Analysis of Gene Network Model of Thyroid Disorder and Associated Diseases: A Bioinformatics Approach

By

Md Kawsar

171-35-201

A thesis submitted in partial fulfillment of the requirement for the degree of Bachelor of Science in Software Engineering.

Department of Software Engineering DAFFODIL INTERNATIONAL UNIVERSITY

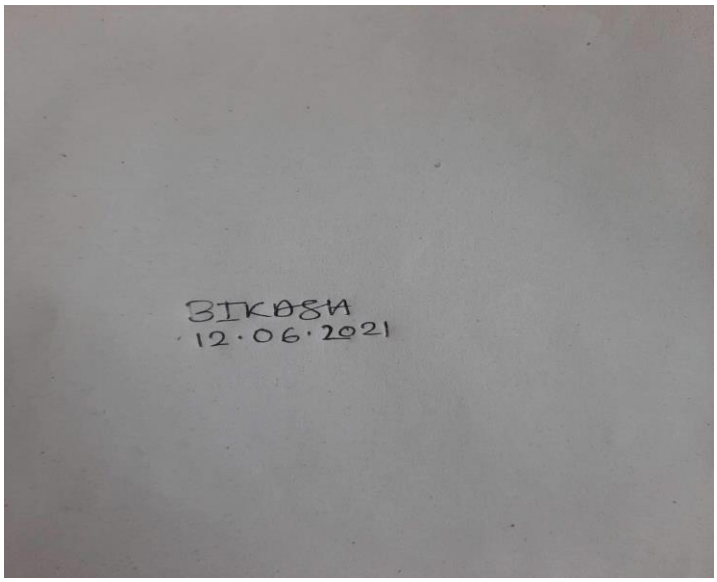
Semester Summer – Year-2021

IM: Dr. Imran Mahmud; SMR: S A M Matiur Rahman; RZ: Raihana Zannat; NH: Nayeem Hasan; SI: Mr. Shariful Islam; SFR: SK. Fazlee Rabby; MA: MarziaAhmed; RM: Md. Rajib Mia.

APPROVAL

This thesis titled "Analysis of Gene Network Model of Thyroid Disorder and Associated Diseases: A Bioinformatics Approach", submitted by Md Kawsar, ID: 171-35-201 to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Software Engineering (SWE) and also published by the Elsevier, Informatics in Medicine Unlocked 2020.

Supervised By



Bikash Kumar Paul

Senior Lecturer

Department of Software Engineering, Permanent Campus

DECLARATION

We hereby declare that we have taken this thesis under the supervision of Bikash Kumar Paul, Lecturer, Department of Software Engineering, Daffodil International University. We also declare that this thesis was submitted on Elsevier by 5 May 2020 and Published on Elsevier, Informatics in Medicine Unlocked by 19 June 2020.

DOI: https://doi.org/10.1016/j.imu.2020.100381

Md Kawsar

Student ID: 171-35-201

Department of Software Engineering

Faculty of Science & Information Technology

Daffodil International University

ACKNOWLEDGEMENT

We had started this thesis with zero knowledge. We had to discover a whole new arena to work on this thesis. Because of having a technological background, understanding the thoughts and theories of Biology and Biotechnology was pretty tough for us. But we became able to jump forward smoothly by the help of our respectable teacher Bikash Kumar Paul. He supervised us on every step. He motivated us to learn new things, provided all the important files and information and taught us how to make the work done. Special thanks go to another honorable teacher Kawsar Ahmed. He spectated our every step and directed us to write down the paper correctly so that it can be published. And by the grace of the almighty, our work was finally published on Elsevier Gene Reports. This couldn't have been possible without them.

We also want to thank our respectable Touhid Bhuiyan Sir, Matiur Rahman Sir, Sayed Asaduzzaman Sir, Nayeem Ahmed Sir and Sharif Ahmed Sir. They were always cordial to us, helped us and motivated us. We are always thankful to them. We would like to express our gratitude towards our seniors, classmates for the time and support they have given to us with all their capabilities. We are grateful to all of them. Thank you so much.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

.

LIST OF ABBREVIATION

TD = Thyroid Disorders

HPB = High Blood Pressure

CKD = Chronic Kidney Disease

PPI = Protein Protein Interaction

TP = Topological Properties

GO = Gene Ontology

GRN = Gene Regulatory Network

Abstract

Chronic Kidney Disease (CKD), High Blood Pressure (HBP) and Thyroid Disorders (TD) are three relational diseases, when human affected by one of them, increases the possibility affected those people by other two diseases. Found large numbers of similarly biological and genetic features among HBP, CKD, and TD. In this investigation, we will find out behind the reason for 3 diseases are related to each other. Identify the common genes among HBP, CKD, and TD and finding the most significant genes. First step of this investigation by reducing the number of reactive genes intersection is obtained using RStudio. After completing this process, we have identified ten genes that are shared by HBP, CKD, and TD. Based on the biological, biochemical, and genetic relationships between common genes, this analysis identifies the most significant hub proteins. We designed Protein-Protein Interactions network, Co-Expression network, Enrichment Analysis, Topological properties analysis, Gene regulatory network, and Physical Interaction network based on our understanding of biological, biochemical, and genetic relationships. This analysis allows us to identify biological and genetic similarities between HBP, CKD, and TD. Finally, we've arrived at our targeted hub proteins.

CHAPTER 1

INTRODUCTION

1.1 Background

The thyroid gland, a butterfly-shaped gland in the front of the neck, is affected by TD. Thyroid hormone metabolism or iodine nutrition deficiencies can cause TD [1]. Thyroid hormone levels in TD are unusually high, which can affect the energy and metabolism of substances (such as metal ions and H2O) in the body [1]. HBP is a long-term condition in which the arterial blood pressure is constantly elevated [2]. High blood pressure is present in most adults if the residual blood pressure is consistently 130/80 or 140/90 mmHg or higher [3]. HBP is classified as either primary (required) or secondary hypertension. Because of genetic features, approximately 90-95 percent of cases are well-defined as primary HBP. Extra body weight, smoking, and alcohol consumption are all risk factors [2]. Affected organs include the kidney, brain, lungs, and heart, with symptoms such as chest pain, drowsiness, and shortness of breath [5]. CKD is the most common kidney disease, in which kidney function gradually declines over a period of months to years. There are no early symptoms of the disease; however, later symptoms may include leg swelling, fatigue, loss of appetite, confusion, and nausea. HBP, heart disease, anemia, and bone disease are all possible indicators [6]. CKD affected 753 million people in 2016 [7]. They consisted of 336 million men and 417 million women [7].

The common genes of the three selected diseases are identified and thoroughly analyzed in this paper. The hub protein responsible for these three diseases is also identified through this analysis. The importance of bioinformatics resources has been demonstrated in several reports on three diseases: TD, HBP, and CKD. Using these properties, we hope to develop a drug for TD, HBP, and CKD. Chemical experiments will be used to further validate the investigated drug.

1.2 Motivation of Research

Dangerous diseases pose a threat to public health. Three dangerous diseases were chosen for examination in this study. The three different disorder diseases were chosen in order to determine their relationships. Furthermore, new drugs will be developed in such a way that they can resist all three diseases simultaneously. When a patient is infected with one disease, the risk is reduced. When a person is infected with more than one of these diseases, the risk is multiplied. The goal of this observation is to find a genetic relationship between these diseases and to analyze the genetic relationship between these two genes. The selected three diseases are Thyroid Disorders (TD), Chronic Kidney Disease (CKD) and High Blood Pressure (HBP).

1.3 Problem Statements

Reviewing the existing research work of Thyroid Disorders (TD), it has been found that there is not much research about risk factors collaborated with PPIN. When the risk factors CKD and HBP in the human there are more possibilities that the human will get TD. If we detect the risk factors earlier it is easier to prevent TD that it is too late. The authors will find out the common genes between the genes of TD and the two particular risk factors including CKD and HBP.

1.4 Research Questions

1. Question 1: What are the risk factors that may be associated with Thyroid Disorders?
2. Question 2: What are the common genes between the genes of Thyroid Disorders and the particular two risk factors?
3. Question 3: How does a PPI Network work on disease identification?
4. Question 4: How can this work be effective in this field?

1.5 Research Objectives

1. To collect the genes of TD, CKD and HBP for analyzing it.
2. To implement an intersection method to find out the common genes of TD, CKD and HBP.
3. To visualize the PPI Network of intersected genes.

1.6 Research Scope

Every year millions of human dies because of TD. So, it is important to find a way to detect it in the early stages so that the patients can survive through taking proper treatment timely. If any Human possesses CKD and HBP in their body, they will have more possibility to be affected by TD. As a few research works are done regarding TD, CKD and HBP there are a lot of research scopes in this area.

1.7 Thesis Organization

In this research, the IEEE referencing system has been used in this document. The paper has been furnished with five chapters which is described below:

Chapter 1: In this chapter, research background, motivation, problem statement, objectives and research scope are given.

