Identification of Key Signaling Pathways and Novel Computational Drug Target for Depression and Coronary Artery Disease

Conference Paper · September 2024

DOI: 10.1109/COMPAS60761.2024.10796501

5 authors, including:

Md. Faruk Hosen
Daffodil International University
17 PUBLICATIONS **47** CITATIONS
SEE PROFILE

Md. Abul Basar
Uttara University
12 PUBLICATIONS **35** CITATIONS
SEE PROFILE

Mst. Farjana Yasmin
Daffodil International University
5 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Monir Morshed
Mawlana Bhashani Science and Technology University
62 PUBLICATIONS **961** CITATIONS
SEE PROFILE

2024 IEEE International Conference on Computing, Applications and Systems (COMPAS), September 25-26, 2024, Bangladesh

Identification of Key Signaling Pathways and Novel Computational Drug Target for Depression and Coronary Artery Disease

Md. Faruk Hosen[a,b], Md. Abul Basar[a,c,*], Mst. Farjana Yasmin[b], Monir Morshed[a], Muhammad Shahin Uddin[a]

[a]*Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Tangail 1902, Bangladesh.*

[b]*Department of Computing and Information System, Daffodil International University (DIU), Ashulia, Dhaka, Bangladesh.*

[c]*Department of Computer Science and Engineering, Uttara University (UU), Holding 77 Beribadh Road, Dhaka 1230, Bangladesh.*

Author's E-mail: farukictmbstu@gmail.com; a.basar@uttarauniversity.edu.bd; farjana16-502@diu.edu.bd; monirmorshed.ict@mbstu.ac.bd; shahin.mbstu@gmail.com

***Abstract*—Psychological disorders, such as anxiety, bipolar disorder, panic disorder, stress, depression, and schizophrenia, are increasingly prevalent worldwide. Among these conditions, depression is particularly notable as one of the most common and debilitating neuropsychiatric disorders. Individuals with depression may be at a higher risk of developing oronary Artery Disease (CAD). Depression can contribute to poor lifestyle choices, such as unhealthy eating, lack of exercise, and smoking, which are risk factors for CAD. The emotional stress and anxiety associated with depression can strain the heart and exacerbate CAD symptoms. The relationship is not one-sided. CAD itself can be a significant source of emotional distress, leading to symptoms of depression and anxiety in affected individuals. Managing both conditions in tandem can be complicated. Treating CAD may involve medications, lifestyle modifications, and potentially surgical interventions. Meanwhile, depression often requires therapy, counseling, and medication. Coordinating care and addressing both conditions simultaneously is crucial. In our study, we investigated the molecular connections between CAD and Depression using GSE98793 and GSE20681 microarray datasets. After preprocessing, we identified key hub genes, including CCT2, SVIL, REPS2, ASPH, and UBC, in the shared ProteinProtein Interaction network. KEGG pathways linked these DEGs to colorectal and cancer pathways. Our next steps involve exploring microRNAs, TFs, and GO analysis. These findings offer promising leads for potential therapies, uniting CAD and Depression under a common molecular framework, advancing our understanding of these conditions.**

***Index Terms*—Bio-informatics, System Biology, Depression, DEGs, Protein-protein interaction, Hub genes, CAD, Drug molecule.**

I. Introduction

Approximately 5% of individuals worldwide suffer from depression, a common mental health condition. In comparison to males, more women than men often experience depression. Suicide can be caused by depression. Mild, moderate, and severe depression can all be successfully treated. Symptoms of depression, a mood disorder, include a protracted feeling of melancholy and loss of interest. Major depressive disorder, also known as clinical depression, affects emotions, thoughts, and behaviors, often resulting in a range of emotional and physical problems. Numerous factors contribute to the development of CAD and aggravate depression [1]. Both CAD and depression are extremely widespread conditions. Furthermore, these situations all share a decreased quality of life and lifespan. When co-occurring with a chronic medical disease, depression has a particularly significant influence on impaired health and is swiftly overtaking other causes of disability in the globe [2]. Two disorders were chosen for investigation in this research. The selection of these two different conditions was made in an effort to highlight their links. The risk decreases considerably when an individual is dealing with just a single health condition. On the other hand, the likelihood rises significantly when an individual has multiple health issues. The aim of this study is to explore the genetic connections between these genes and to uncover a genetic relationship among these disorders. The findings from these studies highlight concerns about the relationship between CAD and the use of high-throughput methods, which have made considerable progress through the analysis of microarray data and the insights gained from it [3]. Significant advancements in high-throughput techniques have been achieved through the analysis of microarray data and the insights provided by that expression dataset. A total of 18 DEGs were identified through gene screening in the GSE23848 and GSE58294 datasets, with both datasets revealing these shared findings. We then built a PPI network to show how intertwined the DEGs are even further. The PPI network's hub genes were located and rated using a degree topological technique. A degree topological

