Bioinformatics Approaches to Identify the Comorbidity Complexities of SARS-CoV-2 Infection with Crohn's Disease

Yue Li\(^1\) and Xiao Han\(^2\)

\(^1\)Department of Emergency and Critical Care Medicine, Jiading District Central Hospital Affiliated Shanghai University of Medicine & Health Sciences, Shanghai, China

\(^2\)Department of Cardiology, Jiading District Central Hospital Affiliated Shanghai University of Medicine & Health Sciences, Shanghai, China

ABSTRACT

Objective: To analyse potential molecular mechanisms and identify potential therapeutic regimens and drugs to treat severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and Crohn’s disease (CD) through bioinformatics and systems biology.

Study Design: Bioinformatics Study.

Place and Duration of the Study: Jiading District Central Hospital Affiliated Shanghai University of Medicine & Health Sciences, Shanghai, China, from May to December 2022.

Methodology: The common differentially expressed genes (DEGs) between CD and SARS-CoV-2 infection were identified using two RNA-seq datasets (GSE147507, GSE153974) extracted from Gene Expression Omnibus (GEO). Subsequently, functional enrichment, pathway analysis, and candidate drug analysis were performed using these DEGs.

Results: In total, 44 DEGs were identified as common between CD and SARS-CoV-2 infection. A protein-protein interaction (PPI) network was constructed, hub genes were identified, and critical modules were determined by means of bioinformatics and combinatorial statistical approaches. Functional and pathway analyses conducted under ontological conditions showed a common association between CD and infection with SARS-CoV-2. The common DEGs were then used to identify coregulatory networks of interactions between transcriptional factors and genes, between proteins and medicines, and between DEGs and miRNAs.

Conclusion: Top 10 hub genes including IL6, CXCL1, CSF2, CXCL2, CXCL5, MMP3, PTGS2, CXCL3, SELE, and LCN2 were identified, which may function as potential candidate targets for SARS-CoV-2 infection. Additionally, the identification of certain promising treatment drugs for patients with SARS-CoV-2 infection and CD was also made.

Key Words: SARS-CoV-2, COVID-19, Drug molecule, Hub gene, Protein-protein interaction (PPI), Gene ontology, Crohn’s disease, Differentially expressed genes.

How to cite this article: Li Y, Han X. Bioinformatics Approaches to Identify the Comorbidity Complexities of SARS-CoV-2 Infection with Crohn’s Disease. J Coll Physicians Surg Pak 2023; 33(10):1093-1099.
affects patients with CD and develop candidate drugs that can alleviate the associated symptoms and complications. As of yet, no bioinformatics study has looked at the impact of SARS-CoV-2 at the molecular level in patients with pre-existing CD.

Disease-specific datasets GSE147507 (SARS-CoV-2) and GSE153974 (CD) were compared from the Gene Expression Omnibus (GEO) database to bioinformatics analysis.

**METHODOLOGY**

The present study utilised microarray and RNA-seq datasets obtained from the GEO database of the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/geo/) to explore the common genetic pathways between SARS-CoV-2 and CD. The research was conducted in Jiading District Central Hospital Affiliated Shanghai University of Medicine & Health Sciences, Shanghai, China. The authors screened the sequences using the following criteria to obtain a high-quality dataset: (1) genome variation profiling by high-throughput sequencing; (2) case-control studies of Homo sapiens; (3) series type; (4) expression profiling by array; (5) comparable conditions for experimental and untreated controls; (6) analysis of gene expression profiling; (7) availability of completely raw and processed microarray data; and (8) the appropriate positioning for RNA extraction.

The SARS-CoV-2 dataset with GEO accession ID GSE147507 was obtained and included high-throughput sequencing using the Illumina HiSeq 500 platform for RNA-seq extractions from lung biopsies of 23 patients (including lung samples derived from a single male COVID19 deceased patient (age: 74 years) with SARS-CoV-2 infection; and 55 healthy controls (including uninfected human lung biopsies were derived from one male (age 72 years) and one female (age 60 years) in the USA. Next, the dataset chosen for analysis, with the accession ID GSE153974, comprised of 57 samples of CD and 17 samples of healthy control individuals in UK. These samples consisted of human ileal biopsies that were preserved in RNA, later and subsequently sequenced using the high-throughput Illumina NextSeq 500 (Homo sapiens) sequencing system. Figure 1(a) shows the summary of the datasets.

