



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Determining crucial genes associated with COVID-19 based on COPD Findings^{*,**}

Pooja Sharma^{a,*}, Anuj K. Pandey^b, Dhruva K. Bhattacharyya^c

^a Department of IT, BIT Sindri, Dhanbad, Jharkhand, 828123, India

^b Department of EE, BIT Sindri, Dhanbad, Jharkhand, 828123, India

^c Department of CSE, Tezpur University, Assam, 784028, India

ARTICLE INFO

Keywords:
COVID-19
SARS
MERS
GeneCard
Pathway

ABSTRACT

Genes act in groups known as gene modules, which accomplish different cellular functions in the body. The modular nature of gene networks was used in this study to detect functionally enriched modules in samples obtained from COPD patients. We analyzed modules extracted from COPD samples and identified crucial genes associated with the disease COVID-19. We also extracted modules from a COVID-19 dataset and analyzed a suspected set of genes that may be associated with this deadly disease. We used information available for two other viruses that cause SARS and MERS because their physiology is similar to that of the COVID-19 virus. We report several crucial genes associated with COVID-19: *RPA2*, *POLD4*, *MAPK8*, *IRF7*, *JUN*, *NFKB1*, *NFKBIA*, *CD40LG*, *FASLG*, *ICAM1*, *LIFR*, *STAT2* and *CCR1*. Most of these genes are related to the immune system and respiratory organs, which emphasizes the fact that COPD weakens this system and makes patients more susceptible to developing severe COVID-19.

1. Introduction

This century has seen outbreaks of different pandemics, particularly those caused by viruses. In 2002, Severe Acute Respiratory Syndrome (SARS) became the first such pandemic. It was caused by a virus with a corona-type outer covering. Hence, such viruses became known as coronaviruses. A special class of coronavirus called SARS-CoV causes SARS. This virus originated in bats, from which it moved to civets and then finally to humans. This disease infected approximately 8,098 people in 29 different countries, and 774 deaths were reported.¹ The disease's infection chain was ruptured by cutting down human-to-human interaction and following strict quarantine measures. This pandemic slowed down in July 2003, and since then, no new cases have appeared.

A similar kind of disease was caused by a coronavirus in April 2012. It was diagnosed from a lung sample of an adult patient in Saudi Arabia. This disease also affected the respiratory system and came to be known as Middle East Respiratory Syndrome (MERS), and the coronavirus that caused it was named MERS-CoV. Unlike SARS, MERS is still circulating. Recently, a new type of disease was detected from clinical samples of

patients in Wuhan, China. These patients showed pneumonia-like symptoms and discomfort in the respiratory organs. The disease came to be known as COVID-19, and the coronavirus that causes it was named SARS-CoV2. Within a month of first being reported, the outbreak of this disease affected more than 100,000 people in nearly 100 countries [13].

SARS-CoV2 has much higher transmission capacity than the other two CoVs, and COVID-19 is affecting the human population at a very alarming rate. The clinical manifestation of COVID-19 is predominantly respiratory symptoms [21]. Most people who are infected with this disease show mild to moderate respiratory dysfunction and recover from the disease within a month. However, the disease proves fatal for people with chronic medical conditions, such as cardiovascular disease, cancer, diabetes, or respiratory disorders.²

COVID-19 also causes severe complications in older age groups and those with Chronic Obstructive Pulmonary Disease (COPD). COPD is a common lung disease that is manifested in two forms. The first is emphysema, which involves the destruction of air sacs in the lungs, thereby disturbing the air flow within the body. The second is chronic bronchitis, which causes inflammation and gradual narrowing of the bronchial tubes, leading to breathing problems. When people with

* Corresponding author.

E-mail address: poojasharma0303@gmail.com (P. Sharma).

¹ <https://www.medicalnewstoday.com/articles/comparing-covid-19-with-previous-pandemics#Severe-acute-respiratory-syndrome>, accessed on 2nd May 2020.

² https://www.who.int/health-topics/coronavirus#tab=tab_1, accessed on 3rd May 2020.

COPD contract COVID-19, they start experiencing discomfort due to breathing problems. Several other complications arise in such cases, which gradually slow down the recovery process. Existing lung damage in severe cases of COPD makes it difficult for patients to fight any kind of foreign invasions by SARS-CoV2. A close analysis of people with COVID-19 suggests that people with COPD have 63% higher risk of being severely affected by the disease as compared to the general population, for which the chance is 33.4%.³ The Chinese Center for Disease Control and Prevention documented the case fatality rate (CFR) of 2.3% among the normal patients infected with COVID-19. The Chinese Center for Disease Control and Prevention documented a case fatality rate (CFR) of 2.3% among normal patients infected with COVID-19. However, the reported CFR of COVID-19 patients with COPD and was 6.3%.⁴

In this work, we analyzed the association of COPD with COVID-19 disease. The analysis was based on the module formation property, by which genes with similar characteristics tend to group together. We also used information available for SARS and MERS viruses because evidence [7] suggests a similarity between these three viruses. We analyzed the association of certain crucial genes identified in our work and investigate their relationships with the deadly COVID-19 disease. We also analyzed the modules obtained from the gene expression values of COVID-19 and discovered a few unknown genes that might be essential in studying COVID-19.