979-8-3315-2976-5/24/$31.00 ©2024 IEEE

technique was used to establish the five top-ranking hub genes utilizing the PPI network's identification of hub genes as a foundation. It has been the subject of several bioinformatics efforts to combine similar DEGs and find particular therapeutic compounds based on those DEGs. In addition, this research includes the evaluation of widespread DEGs, which can help identify other biological pathways and the gene ontology (GO). We suggest a set of predictive pharmaceutical substances that are similar for CAD and depression after identifying important nodes and pathways. Figure 1 depicts the methodology used in this experiment.

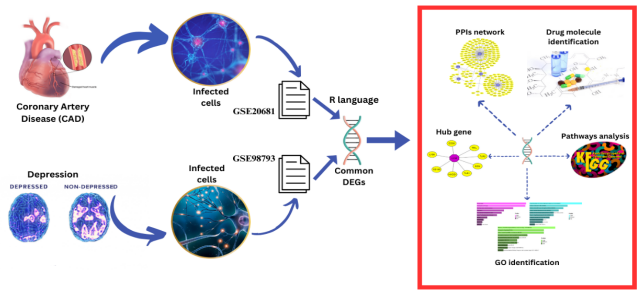

Fig. 1. *The investigation and analytical process is represented by an information flow chart in this figure. The analysis of both healthy and infected cells from GSE183648 and GSE101521 produced sample data. The analysis of these two datasets led to the identification of the most common DEGs. Hub genes, KEGG pathways, the PPI network, and therapeutic compounds were all found by using the common DEGs.*

II. Methodology

A. Dataset Collections

The GEO database served as the source of the data used to create the GSE20681 and GSE98793 datasets. Microarray and high-throughput sequencing datasets are included in the GEO database, which is housed on the NCBI platform. A platform called GPL570 was used to examine this GSE98793 dataset. The National Center for Biotechnology Information (NCBI) is the owner of the Gene Expression Omnibus (GEO) database [4], which uses the dataset identifier GSE98793 as its basis. A variety of molecular data used in study are stored there as a repository. The GPL4133 platform was used to analyze this dataset. The National Center for Biotechnology Information (NCBI)'s Gene Expression Omnibus (GEO) database has the dataset number GSE20681. It gives researchers access to molecular data for their study, including information on gene expression or sequencing.

B. Data preparation techniques, finding DEG locations, and identifying common DEGs in Depression and CAD

The initial step in preparing the data for this study involves extracting differentially expressed genes from the datasets focused on CAD and depression. One may determine the most typical DEGs by using R. An adjusted p-value of 0.01 and an absolute log2 fold change > 1.0 were used during preprocessing for both datasets. Furthermore, the datasets were managed for false discovery rate (FDR) using the Benjamini-Hochberg approach [5]. The DEGs that are the same in both GSE20681 and GSE98793 should be found, and these DEGs should subsequently be represented using a diagram that are called Venn.

C. Protein-protein interaction network

A key area of study in cellular biology is protein-protein interactions (PPIs), which are also seen as being essential to systems biology. The way that proteins carry out their biological tasks can be better understood because to these interactions. PPI networks provide important fresh understanding of protein function. As PPI networks offer a visual depiction of the study of complex organisms, network biology needs their examination. Leveraging the physical interactions of DEG proteins sourced from the IMEx database, we used NetworkAnalyst [6] in our study to construct PPI networks. Software like Cytoscape [7] is used to improve the PPI network's visualization. Understanding the connections between proteins and genes requires the use of expansive network visualizations, such as Cytoscape.

D. Hub gene identification

The nodes, edges, and connections that make up the PPI network are known as hub genes, and the nodes with the greatest degree of connectivity are called nodes.Cytohubba, a newly developed Cytoscape plugin (http://apps.cytoscape.org/apps/cytohubba), utilizes various network characteristics to prioritize and identify key, potential, or target components within a biological network [8]. In this study, the degree topological method is applied to identify active genes. The Cytoscape plugin cytoHubba (http://apps.cytoscape.org/apps/cytohubba) enables the detection of relevant hub genes within the PPI network.

E. Analyzing the pathways

Gene sets with defined chromosomal locations and broad biological activities are the subject of gene set enrichment analysis. Because the KEGG pathway provides a more practical method for researching metabolic pathways, many researchers prefer it over conventional gene annotation techniques. A detailed list of pathways connected to the common genes found in the initial stage is provided by Enrichr [9], an online web tool available at (https://amp.pharm.mssm.edu/Enrichr/). For gene sets that have been assessed over the complete genome, it offers online enrichment analysis.