DESeq2 R package was applied to identify DEGs for the GSE147507 and GSE153974 datasets, with the criteria set as [log fold change] ≥1.0 and false discovery rate (FDR) <0.05. Subsequently, the identification of the common differentially expressed genes (DEGs) was facilitated by the utilisation of Jvenn (http://jvenn.n.toulouse.inra.fr/app/example.html), an online tool for Venn diagram analysis. The biological mechanisms and Reactome signalling pathways of the common differentially expressed genes (DEGs) were analysed, and the top-ranked functional items and pathways were identified using a standardised metric of p-value <0.05.

In order to demonstrate the interconnectedness of SARS-CoV-2 and CD at both a structural and functional level, the differentially expressed genes (DEGs) were uploaded to the STRING database (http://string-db.org/), the goal of which is to evaluate protein-protein interactions (PPI) and integrate the results. The PPI network of overlapping DEGs was generated based on a low confidence score of 0.15. Additionally, the Cytoscape program (v.3.8.0), a commonly employed bioinformatics tool for the visualisation of biological networks and integration of associated data, was utilised in the study to further visualise and investigate the PPI network.

The 10 prominent hub genes in the protein-protein interaction (PPI) network were identified through the application of the Cytohubba MCC technique. Subsequently, their rankings were visually represented using a colour gradient ranging from red to yellow. Additionally, the shortest paths between hub genes were determined by using the proximate neighbourhood ranking features of Cytohubba.

Through the utilisation of the NetworkAnalyst online tool (http://www.networkanalyst.ca/), an exploration was conducted to identify transcriptional factors (TFs) that bind to the shared differentially expressed genes (DEGs). To identify miRNAs that bind to transcripts of target genes and hence, adversely affect protein production by interfering with the messenger RNA (mRNA) stability and translational efficiency, the NetworkAnalyst online tool was employed to conduct an analysis of miRNA-target gene interactions sourced from miRtarBase. Then, the researchers screened for high-degree miRNAs and determined the biochemical processes and characteristics associated with each miRNA.

Enrichr and the Drug Signatures Database (DSigDB) were employed to investigate the molecular composition of drugs by analysing shared differentially expressed genes (DEGs) between SARS-CoV-2 infection and CD. DSigDB is an updated gene set repository for use in gene set enrichment analysis (GSEA) that relates compounds/drugs and their respective target genes.

By utilising the DisGeNET online tool (http://www.disgenet.org/), a comprehensive analysis was conducted to identify multiple chronic conditions associated with the common differentially expressed genes (DEGs).

**RESULTS**

In the dataset of SARS-CoV-2 infection, a total of 1781 differentially expressed genes (DEGs) were identified, comprising 391 downregulated and 1390 upregulated genes. Likewise, 253 DEGs were identified in the CD dataset, among which 4 were downregulated, whereas 249 were upregulated. All the DEGs were identified using the criteria of p-value <0.05 and an absolute log fold change (|logFC|) ≥1. A p-value <0.05 suggested that there is a 5% risk of making a false decision to choose a DEG as per the cutoff threshold. LogFC is a log ratio between gene expression levels and is used for calculating differences in expression levels in two distinct conditions such as control vs. case. Figure 1b shows the volcano plots and heatmap of the identified DEGs in the SARS-CoV-2 infection (Figure 1(bi)) and CD (Figure 1(bii)) datasets. By employing Jvenn, a total of 44 differentially expressed genes (DEGs) that were shared between the two datasets were identified (Figure 1c), indicating that several genes are similar under SARS-CoV-2 infection and CD conditions, and that the pathogenesis may be closely related.
Bioinformatics approaches to identify the comorbidity complexities of SARS-CoV-2 infection with Crohn’s disease

