Identification of Molecular Biomarkers and Key Pathways among Idiopathic pulmonary fibrosis (IPF), Chronic Obstructive Pulmonary Disease (COPD) and Lung cancer

Course Code: CIS 499

Fall 2023

Submitted By

Mst. Farjana Yasmin

ID: 201-16-502

Supervised By

Md. Faruk Hosen

Lecturer, Department of CIS

A thesis submitted fin partial fulfillment of the requirement for the degree of Bachelor of Science in Computing and Information System.

Department of Computing and Information System

DAFFODIL INTERNATIONAL UNIVERSITY

Submission Date: 23 January 2024

APPROVAL

This Project titled "**Identification of Molecular Biomarkers and Key Pathways among Idiopathic pulmonary fibrosis (IPF), Chronic Obstructive Pulmonary Disease (COPD) and Lung cancer**", Submitted by Mst. Farjana Yasmin, ID No:201-16-502 to the Department of Computing & Information Systems, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computing & Information Systems and approved as to its style and contents. The presentation has been held on 23-01-2024.

BOARD OF EXAMINERS

Mr. Md Sarwar Hossain Mollah — **Chairman**
Associate Professor and Head
Department of Computing & Information Systems
Faculty of Science & Information Technology
Daffodil International University

Dr. Shekh Abdullah-Al-Musa Ahmed — **Internal Examiner**
Assistant Professor
Department of Computing & Information Systems
Faculty of Science & Information Technology
Daffodil International University

Md. Nasimul Kader — **Internal Examiner**
Assistant Professor
Department of Computing & Information Systems
Faculty of Science & Information Technology
Daffodil International University

Prof. Mohammad Shorif Uddin — **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

I hereby declare that; this thesis has been done by me under supervision of **Md. Faruk Hosen**, Lecturer, department of Computing and Information System (CIS) of Daffodil International University. I am also declaring that this thesis or any part of there has never been submitted anywhere else for the award of any educational degree like, B.Sc., M.Sc., Diploma or other qualifications.

Supervised By

23-01-2024

Md. Faruk Hosen
Lecturer
Department of CIS
Daffodil International University

Submitted By

Farjana

Name: Mst. Farjana Yasmin
ID: 201-16-502
Department of CIS
Daffodil International University

ACKNOWLEDGEMENT

First and foremost, I extend my sincere thankfulness to the All-Powerful Allah for enabling me to finish this thesis project effectively. This achievement is a testament to His guidance and blessings throughout my academic journey. I would want to sincerely thank my esteemed supervisor, Md. Faruk Hosen, Lecturer in the Department of Computing and Information System. Without his unwavering support and invaluable guidance, this thesis would not have reached its completion. His constant encouragement provided me with the confidence needed to navigate through the complexities of my work. Md. Faruk Hosen's proper direction and mentorship have been instrumental in shaping this thesis, making the process smoother and more comprehensible. I extend my heartfelt gratitude to the entire Department of Computing and Information System at Daffodil International University for fostering an environment of quality education and knowledge. Special appreciation goes to Md. Sarwar Hossain Mollah, Head of the Department. The knowledge acquired from classes during my bachelor's degree was integral to the foundation of this thesis. In addition to my academic mentors, I would want to thank my friends for their constant encouragement and support. Their optimism and camaraderie played a crucial role in overcoming obstacles at every phase of this academic endeavor. Information was carefully obtained for this study from a variety of sources, including publications, newspapers, digital platforms, and other additional sources. I am thankful for the wealth of knowledge these resources provided, contributing significantly to the depth and breadth of this thesis. Once again, I want to express my deepest appreciation to all those who have played a role in this academic journey, shaping my growth and contributing to the successful completion of this thesis.

DEDICATION

In addition to my parents and family, who have all helped me to finish my BSC program, this thesis is dedicated to my Creator. I also give thanks to Almighty God for his guidance, protection, and support. Additionally, I would want to dedicate my effort to my seniors and pals who are truly assisting me in maturing, as well as to my esteemed professors who have honestly assisted me in being competent.

ABSTRACT

Background: By incorporating computer science concepts, bioinformatics is a vital tool in the field of genomics that helps to comprehend complex biological processes. This multidisciplinary discipline uses statistical analysis, data processing, and computer tools to mine biological data for useful insights. Because bioinformatics integrates techniques from several fields, it is essential in many areas, including as genetics, education, and healthcare.

Aim: This study's main objective is to understand the intricate interaction between genetic variants and IPF, COPD, and LC. We aim to investigate the connections between various respiratory disorders by identifying similar pathways, and looking at TF-gene interactions, Gene-miRNA interactions, and Protein-Drug interactions.

Methodology: We carried out a comprehensive analysis of the PPI and PDI networks for IPF, COPD, and LC using a computational technique. Associated genes were taken out of the Finding genes involved using the NCBI gene database and a data mining method that were shared by the three diseases of the lungs.

Result: CFH, ETS1, CCNL1, NEDD9, MSH2, RORA, PMAIP1, SORD, genes shared by IPF, COPD, and lung cancer were identified by the computational model. Through the continued unraveling of a novel pathway, the study shed light on various respiratory disorders as well as possible molecular indicators.

Conclusion: The PPI and PDI networks that were built using the shared genes that were found provide important new information about the molecular makeup of lung cancer, COPD, and IPF. This knowledge is important for medication design as well as for understanding the complex ligand binding that occurs inside the PPI network. Our research advances our knowledge of the intricate molecular mechanisms behind various respiratory illnesses by identifying important pathways and molecular biomarkers.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

NCBI = National Center for Biotechnology Information;

IPF = Idiopathic Pulmonary Fibrosis;

COPD = Chronic Obstructive Pulmonary Fibrosis;

LC = Lung Cancer;

PPIs = Protein-protein interaction;

PDI = Protein drug interaction;

WHO = World Health Organization;

SIF = Simple interaction format;

NSCLCs = Non-small-cell lung cancers;

AC = Adenocarcinoma;

CHAPTER 1: INTRODUCTION

1.1 Background:

Lung cancer is the most frequent cause of cancer-related deaths worldwide. Typically, lung cancer may develop due to mutations in oncogenes leading to in the growth of altered cells that eventually lead to the development of lung tumors [1]. Numerous studies have shown that prevalent diseases including Lung Cancer, IPF and COPD. Three diseases were selected for research in this paper. With the intention of identifying their relationships with each other, three separate diseases were chosen. The risk is significantly lowered when a patient has just one disease. The risk multiplies tenfold when a person has multiple diseases. The purpose of this study is to investigate the genetic connections between these genes and to find a genetic connection between these illnesses. IPF, COPD, and LC are the three diseases. COPD is a severe and only partially reversible illness marked by abnormal inflammatory reactions to environmental pollutants and airway limitation [2]. COPD ranks as the fourth most common cause of death globally [3]. Furthermore, COPD can raise the chance of lung cancer development [4]. Certain epidemiological studies have shown that smokers with COPD have a five-fold increased risk of developing LC in comparison to smokers with normal lung function [5]. The advancement of lung cancer is significantly influenced by poor lung function, and forced expiratory volume in one second (FEV1) has been shown to be a biomarker of smoking-related respiratory risk overall, suggesting a potential connection between LC and COPD [6]. IPF is the other chronic lung condition having age and smoking as danger indicators, however it is distinguished by lung parenchymal scarring on histology and visualization, as well as lung function testing limitations [7]. This severe form of the illness could cause damage to the tissue surrounding the alveoli or the airbags in the lungs [8]. IPF is classified as a potentially cancerous lung disease as people with IPF sometimes acquire concurrent lung cancer, while individuals with IPF have an approximate 3.34-fold increased chance of acquiring LC compared to ordinary people [9]. Even though lung cancer (LC) is thought to be an advanced form of IPF, the histological types of lung cancer linked to IPF are yet unknown, and studies have shown inconsistent findings in this regard [9]. According to a recently released genome sequencing study, IPF and LC have a number of somatic mutations in common [10]. In order to shed light on the relationship between stress and depression and to identify new avenues for treatment, the study makes use of gene-based research. The performance of high throughput approaches has substantially

increased due to analysis of microarray data as well as the information extracted from that expression dataset. Eight common genes identified as shared among datasets were found after analyzing genes from the GSE24206, GSE18842, and GSE76925 datasets. The PPI network will be the subject of our next analysis since it is the key element for the current study. Next, to better show the degree of relatedness between the DEGs, we build a PPI network. Hub genes were identified and prioritized by the PPI network using a degree topological method. In many bioinformatics investigations, similar DEGs have been combined to find specific medicinal compounds based on those DEGs. In addition, examination of common DEGs, particularly is included in this research, may establish gene ontology (GO) and other biological pathways. Microarray data contain molecular information that may be discovered through computer examination, assisting biological researchers. Finding the molecular link among IPF, COPD and LC is the main goal of this research, along with identifying suitable biomarkers in accordance with the findings of gene-based analysis. Differentially expressed genes (DEGs) must be found to be able to ascertain which genes cause IPF, COPD and LC. A constructed PPIs network is then used to demonstrate the DEGs' interaction. To evaluate the fundamental activities of the biological system, KEGG pathway analysis is then carried out. Drug candidates for the shared DEGs among IPF, COPD and LC are suggested following the discovery of hub genes.

1.2 Motivation of the Research:

My study explores the complex interactions between lung cancer, COPD, and IPF, three separate but associated respiratory illnesses. Together, these illnesses represent a substantial burden on global health, as they all add to the complex web of respiratory illnesses that affect millions of people each year. One unusual and mysterious lung condition is IPF, which is defined by increasing scarring of lung tissue that deteriorates lung function and eventually leads to respiratory failure. Contrarily, COPD is a complex illness that includes emphysema and chronic bronchitis. It is mainly brought on by extended exposure to irritant gases or particulate matter, such as tobacco smoke. The chronic, crippling conditions COPD and IPF both have high rates of morbidity and mortality, which puts a tremendous strain on healthcare systems around the globe. One of the more challenging diseases in oncology is lung cancer, which adds to the complexity of the respiratory

environment. Although this disease has several subtypes, the two primary categories are small cell lung cancer and non-small cell lung cancer. Numerous etiological variables, including tobacco use, secondhand smoke, exposure to carcinogens, and genetic predisposition, can result in lung cancer. A wide range of symptoms, from chronic coughing and chest pain to more subtle indicators like inadvertent weight loss, may be a sign of lung cancer. Typical risk markers, metabolic pathways, and—above all—a five-fold elevated chance of developing lung cancer in those with IPF all serve to highlight the complex relationship between these respiratory diseases. This cluster of illnesses necessitates a thorough examination of the molecular foundations that either connect or separate them, opening the door to focused interventions and individualized treatment plans. My thesis takes a multifaceted approach to deciphering the molecular complexities. By utilizing state-of-the-art bioinformatics techniques and genomic analysis, the study seeks to uncover molecular markers, analyze the common genetic landscape, and identify critical pathways that may be targets for future treatments. This study aims to provide important insights that could improve our knowledge of these respiratory diseases and open the door to more efficient diagnostic and treatment approaches that are specific to each condition by clarifying the converging and diverging molecular pathways.