2. Materials and methods

2.1. Dataset used

For our experiment, we used the GSE57148 dataset entitled, *Characterizing gene expression in lung tissue of COPD subjects using RNA-Seq*.⁵ We used the gene expression profile of the 98 COPD subjects obtained from the FPKM normalized form of this dataset. We also analyzed the effectiveness of gene module findings in a GSE157103 dataset for SARS-CoV2 entitled, *Large-scale Multi-omic Analysis of COVID-19 Severity*.⁶ The data set contains 126 samples, of which 100 samples were from those of COVID-19 patients, and 26 samples were from non-COVID patients.

2.2. Method

The gene expression values of 98 COPD samples and 100 COVID-19 samples were used to construct a gene-gene network. We use the WGCNA package [8] in the R platform to build this gene co-expression network based on the correlation between genes. The package provides the ability to choose either a hard or soft threshold using the parameters of the functions *pickSoftThreshold* or *pickHardThreshold* to construct an adjacency matrix representation of the gene network. It has been observed that the algorithm works best with the default parameters using a hard threshold according to the literature [8]. Therefore, we used this approach to construct our required gene networks.

Once the gene coexpression network was constructed, we extracted modules of high cohesiveness or closeness from it. WGCNA uses an unsupervised approach to determine the gene modules. We used a hierarchical clustering approach to extract the modules. The branches from dendrograms of the clustering indicates the modules. Depending on the height at which the tree is cut, we obtain different sets of modules. However, we report modules extracted using the default

³ <https://www.sciencedaily.com/releases/2020/05/200511142146.htm>, accessed on 3rd May 2020.

⁴ <https://www.medicalnewstoday.com/articles/covid-19-and-copd>, accessed on 3rd May 2020.

⁵ <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE57148>, accessed on 4th May 2020.

⁶ <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE157103>, accessed on 24th Sept 2020.

parameters, which were further analyzed for their biological significance.

All extracted modules are not equally significant in terms of biological aspects. The obtained modules are prioritized based on the p-value. We used the Funcassociate tool [2] to calculate the p-value of the gene modules. The tool uses a Monte Carlo simulation method to calculate a gene module's significance. The next step is to find the association of genes in each module with that of causal genes associated with COVID-19. Researchers suggest that COVID-19 has emerged from a virus that resembles the SARS and MERS viruses. Phylogenetic studies of the COVID-19 virus suggest that SARS-CoV2 shares at least 80% similarity with the SARS virus and 50% similarity with the MERS virus [7]. Therefore, we used the causal genes associated with these two diseases reported in GeneCard [14]. GeneCard is a repository that includes information such as causal genes, related diseases, pathways, etc.

Using the list of causal genes, we determined the disease-associated gene modules. Such modules have at least one known causal gene. The members of the disease-associated modules were further analyzed to see whether there is any connection between these benign members and causal genes. In this work, we focused on only the associated pathways, but this approach could be further extended to other phylogenetic features of the disease. A diagrammatic representation of the workflow is shown in Figs. 1 and 2.

3. Experimental results

For the whole experiment, we used the R Package and MATLAB on an HP Z800 workstation with 12 GB of RAM running Windows 64 bit. We performed the gene network construction and module identification on 98 samples from COPD patients obtained from GSE57148. A total of 41 modules were extracted using the WGCNA package with default parameters. We also performed our experiment on the GSE157103 dataset, which contained the expression values of 100 COVID-19 patients. A total of 21 modules were obtained in this case using the WGCNA tool. The modules obtained for both datasets were then tested to determine their biological significance using the *p* – values.

3.1. *p* – value analysis Of gene modules

The biological significance of any module is decided by its functional enrichment. The functional enrichment is given by the *p* – value, which is obtained using the Funcassociate tool. We report the p-value of the top five modules in Tables 1 and 2 for GSE57148 and GSE157103, respectively. Once we had the set of top five modules, we started analyzing the disease-associated modules.

3.2. Disease associated module analysis

A module with at least one disease gene is said to be associated with the disease. Using a list of 51 genes (a combination of SARS and MERS genes), we identified 17 disease-associated modules in GSE 57148 and 9 disease-associated modules in GSE157103. Tables 3 and 4 report the set of disease genes associated with the top five modules for GSE57148 and

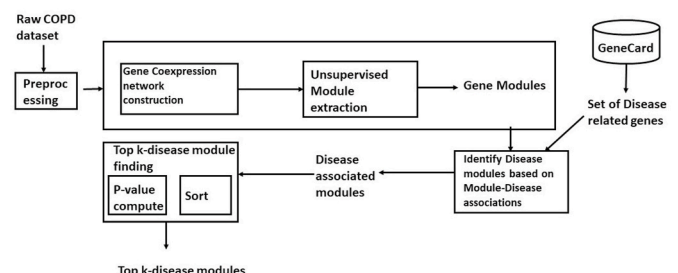


Fig. 1. Workflow for finding disease associated modules.