Chapter 2: This chapter includes discussion of the existing related work and figured out the research gap.

Chapter 3: This chapter contains the research methodology and approaches as it follows for the research.

Chapter 4: This chapter compares the experimented results with existing approaches.

Chapter 5: The research outcome and the limitation of this study is presented here and the direction of the future work of the research has also been guided.

CHAPTER 2

LITERATURE REVIEW

In order to study and apply computational tools to the analysis of enormous biological data sets, such as genetic sequences, cell populations or protein samples, novel assumptions or the exploration of novel biology, computer biology and bio-informatics is an inter-disciplinary discipline. Bioinformatics applies and applies computer science techniques in molecular biology and medicine to solve problems. The success and quicker expansion of this field will mean a dependable, healthy employment potential for the next few years. This paper discusses the PPI network. Protein-protein interactions are very effective biologic interactions, created by interactions including electrostatic strength, hydrogen bonding and biologically driven hydrophobic effects between two or more protein molecules.The topological characteristics are derived with the Cytoscape program based on the PPI network. We believe this prediction as well, however our major goal is to work with the PPI network and analyze a number of models of the biological network. It might thus be another step ahead to this sector following this effort.

CHAPTER 3

RESEARCH METHODOLOGY

Proposed Methodology

There has been a rare use of drug patterns for specific gene identification, implementation of protein-protein interactions (PPIs), pathway analysis, different regulatory networks, and previous studies based on TD, HBP, and CKD. All genes in our studies were derived from the NCBI gene database (https://www.ncbi.nlm.nih.gov/gene/). We used Rstudio (https://rstudio.com/) to identify genes that are shared by three diseases. The common genes are then assigned to NetworkAnalyst (https://www.networkanalyst.ca/) to obtain a drug molecule overview. The application of critical bioinformatics tools aids research in better understanding drug candidates. Using different update tools and technique one by one to find of drug candidates.

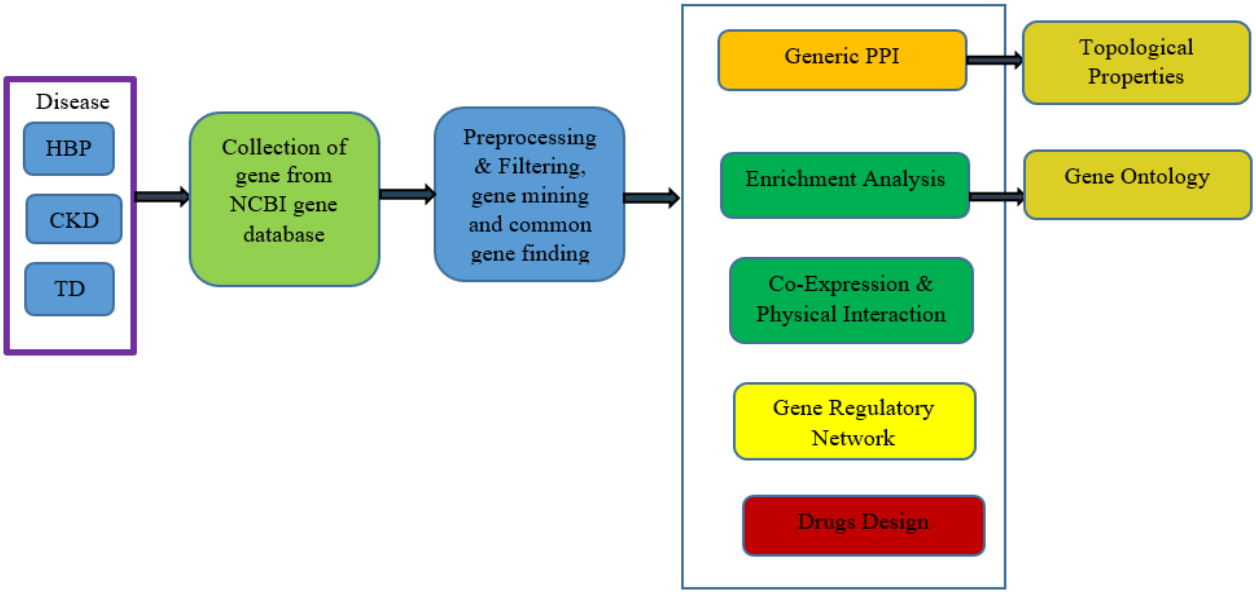


Figure 1: Flowchart of the proposed research Methodology

3.1 Gene Collection

We obtained all of the genes for this study from the NCBI's online gene database (https://www.ncbi.nlm.nih.gov/gene/). They have given us permission to provide accurate information to thousands of genomic databases, including billions of bytes of data. [8]. The NCBI acknowledges effective explanation. [9]. In this analysis, we only need to collect the genes that are responsible for Homo sapiens. Because this study is only for human drug design, these genes are only responsible for Homo sapiens. Download the genes for Homo sapiens based on their weight and import them into Microsoft Excel. Import files are used to look for more common genes.

3.2 Gene intersection and common gene finding

The data mining technique was used in this study to identify eligible data and common genes among HBP, CKD, and TD. To determine the gene intersection between these three diseases, intersected between (HBP-CKD, CKD-TD, TD-HBP and HBP-CKD-TD). The intersection result aids in further research and the identification of common genes. The final common genes between HBP, CKD, and TD, which facilitates effective drug design.

3.3 Generic PPI

Protein functions have many properties, the most important of which is protein-interaction [10]. PPIs define the relationship between two or more proteins, and these collaborations occur as a result of various biochemical and hydrophobic effects, as well as electrostatic forces. The network was disrupted in this analysis due to the discovery of the target genes and the removal of the target genes [11]. Create a PPI network for this study using NetworkAnalyst, an online Bioinformatics tool.

3.4 Topological properties analysis

The graph-based biological network of topological properties (TP) has emerged as one of the most significant innovations [12]. This analysis aids in the understanding of biological networks and the identification of targeted drug proteins [13]. The PPI network is used to collect TP samples for a specific sample [14]. The investigation of TP yielded documentation of the significant gene's degree, clustering coefficient, betweenness centrality, and closeness centrality. In this analysis, we used Cytoscape to determine the TP result.

3.5 Enrichment Analysis: GO terms

Rich investigation is being conducted into the classification of investigational genetic data [15]. Enrichment analysis (EA) is determined by combining several functional systems, the most important of which is gene ontology (GO) [15]. The enrichment analysis for this study was determined using the online Bioinformatics web-based application STRING.

3.6 Co-expression network and Physical interaction network

The co-expression network is widely used to discover gene functionality at the structure-level [16]. Co-expression networks are increasingly being used to investigate gene function at the process level [16]. Co-expressed genes that are regulated by related transcriptional regulatory sequences, functionally linked or complex proteins, or pathway members. Co-expression standards are the constraints on which co-expression network definitions are based, as well as the network used to integrate an unknown identified functional gene [17]. Create a Physical interaction network and a Co-expression network for this study using the online Bioinformatics web-based application GeneMANIA.

3.7 Gene Regulatory Network (GRN)

Gene regulatory networks (GRNs) provide regulatory familiarity between genes, allowing them to build on the various biological functions of genes related to molecular functions and functions of a gene [18]. We created a GRN in this analysis using the online Web-based tool NetworkAnalyst (https://www.networkanalyst.ca/), which generates three types of GRNs. We do three types of work in this section: TF-gene interaction, Gene-miRNA interaction, and TF-miRNA coregulatory.

3.8 Protein-Drug Interactions

Maximizing the efficiency of the drug and reducing the toxicity of the drug are considered as the primary activities of drug design [19]. For this research, determined the enrichment analysis using the online web-based application NetworkAnalyst, which is a Bioinformatics tools.

3.9 Protein-Chemical Interactions

Because combining chemical sets with distinct target genes is a difficult task, calculating protein-chemical interactions is an incredible goal in bioinformatics. The enrichment analysis for this study was determined using the online web-based application NetworkAnalyst, which is a Bioinformatics tool.

CHAPTER 4

RESULT AND DISCUSSION

4.1 Gene Collection

All responsible genes for HPB, CKD, and TD are gathered for this study from the NCBI's online gene database. The National Center for Biotechnology Information (NCBI) is where we look for Homo sapiens genes. There are 29, 471, and 1850 genes collected for HPB, CKD, and TD, respectively.

Table 1. Represent the gene collection of diseases.