1.3 Problem Statement:

- ✓ Idiopathic Pulmonary Fibrosis, or IPF, is an intractable and degenerative respiratory condition that causes the lungs to scar (fibrosis) lacking an established cause.
- ✓ The lung illness known as Chronic Obstructive Pulmonary illness (COPD) is characterized by limited airflow and breathing difficulties that worsen with time.
- ✓ Lung cancer risk can be raised by both IPF and COPD through mechanisms like chronic inflammation and damage to lung tissue.
- ✓ According to the Global cancer statistics 2020: lung cancer affects around 11.6% of all new cancer cases worldwide.
- ✓ The damage caused by IPF and COPD to Lung Cancer patient is not examined significantly in any kind of research so far.
- ✓ According to the references mentioned in this paper, there exists a valid interconnection among IPF, COPD, and Lung Cancer.

1.4 Research Question:

A research question functions as a manageable inquiry into a certain issue or topic. It denotes the beginning of the research process, the first stage in which the direction of the investigation is decided. Choosing an area of study is undoubtedly the most crucial initial stage of the investigation procedure. In order to determine the scope and direction of the research, this step entails exploring a number of questions. In an effort to narrow the research emphasis, the following questions are explored.

- ✓ How should the data set for particular diseases be gathered? Where, also?
- ✓ How should the data set be data mined?
- ✓ How should the PPI be made?
- ✓ What kind of tools needed to get the significant result?
- ✓ How to analyze the Topological Properties?
- ✓ How can the references be added?

1.5 Goals of the Study:

The following are the main objectives of this research paper:

- ✓ Determine which genes are shared by the chosen respiratory conditions.
- ✓ Create a network of protein-protein interactions, or PPIs.
- ✓ Examine the PPI network's topology in detail.
- ✓ Examine any hereditary ties to the illnesses.
- ✓ Investigate miRNA-gene interactions to have a thorough grasp of regulatory mechanisms.

1.6 Study Area:

The regions that the research covers are included in the analysis's range. The following lists the scopes:

- ✓ Obtaining a more extensive dataset.
- ✓ Utilizing R or Python for data mining.
- ✓ PPI network.
- ✓ Analyzing Topological Properties.
- ✓ Heatmap.
- ✓ DEGs.

1.7 Thesis Organization:

Section 2 of the study provides a solid foundation for validating statements and framing the conceptual framework by methodically organizing and synthesizing analyses. This part acts as a pillar, offering crucial proof in favor of the validity of the study. In-depth information about the suggested technique and the chosen strategic course are provided in Section 3. In Section 4, which is devoted to results, detailed results are presented using graphical displays. In Section 5, the study comes to a close with succinct observations that highlight significant discoveries and contributions and capture the importance of the research process in the selected topic.

2.1 Define the Thesis Subject or Field:

In the field of bioinformatics and computational biology work together in an interdisciplinary manner. Large biological datasets, including cell populations, genetic sequences, and protein samples, are analyzed using computer technologies developed and applied in these areas. Bioinformatics, in the context of molecular biology and medicine, leverages computer science approaches to address complicated challenges. The strong expansion in this field indicates a bright future for the workplace, especially for computational biologists who make substantial contributions to develop and conduct studies in the fields of pharmaceuticals, biotechnology, and software for analysis.

2.2 The Path of Investigation:

This study examines the connections among IPF, COPD, and LC within the concept of "Identification of Molecular Biomarkers and Key Pathways among These Three Diseases." Common genes are identified via data mining, and the top 8 weighted genes are highlighted. The study looks at building and analyzing PPI models, Topological Properties, and Co-expression & Physical Interaction in order to provide a thorough knowledge of the common molecular terrain and putatively important.

2.3 Theoretical Framework:

PPIs networks are the main topic of this study, PPIs, or productive interactions between proteins, are investigated by biological mechanisms that include hydrogen bonding, electrostatic forces, and the hydrophobic effect. Using the Cytoscape program, topological features extracted from the PPI network help to explore shared molecular landscapes in COPD, IPF, and LC and offer insights into potential important pathways.

2.4 Resources:

The National Center for Biotechnology Information (NCBI) database is a dependable source of genes for the LC, IPF, and COPD investigations. After downloading the genes and sorting them by weight, the gene retrieval procedure yields 2338 counts for LC, 693 counts for IPF, and 547 count for COPD. Every gene

found has been linked to Homo sapiens, and the analytical resources used in this work were obtained from reliable online sources.

The procedure is detailed in the first figure.

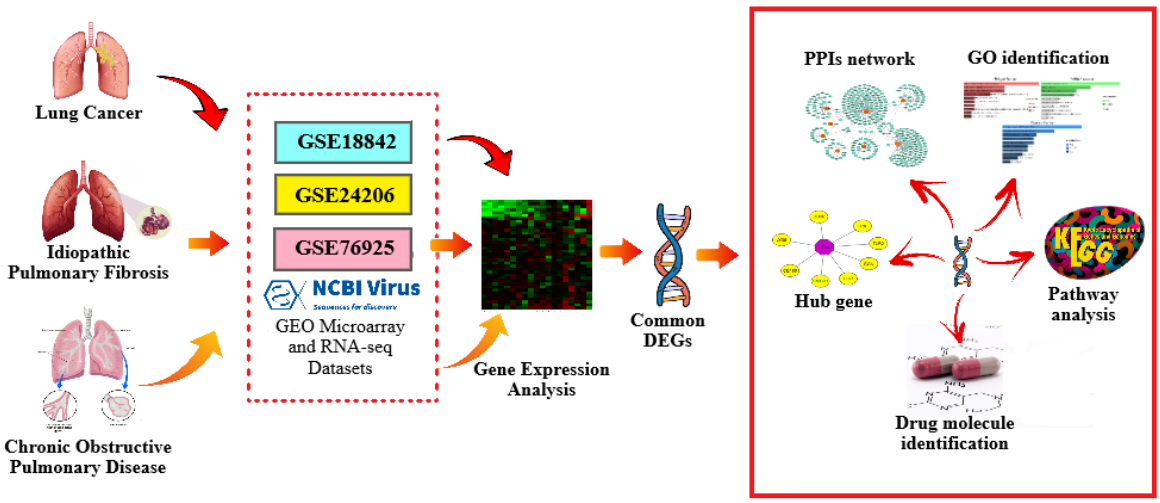


Figure 1: Flowchart of Proposed Methodology

3.1 Dataset Collections:

In research into the causes of lung cancer (LC), IPF, and COPD, reliable data sources are crucial. The NCBI Gene databases are used in this study to obtain genetic data unique to LC, IPF, and COPD. The NCBI gene database is freely available and can be downloaded. It serves as the basis for the collection of responsible genes, which are then subjected to a stringent reprocessing and screening to ensure data integrity before being stored again.

3.2 Preprocessing:

Genes unique to IPF, COPD, and LC are routinely obtained by the NCBI database. Using the name of each disease as a search query, genes are found and downloaded according to weight in ascending order during the gene retrieval process. Everything pertaining to LC, IPF, and COPD is gathered, with an exclusive emphasis on genes linked to humans. After classification, only the genes that are thought to be the cause of human diseases are kept for further investigation.

3.3 Finding DEGs and related DEGs linking lung cancer, COPD, & IPF:

When there is a statistically significant difference in the transcription level across multiple test conditions, a gene is said to be expressed differentially [8]. Finding DEGs for the datasets GSE24206, GSE18842, and GSE76925 is the main goal of this study. The most common DEGs could potentially be found with the use of a programming language for computers called R. To find statistically significant DEGs in each of the datasets, a pair of threshold parameters were used: for up regulated 1 logFC >1.7 , down regulated -1.7 logFC >-1 and an Adjusted p-value0.01. Using the online VENN analysis program venn, the common DEGs of GSE24206, GSE18842, and GSE76925 were obtained.

3.4 Gene ontology and pathway enrichment analysis:

A crucial technique that finds gene sets associated with particular chromosomal locations and functional importance is called gene set enrichment analysis [11]. Understanding metabolic pathways and gene annotations is facilitated by the use of KEGG pathways [12]. An online tool called Enrichr made it easier to analyze pathways for discovered common genes [11]. Gene ontology (GO) terms provide information about molecular activities and gene functions. Three distinct groups are used to classify them: cellular components, molecular functions, along with biological processes [13]. WikiPathways, Reactome, and BioCarta databases were added to KEGG pathways, which are well-known for understanding metabolic pathways, in order to facilitate pathway analysis [14, 15, 16]. The results from these databases were combined by Enrichr [11]. This extensive method used Enrichr to describe biological mechanisms and pathways while examining shared DEGs between illnesses such as IPF, COPD, and lung cancer [17]. KEGG pathways are noteworthy for their ability to capture metabolic processes, which are an essential part of genomic analysis [12]. A standard measure (P value < 0.05) was used to determine the significance degree of the paths that were found [11]. The integration of databases using Enrichr enabled a thorough comprehension of common pathways and biological activities linked to the examined illnesses, providing valuable perspectives on possible interrelated molecular causes.

3.5 Examining the network of interactions between proteins:

PPI interaction is widely acknowledged as the main area of cellular biology study and as a prerequisite for system biology. PPIs shed light on how proteins fulfill their biological functions [18]. PPIs are used to describe the molecular interactions

between two or more proteins that are facilitated by biochemical, hydrophobic, and electrostatic factors. PPI networks offer a wealth of novel knowledge on how proteins operate. With the help of NetworkAnalyst, Using the physical affinities of DEG proteins found in the String database, we constructed PPI networks [19]. Making use of tools like Cytoscpae [https://cytoscape.org/], the PPI network may be seen more clearly. PPI analysis employing topological properties was used to find significantly interacting hub proteins at the degree (greater than 13°).

3.6 Identification of hub genes and module analysis:

In order to comprehend complicated biological systems, it is imperative that hub genes, which are essential components of PPIs networks, be identified [20]. Using Cytoscape and the cytoHubba plugin, the degree topological technique was employed to explore PPI networks and characterize gene linkages [20]. Maximal Clique Centrality (MCC) is a popular tool for highlighting important genes in a network, and Cytohubba offers a variety of network analysis techniques [20]. The PPI network's top 15 hub genes were identified using MCC via Cytohubba, revealing their crucial involvement in the network architecture [20]. Furthermore, the characteristics of Cytohubba allowed for the evaluation of the shortest pathways between hub genes, which improved comprehension of their functional connections within the network architecture [20]. This method helps to understand important genes and how they interact in intricate biological systems.