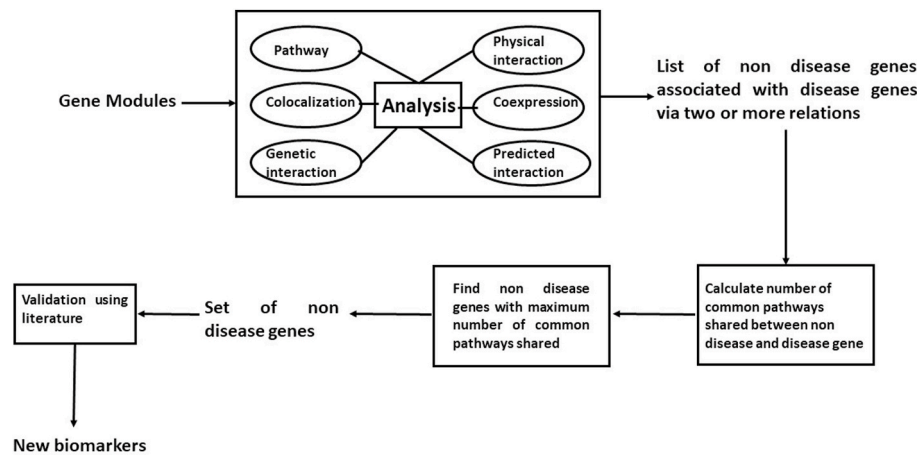


Fig. 2. Workflow for finding crucial genes associated with COVID-19.

Table 1
p – value of top five modules for GSE57148 dataset.

| Module no. | <i>p</i> – value |
|------------|------------------|
| 11 | 1.580e-14 |
| 1 | 2.290e-10 |
| 5 | 5.984e-9 |
| 8 | 2.843e-7 |
| 25 | 3.164e-6 |

GSE157103, respectively. However, we have constrained our analysis to only the top five modules in GSE57148 and the top three modules in GSE157103, which show higher biological significance in terms of the *p* – values.

We chose only three modules from GSE157103 because the last two modules showed much lower *p*-values than the modules found in GSE57148. Modules with more disease genes represent a stronger association with the disease. Therefore, we can say that module 5 is closely associated with the disease as it has 6 disease genes, followed by module 8 and so on in the case of the GSE57148 dataset.

3.3. Association of non disease genes with the disease genes

Several non-disease genes in the modules are correlated with the disease genes. This was established by the fact that they are grouped into the same module according to similarity criteria. We used the GeneMania tool [11] to find out the association of such benign genes with the disease genes. GeneMania uses co-expression data, physical interaction data, genetic interaction data, shared protein domains, co-localization data, and data from pathway and predicted information to determine the association between genes during its network formation.

Information from protein interaction databases such as BioGRID and PathwayCommons were used to determine whether two or more gene products are linked in a network. This is shown by the physical interaction edges in the network. Two or more genes may be shown to be

Table 2
p – value of top five modules for GSE157103 dataset.

| Module no. | <i>p</i> – value |
|------------|------------------|
| 2 | 4.466e-17 |
| 10 | 2.355e-12 |
| 6 | 9.047e-11 |
| 7 | 2.649e-6 |
| 4 | 6.173e-4 |

Table 3
Disease genes found in top five modules for GSE57148 dataset.

| Module no. | No. of disease genes | Disease genes |
|------------|----------------------|---------------------------------------|
| 1 | 4 | IRF3, CTRL, IFNAR1, POLD1 |
| 5 | 6 | CXCL10, MX1, DDX58, CXCL9, CCR5, BAK1 |
| 8 | 4 | CCL2, IL6, MCL1, SOCS3 |
| 11 | 1 | TMPRSS11D |
| 25 | 3 | CCL5, IFNG, STAT1 |

linked in this tool if they share some expression similarity across different conditions in studies associated with gene expression, such as the Gene Expression Omnibus (GEO), which is given by the co-expression edge in the network. Another edge shown as a predicted edge in the network considers information from gene orthologs to decide upon the interaction probability between two or more genes.

GeneMania also explores the reaction possibility of two or more genes in the same pathway by using the information available in databases such as Reactome and PathwayCommons. This is shown by the pathway edges in the network. Furthermore, GeneMania uses data from BioGRID to determine the functional association between two or more genes and is represented by the genetic interaction edges in the network. Another edge parameter is co-localization, which suggests the presence of two or more genes in the same cellular location, thus highlighting their contribution in the network [11]. Table 5 shows these associations for module 1 in the case of the GSE57148 dataset. The association information of the other four modules in GSE57148 is given in Tables 1–5 in the Supplementary file. We have also reported the association information of the top three modules from GSE157103 in Tables 6–8 in the Supplementary file.

