Network and co-expression analysis identifies crucial gene signatures in COVID-19

Devansh Sanghvi (be19b002@smail.iitm.ac.in)
Harish Manoharan (bs19b012@smail.iitm.ac.in)
CS6024- Algorithmic Approaches in Computational Biology (Jan-May 2023)

***Abstract*—This is a 7-page project report prepared to evaluate semester-long research on the topic: Network and co-expression analysis identifies crucial gene signatures in COVID-19 for the course: CS6024. In this report, we elaborate on the theory, methods used, and the results obtained, which were already outlined during the presentation held a few weeks back.**

I. Introduction

This section outlines the motivation, the problem statement, and the objectives we started with for our research.

A. Motivation

COVID-19, caused by the SARS-CoV-2 virus, has had a significant impact on global health and economies since its emergence in late 2019. The virus was first identified in Wuhan, China in December 2019 and has since spread globally. As the pandemic continues to evolve, understanding the molecular mechanisms of the disease is critical to developing effective treatments and vaccines. This project aims to use **bioinformatics techniques** to identify the key genes, regulatory transcription factors, and pathways involved in COVID-19. Traditional methods of identifying hub genes through differential gene expression analysis have limitations, including a lack of consideration for regulation by transcription factors and the complexity of gene expression regulation. Therefore, this project seeks to overcome these limitations by incorporating these factors into gene co-expression network analysis and accounting for multiple interactions involved in gene expression regulation. By doing so, we hope to gain a better understanding of the molecular mechanisms behind COVID-19 and develop more effective treatments and vaccines.

B. Problem Statement

Given the gene expression raw counts dataset for COVID-19 and normal samples, analyze and identify crucial hub genes and regulatory transcription factors. Also, understand how contracting COVID affects gene regulatory networks and the most disturbed pathways.

C. Objectives

The lack of comprehensive understanding of the molecular mechanisms underlying COVID-19 has been a major challenge in developing effective treatments and vaccines. Our research project aims to address this gap.
Our main objectives are:

- Identify key gene modules and hub gene signatures using various packages.
- To identify perturbed pathways associated with COVID-19 using pathway enrichment analysis.
- To perform network analysis using Cytoscape to visualize the interactions between these genes and to identify potential biomarkers.
- Identify the regulatory transcription factors responsible for the behavior of the hub genes
- Identify the most rewired genes according to the DK coefficient value using DYNet in Cytoscape.

II. Existing Paper Critique

This proposal was directly inspired by another research: **Bioinformatics and machine learning approach identifies potential drug targets and pathways in COVID-19.** This the section addresses the theory behind some of the tools used, the summary of the research paper, the results obtained and discusses the strengths and possible weaknesses of the conducted research.

A. Theory for the research paper

The current research paper uses a mix of biological, statistical, and computational terms that could be unfamiliar to the reader. Hence, this section will cover the basics of these concepts in detail. The following concepts are important to understand:

1) **Raw counts dataset (Gene expression levels)**: RNA sequencing (RNA-seq) is a powerful technique that has revolutionized the way we measure gene expression levels. RNA-seq involves the isolation of RNA from a biological sample, followed by the conversion of RNA into cDNA, which can then be sequenced using high-throughput sequencing technology. The output of RNA-seq is a count dataset that shows the number of reads that have been mapped to each gene in the genome. The counts' dataset is generated by aligning the cDNA sequences to a reference genome or transcriptome using alignment algorithms. The number of reads mapped to each gene corresponds to its expression level, which can be used to identify differentially expressed genes between different biological conditions. RNA-seq technology has several advantages over other methods for measuring gene expression, including its high sensitivity, ability to quantify low abundance transcripts, and ability to detect

novel transcripts and isoforms. RNA-seq is a powerful tool for investigating gene expression regulation and can be used to gain insights into the molecular mechanisms of diseases like COVID-19.