3.7 Identification of miRNAs and transcription factors interact with shared DEGs:

TFs are essential for regulating gene expression levels in genetic transcription. Using NetworkAnalyst, we discovered topologically credible TFs from JASPAR that tie to prevalent DEGs. JASPAR provides TF profiles for different species [21], and NetworkAnalyst provides biological insights and extensive analysis of gene expression [22]. Furthermore, the Tarbase and mirTarbase databases were used to investigate miRNAs that target gene interactions [23, 24]. We collected miRNAs associated with shared DEGs for topological analysis by combining miRNA-gene interactions using NetworkAnalyst. Cytoscape made it easier to see TF-gene and miRNA-gene networks, which helped with high-degree miRNA filtering and the development of useful biological hypotheses. Furthermore, the RegNetwork archive

provided linkages for TF-miRNA coregulation [25], providing information on the transcriptional and post-transcriptional regulation of DEGs. Using NetworkAnalyst, the coregulatory network that combined TF and miRNA regulatory.

3.8 Identification of drug signatures:

The focus of current research is on finding therapeutic molecules, which are essential for treating diseases such as IPF, lung cancer, and COPD. Drug compound prediction based on gene expression patterns is aided by DSigDB, which has 22,527 gene sets [26, 27]. Enrichr is the gateway to DSigDB, providing extensive enrichment analysis and gene function visualization [24, 27]. Protein-drug interactions (PDI) can be predicted and possible drug molecules associated with the listed disorders can be found by utilizing Enrichr's access to DSigDB [28, 26, 27]. This method emphasizes how important it is to use gene expression data for drug identification and treatment discovery [2, 27].

4.1 Data Gathering:

NCBI database yields 2338 genes for LC, 693 genes for IPF, and 547 count for COPD as the genes responsible for the targeted disorders. Following sorting and processing, the related genes for Homo sapiens are arranged ascendingly by weight. Table 1 displays the numerical values of the identified genes associated with the related illnesses.

Table 1: The total number of liability genes identified in the NCBI database for specific diseases.

Disease name	Total no. of Homo Sapiens Gene
IPF	693
COPD	547
Lung Cancer	2338

4.2 The identification of DEGs and the location of a gene that is shared by IPF, COPD and Lung Cancer:

We examined the gene expressions in COPD (597 genes), lung cancer (2338 genes), and IPF (693 genes) using R programming, and we found 8 shared DEGs [Figure 2]. Furthermore, an analysis tailored to IPF identified 11 DEGs that were shared by all the datasets [Figure 2]. Using RNA-seq and microarray analysis in R, 2338 genes were identified as being related with COPD, IPF, and lung cancer, of which 293 were up-regulated and 891 down-regulated. Using the DESeq2 and limma software programs for analysis, a comparable method found significant DEGs for COPD and IPF. Understanding the interrelationships and implications of the respiratory illnesses under study is made easier by the identification of both common and unique gene expressions in these disorders. Figure 2 shows the number of genes that are shared by these disorders as a result of the investigation. CFH, ETS1, CCNL1,

NEDD9, MSH2, RORA, PMAIP1, SORD are the top 8weighted genes that were chosen to simplify the data out of 8 common responsible genes found among the chosen disorders. During the intersection process, there are 193 no of common gene between IPF & LC; 84 no of common gene between COPD & LC; 28 no of common gene between IPF & COPD; The Venn diagram for gene counts and common gene ratios is shown in Figure 2. The first step of the inquiry is to gather genes from reliable databases and then apply mining methods. Intersections between two, three, and four diseases undergo thorough scrutiny. The intersection technique takes into account the entire number of genes for LC, IPF, and COPD, verifying the study with Table 2 and Figure 2.

Table 2: The number of genes that all diseases have in common.

Disease	Total no. of gene	Common gene
IPF & LC	2644	193
COPD & LC	2715	84
IPF & COPD	1183	28
IPF & COPD &LC	3578	8

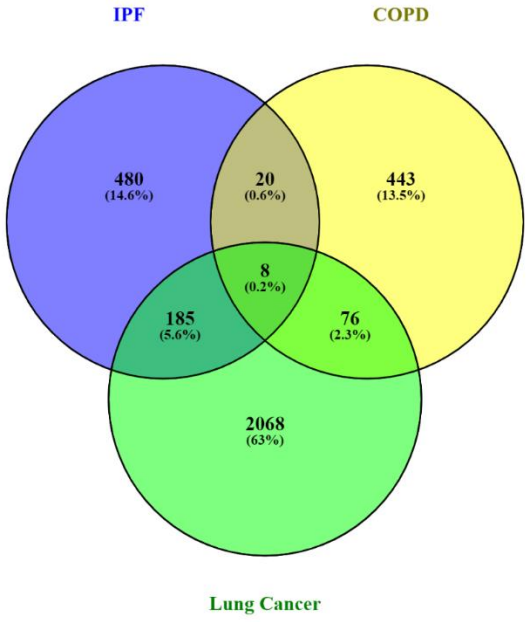


Figure 2: Venn analysis between IPF, LC & COPD.

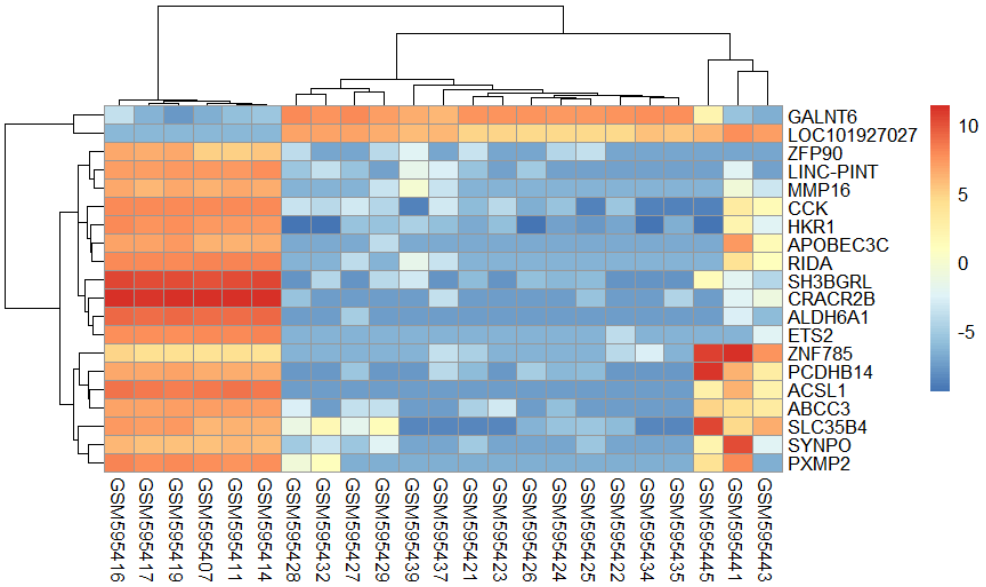


Figure 3A: Gene expression analysis of IPF-infected cells for the main 20 genes with chosen 23 samples using the GSE24206 dataset.

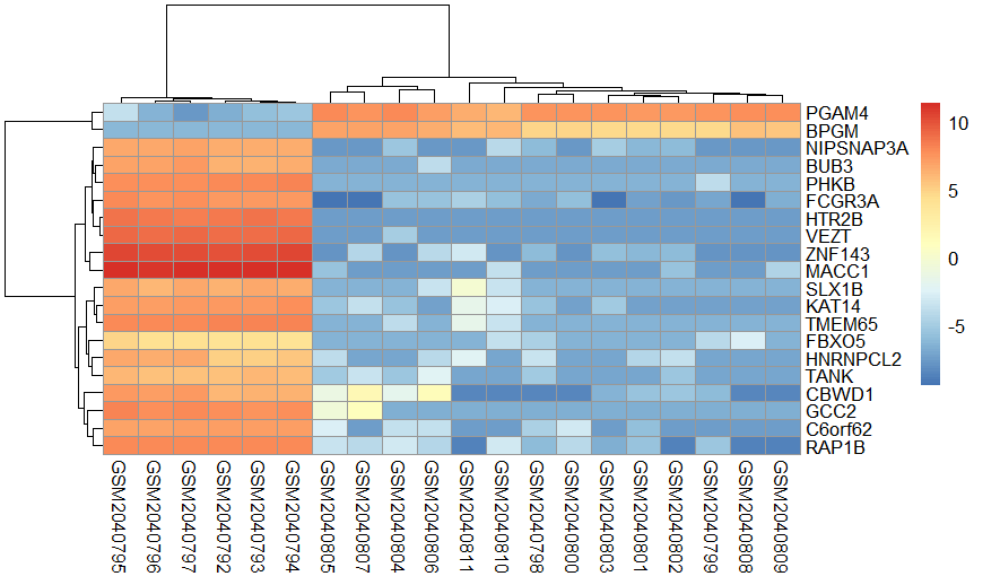


Figure 3B: evaluating the top 20 genes' gene expression in COPD-affected cells as well as choosing 23 samples retrieved from the GSE76925 dataset.

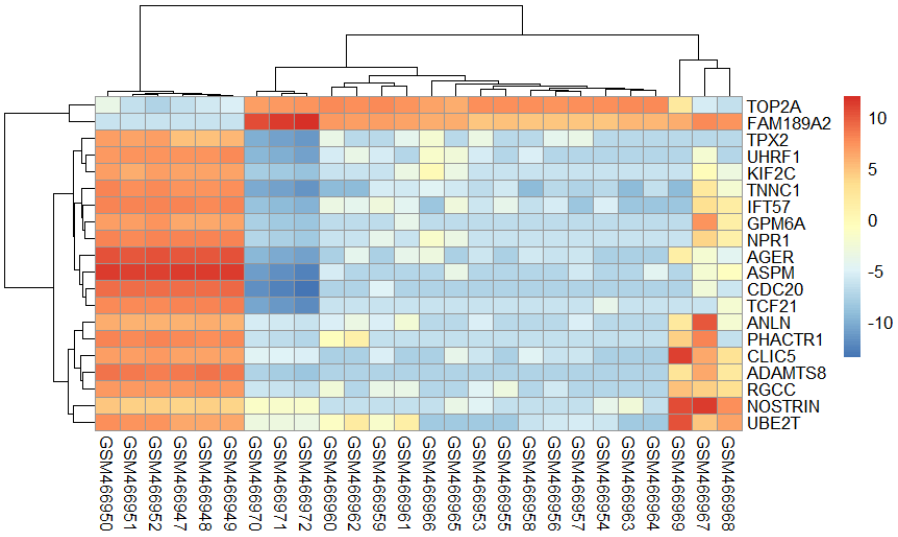


Figure 3C: evaluating the top 20 genes' expression in lung cancer-infected cells as well as choosing 26 samples taken from the GSE18842 dataset.