<table>
<thead>
<tr>
<th>Disease name</th>
<th>GEO accession</th>
<th>GEO platform</th>
<th>Total DEGs count</th>
<th>Up regulated DEGs count</th>
<th>Down regulated DEGs count</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARS-CoV-2</td>
<td>GSE147507</td>
<td>GPL18573</td>
<td>1781</td>
<td>1590</td>
<td>391</td>
</tr>
<tr>
<td>CD</td>
<td>GSE153974</td>
<td>GPL18573</td>
<td>253</td>
<td>249</td>
<td>4</td>
</tr>
</tbody>
</table>

Figure 1: (a) Summary of the geographical characteristics and quantitative analytical measurements of the selected data sets. (b) Discovery of DEGs between SARS-CoV-2 infection and CD. The red dots in the volcano plots in (i) and (ii) represent significant DEGs for SARS-CoV-2 infection and CD, respectively. The criteria for detecting DEGs are $|\log \text{FC}| \geq 1$ and adjusted p-value <0.05. Based on log2 fold changes, the heatmaps show the relationships between the common DEGs of different conditions. DEG, differentially-expressed genes; CD, Crohn’s disease. (c) Venn diagram showing the number of common DEGs between SARS-CoV-2 infection (GSE147507) and CD (GSE153974) datasets.

Figure 2: (a) The PPI network of common DEGs between SARS-CoV-2 infection and CD. The DEGs are denoted by the circular nodes, whereas the corresponding interactions are denoted by the edges. The PPI network consists of 44 nodes and 100 edges. DEGs, differentially expressed genes; CD, Crohn’s disease; PPI, protein-protein interaction. (b) Detection of hub genes. Genes that serve as hubs in the PPI network were found utilising the Cytohubba Cytoscape plugin. The top-ranked 10 hub genes and the molecules they interact with are displayed in red, orange, or yellow. There are 25 nodes in the submodule network, connected by 96 edges. (c) Interactions between the DEGs and TFs. Circles indicate the DEG, while squares represent TFs. 47 nodes and 83 edges were found in the network. TF, transcription factor; DEGs, differentially expressed genes. (d) Interactions between the DEGs and microRNAs (miRNAs). Circles represent miRNAs, and squares represent the genes that interact with them. The network has 19 nodes and 25 edges. DEGs, differentially expressed genes. (e) Reactome 2022 Pathway enrichment analysis of the common DEGs between SARS-CoV-2 infection and CD is presented as bar graphs.
**Figure 2(a)** illustrates the PPI network shared between SARS-CoV-2 infection and CD. The PPI network comprised of 44 nodes and 100 edges. Furthermore, through the use of the Cytoscape plugin Cytohubba, the researchers discovered the hub proteins of the PPI network as depicted in Figure 2(b). Overall, 10 DEGs (22.73%) were considered to be the most influential genes, including IL6, CXCL1, CSF2, CXCL2, CXCL5, MMP3, PTGS2, CXCL3, SELE, and LCN2. These hub genes could help identify drugs for diseases associated with these comorbidities. In order to enhance comprehension of their proximity and connectivity, a submodule network consisting of 25 nodes and 96 edges was constructed utilizing the Cytohubba plugin (Figure 2(b)). The analyses of DEG-TF and DEG-miRNA interaction were conducted with the help of NetworkAnalyst (http://www.networkanalyst.ca/) to determine the transcription and post-transcription regulatory biomolecules controlling the DEGs. The interplay between modulatory TFs and DEGs based on the JASPAR database is shown in Figure 2(c). The TFs YY1, GATA2, FOXL1, and NFKB1 were found to be the most common DEGs between SARS-CoV-2 infection and CD. Specific granules (8 genes) and tertiary granules (7 genes) were found to be the top two GO pathways in the cellular component category. There is evidence that zymogen-granule membrane glycoprotein 2 (GP2) is one of the most frequent targets of autoantibodies in CD. Furthermore, the cytotoxic granule proteins produced by eosinophils are closely associated with CD.