F. Drug compounds identification

An important component of current research is the investigation of medicinal molecules. The DSigDB database, containing 22,527 gene sets, is utilized for the synthesis of therapeutic compounds [10]. For developing drug predictions, DSigDB heavily relies on gene expression-based

datasets. Every gene set is taken into account while analyzing substances. This database offers access to repositories of gene sets using the Enrichr platform, which may be found at https://amp.pharm.mssm.edu/ Enrichr/. Through interactive visualization techniques, these gene sets can help generate enrichment discoveries.

III. Results

A. The identification of DEGs and common genes between CAD and depression.

The genes associated with depression and CAD were analyzed, revealing 18 common DEGs. After filtering the data, the research is now being explored in greater detail. Figure 2 uses a Venn diagram to illustrate the overlap of DEGs across two different datasets.

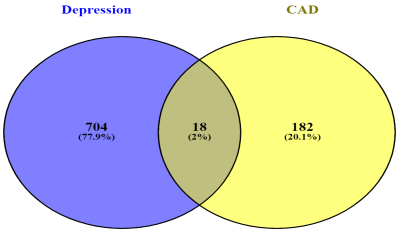

Fig. 2. *This figure visually represents DEGs that are present in both the depression sample dataset and the CAD dataset.. The depression dataset has 704 DEGs, whereas the CAD dataset has 182 DEGs. Interestingly, 18 genes were found to be equivalent, proving a connection between the two datasets.*

B. Protein-protein interaction network

The common DEGs were used as input for NetworkAnalyst. This study will analyze both the PPI network and the hub genes. NetworkAnalyst, a web-based software, and the STRING database are used to generate network diagram SIF files. The PPIs network has 208 nodes, 215 edges, 9 seed nodes, common DEGs, and other nodes. A visual illustration of this PPIs network is shown in Figure 3.

C. The hub genes identification

CCT2 stands out among the network's most crucial hub genes because it has the largest degree count of any DEG. A few more important hub genes include SVIL, REPS2, ASPH, and UBC. As linked nodes in the network, these genes play a crucial role. The Network Analyzer tool in Cytoscape was used to determine the topological characteristics of the top 5 genes. Table I contains the findings of this topological investigation.

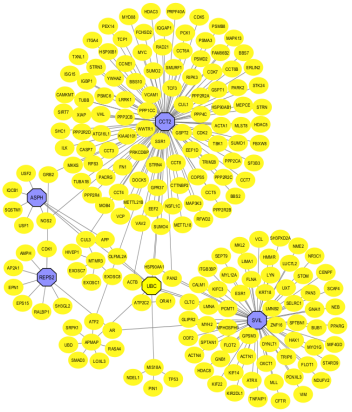

Fig. 3. *This figure illustrates the analysis of protein-protein interaction networks. along with a distinctive cluster made up of the 18 DEGs that are the most similar to one another. Genes showing significant cross-population conservation are clearly highlighted in orange within this network. The PPI network consists of 208 nodes, 215 edges, and 9 seed nodes that also include common DEGs in total.*

TABLE I
WE EXAMINED THE TOPOLOGICAL CHARACTERISTICS OF THE FIVE MOST CRUCIAL HUB GENES USING CYTOSCAPE

Hub gene	*Degree*	*Stress*	*Closeness Centrality*	*Betweenness Centrality*
CCT2	111	81766	0.52941	0.78643
SVIL	67	66514	0.42331	0.52018
REPS2	9	3796	0.32192	0.06832
ASPH	8	5624	0.33768	0.05092
UBC	7	23836	0.48027	0.22233

D. Analysis of pathway

We investigated the KEGG databases to learn more about the cellular mechanisms behind CAD and depression after discovering the DEGs shared by both conditions. To further enhance our research, we utilized the Enrichr online tool, which generates a composite score by integrating the log of the z-score with the p-value. In increasing order of p-values, Figure 4 presents the findings from the KEGG pathway database.

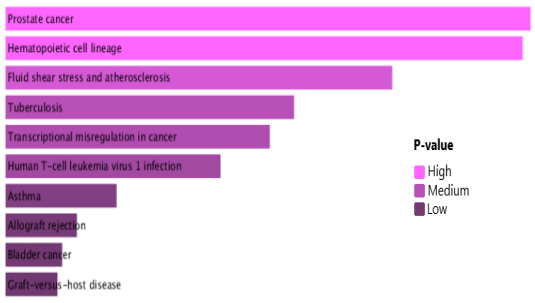

Fig. 4. *KEGG pathway analysis with cell-specific p value information*

E. Drug Compound Identification

The Enrichr online tool is used to retrieve the chemical formulae for drugs from the DSigDB database. The following drugs are recommended based on their individual and adjusted p-values. We give a thorough list of potential treatment compounds for the integrated common DEGs linked to both depression and CAD in Table II. The most potent pharmacological substances created from the most widely used DEGs are included in the table.