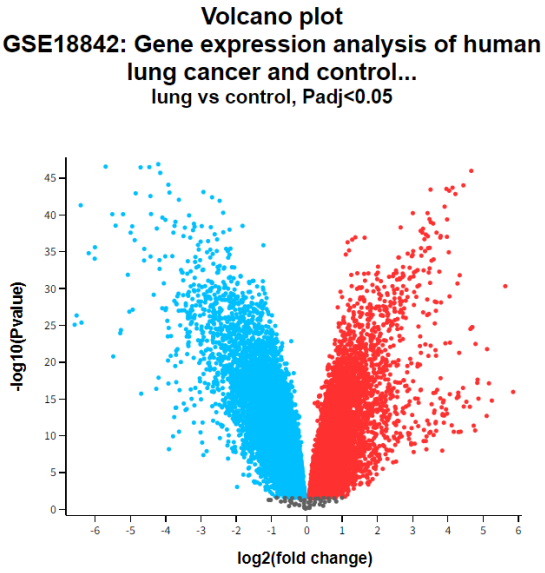


Figure 4A: The up and down-regulated gene expression for GSE18842 is shown by a volcano map.

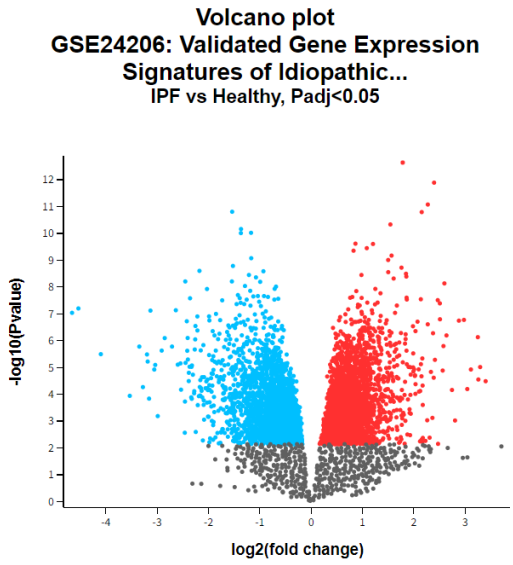


Figure 4B: The up- and down-regulated gene expression for GSE24206 is represented by a volcano plot.

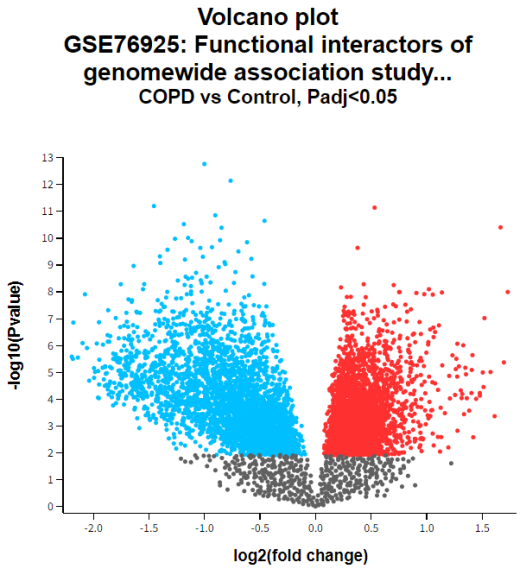


Figure 4C: The GSE76925 control of genes (upregulated and downregulated) is represented by a volcano map.

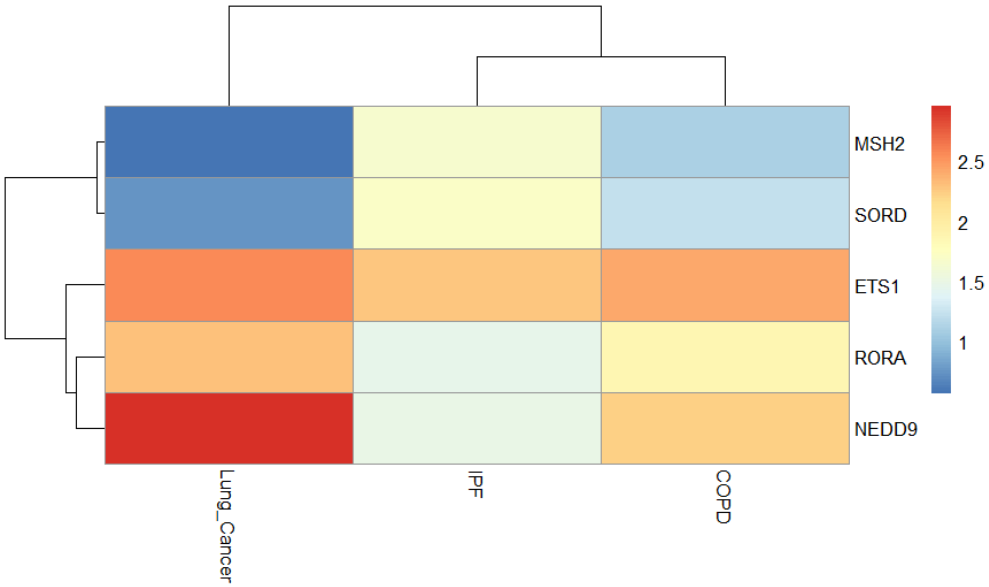


Figure 5: Heat map constructed using the most significant five unconditionally consistent DEGs from the data sets for lung cancer, COPD, & IPF, log fold changes.

4.3 In regard to transcription set enrichment analysis, GO and pathway discovery:

In order to identify shared DEGs between IPF, COPD, and lung cancer, the study carefully used KEGG databases. Enrichr was then used to create composite scores that combined log z-scores and p-values [29]. This made it easier to comprehend how similar molecular processes work. By classifying these DEGs into biological processes, molecular functions, and cellular components, Gene Ontology (GO) analysis provided a thorough understanding of their diverse roles by delving into the functional characteristics and cellular components of these DEGs [29]. Pathway analysis, which included the databases KEGG, WikiPathways, Reactome, and BioCarta, revealed complex biochemical links between various illnesses and gave a more comprehensive framework for their interactions [30]. A comprehensive gene set enrichment analysis was made possible by the Enrichr tool, which also highlighted important GO keywords and pathways. In order to shed light on critical connections among the common DEGs, for example, the results highlighted the importance of critical processes such as metal ion binding in molecular activities, neutrophil chemotaxis, and signaling pathways including IL-17 and TNF [30]. Enrichr's computation of P-values and z-scores resulted in the presentation of thorough results in GO keywords and pathway analyses, which clarified the complex

web of biological processes underlying the disorders under study [38]. Through the use of an integrated methodology, the functional functions and molecular interactions of the discovered DEGs across different pulmonary diseases were comprehensively understood. This resulted in insightful information about prospective therapeutic targets and shared pathways that need additional investigation.

Table 3. The relationship between common genes in GO terms and GO pathways, as well as the corresponding P-values

Category	GO ID	Term	P-value	Genes
GO biological process	GO:0070098	chemokine-mediated signaling pathway	2.12e-06	CXCL8;CXCL12;CXCL1;CCL18
	GO:1990869	cellular response to chemokine	2.81e-06	CXCL8;CXCL12;CXCL1;CCL18
	GO:0031328	positive regulation of cellular biosynthetic process	1.08e-05	CXCL8;IL1B;TYRP1;HBB;CD36
	GO:0045429	positive regulation of nitric oxide biosynthetic process	1.17e-05	IL1B;HBB;CD36
	GO:1904407	positive regulation of nitric oxide metabolic process	1.30e-05	IL1B;HBB;CD36
	GO:0045428	regulation of nitric oxide biosynthetic process	3.88e-05	IL1B;HBB;CD36

	GO:1904645	response to amyloid-beta	5.18e-05	MMP12;MMP13;CD36
	GO:0070486	leukocyte aggregation	7.35e-05	IL1B;RAC2
	GO:0060353	regulation of cell adhesion molecule production	9.44e-05	CXCL8;IL1B
	GO:0043112	receptor metabolic process	1.19e-04	CXCL8;CD36;NSG1
GO molecular function	GO:0008009	chemokine activity	9.54e-07	CXCL8;CXCL12;CXCL1;CCL18
	GO:0042379	chemokine receptor binding	1.34e-06	CXCL8;CXCL12;CXCL1;CCL18
	GO:0045236	CXCR chemokine receptor binding	2.74e-06	CXCL8;CXCL12;CXCL1
	GO:0005125	cytokine activity	8.91e-06	CXCL8;CXCL12;IL1B;CXCL1;CCL18
	GO:0005509	calcium ion binding	2.46e-04	MMP12;MMP13;DSG1;DSC1;DSC2
	GO:0046872	metal ion binding	1.48e-03	MMP12;MMP13;DSG1;DSC1;DSC2
	GO:0004222	metalloendopeptidase activity	8.07e-03	MMP12;MMP13
	GO:0008517	folic acid transmembrane transporter activity	0.008223548	SLC19A2
	GO:0031721	hemoglobin alpha binding	0.008223548	HBB
	GO:0086083	His cell-Purkinje connection participates in the interaction of cells sticky	0.008223548	DSC2

		proteins. myocyte communication		
GO cellular component	GO:0001533	cornified envelope	7.24e-07	DSG1;KRT10;DSC1;DSC2
	GO:0030057	desmosome	2.74e-06	DSG1;DSC1;DSC2
	GO:0070820	tertiary granule	1.48e-04	HBB;DSG1;CXCL1;DSC1
	GO:0030666	endocytic vesicle membrane	2.22e-03	TYRP1;RAC2;CD36
	GO:0101002	ficolin-1-rich granule	3.41e-03	HBB;DSG1;DSC1
	GO:1904724	tertiary granule lumen	3.71e-03	HBB;CXCL1
	GO:0101003	ficolin-1-rich granule membrane	4.55e-03	DSG1;DSC1
	GO:0005884	actin filament	6.28e-03	RAC2;COBL
	GO:0099513	polymeric cytoskeletal fiber	8.52e-03	RAC2;COBL;KRT10
	GO:0005587	collagen type IV trimer	9.86e-03	COL4A1

Table 4: The substantial correlation with the P-values for similar genes found in the KEGG, WikiPathways, Reactome, and BioCarta databases

Databases	**Pathways**	**P-value**	**Genes**
	Malaria	1.79e-08	CXCL8;IL1B;HBB;HBA2;CD36