In this study, 44 DEGs shared similar expression patterns in CD and SARS-CoV-2 infection. Within the biological process, the cytokine-mediated signalling pathway (16 genes) and cellular response to cytokine stimulus (13 genes) ranked the highest. Cytokines are released by immune and non-immune cells and cause inflammation and tissue damage in the gut of patients with CD. It is known that IBD-related inflammation occurs due to faulty production and activity of anti-inflammatory cytokines. Inflammation can be managed with anti-cytokine therapy that targets the immune system to restore intestinal barrier function. In patients with SARS-CoV-2 infection, a cytokine storm prevents further virus transmission and induces secondary tissue damage by secreting large amounts of inflammatory mediators. The growing evidence indicates that SARS-CoV-2 infection severity is associated with a cytokine storm, which is also a critical cause of death.

**Table I**: Medicine candidates in treating SARS-CoV-2 infection and Crohn's disease (CD).

<table>
<thead>
<tr>
<th>Name</th>
<th>p-values</th>
<th>Chemical formula</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptidoglycan CTD 00008490</td>
<td>6.90E-11</td>
<td>C_{60}H_{78}NO_{32}</td>
<td><img src="structure1.png" alt="Structure" /></td>
</tr>
<tr>
<td>1-Nitropyrene CTD 00001569</td>
<td>7.24E-11</td>
<td>C_{8}H_{10}NO_{2}</td>
<td><img src="structure2.png" alt="Structure" /></td>
</tr>
<tr>
<td>Sulodexidil PC3 UP</td>
<td>1.50E-09</td>
<td>C_{10}H_{20}NOS</td>
<td><img src="structure3.png" alt="Structure" /></td>
</tr>
<tr>
<td>Simvastatin and Niacin</td>
<td>1.98E-07</td>
<td>C_{4}H_{17}Od</td>
<td><img src="structure4.png" alt="Structure" /></td>
</tr>
</tbody>
</table>

Researchers identified 10 potential drug molecules for SARS-CoV-2 infection and CD using Enrichr based on DSigDB. As per the p-value, Table I lists the top four candidate compounds, which may be effective for targeting the common DEGs.

**DISCUSSION**

In this study, 44 DEGs shared similar expression patterns in CD and SARS-CoV-2 infection. Within the biological process, the cytokine-mediated signalling pathway (16 genes) and cellular response to cytokine stimulus (13 genes) ranked the highest. Cytokines are released by immune and non-immune cells and cause inflammation and tissue damage in the gut of patients with CD. It is known that IBD-related inflammation occurs due to faulty production and activity of anti-inflammatory cytokines. Inflammation can be managed with anti-cytokine therapy that targets the immune system to restore intestinal barrier function. In patients with SARS-CoV-2 infection, a cytokine storm prevents further virus transmission and induces secondary tissue damage by secreting large amounts of inflammatory mediators. The growing evidence indicates that SARS-CoV-2 infection severity is associated with a cytokine storm, which is also a critical cause of death.

Specific granules (8 genes) and tertiary granules (7 genes) were found to be the top two GO pathways in the cellular component category. There is evidence that zymogen-granule membrane glycoprotein 2 (GP2) is one of the most frequent targets of autoantibodies in CD. Furthermore, the cytotoxic granule proteins produced by eosinophils are closely associated with CD. In SARS-CoV-2 infection, platelet granule release is uncoupled from integrin activation, suggesting that GPIIb/IIIa can be therapeutically targeted.

In terms of biological processes, cytokine activity (11 genes) and CXCR chemokine receptor binding (5 genes) were among the most prominent GO terms in the present analysis. The inflammation that led to tissue damage in CD is triggered by cytokines and the severity of SARS-CoV-2 infection was related to the cytokine storm. Moreover, treatment with antibodies against chemokine ligand (CXCL) 13 (the ligand of CXCR5) reduced the severity of colitis.