Name	All genes for disease	*Homo sapiens* genes for disease
Chronic Kidney Disease (CKD)	**697**	**471**
High Blood Pressure (HBP)	**51**	**29**
Thyroid Disorder (TD)	**1909**	**1850**

4.2 Gene intersection and common gene finding

For this study, we used Rstudio to identify genes that are linked to HBP, CKD, and TD. The interactions for HBP, CKD, and TD are as follows: (HBP-CKD, CKD-TD, TD-HBP and HBP-CKD-TD). The intersected genes for HBP-CKD, CKD-TD, and TD-HBP are 15, 193, and 15, respectively. The intersection is used to identify the common gene among three diseases (HBP-CKD-TD). We discovered a total of ten common genes from the intersection of HBP, CKD, and TD. TNF, APOE, IL6, MTHFR, ACE, IL10, CRP, FTO, AGTR1, and CST3 are examples of common genes. To display the results, we used Interactivenn

(http://www.interactivenn.net/index2.html).

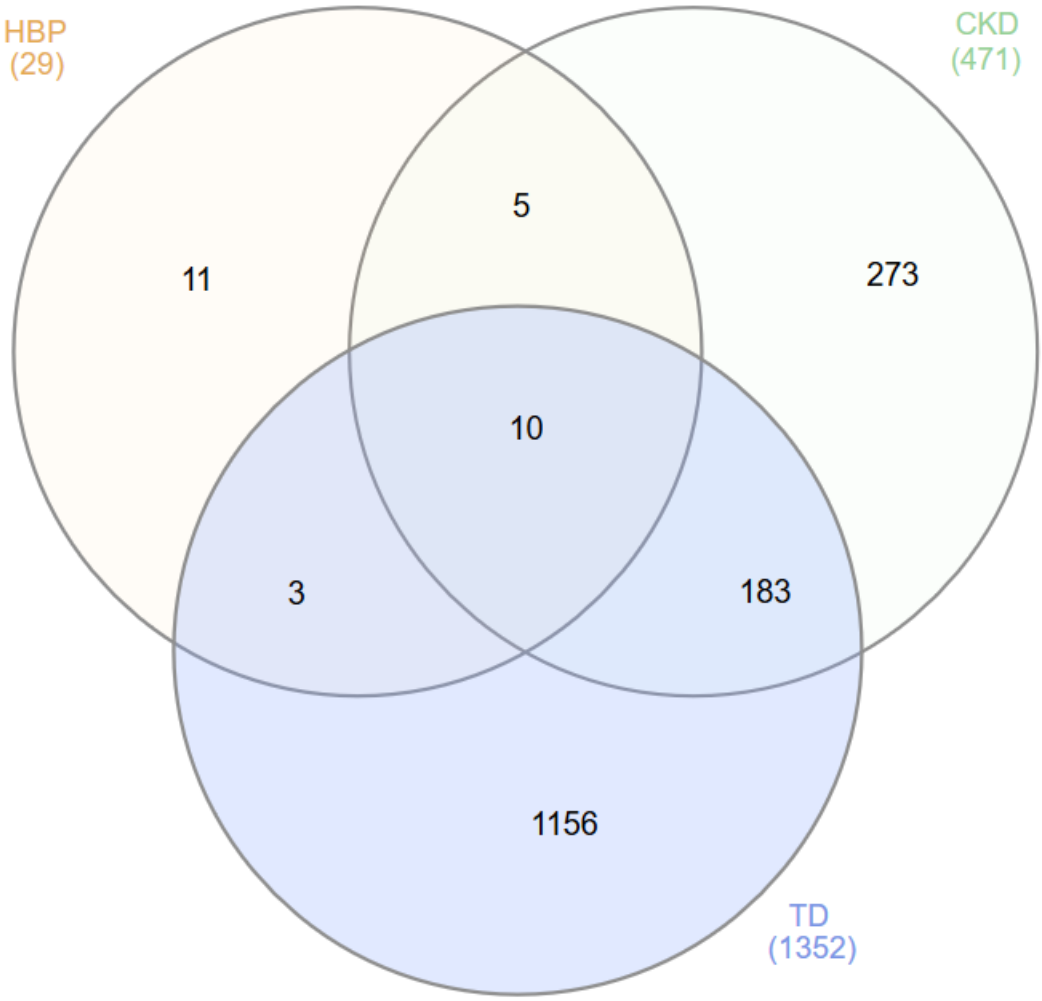


Figure 2: Venn diagram showing the intersection result.

Table 2. Represent the intersection result of HBP-CKD, CKD-TD, TD-HBP and HBP-CKD-TD.

Disease name	Total gene	Common Gene
HBP & CKD	**500**	**15**
CKD & TD	**2321**	**193**
TD & HBP	**1879**	**13**
HBP, CKD & TD	**2350**	**10**

4.3 Generic PPI

NetworkAnalyst is a comprehensive online web-based tool that enables association researchers to use a consistent web interface for both complex and simple meta-testing of gene expression data [20]. NetworkAnalyst provides gene expression data that is related to the Network of Protein-Protein Interactions (PPI) [21]. NetworkAnalyst lists ten common genes for Chronic Kidney Disease (CKD), High Blood Pressure (HBP), and Thyroid Disorders (TD). Figure 3 depicts the NetworkAnalyst output.

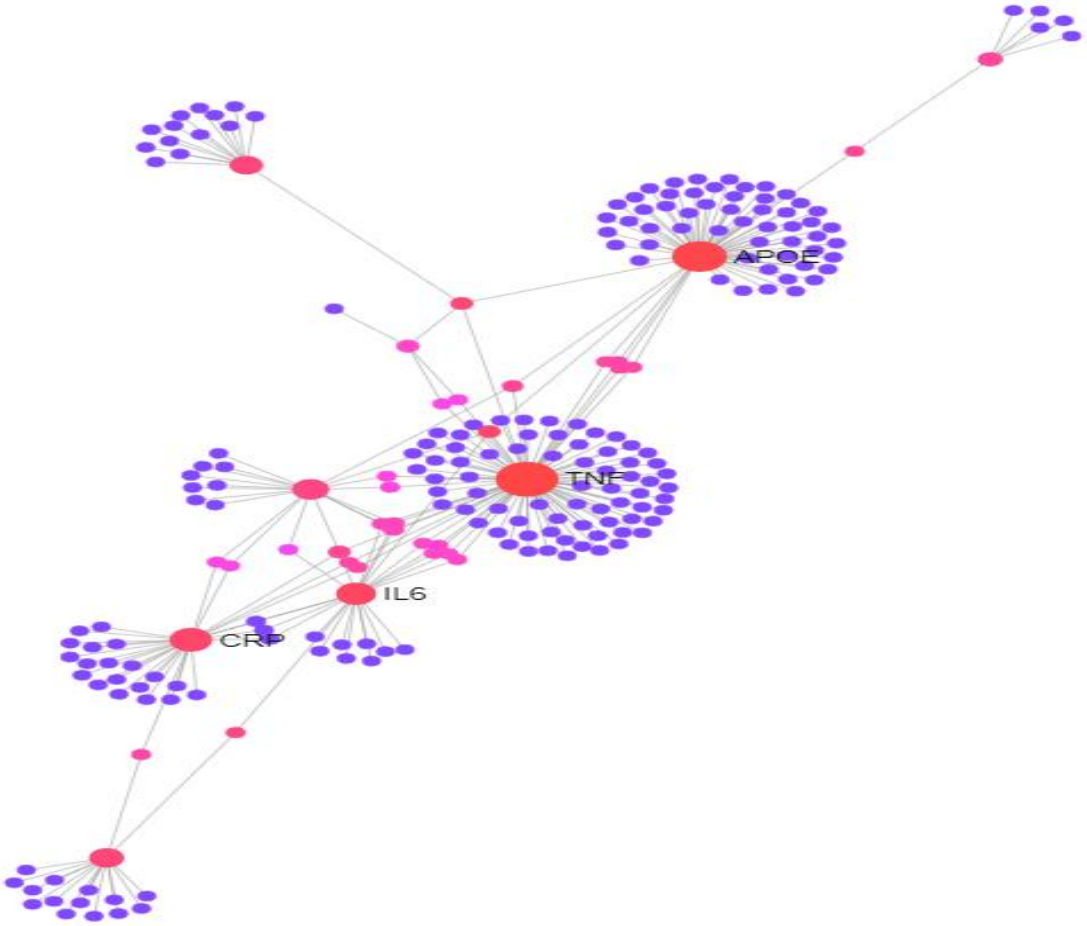


Figure 3: This graph depicts the PPI for ten common genes. The edges represent interactions between two genes, and each node represents a gene. The larger nodes represent a greater number of edges, and the red color represents the network's most important nodes. This network has 232 nodes, 265 edges, and 9 seed nodes.

4.4 Topological property

A specific protein, betweenness, cluster coefficient, topological coefficient, and including degree are all present in the PPI network. The collected SIF file is set into Cytoscape after downloading the PPI network from NetworkAnalyst for topological investigation. Table 3 shows the values of topological properties for nine seeds, while Figures 4 and 5 show graphical representations of the topological properties.

Table 3. Topological properties for 9 seed node genes using Cytoscape.