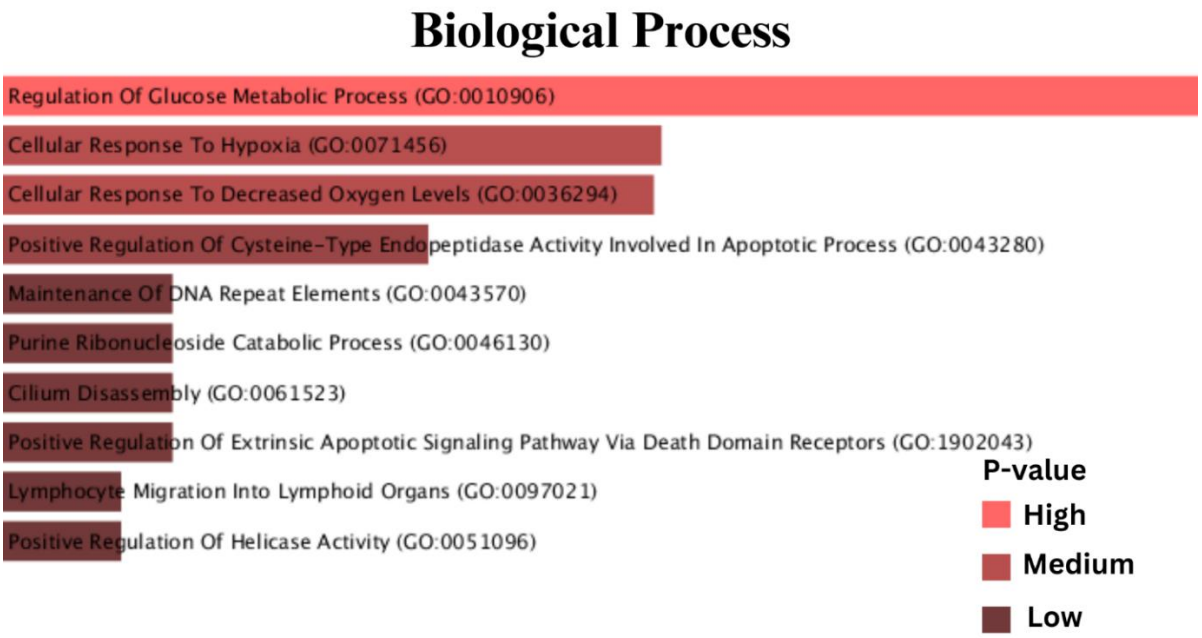
KEGG	Chemokine signaling pathway	1.48e-05	CXCL8;CXCL12;RAC2;CXCL1;CCL18
	Rheumatoid arthritis	1.62e-05	CXCL8;CXCL12;IL1B;CXCL1
	IL-17 signaling pathway	1.69e-05	CXCL8;MMP13;IL1B;CXCL1
	Interaction of viral proteins with cytokines and cytokine receptors	2.15e-05	CXCL8;CXCL12;CXCL1;CCL18
	Amoebiasis	2.33e-05	CXCL8;COL4A1;IL1B;CXCL1
	NF-kappa B signaling pathway	2.51e-05	CXCL8;CXCL12;IL1B;CXCL1
	African trypanosomiasis	3.06e-05	IL1B;HBB;HBA2
	Yersinia infection	7.39e-05	CXCL8;IL1B;RAC2;PIP5K1B
	Legionellosis	1.13e-04	CXCL8;IL1B;CXCL1
	Spinal Cord Injury WP2431	3.61e-08	MMP12;BTG2;CXCL8;COL4A1;IL1B;CXCL1
	IL-18 signaling pathway WP4754	4.87e-06	BTG2;CXCL8;MMP13;IL1B;CD36;CCL18
	Inhibition of inherent immunity and cell-specific immune reaction in SARS-CoV-2 WP5039	1.74e-04	CXCL8;CXCL12;CXCL1
	WP4891, the COVID-19 adverse outcome route	2.74e-04	CXCL8;IL1B

WikiPathways	Allograft Rejection WP2328	4.22e-04	CXCL8;CXCL12;IL1B
	LTF danger signal response pathway WP4478	0.000443589	CXCL8;IL1B
	Overview of nanoparticle effects WP3287	0.000443589	CXCL8;COL4A1
	Senescence and Autophagy in Cancer WP615	0.000684097	CXCL8;IL1B;CXCL1
	Cytokines and Inflammatory Response WP530	0.000837014	IL1B;CXCL1
	Bradleykinin and RAS pathways in COVID-19 WP4969	0.001042394	CYP24A1;IL1B
Reactome	R-HSA-6783783 Interleukin-10 Communication	5.54e-05	CXCL8;IL1B;CXCL1
	R-HSA-380108 Chemotherapy Sensors Bind The chemicals	1.07e-04	CXCL8;CXCL12;CXCL1
	R-HSA-6785807 Interleukin-4 with The interleukin-13 Interaction	7.23e-04	CXCL8;IL1B;CD36
	Decomposition of Keratin R-HSA-1442490	1.98e-03	MMP12;MMP13
	Scavenger transporters' binding and uptake of ligands (R-HSA-2173782)	2.08e-03	COL4A1;CD36
	Collagen Fibril Assemblies and Other Multimeric Structures R-HSA-	3.98e-03	MMP13;COL4A1

	2022090Multimeric Structures R-HSA-2022090		
	Peptide Ligand-Binding Receptors R-HSA-375276	4.07e-03	CXCL8;CXCL12;CXCL1
	Keratinization R-HSA-6805567	4.81e-03	KRT19;DSG1;KRT10
	Cytochrome P450 - Arranged By Substrate Type R-HSA-211897	5.15e-03	CYP24A1;CYP3A5
	Innate Immune System R-HSA-168249	6.31e-03	MMP12;IL1B;RAC2;DSG1;CXCL1;CD36
BioCarta	Histoplasm's Chaperone Humans (h) h ahsp Pathway	0.000203595	HBB;HBA2
	Nontypeable Hemophilus influenzae activates NFkB Humans (Sapiens) in the Pathway	0.001042394	CXCL8;IL1B
	Apoptosis Induced by TSP-1 in Human Microvascular Endothelial Cells h tsp1Pathway	0.011494589	CD36
	CCR5 Signaling in Macrophage Homo sapiens h Ccr5Pathway is Uncaring to Pertussis Toxin	0.01475517	CXCL12
	BTG family proteins and the control of the cell cycle The human species h btg2Pathway	0.01475517	BTG2

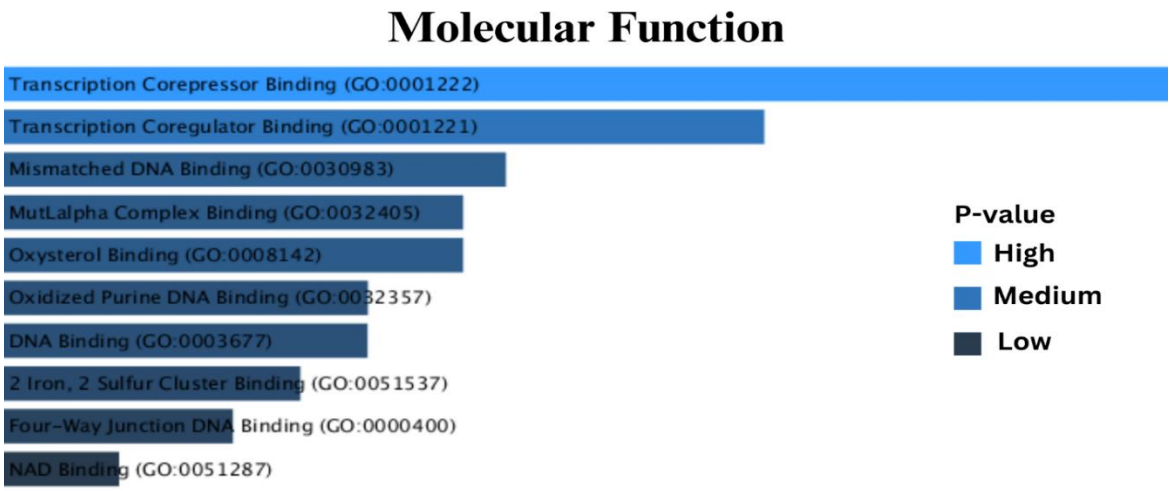
	Tubby Proteins in G-Protein Communication Tubby Walkway h Tubby humans	0.01638 1548	CXCL12
	CXCR4 Signaling Pathway Homo sapiens h cxcr4Pathway	0.01800 5323	CXCL12
	Regulators of Bone Mineralization Homo sapiens h npp1Pathway	0.01800 5323	COL4A1
	Activation of PKC through G-protein coupled receptors Homo sapiens h pkcPathway	0.01800 5323	CXCL12
	Platelet Amyloid Precursor Protein Pathway Homo sapiens h plateletAppPathway	0.02286 1069	COL4A1

(A)


Biological Process
Regulation Of Glucose Metabolic Process (GO:0010906)
Cellular Response To Hypoxia (GO:0071456)
Cellular Response To Decreased Oxygen Levels (GO:0036294)
Positive Regulation Of Cysteine-Type Endopeptidase Activity Involved In Apoptotic Process (GO:0043280)
Maintenance Of DNA Repeat Elements (GO:0043570)
Purine Ribonucleoside Catabolic Process (GO:0046130)
Cilium Disassembly (GO:0061523)
Positive Regulation Of Extrinsic Apoptotic Signaling Pathway Via Death Domain Receptors (GO:1902043)
Lymphocyte Migration Into Lymphoid Organs (GO:0097021)
Positive Regulation Of Helicase Activity (GO:0051096)
P-value
High
Medium
Low

\

(B)


Molecular Function
Transcription Corepressor Binding (GO:0001222)
Transcription Coregulator Binding (GO:0001221)
Mismatched DNA Binding (GO:0030983)
MutLalpha Complex Binding (GO:0032405)
Oxysterol Binding (GO:0008142)
Oxidized Purine DNA Binding (GO:0032357)
DNA Binding (GO:0003677)
2 Iron, 2 Sulfur Cluster Binding (GO:0051537)
Four-Way Junction DNA Binding (GO:0000400)
NAD Binding (GO:0051287)
P-value
High
Medium
Low

(C)