The IL-8-CXCR-1/-2 axis targeting suppressed the activation of neutrophils, NETosis, degranulation, and IL-8 production, which led to a reduced number of SARS-CoV-2 spike protein-induced, ACE2-mediated microthromboses in the lungs.
Ten hub proteins, namely IL6, CXCL1, CSF2, CXCL2, CXCL5, MMP3, PTGS2, CXCL3, SELE, and LCN2, were found to be involved in the two diseases based on the PPI network.

IL-6 is a pleiotropic cytokine involved in immune system regulation among many other extensive functions. As a powerful pro-inflammatory cytokine, IL-6 is an important component involved in the response of the immune system to pathogens and acute stress. Cancer, inflammation, and autoimmunity are pathologically related to IL-6 pathways. In SARS-CoV-2 infection, exploratory studies have suggested that a significant increase in IL-6 levels is associated with adverse outcomes. Hence, the inhibition of IL-6 may be one of the most promising therapeutic targets for patients with SARS-CoV-2 infection with dysregulated host responses. Meanwhile, CXCL1, CXCL2, and CXCL5 are the proinflammatory cytokines responsible for CD inflammatory responses. A correlation between MMP-3 and MMP-9 concentrations and CD severity suggests that they may be useful in evaluating CD severity, while MMP3 is currently being investigated for its potential as a biomarker as well as a drug target in SARS CoV-2 infection. LCN2 is a non-invasive biomarker that correlates with intestinal inflammation. Moreover, there is potential for neutrophil-related LCN2 to serve as a new biomarker to anticipate the severity of SARS-CoV-2 infections, and it may also be used for new treatment targets. Thus, these hub genes can potentially function as disease biomarkers or as novel therapeutic targets, if the biological knowledge gained on SARS-CoV-2 infection can be verified.

Subsequently, the researchers proceeded to establish associations among DEGs, TFs, and miRNAs. The different types of digestive system disorders appeared to be associated with the TFs identified (YY1, GATA2, FOXL1, NFKB1, MEF2A, CREB1, FOXC1, STAT3, and GATA3). Moreover, several of the identified miRNAs are implicated in lung cancer (e.g., hsa-mir-1-3p, hsa-mir-138-5p, hsa-mir-26a-5p), hepatocellular carcinoma (e.g., hsa-mir-26a-5p, hsa-mir-124-3p, hsa-mir-1-3p) and osteosarcoma (e.g., hsa-mir-335-5p, hsa-mir-26a-5p). Therefore, these molecules are worth further testing as new therapeutic targets for treating the symptoms of SARS-CoV-2 infection. In essence, these transcription factors and microRNAs target proteins to influence disease progression. For instance, there has been a research to demonstrate that hsa-mir-335-5p and hsa-mir-98-5p target IL6 and hsa-mir-1-3p and hsa-mir-98-5p target CXCL1. Additionally, four miRNAs have been predicted (hsa-mir-1-3p, hsa-mir-335-5p, hsa-mir-124-3p, and hsa-mir-155-5p) to be associated with different CD-related genes.

The association between DEGs and different diseases was predicted using gene-disease analysis. Different illnesses have been linked to SARS-CoV-2 infection, such as mouth neoplasms, pancreatic neoplasms, colorectal neoplasms, bladder neoplasms, kidney neoplasms, esophageal neoplasms, and stomach neoplasms. A study has also shown a close correlation between SARS-CoV-2 and cancer risk, mortality, and severe illnesses. Tumour cells can be stimulated to switch metabolisms by using SARS-CoV-2 to initiate metabolic modes with higher production efficiency, which would promote tumour progression and facilitate SARS-CoV-2 replication. An alternative study suggested that tumour suppressor genes are inhibited by SARS-CoV-2, possibly promoting cancer onset. Additionally, a correlation between SARS-CoV-2 and autoimmune diseases was also discovered. Several studies indicated that patients with systemic autoimmune disease have a higher chance of developing SARS-CoV-2 infection.