2) **Differentially expressed Genes (DE Genes)**: Differential gene expression analysis is a widely used method to identify genes that are significantly altered between different biological conditions. Differentially expressed genes play a crucial role in understanding the underlying molecular mechanisms of various biological processes, including diseases like COVID-19. In recent years, RNA sequencing (RNA-seq) technology has become a popular tool for measuring gene expression levels, which generates a counts dataset of the number of reads mapped to each gene in the genome. One of the widely used R packages for analyzing RNA-seq data is DESeq2, which uses a negative binomial distribution-based model to estimate variance and differential expression. DESeq2 allows for normalization and adjustment for technical variability within the dataset, which is important for accurately identifying differentially expressed genes. By identifying differentially expressed genes with DESeq2, we can gain insights into the underlying molecular mechanisms of biological processes and diseases, such as COVID-19. When identifying the DE genes from the raw counts' dataset using the DESeq2 Package from R, these terms came across us:
   2.1. **P adjusted values**: This statistical measure is used to control for multiple comparisons in differential gene expression analysis. It represents the probability of observing a difference in expression between groups by chance after considering the number of tested genes.
   2.2. **Log Fold Change**: This is a measure of the difference in expression between two groups of samples, expressed as the log2 ratio of the mean expression values. A positive log fold change indicates upregulation, while a negative log fold change indicates downregulation.
   2.3. **Wald Test**: The Wald test is a statistical test used in differential gene expression analysis to determine if the log fold change between two conditions significantly differs from zero. It is based on the normal distribution and calculates p-values and confidence intervals.
   2.4. **Upregulated and downregulated genes**: Upregulated genes are those that show higher expression levels in one condition compared to another, while downregulated genes are those that show lower expression levels. These genes can be identified by calculating the log fold change and significance values using statistical tests such as the Wald test and adjusting the p-values for multiple hypothesis testing.
3) **Weighted Gene Coexpression Network Analysis**: Weighted Gene Co-expression Network Analysis (WGCNA) is a widely used bioinformatics tool for identifying co-expression modules and hub genes from transcriptomic data. The main purpose of WGCNA is to identify groups of co-expressed genes across multiple samples, which can provide insights into the underlying biological processes and molecular mechanisms of complex diseases like COVID-19. The inputs to WGCNA are typically gene expression data, usually in the form of a counts dataset, and sample metadata that describes the biological conditions being compared. The outputs of WGCNA include a hierarchical clustering tree that shows the relationship between different co-expression modules and hub genes, which are highly connected genes that play a central role in the co-expression network. By analyzing the results of WGCNA, we can infer the biological functions of different co-expression modules, identify key genes involved in disease progression, and develop targeted therapies to treat complex diseases like COVID-19.
4) **Protein protein interaction (PPI) Network**: The physical connections between two or more proteins that take place in a living cell are known as protein-protein interactions (PPIs). Nearly all biological functions, such as signal transduction, gene control, and metabolic pathways, depend on these interactions. Experimental methods like yeast two-hybrid experiments, co-immunoprecipitation, and affinity purification followed by mass spectrometry can be used to identify PPIs. In addition to other computational techniques, bioinformatics tools can be used to predict PPIs based on sequence and structural similarities. PPIs can be used to build protein-protein interaction networks after they have been found, which can reveal information about the interactions between various proteins and biological processes. These networks can be used to forecast the consequences of genetic changes on protein function and discover prospective therapeutic targets. Understanding PPIs is critical for advancing our understanding of cellular processes and developing targeted therapies for diseases like cancer and COVID-19.