Name	Degree value	Betweenness Centrality value	Closeness Centrality value	Clustering Coefficient value	Topological Coefficient value
TNF	**92**	**0.61554188**	**0.42541436**	**0.0**	**0.05797101**
APOE	**62**	**0.46389722**	**0.37745098**	**0.0**	**0.03494624**
IL6	**24**	**0.24503298**	**0.35106383**	**0.01449275**	**0.06944444**
ACE	**4**	**0.00982496**	**0.27467301**	**0.0**	**0.41666667**
IL10	**10**	**0.08788798**	**0.32038835**	**0.0**	**0.25**
CRP	**29**	**0.19068908**	**0.32307692**	**0.00985222**	**0.04980843**
FTO	**14**	**0.10961792**	**0.28136419**	**0.0**	**0.07142857**
AGTR1	**16**	**0.1178242**	**0.22362052**	**0.0**	**0.0625**
CST3	**6**	**0.04291361**	**0.22**	**0.0**	**0.16666667**

(a) (b)

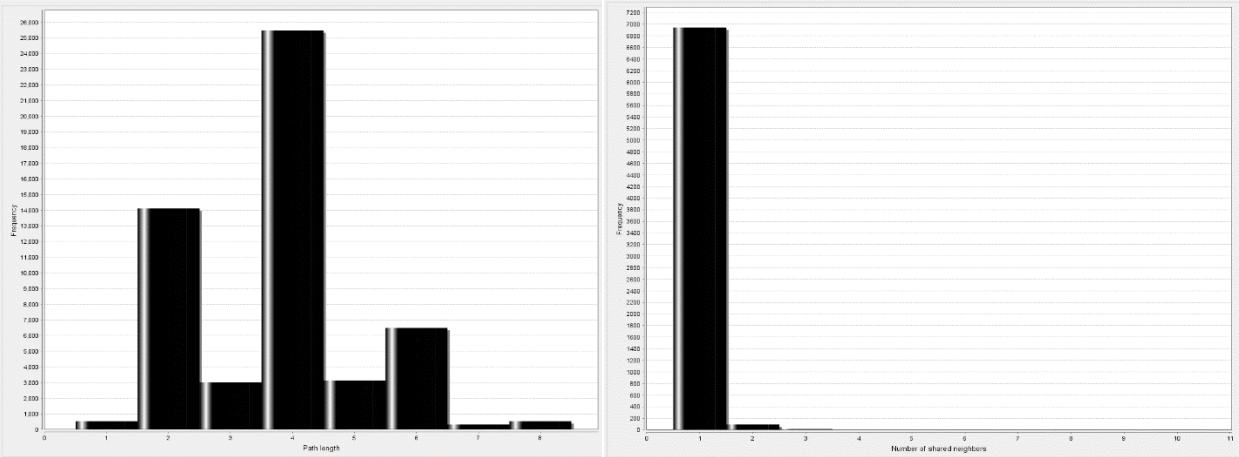
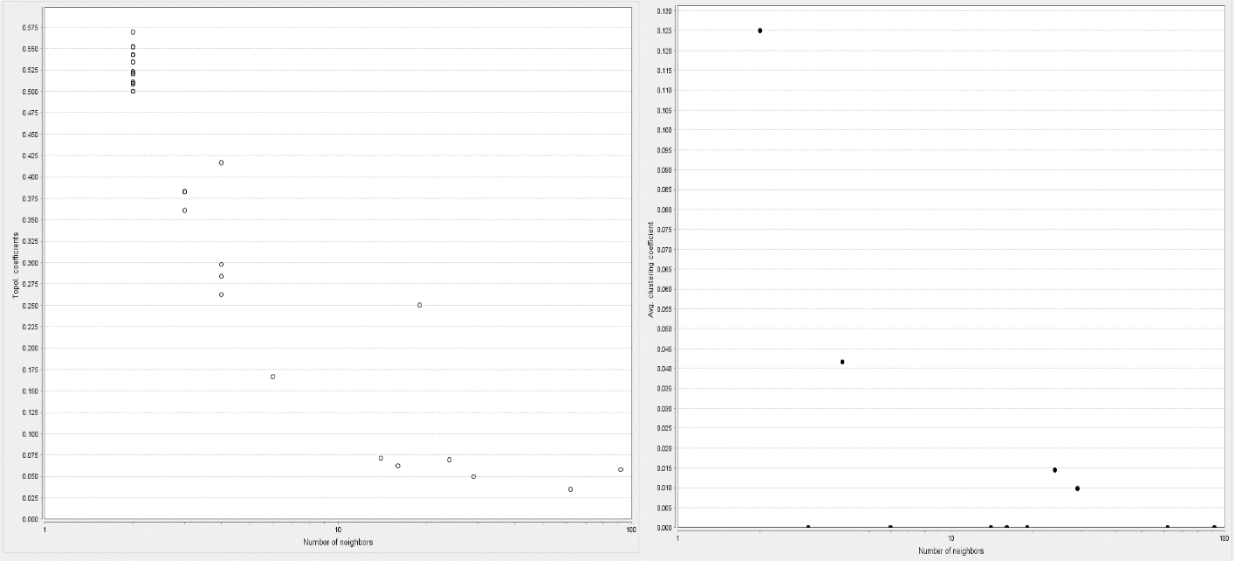


Figure 4: (a) The length of the path in relation to the frequency is depicted in the graph. The X and Y axes Represents the value of path length and frequency, respectively. (b) The graph depicts the frequency distribution of the shared neighbor. The X and Y axes Represents the distribution and frequency of the value of shared neighbors.

(a) (b)



(c) (d)

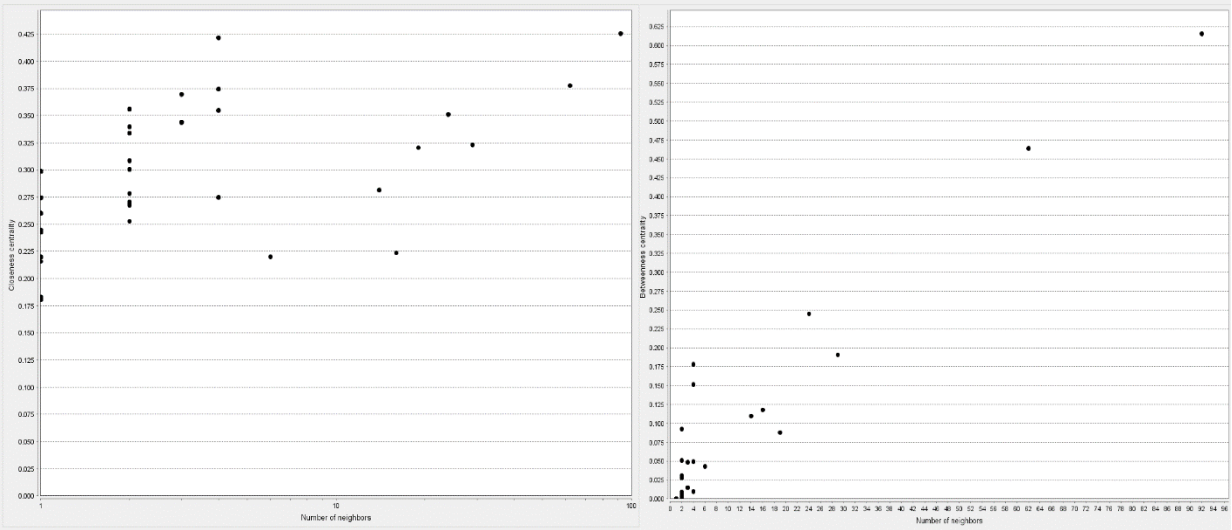


Figure 5: (a) The graph indicates the number of neighbors in relation to the Topological Coefficient. The X and Y axes Represents the value of the number of neighbors and the Topological Coefficient. (b) The graph shows the number of neighbors in relation to the Cluster coefficient. The X and Y axes Represents the value of the number of neighbors and the Cluster coefficient, respectively. (c) The graph depicts the number of neighbors in relation to the centrality of Closeness. The X and Y axes Represents the value of the number of neighbors and the centrality of closeness. (d) The graph refers to the number of neighbors relative to the Betweenness centrality. X-axis and Y-axis Respectively represents value of number of neighbors and Betweenness centrality.

4.5 Enrichment Analysis: GO terms

In pathway network analysis and gene expression analysis, gene set enrichment analysis is required [22]. A primary term in enrichment analysis is gene ontology (GO). We used the online web-based tool STRING for this GO analysis. The results of the analysis are shown in Table 4.

Table 4. Enrichment analysis for 10 common genes.