3.4. Common pathways associated with disease and non-disease genes

A biological pathway is any sequence of events (interactions) between cellular molecules resulting in the formation of new products, activity, etc. Two or more genes that share some common pathway are considered to be significant in their proper functioning. Therefore, we considered this concept to analyze the contribution of such genes. The

Table 4
Disease genes found in top five modules for GSE157103 dataset.

| Module no. | No. of disease genes | Disease genes |
|------------|----------------------|--|
| 2 | 6 | SH2D3A, IRF3, CCL5, GPT, IFNL1, POLD1 |
| 10 | 3 | DDX58, TMPRSS2, STAT1 |
| 6 | 3 | CCL8, IL10, IL2 |
| 7 | 0 | * No disease gene found in this module |
| 4 | 1 | IL10 |

Table 5
Association of disease gene with other genes in the module 1 for GSE57148 dataset.

| Disease gene(s) in module | Members of module (except disease gene) | Physical interaction | Co-expression | Predicted interaction | Path way | Genetic interaction | Colocalization |
|---------------------------|---|----------------------|---------------|-----------------------|----------|---------------------|----------------|
| POLD1 | POLD4 | Yes | Yes | Yes | Yes | - | - |
| | NCAP G2 | - | Yes | - | - | - | - |
| | POLD3 | - | Yes | - | - | - | - |
| | TRAF2 | - | Yes | - | - | - | - |
| | IRF3 | - | Yes | - | - | - | - |
| | MAP3K 14 | - | Yes | - | - | - | - |
| | RPA2 | - | Yes | - | - | - | - |
| IRF3 | MAP2 K2 | - | Yes | - | - | - | - |
| | TRADD | - | Yes | - | - | - | - |
| | TRAF2 | Yes | Yes | - | - | - | - |
| | POLD1 | - | Yes | - | - | - | - |
| | MAPK8 | Yes | Yes | - | - | - | - |
| | HRAS | - | Yes | - | - | - | - |
| IFNAR1 | EP300 | Yes | - | - | - | - | - |
| | EP300 | Yes | - | - | - | - | - |

pathway information of member genes in each module was obtained using the DAVID tool [15]. The number of common pathways associated with disease genes and other genes for modules 1, 5, 8, and 25 is reported in Tables 6–9 for GSE57148 dataset. These tables report only entries that share a minimum of 4 common pathways with the disease genes in the case of GSE57148. Tables 10–12 report the number of common pathways for the top three modules obtained from the GSE157103 dataset. Since none of the member genes shared more than 3 common pathways with the disease gene in the case of modules obtained from GSE157103, we reported all its member genes along with the number of common pathways.

The results show that the non-disease genes share a high percentage of pathways with those of the disease genes in the module. Moreover, we know that members of the same module share high similarity among themselves, which is also established by calculating their p-value. Hence, we carried out further analyses of these non-disease genes given in Tables 6–12 to find their association with COVID-19.

3.5. Some crucial genes associated with COVID-19

The inherent property of gene modules suggests a functionally similar nature of each member within the module. The member genes work in coordination to perform similar functions. Therefore, we can assume that the non-disease genes within the diseased module are also responsible for causing the disease. The results thus far show that several non-disease genes might be related to disease genes and contribute to aggravating the disease condition. Table 13 reports the top non-disease genes that share the maximum number of common pathways with disease genes. These genes are found to be common in our findings. These

Table 6
Common pathways along with genes related to disease gene in Module 1 for GSE57148.

| Disease gene | Related genes in module | No. of common pathways |
|--------------|-------------------------|------------------------|
| POLD1 | POLD4 | 9 |
| | RPA2 | 4 |
| IRF3 | IKBKB | 8 |
| | MAPK8 | 8 |
| | TRAF2 | 5 |
| | EIF2AK3 | 4 |
| | EP300 | 4 |
| | TRADD | 4 |
| | IFNAR1 | 7 |
| IFNAR1 | MAPK8 | 6 |
| | MAP2K2 | 5 |
| | EIF2AK3 | 4 |
| | EP300 | 4 |
| | HRAS | 4 |
| | RAC1 | 4 |

Table 7
Common pathways along with genes related to disease gene in Module 5 for GSE57148.

| Disease gene | Related genes in module | No. of common pathways |
|--------------|-------------------------|------------------------|
| DDX58 | IKBKE | 7 |
| | IRF7 | 7 |
| | EIF2AK2 | 5 |
| | IFIH1 | 5 |
| | MYD88 | 5 |
| | PIK3CD | 5 |
| | PIK3R5 | 5 |
| | STAT2 | 5 |
| | IL12B | 4 |
| | IRF9 | 4 |
| | OAS1 | 4 |
| | OAS2 | 4 |
| | OAS3 | 4 |
| | CXCL10 | IKBKE |
| IL12B | | 4 |
| IRF7 | | 4 |
| PIK3CD | | 4 |
| PIK3R5 | | 4 |

genes were further analyzed to determine their association with COVID-19. Some of the associative genes along and their role in COVID-19 are discussed.