B. Bioinformatics Pipeline used in the paper

The above research paper aims to identify hub gene signatures, key modules, and perturbed pathways due to COVID-19. It uses the traditional Bioinformatics pipeline shown in the diagram below[1] (Fig 1), which is taken directly from the paper. The following steps (in brief) were taken during the research:

1) The research involved utilizing two RNA-sequencing PBMC datasets of SARS-CoV-2, with one dataset used for creating the weighted gene coexpression network and the other for validating the hub genes identified in the first dataset.
2) After analyzing the first dataset, approximately 1520 differentially expressed genes were selected for further network analysis.
3) The 1520 genes were used to create and analyze the weighted gene coexpression network, and key modules

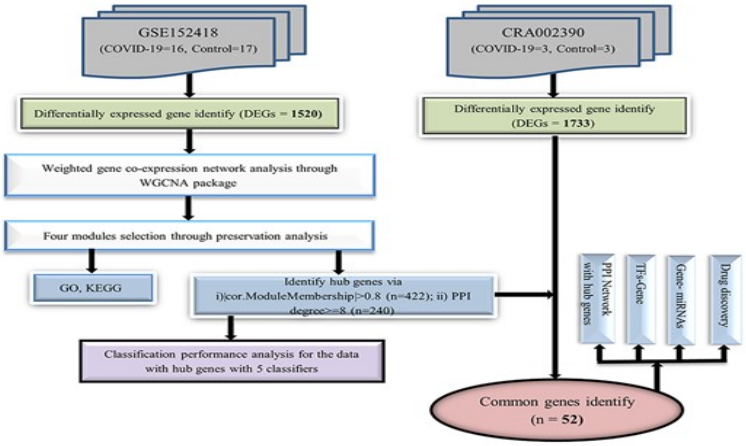

Fig. 1. Bioinformatics pipeline used in the reference research paper [1]

were identified. DAVID tools were then used to determine biological insights.

4) The PPIhub and MMhub were utilized to identify hub gene signatures, which were then shortlisted by selecting only the genes common in the PPI network and module membership network.
5) The hub gene signatures that were differentially expressed were validated using machine learning models.
6) The analysis aimed to identify transcriptional regulators of hub gene signatures and candidate drugs using the inverse of hub gene signature.

C. Results from the current research paper

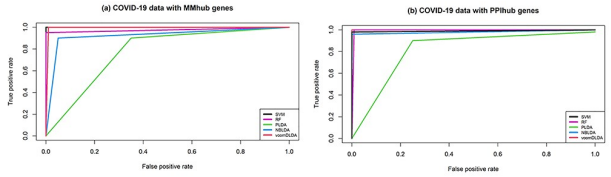

Fig. 2. Receiver operating curve (ROC) plot of the five classifier performance based on (A) accuracies, (B) AUC.

1) We identified 52 common genes using module membership statistics and PPI networks from the 1733 DEGs detected in the RNA-Seq dataset CRA002390.
2) From these 52 genes, we identified 10 hub genes through PPI network analysis, including PLK1, AURKB, AURKA, CDK1, CDC20, KIF11, CCNB1, KIF2C, DTL, and CDC6.
3) Additionally, we found TFs, including FOXC1, GATA2, YY1, E2F1, NFIC, FOXL1, and SRF, that potentially regulate these hub genes.
4) We also identified several repurposed drugs, including naproxol, teniposide, amsacrine, BRD-K68548958, and palbociclib, as top candidates for COVID-19 treatment.
5) Further wet-lab experiments are necessary to confirm these findings and their potential for clinical application.