The curative radical therapy of SARS-CoV-2 is being explored in greater detail in this study. In the present investigation, the researchers anticipated the identification of four potential pharmaceutical compounds for the treatment of SARS-CoV-2. Peptidoglycan was identified, which acts against enveloped viruses by disrupting virion integrity. A study has shown that enterovirus inactivation can be affected by lipopolysaccharide or peptidoglycan of bacterial origin. 1-NITROPYRENE can cause apoptosis and pyroptosis in cells. A form or forms of cytochrome P450 are selectively induced by 1-NITROPYRENE to facilitate the metabolism of the substance. 1-NITROPYRENE induces macrophage synthesis of proinflammatory cytokines including, interleukin 1 (IL-1), IL-6, and TNF, which exhibit proinflammatory responses. Therefore, there is a possibility that 1-NITROPYRENE may enhance virus inactivation and lead to an immune response to help patients fight SARS-CoV-2. An antithrombotic drug, sulodictil, as a potential SARS-CoV-2 therapeutic agent, has been reported as an antifungal agent, especially against Candida albicans. This drug may play a significant role in patients infected with SARS-CoV-2 complicated with fungal infections and pulmonary embolism.

Furthermore, simvastatin and niacin were approved by the FDA to lower lipids and may enhance patients' recovery outcome in SARS-CoV-2 infection. The nucleocapsid protein (NP) of SARS-CoV-2 strongly activates human endothelial cells via the MAPK and TLR2/NF-κB pathways, while simvastatin significantly inhibits endothelial activation induced by NP, thus suppressing virus cell entry and inflammatory cytokine production. To date, there has been no comprehensive examination of the molecular-level effects of SARS-CoV-2 on individuals with pre-existing CD in both local and international bioinformatics research. The findings of this study lend support to future pharmaceutical studies on the candidate drugs to treat patients with CD and SARS-CoV-2 infection. However, notwithstanding our diligent endeavours, this study does possess certain limitations. Insufficient sample size and the different selection criteria of the datasets may hinder the identification of common DEGs in disease studies, thereby potentially limiting the comprehensive capture of all necessary disease-associated genes. Therefore, an additional investigation is necessary to comprehensively...
evaluate the biological significance of the potential target candidates identified in this study. It is worth mentioning that the above results were obtained by applying bioinformatics analyses without experiments or clinical trials. Therefore, further experiments are necessary to verify the safety and efficacy of the potential medicines against CD and SARS-CoV-2 infection.

**CONCLUSION**

The top 10 hub genes were identified, namely IL6, CXCL1, CXCL2, CXCL5, MMP3, PTGS2, CXCL3, SELE, and LCN2 between CD and SARS-CoV-2 infection. Moreover, the analysis identified the TFs and miRNA that play an important role in SARS-CoV-2 infection, thus highlighting potential risk factors. Five potential drugs were also predicted: Peptidoglycan (a component of bacterial cell walls), 1-NITROPYRENE, sulocidil, simvastatin, and niacin, which may lead to new therapeutic targets and the development of SARS-CoV-2 infection vaccines.

**ACKNOWLEDGEMENT:**

This work was supported by the Scientific Research Foundation of Shanghai University of Medicine and Health Sciences (SSF-21-16-001) and Shanghai University of Medicine and Health Sciences Clinical Research Centre for Metabolic Vascular Diseases Project (20MC2020004).

**ETHICAL APPROVAL:**

The study was approved by the Ethics Committee of Jiading District Central Hospital Affiliated Shanghai University of Medicine & Health Sciences, Shanghai, China (Approval No. XHEC-C-2022-12).

**PATIENTS’ CONSENT:**

Written informed consents were obtained from patients to publish the data.

**COMPETING INTEREST:**

The authors declare no potential competing interest with respect to the research, authorship, or publication of this article.

**AUTHORS’ CONTRIBUTION:**

YL: Manuscript writing and data analysis.

XH: Subject design and data review.

All authors have approved the final version of the manuscript to be published.

**REFERENCES**