- RPA2*- Coronavirus is a class of RNA virus that causes DNA damage to enter into the host body. Xu et al. [18] showed that coronavirus infectious bronchitis virus (IBV) is replicated by ATR signaling activation, which is associated with the phosphorylation of substrates like RPA2 on Ser.
- POLD4*- People with COPD have a reduced production of POLD4, which leads to an increase in the number of karyomere-like cells [6]. This leads to genomic instability in the long run. The cells keep growing despite the reduction in POLD4 until the DNA damage reaches a threshold level and manifests disease symptoms in patients.
- MAPK8*- The MAPK8 gene is also known as JNK,⁷ and has been already established as being affected by viral infections. Fung suggests that JNK/MAPK8 is involved in coronavirus–host interaction according to overexpression experiments [5]. The MAPK8 gene belongs to the family of mitogen-activated protein kinase (MAPK) genes and is responsible for the translation of signals that regulate stress conditions, growth factors, and pro-inflammatory cytokines. Cytokines are responsible for producing signals associated with cell

⁷ <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=5599>, accessed on 18th May 2020.

Table 8
Common pathways along with genes related to disease gene in Module 8 for GSE57148.

| Disease gene | Related genes in module | No. of common pathways | |
|--------------|-------------------------|------------------------|---|
| IL6 | NFKB1 | 22 | |
| | JUN | 13 | |
| | NFKBIA | 13 | |
| | FOS | 10 | |
| | TNFRSF1A | 9 | |
| | CDKN1A | 7 | |
| | CDK2 | 6 | |
| | ICAM1 | 6 | |
| | MYC | 6 | |
| | TICAM1 | 6 | |
| | CSF1 | 5 | |
| | CXCL1 | 5 | |
| | CXCL2 | 5 | |
| | CXCL3 | 4 | |
| | ITGA5 | 4 | |
| | SOCS3 | NFKB1 | 9 |
| | | TNFRSF1A | 8 |
| NFKBIA | | 7 | |
| JUN | | 5 | |
| FOS | | 4 | |
| CCL2 | SOCS2 | 4 | |
| | NFKB1 | 6 | |
| | NFKBIA | 6 | |
| | JUN | 5 | |
| | TNFRSF1A | 5 | |
| | CXCL1 | 4 | |
| | CXCL2 | 4 | |
| MCL1 | FOS | 4 | |
| | ICAM1 | 4 | |
| | ICAM1 | 4 | |
| | ICAM1 | 4 | |

Table 9
Common pathways along with genes related to disease gene in Module 25 for GSE57148.

| Disease gene | Related genes in module | No. of common pathways |
|--------------|-------------------------|------------------------|
| IFNG | FASLG | 10 |
| | CD28 | 7 |
| | CD40LG | 6 |
| STAT1 | FASLG | 5 |
| | BCL2 | 4 |
| CCL5 | FASLG | 4 |

Table 10
Common pathways along with genes related to disease gene in Module 2 for GSE157103.

| Disease gene | Related genes in module | No. of common pathways |
|--------------|-------------------------|------------------------|
| GPT | NDUFA13 | 1 |
| | NDUFA2 | 1 |
| | BCAT2 | 1 |
| | HYKK | 1 |
| | BCKDHA | 1 |
| | HYI | 1 |

proliferation, inflammation, wound healing, and cell migration [4]. A person with a severe form of COVID-19 produces a very high influx of pro-inflammatory cytokines, such as IL-6, IL-1, TNF- α and interferon [12]. This phenomenon is commonly known as a cytokine storm, which ultimately leads to apoptosis of the epithelial and endothelial cells in the body, thereby deteriorating the recovery chances of the patient [16].

d) *IRF7*- The *IRF7* gene is one of the active genes associated with the p-53-directed production of type I interferons, which enhance the body's natural immunity [19]. Yuan et al. revealed that coronaviruses escape the human body's innate response to fight viral infections by inhibiting the p-53 signaling pathway. This is possible

Table 11
Common pathways along with genes related to disease gene in Module 6 for GSE157103.

| Disease gene | Related genes in module | No. of common pathways |
|--------------|-------------------------|------------------------|
| IL10 | CNTF | 2 |
| | IFNL2 | 2 |
| | LIFR | 2 |
| | CXCR4 | 1 |
| | IRS2 | 1 |
| | CXCL8 | 1 |
| | CCL3L3 | 1 |
| IL2 | CNTF | 2 |
| | IFNL2 | 2 |
| | LIFR | 2 |

Table 12
Common pathways along with genes related to disease gene in Module 10 for GSE157103.