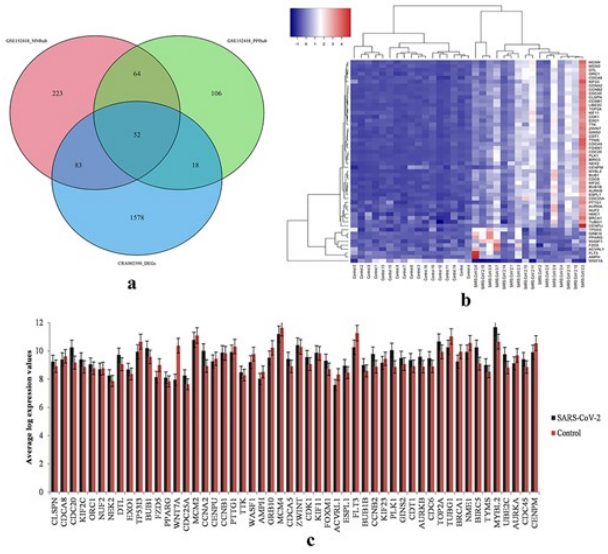

Fig. 3. Hub gene expression profiles. (A) Venn diagram of common hub genes identified among the hub genes of GSE152418 identified via MM scores and PPI and the DEGs of CRA002390 data. (B) Heatmap of hub genes of GSE152418 dataset. (C) Bar chart of the log expression values of 52 common hub genes in the GSE152418 dataset.

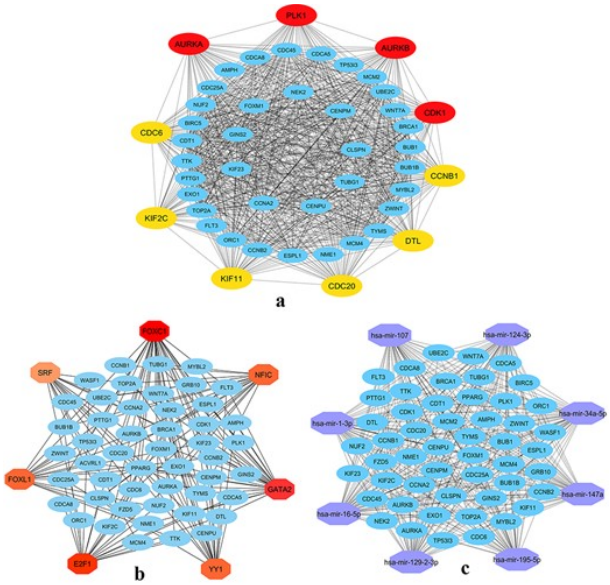

Fig. 4. Network construction. (A) PPI network of the 52 common hub genes of COVID-19 data, (B) TFs-Gene interaction network of the 52 common hub genes, (C) gene-miRNAs interaction for the common hub genes of COVID-19.

6) These results provide new insights into COVID-19 pathogenesis and molecular targets for novel interventions.
7) Figure 2 shows the performances of the classifiers based on two measurements.
8) Figure 3 shows the Hub gene expression profiles
9) Figure 4 shows the networks constructed

D. Strengths of the existing paper

- Detects co-expression patterns: WGCNA uses a weighted correlation network approach to identify highly co-expressed genes and group them into modules
- Resistant to noise: WGCNA is robust to noise and can handle any missing data, making it suitable for large-scale datasets
- Modular structure: The modules identified by WGCNA have a modular structure, with genes within a module showing high co-expression and functional coherence.
- The research implemented gene coexpression module analysis that provides key gene modules rather than finding DEGs. This is a reliable approach to finding the hub genes.
- The research used two independent analyses to identify hub genes, specifically module membership significance analysis and PPI analysis to further increase the reliability of the research.

E. Weaknesses of the existing paper

- This study relied on bioinformatics analysis without functional studies in wet lab.
- Loss of information: The weighted gene coexpression network was made only on differentially expressed genes, non-differentially expressed genes that may be important for network connectivity or regulation can be overlooked, leading to a loss of information.
- By using a subset of genes, the level of noise in the network could be increased, which could impact the reliability of hub gene identification.
- Module merging: WGCNA can merge similar modules into a single module, which can reduce the granularity of the analysis and potentially miss important gene co-expression relationships.
- Lack of directionality: WGCNA does not infer directed edges, so it cannot distinguish between causal and non-causal relationships among genes. This can make it difficult to infer regulatory relationships within the network. WGCNA can only consider positive correlations.
- Even though the differentially expressed genes may be involved with a pathway, it remains a challenge to assess how the activation/ inhibition of a pathway worsens the disease.
- Furthermore, the differential gene expression approach ignores the effect of multiple interactions involved with gene expression regulation, a key factor to phenotype determination.
- Transcription factors (TFs) are central in expressing smooth muscle contractile proteins and determining phenotypes. As mentioned above, it is difficult to gain insight into gene regulation by TFs from differential expression analysis alone.