Biological Process (GO)			
GO-term	Description	Count in gene set	False discovery
GO:1905952	regulation of lipid localization	6 of 129	3.20e-08
GO:1905953	negative regulation of lipid localization	4 of 40	3.74e-06
GO:0010883	regulation of lipid storage	4 of 42	3.74e-06
GO:0050793	regulation of developmental process	9 of 2416	2.44e-05
GO:0042127	regulation of cell population proliferation	8 of 1594	2.48e-05

Molecular Function (GO)			
GO-term	Description	Count in gene set	False discovery
GO:0031711	bradykinin receptor binding	2 of 2	0.00018
GO:0005102	signaling receptor binding	7 of 1513	0.00018
GO:0071813	lipoprotein particle binding	2 of 25	0.0028
GO:0050750	low-density lipoprotein particle receptor binding	2 of 22	0.0028
GO:0005125	cytokine activity	3 of 216	0.0029
Cellular Component (GO)			
GO-term	Description	Count in gene set	False discovery
GO:0005615	extracellular space	7 of 1134	2.93e-05
GO:0005788	endoplasmic reticulum lumen	3 of 299	0.0128
GO:0005768	endosome	4 of 876	0.0174

4.6 Co-expression and Physical interaction network

We used GeneMANIA to create networks of physical interaction and co-expression for ten common genes for this analysis. Figures 6(a) and 6(b) depict the Co-expression network and the Physical Interaction network, respectively.

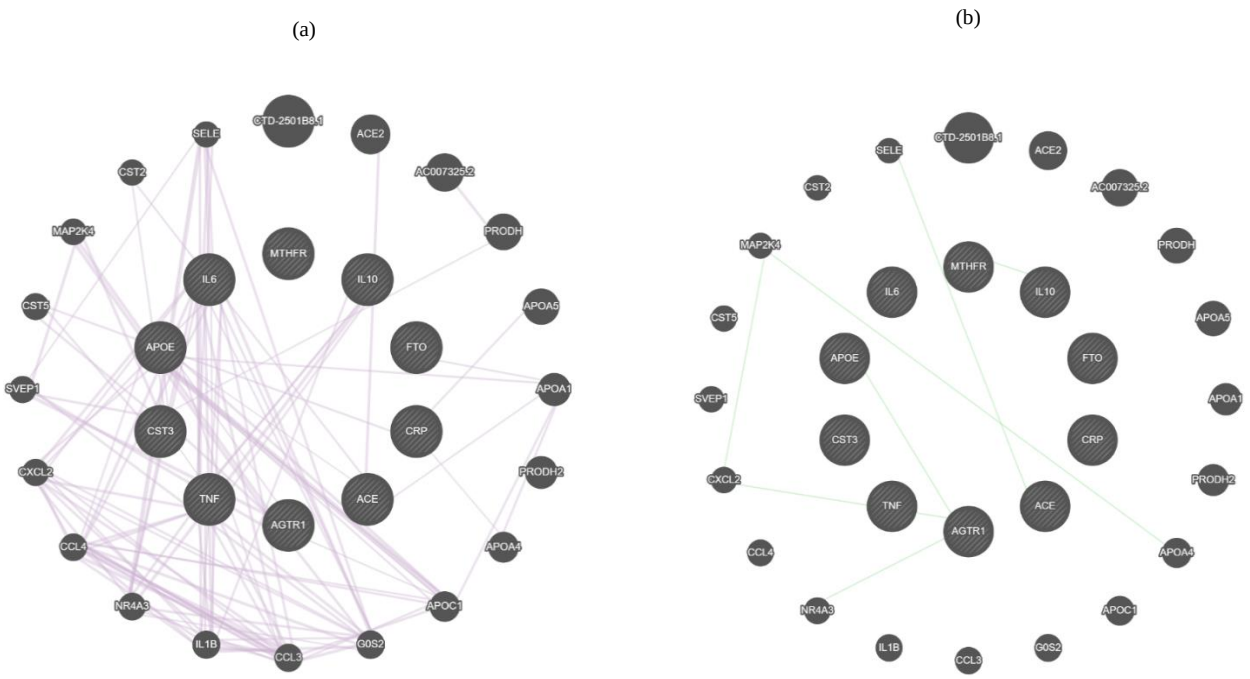


Figure 6: Edges represent interactions between two genes and each node represent a gene. (a) Represents the co-expression network, (b) represents the physical interaction network among TNF, APOE, IL6, MTHFR, ACE, IL10, CRP, FTO, AGTR1 and CST3 genes.

4.7 Gene Regulatory Network (GRN)

For this study, we created a TF-gene interaction network, a TF-miRNA co-regulatory network, and a Gene-miRNA interaction network. The Gene Regulatory networks depicted in Figures 7, 8, and 9 are the Gene-miRNA interaction network, the TF-gene interaction network, and the TF-miRNA co-regulatory network, respectively.

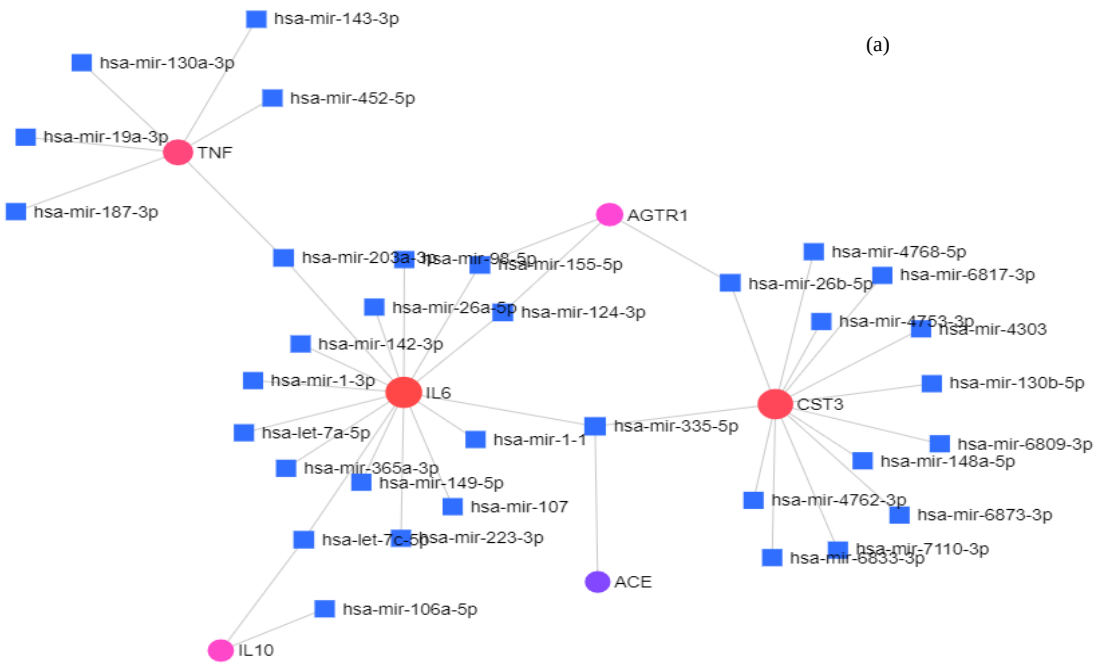


Figure 7(a): This graph depicts a network of gene-miRNA interactions for ten common genes associated with HBP, CKD, and TD. Each node represents a gene, and the edges represent interactions between two genes. Seed nodes are represented by the circle node. This network consists of 39 nodes, 40 edges, and 6 seed nodes (TNF, IL6, MTHFR, ACE, IL10, and CST3). The most significant hub-proteins of the network based on one-degree value are IL6 and CST3. These two genes account for 67.5 percent of the network's edges.

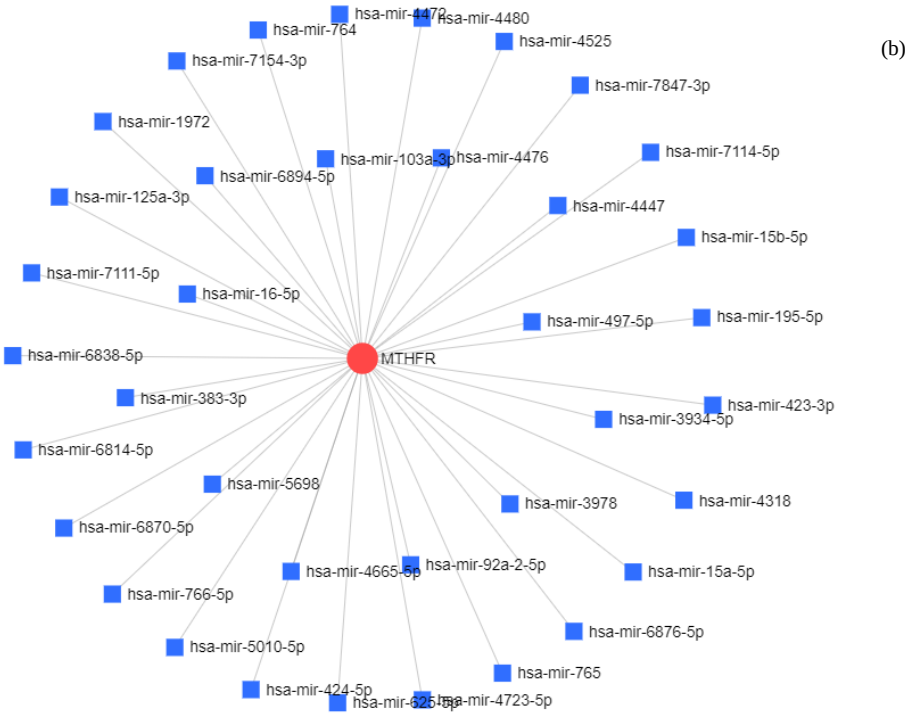


Figure 7(b): This graph depicts a network of gene-miRNA interactions for ten common genes associated with HBP, CKD, and TD. Each node represents a gene, and the edges represent interactions between two genes. Seed nodes are represented by the circle node. This network consists of 38 nodes, 37 edges, and one (MTHFR) seed node. MTHFR is the network's most significant hub-protein in terms of degree value.