| Disease gene | Related genes in module | No. of common pathways |
|--------------|-------------------------|------------------------|
| STAT1 | STAT2 | 3 |
| | CCR1 | 1 |
| | TSLP | 1 |
| DDX58 | AIM2 | 2 |

Table 13
Maximum number of common pathways shared between non disease gene and disease gene.

| Dataset | New biomarkers | No. of common pathways shared with disease gene | New biomarkers | No. of common pathways shared with disease gene |
|-----------|----------------|---|----------------|---|
| GSE57148 | POLD4 | 9 | MX1 | 2 |
| | RPA2 | 4 | CXCL9 | 3 |
| | TRAF2 | 5 | IRF7 | 7 |
| | IKBKE | 3 | EIF2AK2 | 4 |
| | MAPK8 | 8 | IRF9 | 4 |
| | EP300 | 4 | CXCL3 | 4 |
| | IL12B | 3 | OAS2 | 3 |
| | IFIH1 | 5 | CXCL2 | 5 |
| | OAS3 | 4 | ICAM1 | 6 |
| | PIK3CD | 4 | FOS | 3 |
| | CDKN1A | 3 | BCL2 | 4 |
| | OAS1 | 4 | JUN | 13 |
| | MYD88 | 5 | NFKB1 | 22 |
| | CCR5 | 2 | NFKBIA | 13 |
| GSE157103 | CXCL1 | 5 | CD40LG | 6 |
| | FASLG | 10 | TRADD | 4 |
| | EP300 | 3 | PIK3R5 | 3 |
| | MAPK8 | 4 | PIK3R5 | 4 |
| | STAT2 | 3 | LIFR | 2 |
| | CCR1 | 1 | | |

due to the presence of a protein-like structure called papain-like proteases (PLPs) around the corona of the COVID-19 virus. This protein is known to suppress the inborn immunity of any individual.

e) *JUN*- This gene is associated with inducing cAMP signaling pathway simulation, which directly modulates the MAPK signaling pathways. This pathway regulates the entire process of cytokine production in human tissues [3], thereby preventing the occurrence of cytokine storm, a characteristic feature of COVID-19. Studies show that antiviral drugs such as mercaptopurine and melatonin can be used as potential drugs to target the human coronavirus as they block the c-Jun signaling pathway [22].

f) *NFKB1* & *NFKBIA*- *NFKB1* and *NFKBIA* are in a class of proteins that are homologous to retroviral proteins. They are involved in the regulation of gene expression by acting as transcription factors via the NFKB signaling pathway [9,20]. This pathway is involved in regulating the stress response and natural immunity in humans.

Mutation in the transcription process of this protein has been associated with cancer and autoimmune diseases.⁸

Immunocompromised people are more susceptible to coronavirus as their immune systems are unable to fight back any foreign invasions. Various types of cancers such as leukemia and lymphoma are associated with weak immune systems. The reason is that these two types of cancers are associated with bone marrow and white blood cells, respectively, which are known to fight against infections. However, when such patients undergo chemotherapy, the target is to wipe out these cancer-affected cells, leading to a weakened immune system. This makes such people easy targets for COVID-19.⁹

Severe cases of COVID-19 are found to attack the gas-exchanging units of the lungs and invade the type II cells of alveolar units. The SARS-CoV2 virus propagates within these alveolar units, leading to the release of large units of poisonous viral particles. As a result, the innate cells undergo apoptosis and finally die [10]. COVID-19 is found to be more severe in older people for the same reason as in those with weakened immune systems.

- g) *CD40LG*- Rare mutations in the CD40LG gene are associated with X-linked hyper-IgM syndrome [1], a rare class of immunodeficiency disorder¹⁰. People affected by this disease do not produce enough antibodies to fight any kind of infections, making them more susceptible towards coronavirus attacks.
- h) *FASLG*- The FAS/FASLG gene is linked to immune system regulation and cytotoxic T lymphocyte-induced cell death.¹¹ Any deviation from the normal immune-response in the body makes a person prone to foreign infections.
- i) *ICAM1*- Allergic reactions occurring in the body show the presence of soluble ICAM1 in the nasal epithelial cells. This protein is also found in high proportions among the bronchial asthma patients [17]. Increased ICAM1 expression among cells supports viral replication. The presence of this protein provides a favorable environment within the nose for the coronavirus to enter and survive in the human body.
- j) *LIFR*- Smoke from cigarettes reduces the expression of the LIF gene and that of its receptor, the LIFR gene. It has been observed that reduced LIF expression in COPD patients leads to worsening health conditions causing lung damage [23] associated with COVID-19.
- k) *STAT2*- Boudewijns et al.'s work [24] suggest a dual role of the STAT2 gene. The expression of STAT2 disrupts the spread of viral infection in the body. On the other hand, it is also associated with severe lung injury, which would deteriorate the health condition of any person suffering from lung infection.
- l) *CCR1*- Cross examination of the pathological samples obtained from COVID-19 patients shows the involvement of a hyperactive immune system apart from other symptoms [25]. Certain studies suggest that inhibition of chemokine receptors such as CCR1 and CCR5 may overcome the problem of immune hyperactivation in COVID-19 patients.¹²

⁸ <https://www.novusbio.com/antibody-news/antibodies/nfkb-and-p62-both-activate-and-regulate-inflammation>, accessed on 20th May 2020.