III. Research Conducted

With this goal in mind, we explored other tools and methods that could be used to achieve the objectives we had set out. Many of the tools, as mentioned before had to be read and are not common to every reader. Hence, we will first review the theory for the new methods used. We will then have a look at the pipeline we used for our research and compare the results with those of the current research. In the end, we'll compare the two bioinformatics pipelines and conclude.

A. Theory for conducted research

The concepts and tools that need to be known before reading further are:

1) **Regulatory Impact Factors (RIF)**:
   1.1. **Transcription Factors**: Transcription factors (TFs) play a crucial role in gene expression regulation, acting as molecular switches that bind to specific DNA sequences and activate or repress gene transcription. Identifying key TFs involved in a biological process can provide valuable insights into the regulatory networks that control gene expression and can help unravel the underlying molecular mechanisms. TFs can be identified through various methods, including gene expression analysis, chromatin immunoprecipitation followed by sequencing (ChIP-seq), and motif enrichment analysis. By identifying the key TFs involved in a specific biological process, researchers can better understand the regulatory networks that control gene expression and develop targeted disease interventions.
   1.2. **RIF Score 1**: The RIF score one measures the change in downstream gene expression resulting from the perturbation of a transcription factor (TF) while accounting for changes in the expression of the TF itself. A high RIF score 1 indicates that the TF has a strong regulatory impact on downstream targets.
   1.3. **RIF Score 2**: The RIF score 2, on the other hand, measures the change in TF expression resulting from the perturbation of a downstream gene while accounting for changes in the expression of that downstream gene. A high RIF score 2 indicates that the downstream gene is a strong regulator of the TF.
   1.4. **Differential Wiring (DW)**: Differential wiring, which is calculated by subtracting the absolute values of RIF score 1 and RIF score 2, indicates the directionality of the regulatory relationship between a TF and its downstream targets. A positive differential wiring value indicates that the TF has a stronger effect on its downstream targets than the downstream targets have on the TF. Conversely, a negative differential wiring value indicates that the downstream targets have a stronger effect on the TF than the TF has on its downstream targets.
2) **Partial correlation with Information Theory (PCIT)**: PCIT stands for Partial Correlation and Information Theory. It is a statistical method to identify direct and indirect associations between variables, such as genes, in large datasets.

2.1. **The steps in PCIT include:**

i) Calculating the correlation matrix of the input data
ii) Estimating partial correlation coefficients by adjusting for all other variables in the dataset
iii) Thresholding the partial correlation matrix to create an adjacency matrix
iv) Calculating the mutual information matrix from the adjacency matrix
v) applying a data processing inequality algorithm to obtain a final network.

2.2. **Input for PCIT**: The input for PCIT is typically a gene expression dataset in the form of a matrix with samples as columns and genes as rows.

2.3. **Output from PCIT**: The outputs from PCIT include a network or graph, where nodes represent genes and edges represent the strength and direction of their association. Additional outputs may include statistics on node centrality and module identification.

2.4. **What can we infer?**: From PCIT results, we can infer the direct and indirect relationships between genes in the input dataset. This can be useful for identifying key genes and pathways involved in various biological processes and diseases. Additionally, the network can be used for further analysis and modeling, such as identifying hub genes and predicting gene functions.