TABLE II
ESTIMATES OF POSSIBLE THERAPEUTIC DRUGS BASED ON DEGS COMMONLY RELATED TO CAD AND DEPRESSION.

Name of drugs	*p-value*	*Adjusted p-value*	*Genes*
rimexolone HL60 UP	0.000191392	0.02506222	TPST1; IL1R2
letrozole HL60 UP	0.000226705	0.02506222	TPST1; IL1R2
flunisolide HL60 UP	0.000245466	0.02506222	TPST1; IL1R2
halcinonide HL60 UP	0.000264963	0.02506222	TPST1; IL1R2
fludroxycortide HL60 UP	0.000285192	0.02506222	TPST1; IL1R2

IV. Discussion

While depression and CAD may appear to be unrelated, they really have intricate relationships since depression may get worse due to CAD and vice versa, which emphasizes the value of holistic treatment methods. Genes in GSE98793 and GSE20681 were first filtered, and 18 DEGs that displayed similarity were eventually found. The PPIs network under investigation was then examined, and hub genes were discovered through analysis. The top 5 hub genes were identified by the study's findings. A notable discovery was made after using Enrichr to examine differentially expressed genes and KEGG pathway assignments for substantially regulated genes. One of the most important terms in the study was the importance of the inflammatory response. This shows that the inflammatory response is fundamentally involved in the pathogenesis of depression. A number of bioinformatics research have been carried out to group similar DEGs and identify particular pharmacological compounds linked to these DEGs. The goal is to find medications that could be useful in the field of therapeutic research for both depression and coronary artery disease.

V. Conclusion

The study of CAD and depression had never included transcriptome analysis. Instead, we employed a systems biology and functional enrichment approach to pinpoint the crucial proteins and pathways relevant to both diseases.Central to this approach and foundational to systems biology is the gene filtering process, which is conducted through various bioinformatics analyses. For depression and CAD, respectively, we looked at the microarray datasets GSE98793 and GSE20681. We found up-regulated and down-regulated genes in both datasets that were differentially expressed via analysis. We created PPI networks using common DEGs, and based on the degree values within these networks, we chose 5 hub genes (CCT2, SVIL, REPS2, ASPH, and UBC). The hub genes are designated as particularly important in connection to both conditions by the existence of these same DEGs, and we utilized them as a basis to find efficient therapeutic drug compounds. The top five biomarkers have been identified as unique biological indicators that could assist in developing treatments for CAD and depression. Thus, we provided recommendations for both illnesses' earlier, more accurate diagnosis and therapy.

ACKNOWLEDGMENT

Everyone engaged in this study is excited to provide their time and knowledge. The fact that this writing hasn't been published and isn't being considered for publication anywhere should be noted.

References

[1] Serrano Jr, C.V., Setani, K.T., Sakamoto, E., Andrei, A.M. and Fraguas, R., 2011. Association between depression and development of coronary artery disease: pathophysiologic and diagnostic implications. Vascular health and risk management, pp.159-164.

[2] Davidson, K.W., 2012. Depression and coronary heart disease. International Scholarly Research Notices, 2012.

[3] Basar, M.A., Hosen, M.F., Paul, B.K., Hasan, M.R., Shamim, S.M. and Bhuyian, T., 2023. Identification of drug and protein-protein interaction network among stress and depression: a bioinformatics approach. Informatics in medicine unlocked, 37, p.101174.

[4] Clough, E. and Barrett, T., 2016. The gene expression omnibus database. Statistical Genomics: Methods and Protocols, pp.93-110.

[5] Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological), 57(1), pp.289-300.

[6] Zhou, G., Soufan, O., Ewald, J., Hancock, R.E., Basu, N. and Xia, J., 2019. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic acids research, 47(W1), pp.W234-W241.

[7] Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research, 13(11), pp.2498-2504

[8] Chin, C.H., Chen, S.H., Wu, H.H., Ho, C.W., Ko, M.T. and Lin, C.Y., 2014. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC systems biology, 8(4), pp.1-7.

[9] Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A. and McDermott, M.G., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic acids research, 44(W1), pp.W90-W97.

[10] Yoo, M., Shin, J., Kim, J., Ryall, K.A., Lee, K., Lee, S., Jeon, M., Kang, J. and Tan, A.C., 2015. DSigDB: drug signatures database for gene set analysis. Bioinformatics, 31(18), pp.3069-3071