⁹ <https://theconversation.com/what-does-it-mean-to-be-immunocompromised-and-why-does-this-increase-your-risk-of-coronavirus-135200>, accessed on 23rd May 2020.

¹⁰ <https://rarediseases.org/rare-diseases/hyper-igm-syndrome>, accessed on 23rd May 2020.

¹¹ <https://ghr.nlm.nih.gov/gene/FASLG>, accessed on 23rd May 2020.

¹² <https://frontlinegenomics.com/study-finds-chemokine-receptor-inhibition-can-overcome-immune-hyperactivation-in-covid-19-patients/> accessed on 04.10.2020.

4. Discussion

In this work, we analyzed gene modules obtained from a co-expression network for both COPD patients and COVID-19 patients. Our analysis revealed that gene modules obtained from COPD samples were more biologically significant in terms of the p-value than those obtained from COVID-19 samples. Moreover, it has been found that COVID-19 affects the human respiratory tract, and since COPD is an untreatable respiratory disorder, we tried to identify the susceptibility of genes found in COPD patients with respect to COVID-19.

During our analysis, we came across a number of crucial genes that might be directly or indirectly involved in the worsening of symptoms of COVID-19. We also analyzed the expression values of COVID-19 patients and identified a few crucial genes that might be associated with the disease. The genes found could be used by biologists and drug designers to design target-specific drugs that would limit the severe complications arising in COVID-19.

Conflict of interests

The authors declare that they have no competing interest.

Acknowledgement

This work has been supported by All India Council for Technical Education (AICTE- under Ministry of HRD) Collaborative Research Scheme (CRS) under TEQIP..

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compbiomed.2020.104126>.

References

- [1] C. Aloui, C. Sut, A. Prigent, J. Fagan, F. Cognasse, V. Granados-Herbecin, R. Touraine, B. Pozzetto, M. Aouni, C. Fendri, M. Hassine, Are polymorphisms of the immunoregulatory factor CD40LG implicated in acute transfusion reactions? *Sci. Rep.* 4 (2014) 7239.
- [2] G.F. Berriz, O.D. King, B. Bryant, C. Sander, F.P. Roth, Characterizing gene sets with FuncAssociate, *Bioinformatics* 18 (2003) 2502–2504.
- [3] G.M. Fimia, P. Sassone-Corsi, Cyclic AMP signalling, *J. Cell Sci.* 114 (11) (2001) 1971–1972.
- [4] H.F. Florindo, R. Kleiner, D. Vaskovich-Koubi, R.C. Acúrcio, B. Carreira, E. Yeini, G. Tiram, Y. Liubomirski, R. Satchi-Fainaro, Immune-mediated approaches against COVID-19, *Nat. Nanotechnol.* (2020) 1–6.
- [5] T.S. Fung, D.X. Liu, Activation of the c-Jun NH 2-terminal kinase pathway by coronavirus infectious bronchitis virus promotes apoptosis independently of c-Jun, *Cell Death Dis.* 8 (12) (2017) 1–3.
- [6] Q.M. Huang, T. Akashi, Y. Masuda, K. Kamiya, T. Takahashi, M. Suzuki, Roles of POLD4, smallest subunit of DNA polymerase δ , in nuclear structures and genomic stability of human cells, *Biochem. Biophys. Res. Commun.* 391 (1) (2010) 542–546.
- [7] C.C. Lai, T.P. Shih, W.C. Ko, H.J. Tang, P.R. Hsueh, Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges, *Int. J. Antimicrob. Agents* (2020) 105924.
- [8] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, *BMC Bioinf.* 9 (1) (2008) 559.
- [9] L. Li, Z.T. Zhang, Genetic Association between NFKBIA and NFKB1 Gene Polymorphisms and the Susceptibility to Head and Neck Cancer: a Meta-Analysis, *Disease markers*, 2019, <https://doi.org/10.1155/2019/6523837>, 2019.
- [10] J. Mason Robert, Pathogenesis of COVID-19 from a cell biology perspective, *Eur. Respir. J.* 55 (2020) 2000607, <https://doi.org/10.1183/13993003.00607-2020>.
- [11] S. Mostafavi, D. Ray, D. Warde-Farley, C. Grouios, Q. Morris, GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function, *Genome Biol.* 9 (S1) (2008) S4.
- [12] D. Ragab, H. Salah Eldin, M. Taeimah, R. Khattab, R. Salem, The COVID-19 cytokine storm; what we know so far, *Front. Immunol.* 11 (2020) 1446.
- [13] A. Remuzzi, G. Remuzzi, COVID-19 and Italy: what next? *Lancet* (2020) [https://doi.org/10.1016/S0140-6736\(20\)30627-9](https://doi.org/10.1016/S0140-6736(20)30627-9).
- [14] M. Safran, I. Solomon, O. Shmueli, M. Lapidot, S. Shen-Orr, A. Adato, U. Ben-Dor, N. Esterman, N. Rosen, I. Peter, T. Olander, GeneCards™ 2002: towards a complete, object-oriented, human gene compendium, *Bioinformatics* 18 (11) (2002) 1542–1543.
- [15] B.T. Sherman, Q. Tan, J.R. Collins, W.G. Alvord, J. Roayaei, R. Stephens, M. W. Baseler, H.C. Lane, R.A. Lempicki, The DAVID Gene Functional Classification