2.5. **Formulae used in the two steps:**

i) **Step 1- Partial Correlations:** For every trio of genes in x, y, and z, the three first-order partial correlation coefficients are computed by:

$$r_{xy.z} = \frac{r_{xy} - (r_{xz} \cdot r_{yz})}{\sqrt{(1 - r_{xz}^2) \cdot (1 - r_{yz}^2)}} \quad (1)$$

ii) **Step 2- Information theory:** The DPI theorem of Information Theory obtains the tolerance level (ϵ) for capturing significant associations among every trio of genes. This theorem states that no data manipulation can improve the inference that can be made from it. In order to calculate the tolerance level, **the average ratio of partial** to direct correlation is determined as follows:

$$\epsilon = (\frac{r_{xy.z}}{r_{xy}} + \frac{r_{xz.y}}{r_{xz}} + \frac{r_{yz.x}}{r_{yz}}) \quad (2)$$

In the context of our network reconstruction, **a connection between genes x and y is discarded if:**

$$|r_{xy}| < |\epsilon \cdot r_{xz}| and |r_{xy}| < |\epsilon \cdot r_{yz}| \quad (3)$$

In order to establish a connection between a pair of genes x and y in the gene co-expression network, we need to confirm the significance of their association. This is done by repeating the process for each of the remaining genes (denoted by z) in the network. First, we calculate the partial correlation between genes x and y while controlling for the effect of gene z. Next, we compute the direct correlation between genes x and y. If the partial correlation is significantly different from the direct correlation, we consider the association between genes x and y significant, and a connection is established between them in the gene co-expression network. This process is repeated for all the remaining genes in the network to ensure the validity of the connections.

3) **Differential Gene Correlation Analysis (DGCA)**: DGCA stands for Differential Gene Co-expression Analysis, which is a method used to identify differentially co-expressed gene pairs between two groups of samples.

3.1. **Steps in DGCA:**

i) Preprocessing of the gene expression data and normalization.
ii) Identification of differentially expressed genes (DEGs) between the two groups of samples using methods such as t-test, DESeq2, edgeR, etc.
iii) Construction of co-expression networks for each group of samples using correlation or partial correlation methods.
iv) Calculation of differential co-expression values for each gene pair between the two groups of samples.
v) Identification of differentially co-expressed gene pairs based on statistical significance testing.

3.2. **Input in DGCA:** The input required for DGCA is gene expression data from two groups of samples.

3.3. **Output from DGCA:** The outputs obtained from DGCA are the differentially co-expressed gene pairs with their corresponding statistical measures such as p-values, false discovery rate (FDR), fold change, etc.

3.4. **Inferences from DGCA:** From the DGCA results, we can infer the gene pairs that show differential co-expression patterns between the two groups of samples. These gene pairs can provide insights into the biological processes that are differentially regulated between the two groups and can serve as potential biomarkers or therapeutic targets.

4) **DYNet**: DYNet is a Cytoscape app providing dynamic network visualization for time-course data. It lets the user visualize how the network evolves over time by showing edge and node attributes at different time points. It can be used for various applications, such as gene expression, protein-protein interaction, and signaling pathway analysis. DYNet also offers several layout algorithms to help intuitively visualize complex network structures. The app is easy to use and highly customizable, with features such as node clustering and coloring based on attributes. Overall, DYNet is a valuable tool for exploring and interpreting dynamic network data in a user-friendly and interactive way.

5) **Gene and pathway enrichment analysis**: Gene and pathway enrichment analysis is a widely used bioinformatics tool that identifies the significant biological pathways and gene sets involved in a particular phenotype

or disease. It works by taking a list of differentially expressed genes and identifying the overrepresented pathways or gene sets in this list. The input data can be obtained from various high-throughput technologies, such as RNA-Seq or microarray, and is usually analyzed using tools such as DAVID or Enrichr. The analysis output includes the significantly enriched gene sets or pathways along with the associated statistical measures, such as p-values or false discovery rates. This analysis provides insights into the underlying biological mechanisms involved in a phenotype or disease, thereby aiding in identifying potential therapeutic targets.

