 10.6, and a clustering coefficient of 0.341, showing metabolic pathway disruptions. Both networks had more interactions than predicted, which highlighted their biological relevance in diabetic obesity. To draw the networks, cytoscape, along with its MCODE plugin, were used on PPIs of the DEGs to extract from the STRING database. Thus, crucial nodes and hubs in diabetic obesity were pinpointed. This up-regulated network comprising 347 nodes and 2,106 edges underlined activation of the immune system along with inflammation with TYROBP being key (Fig 9). A downregulated network consisting of 477 nodes and 2,538 edges with RACK1 as

central node highlighted its importance (Fig 10). MCODE analysis provided these clusters with biological significance, which shed light on the molecular mechanisms of diabetic obesity.

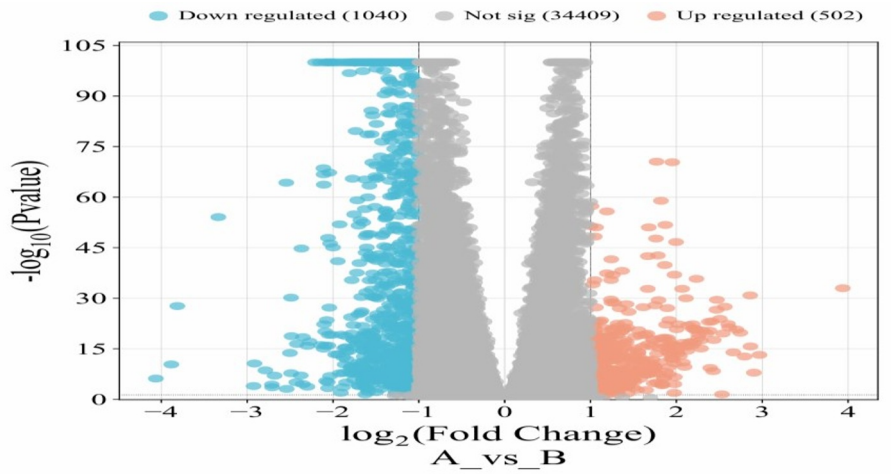

Fig. 2. Volcano plot of Differential Expressed Genes.