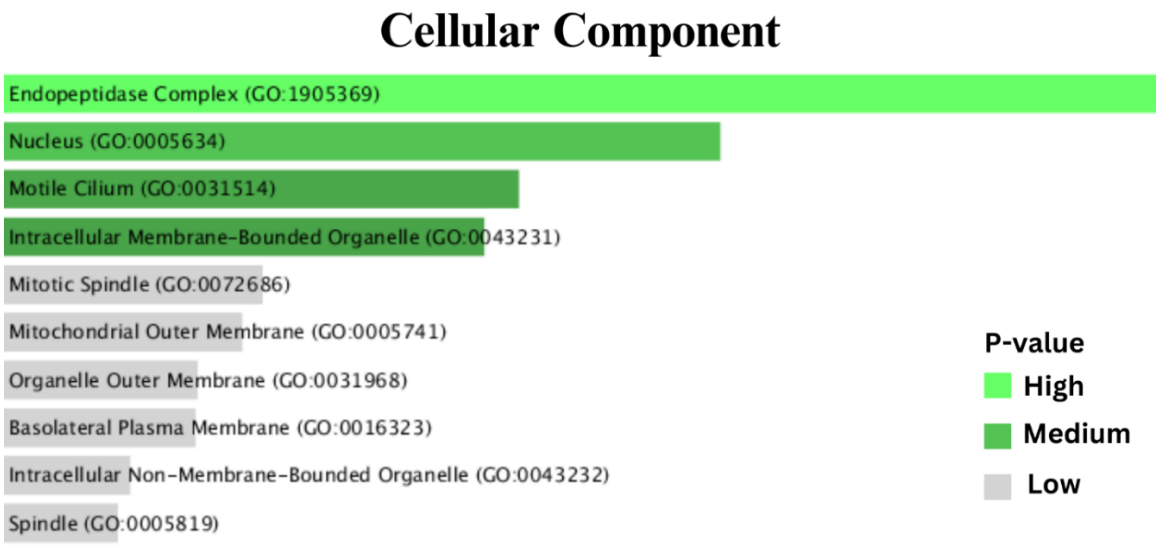
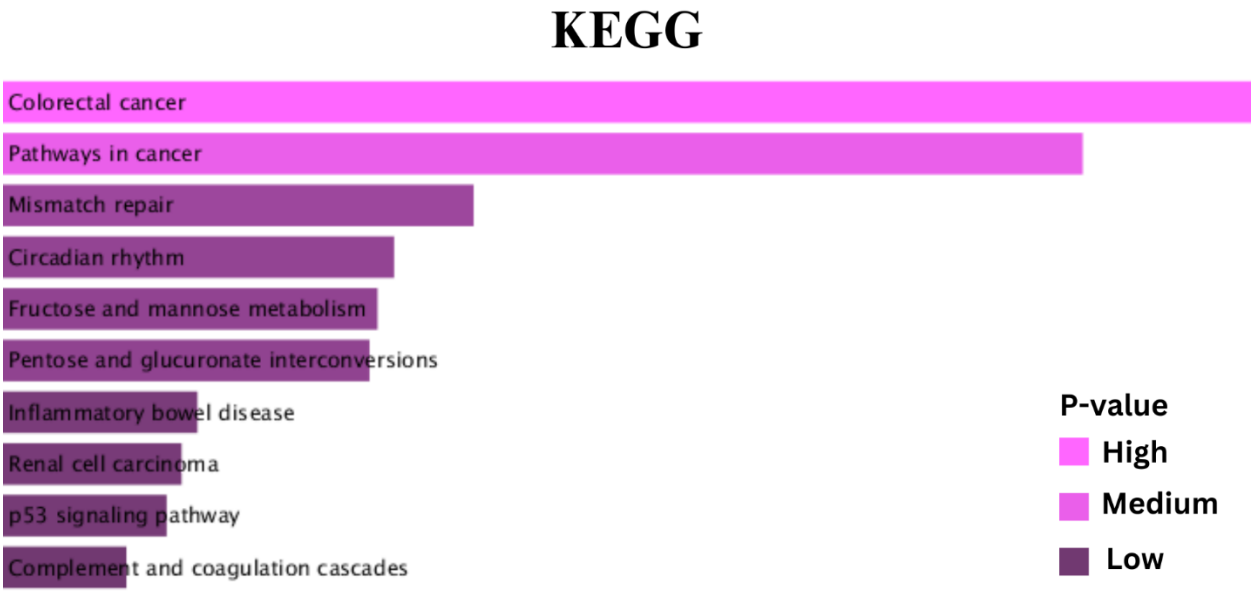
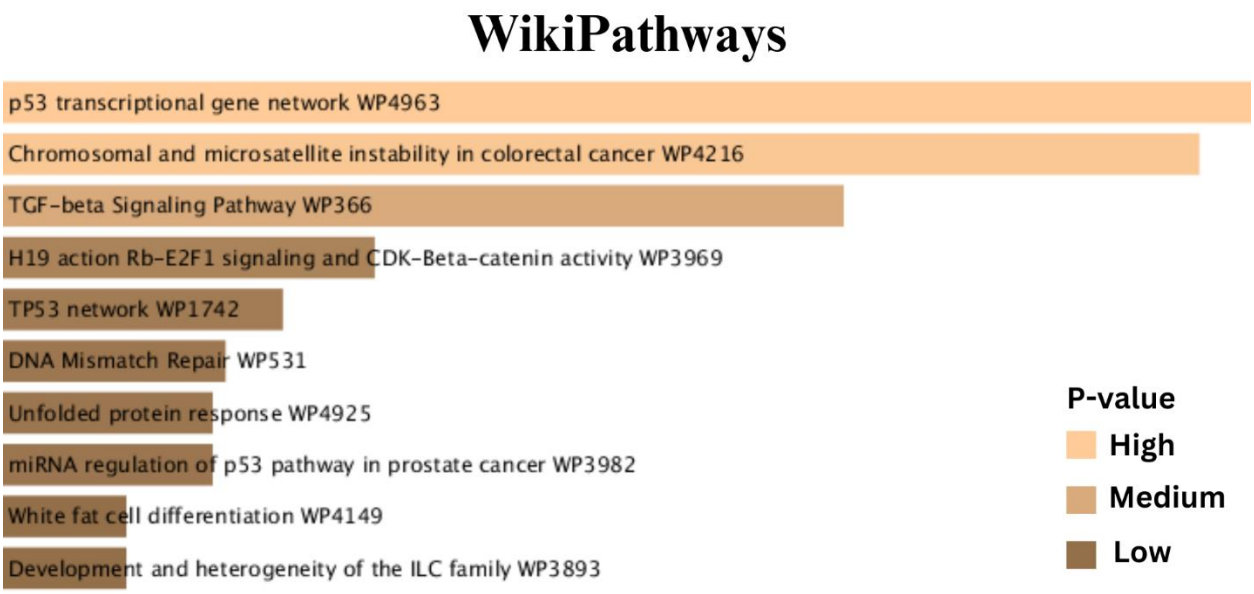


Figure 6: GO keywords linked to biological processes (A), molecular functions (B), and cellular components (C) were extracted using P-values.

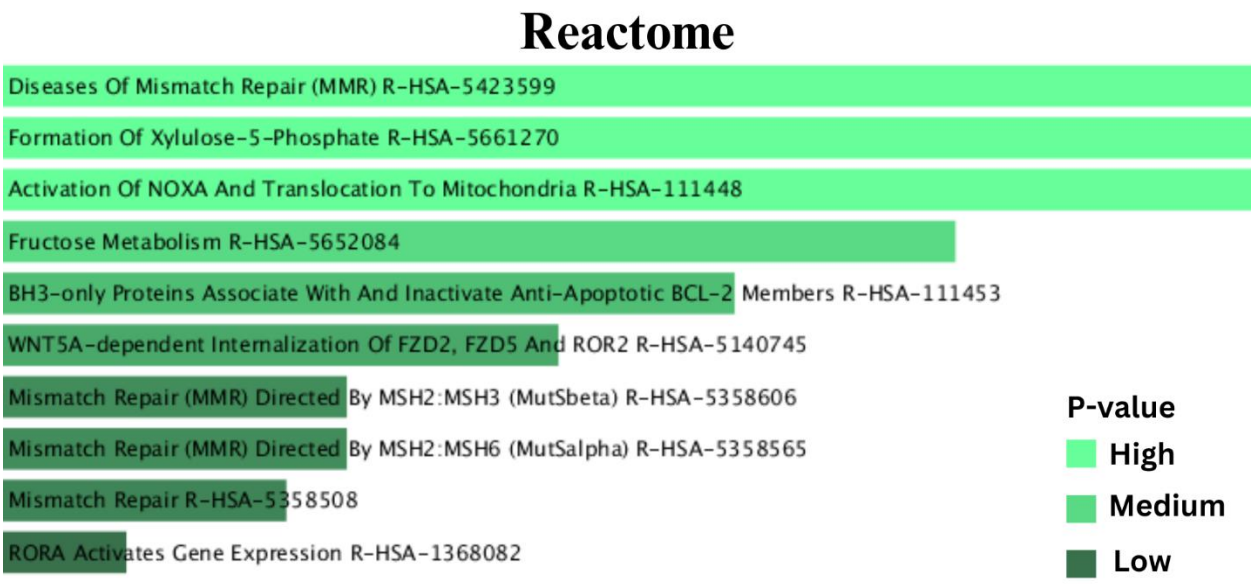
(A)



(B)


WikiPathways
p53 transcriptional gene network WP4963
Chromosomal and microsatellite instability in colorectal cancer WP4216
TGF-beta Signaling Pathway WP366
H19 action Rb-E2F1 signaling and CDK-Beta-catenin activity WP3969
TP53 network WP1742
DNA Mismatch Repair WP531
Unfolded protein response WP4925
miRNA regulation of p53 pathway in prostate cancer WP3982
White fat cell differentiation WP4149
Development and heterogeneity of the ILC family WP3893
P-value
High
Medium
Low

(C)


Reactome
Diseases Of Mismatch Repair (MMR) R-HSA-5423599
Formation Of Xylulose-5-Phosphate R-HSA-5661270
Activation Of NOXA And Translocation To Mitochondria R-HSA-111448
Fructose Metabolism R-HSA-5652084
BH3-only Proteins Associate With And Inactivate Anti-Apoptotic BCL-2 Members R-HSA-111453
WNT5A-dependent Internalization Of FZD2, FZD5 And ROR2 R-HSA-5140745
Mismatch Repair (MMR) Directed By MSH2:MSH3 (MutSbeta) R-HSA-5358606
Mismatch Repair (MMR) Directed By MSH2:MSH6 (MutSalpha) R-HSA-5358565
Mismatch Repair R-HSA-5358508
RORA Activates Gene Expression R-HSA-1368082
P-value
High
Medium
Low

(D)

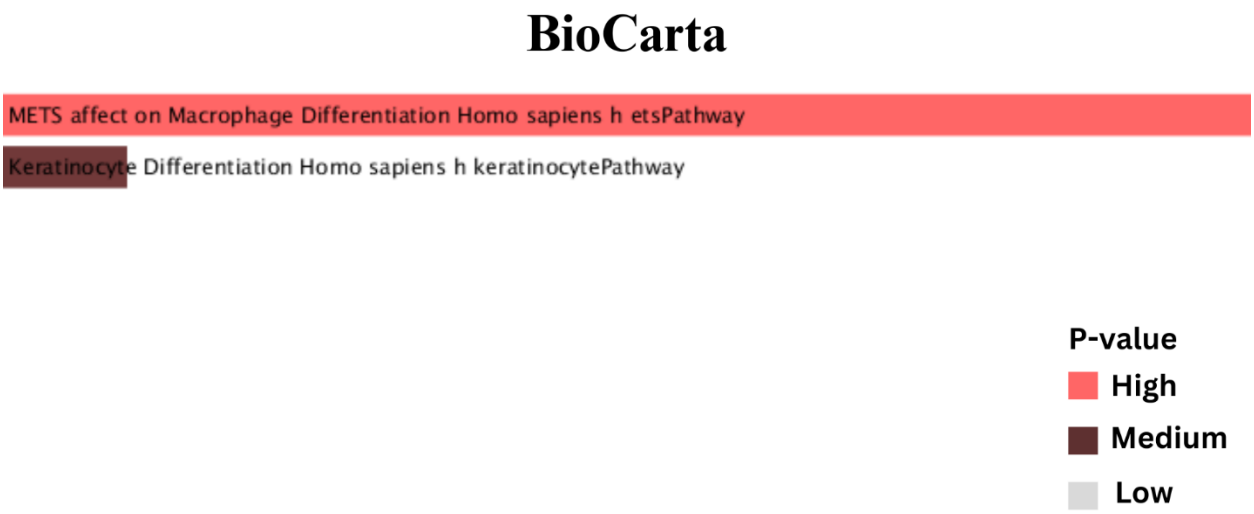


Figure 7: Pathway analysis results through KEGG (A), WikiPathways (B), Reactome(C) and BioCarta (D) databases according to P-values.