(c)

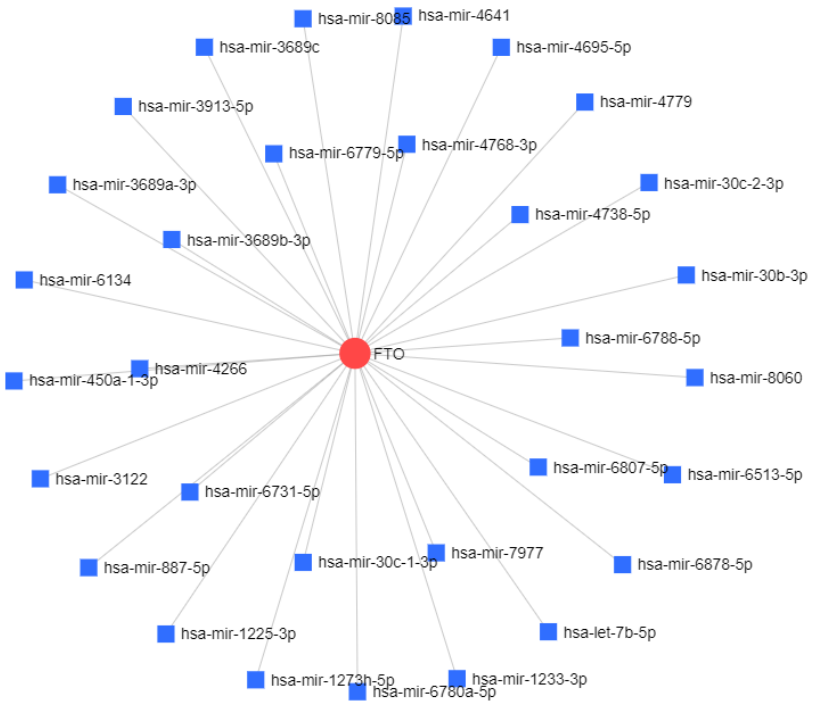


Figure 7(c): This graph depicts a network of gene-miRNA interactions for ten common genes associated with HBP, CKD, and TD. Each node represents a gene, and the edges represent interactions between two genes. Seed nodes are represented by the circle node. This network consists of 32 nodes, 31 edges, and one (FTO) seed node. FTO is the network's most important hub-protein in terms of degree value.

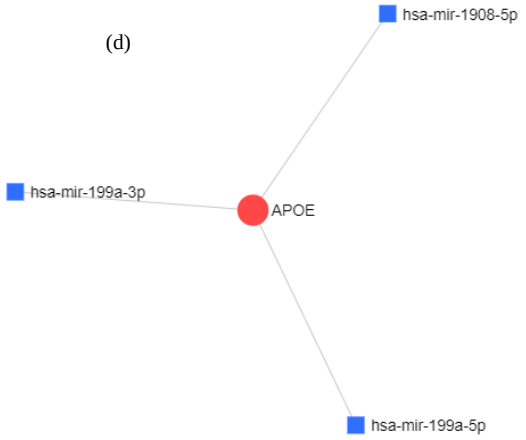


Figure 7(d): This graph depicts a network of gene-miRNA interactions for ten common genes associated with HBP, CKD, and TD. Each node represents a gene, and the edges represent interactions between two genes. Seed nodes are represented by the circle node. This network consists of four nodes, three edges, and one (APOE) seed node. Based on degree value, APOE is the network's most significant hub-protein.

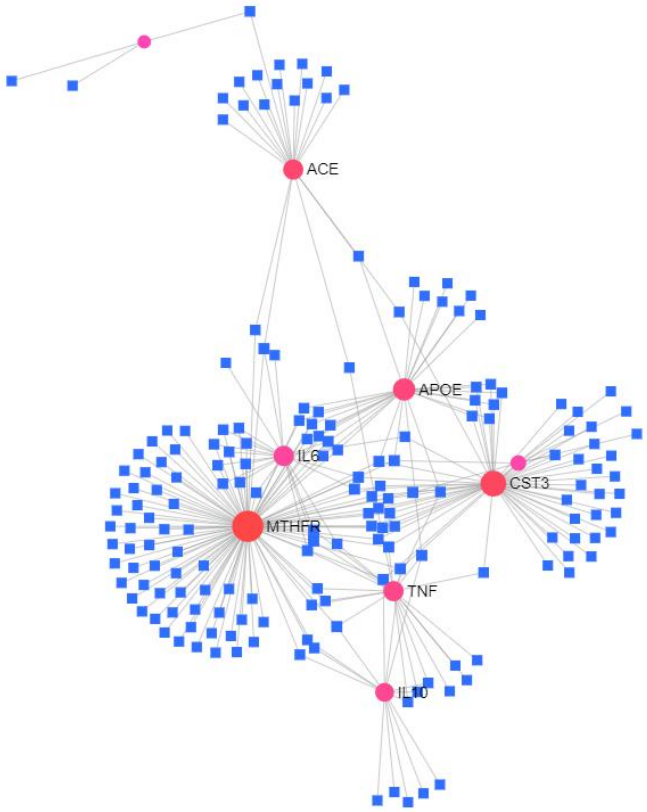


Figure 8: This graph represents TF-gene Interaction's network for 10 common genes for HBP, CKD and TD. Edges represent interactions between two genes and each node represent a gene. The circle node represents seeds nodes. This network has 193 nodes, 269 edges and 9(TNF, APOE, IL6, MTHFR, ACE, IL10, CRP, FTO and CST3) seed nodes. TNF and IL6 are most significant

hub-protein of the network based on degree value. These two genes make up 57.25% edges of the total network.

(a)

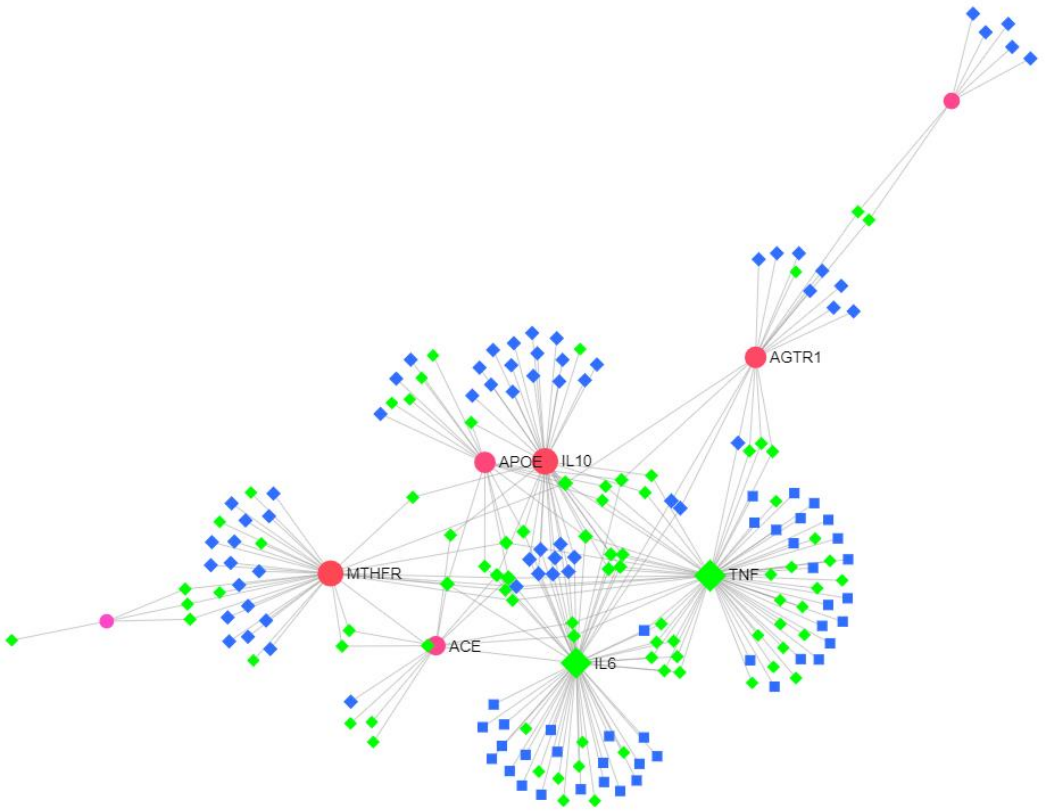


Figure 9(a): This graph depicts subnetwork 1 of the TF-miRNA Coregulatory network for ten common genes associated with HBP, CKD, and TD. Each node represents a gene, and the edges represent interactions between two genes. This network has 191 nodes, 257 edges, and 9 seed nodes (TNF, APOE, IL6, MTHFR, ACE, IL10, CRP, AGTR1, and CST3). TNF and IL6 are the network's most important hub proteins in terms of degree value. These two genes account for 48.24 percent of the network's edges.