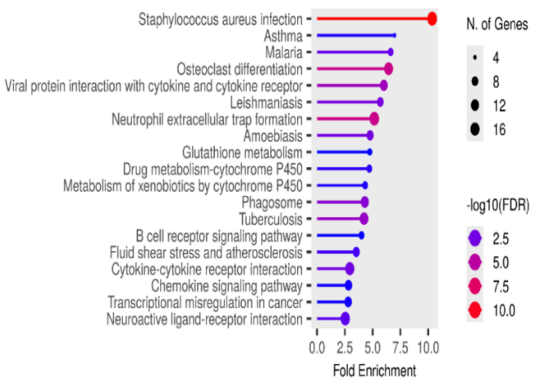

Fig. 3. Fold enrichment pathway for upregulated genes

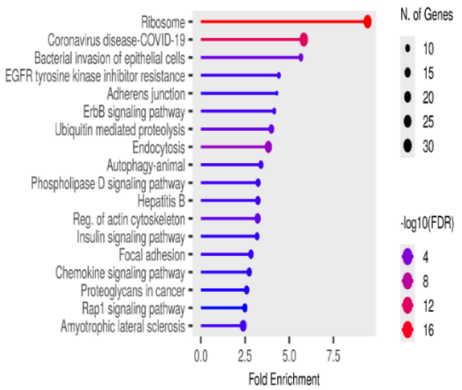

Fig. 4. Fold enrichment pathway for downregulated genes

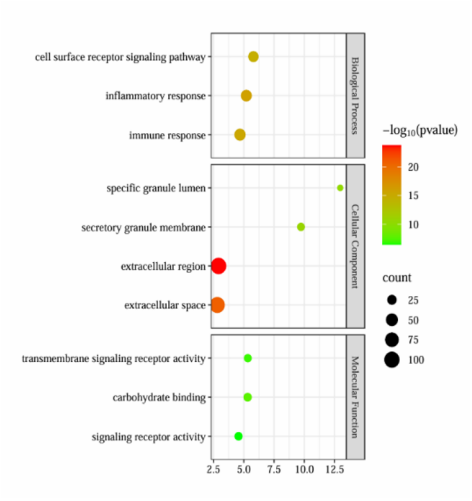

Fig. 5. KEGG Pathway Enrichment upregulated genes

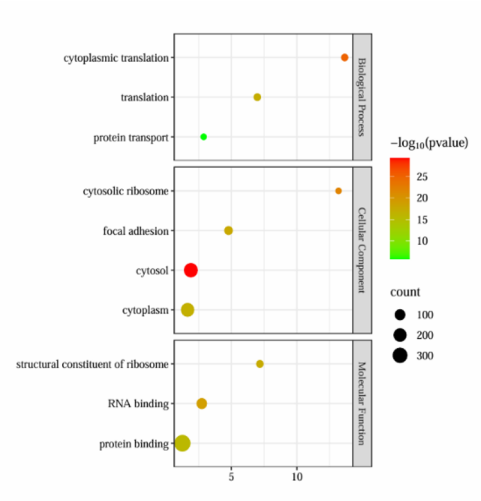

Fig. 6. KEGG Pathway Enrichment for downregulated genes

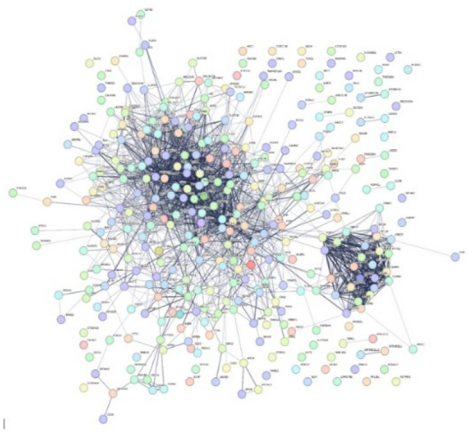

Fig. 7. PPI Network of upregulated genes

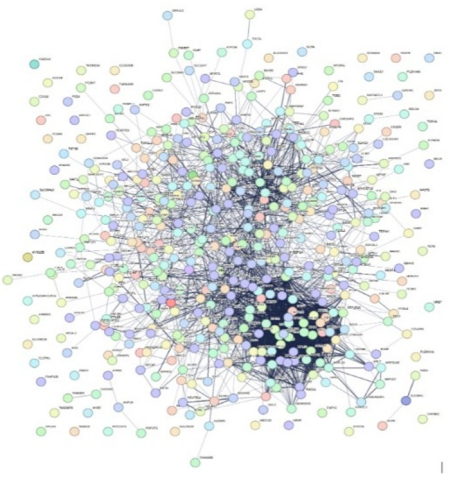

Fig. 8. PPI Network of downregulated genes

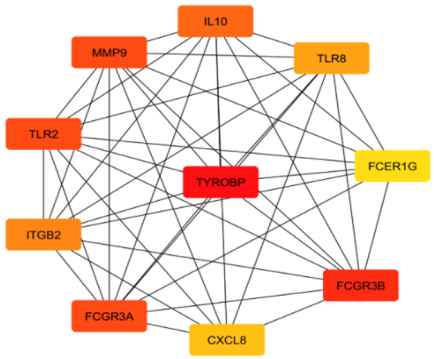

Fig. 9. The Cluster 1(TYROBP)

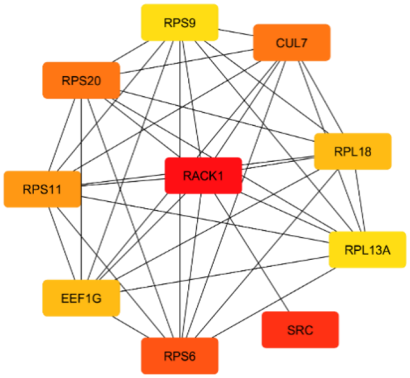

Fig. 10. The Cluster 2(RACK1)