4.4 PPIs network for module analysis:

The PPI network was built from shared DEGs using NetworkAnalyst and STRING, which facilitated the study of hub genes and interaction pathways. This network, shown in Figure 3, was represented as 95 nodes and 94 edges in Cytoscape. In addition, 5 significant hub genes, ETS1, MSH2, SORD, RORA, and NEDD9, were discovered by analysis of the PPI network in Cytoscape and may be used as biomarkers or therapeutic targets for the disorders under investigation. To improve comprehension of these hub genes' interactions within the broader network, a submodule network was built to provide insights into the proximity and connectedness of these genes. The PPI analysis, which included 95 edges and 94 nodes, supported the investigation's emphasis on finding pharmacological substances and possible treatment approaches for IPF and lung cancer.

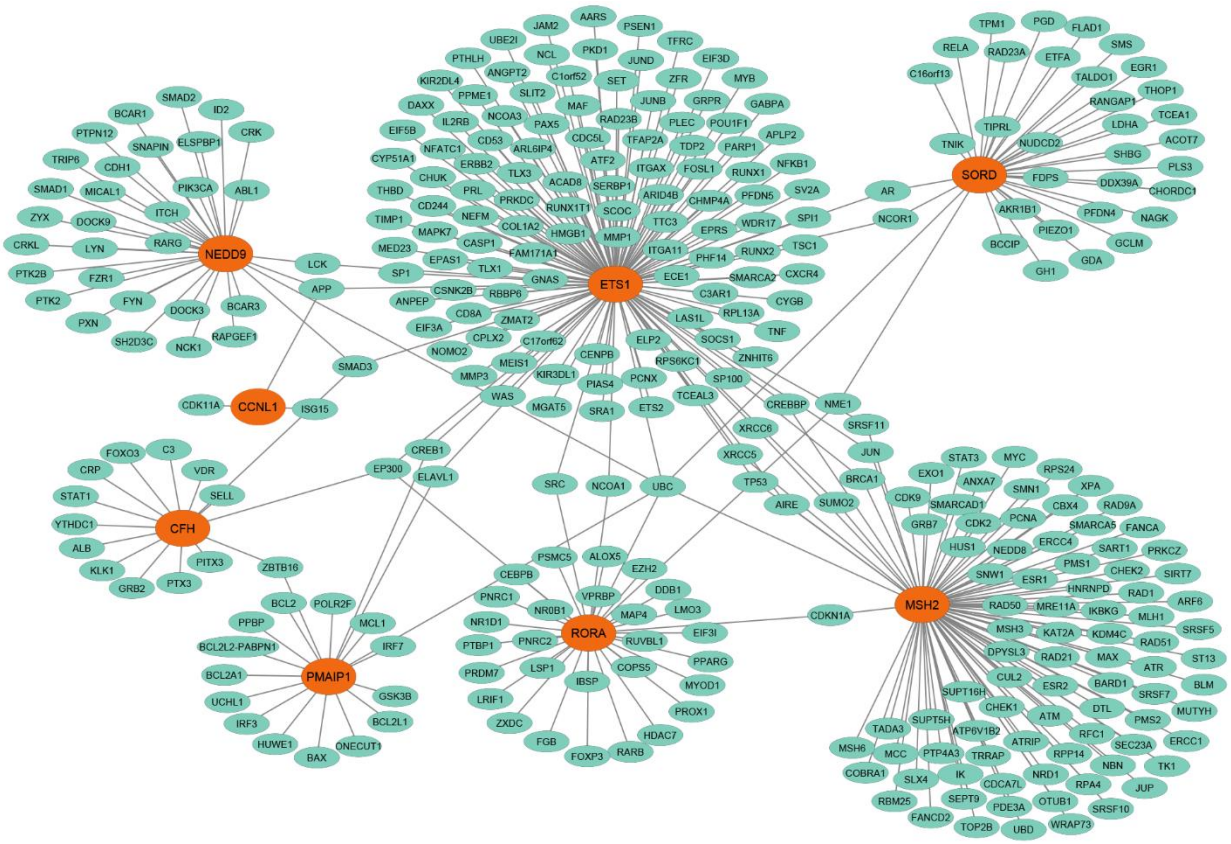


Figure 8: PPI network for identified concordant DEGs shared by IPF, COPD, and LC

4.5 Hub gene identification and module analysis to provide recommendations for treatment:

To comprehend the PPI network's connection and potential biological significance, hub genes must be identified. With Cytohubba in Cytoscape, top hub genes such as CFH, MSH2, SORD, and others were found. The degree ratings of these genes, which show the intensity of interactions within the PPI network, were used to select them. [Figure 3]. Table 3, which showcases the properties of the genes and their connection, displays the topological characteristics of the PPI network as determined by analysis using Network Analyzer in Cytoscape. In addition, analysis of the PPI network revealed highly interconnected nodes that could be hub genes essential for researching therapeutics and comprehending diseases. These results emphasize the value of hub gene analysis in understanding intricate networks and locating putative biomarkers for focused treatment interventions in medical conditions.

Table 5: The top five hub genes' topological results are displayed in this table, with a network density of 0.274, a network width of 3, and a network radius of 2.

Hub gene	Degree	Stress	Closeness centrality	Betweenness centrality
ETS1	133.0	394372.0	204.16667	72812.13461
MSH2	95.0	305904.0	176.46667	51293.2645
SORD	35.0	49354.0	136.46667	20606.21815
RORA	34.0	59786.0	137.65	19531.40852
NEDD9	33.0	48032.0	137.5	19569.67527

4.6 TF-gene interactions:

NetworkAnalyst was used to gather TF-gene interactions. The TF-genes were discovered for the common DEGs (CFH, ETS1, CCNL1, NEDD9, MSH2, RORA, PMAIP1, SORD). Figure 8 shows how TF regulators interact with the typical DEGs. The network has 95 nodes and 94 nodes.

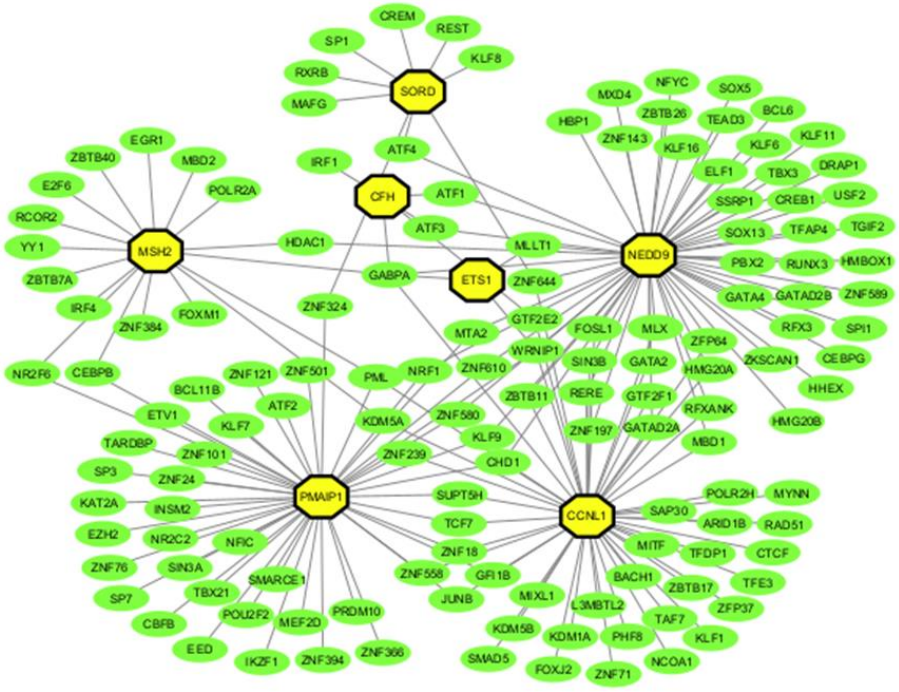


Figure 9: TF-gene network of interaction founded on analogous DEGs. Common DEGs were highlighted by yellow-colored nodes, and TF genes were marked by light green-colored nodes.

4.7 The regulatory network of TF-miRNA:

To construct the TF-miRNA coregulatory network, use NetworkAnalyst. The way that TFs and miRNAs interact with common DEGs is shown by the TF-miRNA coregulatory network study. This interaction may lead to the regulation of the expression of the DEGs. There are 101 nodes and 131 edges in the TF-miRNA coregulatory network. Thirty-nine miRNAs and fifty-three TF-genes have been interacting with the common DEGs. Figure 9 depicts the TF-miRNA coregulatory network.

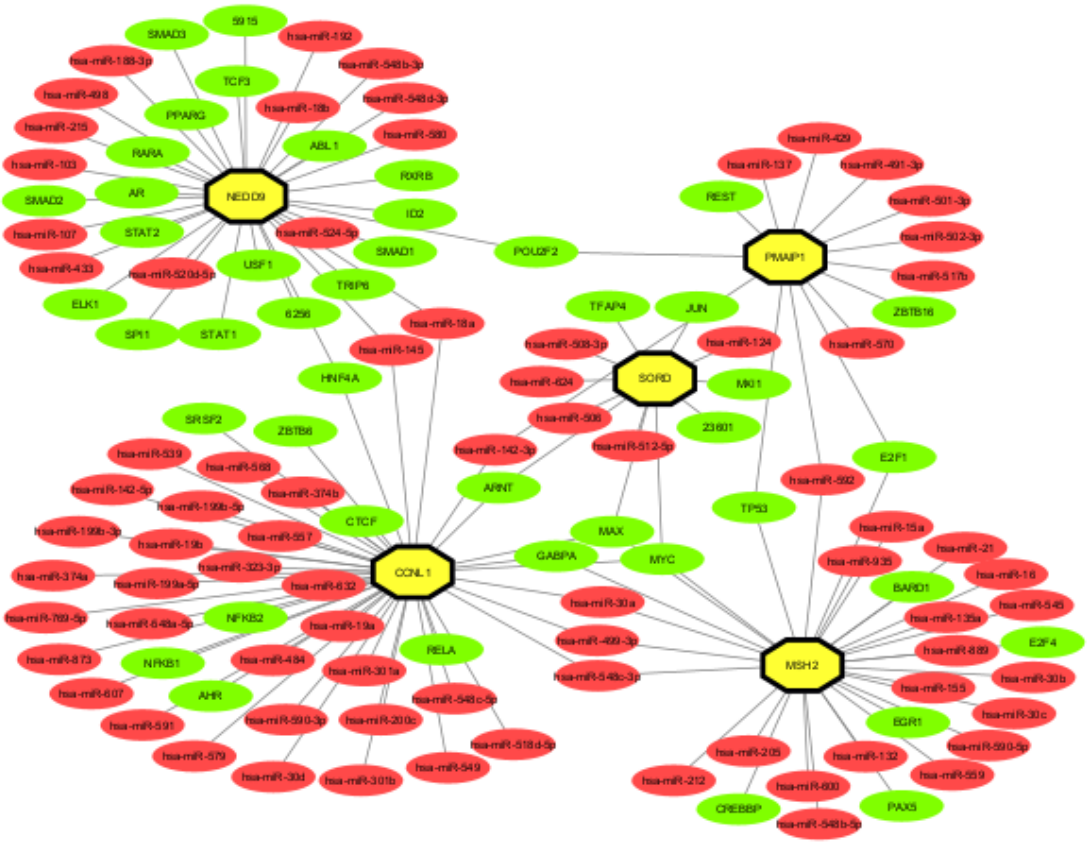


Figure 10: The TF-miRNA regulatory network represented. The yellow-colored nodes stand for miRNA, whereas the rest of the nodes show DEGs.