- Tool: a novel biological module-centric algorithm to functionally analyze large gene lists, *Genome Biol.* 8 (9) (2007) R183.
- [16] Y. Tang, J. Liu, D. Zhang, Z. Xu, J. Ji, C. Wen, Cytokine storm in COVID-19: the current evidence and treatment strategies, *Front. Immunol.* 11 (2020) 1708.
- [17] B. Winther, E. Arruda, T.J. Witek, S.D. Marlin, M.M. Tsianco, D.J. Innes, F. G. Hayden, Expression of ICAM-1 in nasal epithelium and levels of soluble ICAM-1 in nasal lavage fluid during human experimental rhinovirus infection, *Arch. Otolaryngol. Head Neck Surg.* 128 (2) (2002) 131–136.
- [18] L.H. Xu, M. Huang, S.G. Fang, D.X. Liu, Coronavirus infection induces DNA replication stress partly through interaction of its nonstructural protein 13 with the p125 subunit of DNA polymerase δ , *J. Biol. Chem.* 286 (45) (2011) 39546–39559.
- [19] L. Yuan, Z. Chen, S. Song, S. Wang, C. Tian, G. Xing, X. Chen, Z.X. Xiao, F. He, L. Zhang, p53 degradation by a coronavirus papain-like protease suppresses type I interferon signaling, *J. Biol. Chem.* 290 (5) (2015) 3172–3182.
- [20] G.L. Zhang, Y.F. Zou, X.L. Feng, H.J. Shi, X.F. Du, M.H. Shao, Y. Gu, Q. Zhou, Association of the NFKBIA gene polymorphisms with susceptibility to autoimmune and inflammatory diseases: a meta-analysis, *Inflamm. Res.* 60 (1) (2011) 11–18.
- [21] Y.Y. Zheng, Y.T. Ma, J.Y. Zhang, X. Xie, COVID-19 and the cardiovascular system, *Nat. Rev. Cardiol.* 17 (5) (2020) 259–260.
- [22] Y. Zhou, Y. Hou, J. Shen, Y. Huang, W. Martin, F. Cheng, Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2, *Cell Discovery* 6 (1) (2020) 1–8.
- [23] Justin Poon, Michael Campos, Robert F. Foronjy, Sridesh Nath, Gayatri Gupta, Christopher Railwah, Abdoulaye J. Dabo, Nathalie Baumlin, Matthias Salathe, Patrick Geraghty, Cigarette smoke exposure reduces leukemia inhibitory factor levels during respiratory syncytial viral infection, *Int. J. Chronic Obstr. Pulm. Dis.* 14 (2019) 1305.
- [24] Robbert Boudewijns, Hendrik Jan Thibaut, Suzanne JF. Kaptein, Rong Li, Valentijn Vergote, Laura Seldeslachts, De Keyzer, Carolien, Sapna Sharma, Sander Jansen, Van Weyenbergh, Johan, STAT2 Signaling as Double-Edged Sword Restricting Viral Dissemination but Driving Severe Pneumonia in SARS-CoV-2 Infected Hamsters, *BioRxiv*, 2020.
- [25] Jixin Zhong, Jungen Tang, Cong Ye, Dong, Lingli the immunology of COVID-19: is immune modulation an option for treatment? *The Lancet Rheumatology* (2020).

Pooja Sharma received her Ph.D. in Computer Science from Tezpur University in 2018. Her research interests include Machine Learning, Bioinformatics and Data mining.

Anuj K Pandey is pursuing his Ph.D. in the Department of Electrical Engineering at NIT Durgapur, India. His research interest includes Power Electronics, Optimization, Soft Computing and Artificial Intelligence.

Dhruba K Bhattacharyya received his Ph.D. in Computer Science from Tezpur University in 1999. His research areas include Data Mining, Network Security and Content-based Image Retrieval. Prof. Bhattacharyya has published 250 + research papers in leading international journals and conference proceedings. In addition, Dr Bhattacharyya has written and edited many books.