(b)

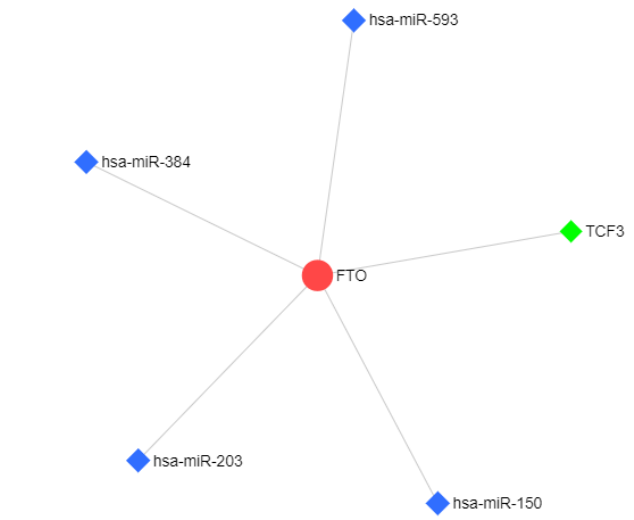


Figure 9(b): This graph depicts subnetwork2 of the TF-miRNA Coregulatory network for ten common genes associated with HBP, CKD, and TD. Each node represents a gene, and the edges represent interactions between two genes. Seed nodes are represented by the circle node. This network has six nodes, five edges, and one (FTO) seed node. FTO is the network's most important hub-protein in terms of degree value.

4.8 Protein-Drug Interactions

In this analysis, Protein-chemical interaction network scheme using NetworkAnalyst. For this research, we gathered suggested drug from Comparative Toxicogenomics database.

(a)

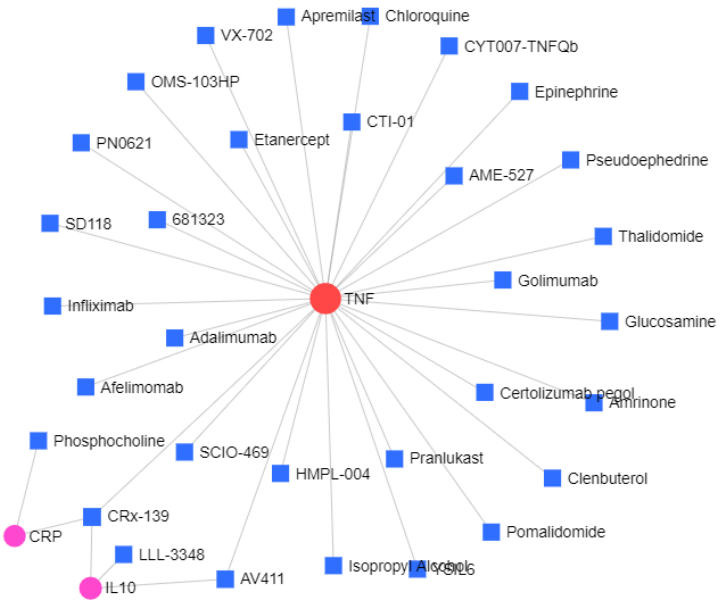


Figure 10(a): This graph depicts subnetwork1 of the Protein-drug Interactions for ten common genes associated with HBP, CKD, and TD. Each node represents a gene, and the edges represent interactions between two genes. Seed nodes are represented by the circle node. This network has 35 nodes, 35 edges, and three seed nodes (TNF, CRP, IL10). TNF is the network's most significant hub-protein in terms of degree value. TNF accounts for 85.71 percent of the total network's edges.

(b)

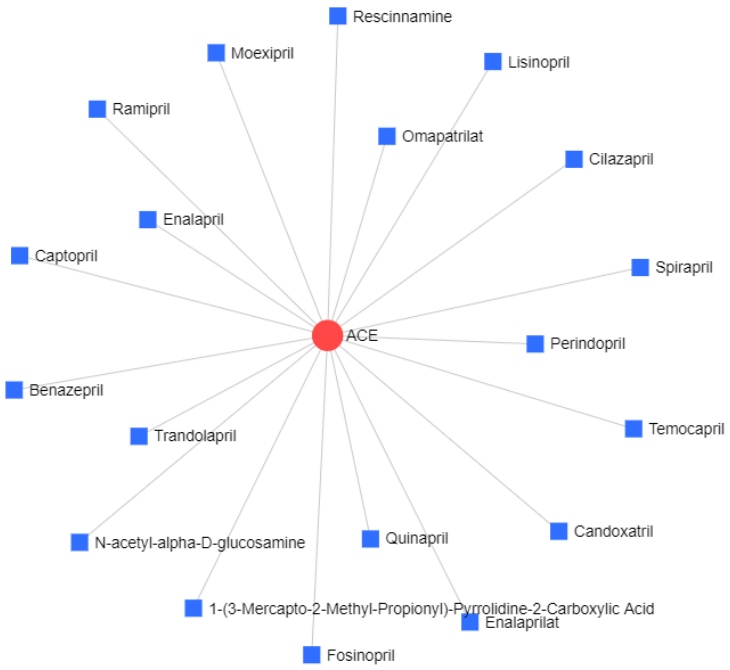


Figure 10(b): This graph depicts the Protein-drug subnetwork2. Interactions for ten common genes associated with HBP, CKD, and TD. Each node represents a gene, and the edges represent interactions between two genes. Seed nodes are represented by the circle node. This network consists of 20 nodes, 19 edges, and one (ACE) seed node. Based on degree value, ACE is the network's most significant hub-protein.

(c)

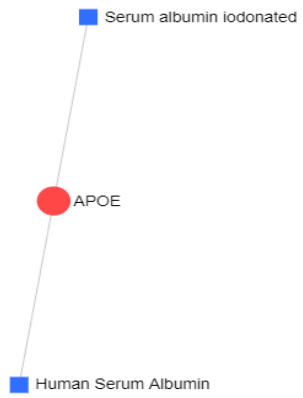


Figure 10(c): This graph depicts subnetwork3 of Protein-drug Interactions for ten common genes associated with HBP, CKD, and TD. Edges represent interactions between two genes, whereas nodes represent individual genes. Seed nodes are represented by the circle node. This network consists of three nodes, two edges, and one (APOE) seed node. Based on degree value, APOE is the network's most significant hub-protein.

(d)

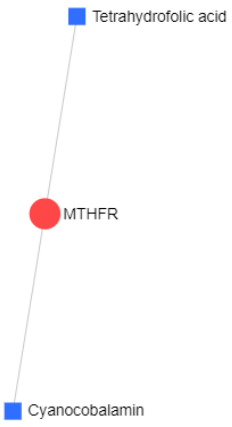


Figure 10(d): This graph depicts subnetwork4 of Protein-drug Interactions for ten common genes associated with HBP, CKD, and TD. Each node represents a gene, and the edges represent interactions between two genes. Seed nodes are represented by the circle node. This network consists of three nodes, two edges, and one (MTHFR) seed node.

4.9 Protein-Chemical Interactions

In this analysis, Protein-chemical interaction network scheme using NetworkAnalyst. For this research, we gathered suggested drug from DisGeNET database.