4.8 Topological properties:

Using the Cytoscape tool's Network Analyzer program, topological features are estimated with an emphasis on the Protein-Protein Interaction (PPI) network. An overview of the topological characteristics of the top 5 genes exposed to liability is given in Table 3, which also shows the appropriate topological graph for the PPI network. A measure of an object's ability to send information over a network efficiently is called closeness centrality. The distances between nodes with higher proximity centrality scores and all other nodes are smaller. the average clustering coefficient, representing the degree of average clustering among nodes in the network. Betweenness centrality quantifies the influence of a vertex on the flow of information. the topological coefficient, a quantifiable evaluation of how strongly a node connects neighbors. Neighbors for nodes with a topological coefficient of 0 are either one or none.

Table 6: Using the Cytoscape tool, the top 5 responsible genes from the PPI network's topological properties

Hub gene	Degree	Stress	Closeness centrality	Betweenness centrality
ETS1	133.0	394372.0	204.16667	72812.13461
MSH2	95.0	305904.0	176.46667	51293.2645
SORD	35.0	49354.0	136.46667	20606.21815
RORA	34.0	59786.0	137.65	19531.40852
NEDD9	33.0	48032.0	137.5	19569.67527

4.9 Drug Compound Identification:

The drug compounds are extracted from the DSigDB database using the online Enrichr program. The p-value and modified p-value were used to predict the following recommended medications. The following table lists common DEGs that are present in lung cancer, COPD, and IPF and may be used as therapeutic agents. TABLE 5 shows the most effective drug compounds based on the most widely used DEGs.

TABLE 7: THE COMMON DEGS OF IPF, COPD, AND LUNG CANCER-RELATED DRUG SUGGESTED COMPOUNDS

Name of drugs	P-value	Adjusted P-value	Genes
astemizole MCF7 UP	0.0000345	0.009510477	NEDD9; PMAIP1; CCNL1
calmidazolium MCF7 UP	0.0000401	0.009510477	NEDD9; PMAIP1; CCNL1
ivermectin MCF7 UP	0.0000926	0.009510477	NEDD9; PMAIP1
DICHLOROMETHANE CTD 00006313	0.0001631	0.009510477	NEDD9; PMAIP1
ETHYLBENZENE CTD 00000178	0.0001699	0.009510477	NEDD9; PMAIP1

5.1 Findings:

An overview of four disorders is given in this study: COPD, IPF, and LC. Protein-Protein Interaction (PPI) and GO and pathway finding are two important aspects that are carefully examined. Choosing the right datasets, variables, algorithms, and formatting techniques for data mining effectively is part of the modeling process. Comprehending the genetic connections among related illnesses is essential for formulating therapeutic recommendations. By mapping inter-genes to PPI, a cross-discussion sub pathway is created in this study, illustrating the biological relationship between the chosen disorders. Drug design for the chosen disorders is influenced by PPI networks, co-expressions, and topological characteristics. Topological features are produced to depict genes related to each of the four diseases using Cytoscape. Gene MANIA ensures a common drug design method by establishing gene co-expression and physical interaction. Understanding Protein-Drug Interaction (PDI) is aided by this work, which also advances systems biology and bioinformatics research. The goals of the research align with the selection of disorders, data collection, and identification of common genes, protein-protein interaction, topological analysis, TF-gene interaction, gene-miRNA interaction, and protein-drug interaction.

5.2 Recommendation for Future works:

Comprehending the relationship between mutated genes in many diseases is essential for creating drugs that address several ailments. Achieving treatment objectives depends on disentangling the ties between genes and related illnesses. With strong assistance from bioinformatics tools, this study methodically carries out a variety of investigations, allowing for extensive calculations in the field. When bioinformatics software is used, thorough analyses are made easier. Scholars striving for even greater progress can investigate the development of standardized medications for illnesses by utilizing microarray or microRNA datasets for comprehensive examination.

REFERENCES

[1] Amirhossein Fathinavid [a b], Zaynab Mousavian [c,] , Identifying common signatures and potential therapeutic biomarkers in COPD and lung cancer using miRNA-mRNA co-expression networks.

[2] Yang Ian A, Relan Vandana, Wright Casey M, Davidson Morgan R, Sriram Krishna B, Savarimuthu Francis Santiyagu M, Clarke Belinda E, Duhig Edwina E, Bowman Rayleen V, Fong Kwun M. Common pathogenic mechanisms and pathways in the development of COPD and lung cancer. Expert Opin Ther Targets 2011;15(4):439–56.

[3] Zhang Feng, Chen Xia, Ke Wei, Liu Daoming, Xu Xiaodong, Zhang Xing, Shi Hong. Identification of key transcription factors associated with lung squamous cell carcinoma. Med Sci Monit 2017.

[4] Gagnat Ane Aamli, Gjerdevik Miriam, Lie Stein Atle, Gulsvik Amund, Bakke Per, Nielsen Rune. Acute exacerbations of COPD and risk of lung cancer in COPD patients with and without a history of asthma. Eur Clin Respir J 2020;7(1): 1799540.

[5] Wang H, Yang L, Zou L, Huang D, Guo Y, Pan M, et al. (2012) Association between Chronic Obstructive Pulmonary Disease and Lung Cancer: A Case-Control Study in Southern Chinese and a Meta-Analysis. PLoS ONE 7(9): e46144. https://doi.org/10.1371/journal.pone.0046144

[6] Young RP, Hopkins R, Eaton TE (2007) Forced expiratory volume in one second: not just a lung function test but a marker of premature death from all causes. Eur Respir J 30: 616–622.

[7] Ghosh, A.J., Hobbs, B.D., Yun, J.H. *et al.* Lung tissue shows divergent gene expression between chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. *Respir Res* **23**, 97 (2022). https://doi.org/10.1186/s12931-022-02013-w

[8] S M Hasan Mahmud and others, Bioinformatics and system biology approach to identify the influences of SARS-CoV-2 infections to idiopathic pulmonary fibrosis and chronic obstructive pulmonary disease patients, *Briefings in Bioinformatics*, Volume 22, Issue 5, September 2021, bbab115, https://doi.org/10.1093/bib/bbab115

[9] Leng, D., Yi, J., Xiang, M. *et al.* Identification of common signatures in idiopathic pulmonary fibrosis and lung cancer using gene expression modeling. *BMC Cancer* **20**, 986 (2020). https://doi.org/10.1186/s12885-020-07494-w

[10] Szklarczyk et al., "The STRING database in 2017: Qualitycontrolled protein-protein association networks, made broadly accessible," Nucleic Acids Res, vol. 45, no. D1, pp. D362–D368, 2017, doi: 10.1093/nar/gkw937

[11] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. ProcNatl Acad Sci USA 2005;102:15545–50. doi: 10.1073/pnas. 0506580102.

[12] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," 2000. [Online]. Available: http://www.genome.ad.jp/kegg/

[13] Doms A, Schroeder M. GoPubMed: exploring PubMed with the gene ontology. Nucleic Acids Res 2005;33(suppl_2): W783–6.

[14] Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28(1):27–30.

[15] Slenter DN, Kutmon M, Hanspers K, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. Nucleic Acids Res 2018;46(D1): D661–7.

[16] Fabregat A, Jupe S, Matthews L, et al. The reactome pathway knowledgebase. Nucleic Acids Res 2018;46(D1):D649–55.

[17] Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016;44:W90–7. doi: 10.1093/nar/gkw377.

[18] M. Abul Basar, M. Faruk Hosen, M. Al Amin, N. Islam Bithi and B. Kumar Paul, "A bioinformatics and system biology technique to identify candidate biomarkers and functional pathways among stress and depression," 2022 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), Naya Raipur, India, 2022, pp. 143-146, doi: 10.1109/WIECON-ECE57977.2022.10151170.

[19] 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.

[20] 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.

[21] Khan A, Fornes O, Stigliani A, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res 2018;46:D260–6. doi: 10.1093/nar/gkx1126.

[22] Xia J, Gill EE, Hancock REW. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. Nat Protoc 2015;10:823–44. doi: 10.1038/nprot.2015.052

[23] Sethupathy P, Corda B, Hatzigeorgiou AG. TarBase: a comprehensive database of experimentally supported animal microRNA targets. RNA 2006;12:192–7. doi: 10.1261/rna.2239606.

[24] Sheng-Da H, Lin FM, Wu WY, et al. MiRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res 2011;39:D163–9. doi: 10.1093/nar/gkq1107.

[25] Liu ZP, Wu C, Miao H, et al. RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. Database 2015;2015:bav095.

[26] Yoo M, Shin J, Kim J, et al. DSigDB: drug signatures database for gene set analysis. Bioinformatics 2015;31:3069–71. doi:10.1093/bioinformatics/btv313.

[27] Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 2013;14(1):128.

[28] Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016;44:W90–7. doi: 10.1093/nar/gkw377.

[29] Carbon S, Dietze H, Lewis SE, et al. Expansion of the gene ontology knowledgebase and resources: the gene ontology consortium. Nucleic Acids Res 2017;45:D331–8. doi:10.1093/nar/gkw1108.

[30] Wittig U, De Beuckelaer A. Analysis and comparison of metabolic pathway databases. Brief Bioinform 2001;2:126–42. doi: 10.1093/bib/2.2.126.

PLAGIARISM REPORT

201-16-502

ORIGINALITY REPORT

13%	8%	11%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	academic.oup.com Internet Source	1%
2	Md Abul Basar, Md Faruk Hosen, Bikash Kumar Paul, Md Rakibul Hasan, S.M. Shamim, Touhid Bhuyian. "Identification of drug and protein-protein interaction network among stress and depression: A bioinformatics approach", Informatics in Medicine Unlocked, 2023 Publication	1%
3	Md. Abul Basar, Md. Faruk Hosen, Md. Al Amin, Nasima Islam Bithi, Bikash Kumar Paul. "A bioinformatics and system biology technique to identify candidate biomarkers and functional pathways among stress and depression", 2022 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), 2022 Publication	1%
4	www.biorxiv.org Internet Source	