B. Bioinformatics pipeline used for the conducted research

The approach is depicted by **Fig 5**. The steps are also mentioned in detail below:

1) To achieve our objectives, we first collected gene expression data from COVID-19 patients and healthy individuals. The transcriptomic dataset we used was **GSE152418** with about 60,683 genes. It has 17 normal and 16 COVID-19-affected samples.
2) Genes with low expression (gene counts for all samples less than 100) were excluded, and the data was normalized. After this step, we did three independent analyses:
   - Through differential gene expression analysis using the DESeq2 package in R, we identified DEGs.
   - Transcription factors play a vital regulatory role in controlling gene expression. We used regulatory impact factor (RIF1 and RIF2) analysis to assign a score and identify the transcription factors regulating the expression of DEGs.
   - We created two networks based on the partial correlation and information theory (PCIT) algorithm to identify differences in gene co-expression profiles between the conditions. After this, we subsetted the transcription factors, the DEGs from the two networks created through PCIT.
3) Using the two networks, we found the hub genes for both the case: control and the COVID-19 affected.
4) We also made a central reference network from the two samples. Based on differential connectivity (DK), we identified the most rewired genes between cases and controls.
5) Independently, we planned to perform Differential co-expression network analysis using the DGCA R. package. However, due to computational restrictions, we could not run the code.
6) Further, through enrichment of the particular network and functional enrichment analysis using ClueGO, we can find the pathways that are most affected due to COVID-19

C. Results

1) **Hub Genes** - Genes with high degree and betweenness centrality were marked as hub genes. The marked hub genes are given in Table I

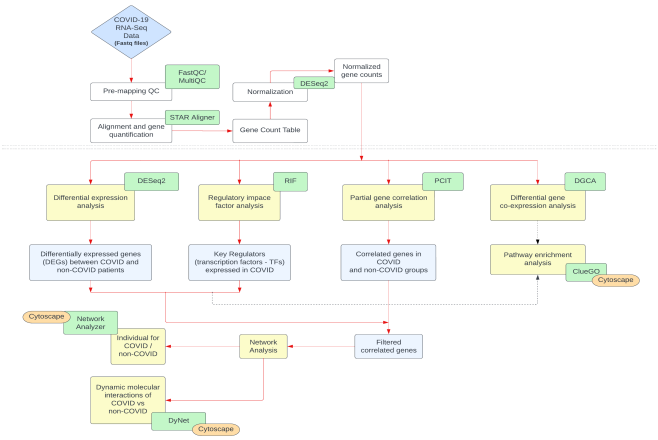

Fig. 5. Proposed bioinformatics pipeline

TABLE I
HUB GENES ALONG WITH THEIR DEGREE AND BETWEENNESS CENTRALITY

Gene	Degree	Betweenness Centrality
ENSG00000100162	31	0.22
ENSG00000101162	47	0.13
ENSG00000113140	47	0.08
ENSG00000077152	29	0.06

2) **Regulatory Transcription Factors** - RIF analysis was used to find the most regulatory transcription factors (Table II).
3) **Differential Connectivity (Dynet)** - Differential Connectivity values were calculated for each gene and top genes were considered (Table III). **Fig 6**. shows differential connected nodes and edges (Green nodes and edges - healthy, Red nodes and edges - COVID-19). Differential Connectivity values were calculated for each gene
4) **Final gene list** - The Regulatory Transcription Factors and the top genes based on Dk values were combined to form the final list of 19 genes
5) **Validation** - To validate our results, a machine learning classifier was built to predict COVID or healthy only based on the gene expression values of the chosen genes. The model was trained on the same dataset the analysis was done on, but tested on the validation dataset - CRA002390. The accuracies of different classifiers are shown in Table IV

TABLE II
REGULTORY TRANSCRIPTION FACTORS ALONG WITH THEIR RIF SCORES

Gene	RIF1	RIF2
ENSG00000179528	2.04	0.17
ENSG00000153234	-1.22	-2.10
ENSG00000114315	-2.8	-1.7
ENSG00000029153	-0.4	-2.12
ENSG00000197279	-1.5	-2.10
ENSG00000095794	-2.32	-1.94
ENSG00000167074	2.01	0.44