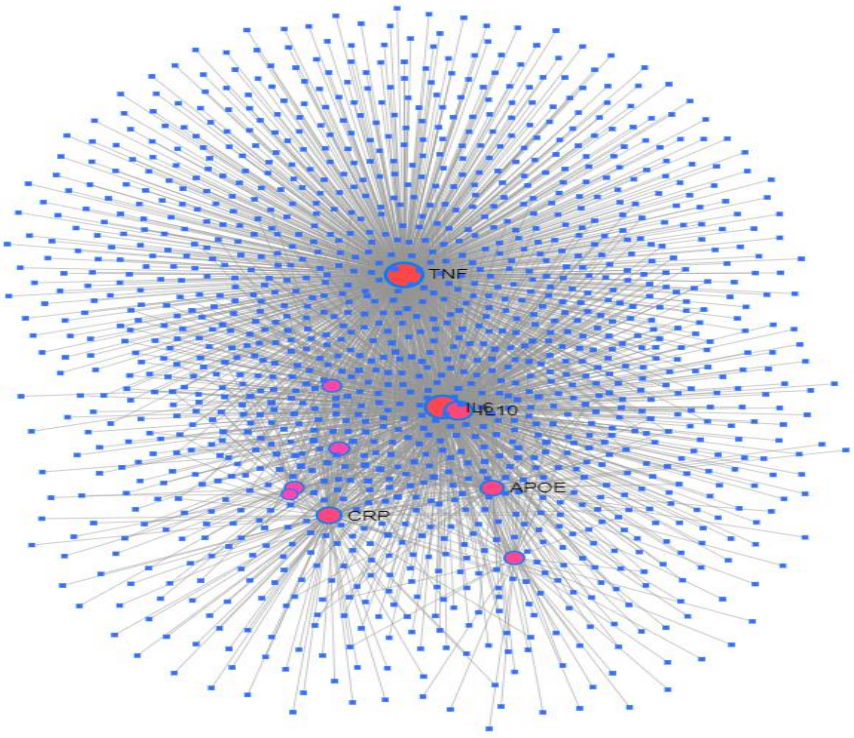


Figure 11: This graph depicts Protein-Chemical Interactions for ten common genes associated with HBP, CKD, and TD. Each node represents a gene, and the edges represent interactions between two genes. The larger node size represents a greater number of edges, and the red color represents the network's most important nodes. The network has 1331 nodes, 2226 edges, and 10 seed nodes (TNF, APOE, IL6, MTHFR, ACE, IL10, CRP, FTO, AGTR1, and CST3). TNF and IL6 are the network's most important hub proteins in terms of degree value. These two genes account for 72.46 percent of the network's edges.

4.10 Summary

Our primary goal was to find out if there was a genetic link between these diseases. For this reason, we collect genes from NCBI and finding extract common genes from these genes. The number of genes was quite large after collecting them from the NCBI gene database. As a result, the pre-processing, filtering, and gene mining phases must be followed. TNF, APOE, IL6, MTHFR, ACE, IL10, CRP, FTO, AGTR1, and CST3 are the ten most common genes discovered. We create a PPI network and analyze it with a cytoscape to find the most important genes among common genes. To better understand the biological process, we examined the topological property PPI network and discovered a relationship between cluster coefficients, betweenness, degree, number of neighbors, and so on. Finding the four most significant genes (TNF, APOE, CRP, and IL6) based on the degree value. To understand how different biological and biomedical works inside a cell and how it interacts with other genes in the cell, we are analyzing several of these genes above. These various analyses help us to find most significant three (TNF, IL6 and APOE) hub proteins.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Findings and contributions

TNF, APOE, IL6, MTHFR, ACE, IL10, CRP, FTO, AGTR1 and CST3 were identified as common genes between HBP, CKD, and TD. Broad bioinformatics research is being used to classify collective genes from three dissimilar relative diseases, which will aid in the classification of potential drugs for drug design and discovery processes. The number of genes was reduced in the following process, which aided in identifying the specific genes associated with HBP, CKD, and TD. The PPI network identified significant proteins that enable effective protein-drug interaction analysis. Gene regulatory networks (GRNs) were created to facilitate the investigation of genomic programs. Protein-Chemical Interaction and Protein-Drug Interaction network specifies the collaboration of chemical compounds and drug compounds. At last, we get three most significant genes (TNF, IL6 and APOE).

5.2 Recommendations for future works

It is theoretically shown that when a human has the risk factors including CKD, HBP she has an increased percent that she might get TD. In future, it can be examined practically with the human who already have CKD and HBP. If the risk to get TD found earlier than the patient takes the proper treatment earlier.

References

[1] Poulter, N. R., Prabhakaran, D., & Caulfield, M. (2015). *Hypertension. The Lancet, 386(9995), 801–812.* doi:10.1016/s0140-6736(14)61468-9 .

[2] Poulter, N. R., Prabhakaran, D., & Caulfield, M. (2015). *Hypertension. The Lancet, 386(9995), 801–812.* doi:10.1016/s0140-6736(14)61468-9 .

[3] Carretero, O. A., & Oparil, S. (2000). *Essential Hypertension : Part I: Definition and Etiology. Circulation, 101(3), 329–335.* doi:10.1161/01.cir.101.3.329 .

[4] Marik, P. E., & Varon, J. (2007). Hypertensive Crises. Chest, 131(6), 1949–1962. doi:10.1378/chest.06-2490 .

[5] Ketteler, M., Block, G.A., Evenepoel, P., Fukagawa, M., Herzog, C.A., McCann, L., Moe, S.M., Shroff, R., Tonelli, M.A., Toussaint, N.D. and Vervloet, M.G., 2018. Diagnosis, evaluation, prevention, and treatment of chronic kidney disease–mineral and bone disorder: Synopsis of the kidney disease: improving global outcomes 2017 clinical practice guideline update. *Annals of internal medicine*, *168*(6), pp.422-430.

[6] Bikbov, B., Perico, N. and Remuzzi, G., 2018. on behalf of the GBD Genitourinary Diseases Expert Group. Disparities in Chronic Kidney Disease Prevalence among Males and Females in 195 Countries: Analysis of the Global Burden of Disease 2016 Study. *Nephron*, *139*(4), pp.313-318.

[7] Zhang, F., Liu, N., Wang, X., Zhu, L. and Chai, Z., 2004. Study of trace elements in blood of thyroid disorder subjects before and after 131 I therapy. *Biological trace element research*, *97*(2), pp.125-133.

[8] Lomax, J. and McCray, A.T., 2004. Mapping the gene ontology into the unified medical language system. *International Journal of Genomics*, *5*(4), pp.354-361.

[9] Cohen, B., Oren, M., Min, H., Perl, Y. and Halper, M., 2008. Automated comparative auditing of NCIT genomic roles using NCBI. Journal of biomedical informatics, 41(6), pp.904-913.

[10] Wang, X., Ren, X., Li, B., Yue, J. and Liang, L., 2012. Applying modularity analysis of PPI networks to sequenced organisms. Virulence, *3*(5), pp.459-463.

[11] Jeong, H., Mason, S.P., Barabási, A.L. and Oltvai, Z.N., 2001. Lethality and centrality in protein networks. *Nature*, *411*(6833), pp.41-42.

[12] Assenov, Y., Ramírez, F., Schelhorn, S.E., Lengauer, T. and Albrecht, M., 2008. Computing topological parameters of biological networks. *Bioinformatics*, *24*(2), pp.282-284.

[13] Zhu, M., Gao, L., Li, X., Liu, Z., Xu, C., Yan, Y., Walker, E., Jiang, W., Su, B., Chen, X. and Lin, H., 2009. The analysis of the drug–targets based on the topological properties in the human protein–protein interaction network. *Journal of drug targeting*, *17*(7), pp.524-532.

[14] Ding, Y., Tang, J. and Guo, F., 2016. Identification of protein–protein interactions via a novel matrix-based sequence representation model with amino acid contact information. *International journal of molecular sciences*, *17*(10), p.1623.

[15] Alexeyenko, A., Lee, W., Pernemalm, M., Guegan, J., Dessen, P., Lazar, V., Lehtiö, J. and Pawitan, Y., 2012. Network enrichment analysis: extension of gene-set enrichment analysis to gene networks. *BMC bioinformatics*, *13*(1), p.226.

[16] Zhang, B. and Horvath, S., 2005. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology*, *4*(1).

[17] Tieri, P., Farina, L., Petti, M., Astolfi, L., Paci, P., & Castiglione, F. (2018). Network Inference and Reconstruction in Bioinformatics. Refrence Module in Life Sciences.

[18] Yu, D., Lim, J., Wang, X., Liang, F. and Xiao, G., 2017. Enhanced construction of gene regulatory networks using hub gene information. BMC bioinformatics, 18(1), p.186.

[19] Khatri, P., Sirota, M. and Butte, A.J., 10 Years of Pathway Analysis: Current Approaches and Outstanding Challenges-Supplementary Notes.

[20] Xia, J., Gill, E.E. and Hancock, R.E., 2015. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. Nature protocols, 10(6), p.823.

[21] 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.

[22] Cha, S., Imielinski, M.B., Rejtar, T., Richardson, E.A., Thakur, D., Sgroi, D.C. and Karger, B.L., 2010. In situ proteomic analysis of human breast cancer epithelial cells using laser capture microdissection: annotation by protein set enrichment analysis and gene ontology. *Molecular & Cellular Proteomics*, *9*(11), pp.2529-2544.