4 Discussion

Highly significant disparate expression profiles of the expression in the obese diabetic and non-diabetic subjects appeared under the differential gene expression that included 509 upregulated, 33,885 down-regulated genes among which 7 up-regulated were and 1,559 uncharacterized down-regulated. Up-regulated genes with Log2 FC >1 indicate that they increased biological activity in pathways possibly underlying diabetic obesity, while genes with Log2FC < 1 are associated with pathway suppression. GO enrichment analysis categorized the DEGs under cellular components, molecular functions, and biological processes. Membrane-associated structures were some of the notable terms in the category of cellular components. MF terms highlighted receptor activity and signal transduction. BP terms revealed involvement in immune response regulation and metabolic processes, indicating their potential roles in obesity and diabetes pathophysiology. Volcano plot visualizations using SRP plots have highlighted significant DEGs, thus helping us identify key pathways through the analysis of KEGG. Fold enrichment analysis with Shiny has illustrated specific functional enrichments of upregulated and downregulated genes that emphasize differential pathway activation. The PPI network based on STRING database and visualized using Cytoscape has given a comprehensive view of molecular interactions among DEGs. MCODE clustering identified densely connected regions, and TYROBP and RACK1 were the central hubs. TYROBP is associated with immune signaling, and RACK1, implicated in kinase activation, outline the immune and signaling pathways in diabetic obesity. These results indicate that TYROBP and RACK1, and their interacting genes, are crucial components of the molecular etiology of diabetic obesity. Such Clusters represent fascinating potential therapeutic targets for interventions with the metabolic dysregulations of diabetes and obesity. This integrated approach, in particular, demonstrates the utility of multi-platform bioinformatics tools in better understanding complex disease mechanisms and discovery of novel candidate genes.

5 Conclusion

This study analyzed data from GEO about RNA-Seq to compare gene expression in diabetic and nondiabetic obese, suggesting 509 genes up-regulated and 33,885 downregulated. The upregulated genes were associated with the activation of the immune response, and the genes downregulated mirrored metabolic dysregulation. Further, GO and KEGG pathway analyses

confirmed that the top dysregulated pathways were immune regulation, signaling, and metabolism. STRING-based PPI networks in Cytoscape show that TYROBP is a central node of the immune regulation network and RACK1 in the metabolic regulation network as the central node. The results of MCODE clustering enhanced the potential biomarkers or therapeutic targets. It integrates bioinformatics to the clinical applicability of molecular research. With the approach, one can really do personalized treatment of diabetic obesity.

Acknowledgement

This study does not involve human participants, animals, or sensitive data requiring ethical approval or consent to participate. The data and materials used in this study were obtained from freely available online tools and resources, and can be accessed through the respective web platforms. The authors declare that they have no competing interests. No specific funding was received for this research work.

List of Abbreviations

GEO: Gene Expression Omnibus; DAVID: Database for Annotation, Visualization and Integrated Discovery; GO: Gene Ontology; CC: Cellular Component; MF: Molecular Function; BP: Biological Process; Log2FC: Log2 Fold Change; KEGG: Kyoto Encyclopedia of Genes and Genomes; PPI: Protein-Protein Interaction; SRP: Signal Recognition Particle; MCODE: Molecular Complex Detection (a Cytoscape plugin); RACK1: Receptor for Activated C Kinase; TYROBP: Transmembrane Immune Signaling Adaptor Protein.

References

1. Alberto Luiz P. Reyes, GENAVi: A shiny web application for gene expression normalization, analysis, and visualization, BMC Genomics (2019). DOI:10.1186/s12864-019-5909-2
2. L. P. & D. J. D. G., Meta-analysis of lean and obese RNA-seq datasets to identify genes targeting obesity, Biomedical Informatics (2023). DOI:10.1016/j.bmcbio.2023.101423
3. Brad T. Sherman, DAVID: A web server for functional enrichment analysis and functional annotation of gene lists, Nucleic Acids Research, vol. 50 (2022). DOI:10.1093/nar/gkaa1115
4. P. B. R. Babu, Prediction of anti-microtubular target proteins of tubulins and their interacting proteins using Gene Ontology tools, Genetic Engineering and Biotechnology (2023). DOI:10.1080/13102818.2023.2046820
5. Doudou Tang, SRplot: A free online platform for data visualization and graphing, PLOS ONE (2023). DOI:10.1371/journal.pone.0280457
6. Paul Shannon, Cytoscape: A software environment for integrated models of biomolecular interaction networks, Bioinformatics (2001). DOI:10.1093/bioinformatics/17.13.133