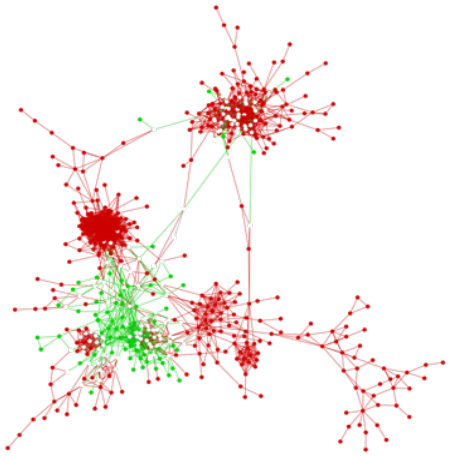

Fig. 6. Combined Network. Red - COVID, Green - healthy

TABLE III
GENES ALONG WITH THEIR DIFFERENTIAL CONNECTIVITY (DK) VALUES

Gene	Dk value
ENSG00000153234	-0.83
ENSG00000095794	-0.75
ENSG00000064601	-.71
ENSG00000119508	-0.68
ENSG00000112149	-0.66
ENSG00000197279	-0.66
ENSG00000130844	-0.65
ENSG00000118503	-0.62
ENSG00000124466	-0.58
ENSG00000177535	-0.57
ENSG00000073111	0.71
ENSG00000105011	0.78
ENSG00000121152	0.89
ENSG00000137807	0.93

IV. CONCLUSION

A. Comparison of results

While the related paper was able to get 99% accuracy with 52 genes, we were able to get 95% with just 19 genes. The accuracy of the model training on the top 19 genes from the related paper gives only about 85% accuracy on the validation data. We were able to attain decent accuracy with a much smaller set of genes.

B. Predicted Reason for Differences/Similarities

Different sets of analyses were done and merged to form the final list of genes. We felt that the hub genes alone (as done in the related paper) were not an accurate representation of the genes which affected it the most. It completely misses out on the genes that are highly differentially expressed. We also believed that the regulatory transcription factors to play an important role in the biological context.

TABLE IV
ACCURACY ON VALIDATION DATA

Model	Accuracy on Validation data
SVM	0.93
Logistic Regression	0.86
Random Forest	0.96

C. Future Work

While we had planned to implement DGCA (Differential co-expression network analysis) - it was found to be computationally infeasible on limited laptop memory. Doing this analysis could have reinforced the final list of genes we obtained and possibly add other genes to the list increasing the validation accuracy.

Gene and Pathway Enrichment Analysis - Analysis can be done using ClueGo App on Cytoscape to understand the biology and the pathways that these genes affect.

REFERENCES

[1] "Bioinformatics and machine learning approach identifies potential drug targets and pathways in COVID-19.", academic.oup.com, 2021 [Online]. Available: **Link**

[2] A Novel Method to Identify the Differences Between Two Single Cell Groups at Single Gene, Gene Pair, and Gene Module Levels, NCBI Website, 2021 [Online]. Available: **Link**.

[3] ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks, NCBI Website, 2009 [Online]. Available: **Link**

[4] Integrative transcriptomics analysis of lung epithelial cells and identification of repurposable drug candidates for COVID-19, www.sciencedirect.com, 2020 [Online]. Available: **Link**

[5] Gene Expression in Asthmatic Airway Smooth Muscle, 2007 [Online]. Available: **Link**

[6] "DyNet", https://apps.cytoscape.org/apps/dynet, 2016 [Online]. Available: **Link**

[7] "Analyzing Regulatory Impact Factors and Partial Correlation and Information Theory", http://www.bioconductor.org/packages/release/bioc/vignettes/CeTF/inst/doc/CeTF.html, 2023 [Online]. Available: **Link**