OXFORD

*Briefings in Bioinformatics*, 00(0), 2021, 1–15

**doi: 10.1093/bib/bbab003**
Case Study

# Diseasome and comorbidities complexities of SARS-CoV-2 infection with common malignant diseases

Md. Shahriare Satu, Md. Imran Khan, Md. Rezanur Rahman, Koushik Chandra Howlader, Shatabdi Roy, Shuvo Saha Roy, Julian M.W. Quinn and Mohammad Ali Moni

Corresponding author: Mohammad Ali Moni, WHO Collaborating Centre on eHealth, UNSW Digital Health, School of Public Health and Community Medicine, Faculty of Medicine, University of New South Wales, Sydney, Australia; E-mail: m.moni@unsw.edu.au

## Abstract

With the increasing number of immunoinflammatory complexities, cancer patients have a higher risk of serious disease outcomes and mortality with SARS-CoV-2 infection which is still not clear. In this study, we aimed to identify infectome, diseasome and comorbidities between COVID-19 and cancer via comprehensive bioinformatics analysis to identify the synergistic severity of the cancer patient for SARS-CoV-2 infection. We utilized transcriptomic datasets of SARS-CoV-2 and different cancers from Gene Expression Omnibus and Array Express Database to develop a bioinformatics pipeline and

**Md. Shahriare Satu** received his B.Sc. and M.Sc. degrees in information technology from Janagirnagar University in 2015 and 2017, respectively. He is currently working as a lecturer at the Department of Management Information Systems, Noakhali Science and Technology University, Noakhali, Bangladesh. From March 2016 to 2 December 2018, he worked as a lecturer at the Department of Computer Science and Engineering, Gono Bishwabidyalay, Bangladesh. His research interest includes data mining, health informatics and big data analytics.

**Md. Imran Khan** received his B.Sc. in computer science and engineering from Gono Bishwabidyalay in 2019. His research interest includes machine learning and bioinformatics.

**Md. Rezanur Rahman** is a lecturer (on study leave) Khwaja Yunus Ali University, Bangladesh. He is a biotechnologist and bioinformatician by training. He received MSc and a BSc in biotechnology and genetic engineering with outstanding academic results holding first position in the faculty in both degrees from Islamic University, Bangladesh. For the highest CGPA in the faculty, he received the most prestigious award "Prime Minister Gold Medal 2017" from Prime Minister Sheikh Hasina. He has published 26 research articles in internationally reputed journals. His research focuses on to understand the molecular mechanism of complex diseases utilizing transcriptomics, systems biology and bioinformatics.

**Koushik Chandra Howlader** was born in Jhalakathi, Bangladesh, in 1991. He received his B.Sc. and M.Sc. degree in information technology from Jahangirnagar University, Bangladesh. Then he took his internship program from Infosys Ltd., India. After coming back to India, he joined as a junior Java Programmer in the IBCS-Primax Software Ltd., Dhaka. Currently, he is working as an assistant professor at the Department of Computer Science and Telecommunication Engineering at Noakhali Science and Technology University, Bangladesh. From 27 December 2015 to 26 December 2017, he worked as lecturer in the same department Noakhali Science and Technology University. His research interest includes data mining, bioinformatics, machine learning, etc.

**Shatabdi Roy** received her B.Sc. in computer science and telecommunication engineering from Noakhali Science and Technology University in 2019, respectively. Her research interest includes bioinformatics and bio-medical engineering.

**Shuvo Saha Roy** received his B.Sc. in computer science and telecommunication engineering from Noakhali Science and Technology University in 2019, respectively. His research interest includes bioinformatics.

**Julian M. W. Quinn** received the PhD from the University of Oxford, UK, in 1992, then moved to Australia for postdoctoral training in bone, joint and cancer biology at the St Vincent's Institute of Medical Research as a senior research fellow since 2014. His interests are in applications of biostatistics and bioinformatics, and now he works as a surgical research officer at the Surgical Education and Research Training Institute at Royal North Shore Hospital, Sydney Australia.

**Mohammad Ali Moni** is a research fellow and conjoint lecturer at the University of New South Wales, Australia. He received his PhD degree in clinical bioinformatics and machine learning from the University of Cambridge. His research interest encompasses artificial intelligence, machine learning, data science, health informatics and clinical bioinformatics.

**Submitted:** 13 October 2020; **Received (in revised form):** 25 December 2020

software tools to analyze a large set of transcriptomic data and identify the pathobiological relationships between the disease conditions. Our bioinformatics approach revealed commonly dysregulated genes (MARCO, VCAN, ACTB, LGALS1, HMOX1, TIMP1, OAS2, GAPDH, MSH3, FN1, NPC2, JUND, CHI3L1, GPNMB, SYTL2, CASP1, S100A8, MYO10, IGFBP3, APCDD1, COL6A3, FABP5, PRDX3, CLEC1B, DDIT4, CXCL10 and CXCL8), common gene ontology (GO), molecular pathways between SARS-CoV-2 infections and cancers. This work also shows the synergistic complexities of SARS-CoV-2 infections for cancer patients through the gene set enrichment and semantic similarity. These results highlighted the immune systems, cell activation and cytokine production GO pathways that were observed in SARS-CoV-2 infections as well as breast, lungs, colon, kidney and thyroid cancers. This work also revealed ribosome biogenesis, wnt signaling pathway, ribosome, chemokine and cytokine pathways that are commonly deregulated in cancers and COVID-19. Thus, our bioinformatics approach and tools revealed interconnections in terms of significant genes, GO, pathways between SARS-CoV-2 infections and malignant tumors.

**Key words:** comorbidities; COVID-19; cancers

## Introduction

Coronavirus disease-19 (COVID-19) caused by the SARS-CoV-2 virus has become a global crisis where the World Health Organization (WHO) declared it as a pandemic on 11 March 2020 [1]. This virus initially creates a respiratory illness that can spread rapidly. In addition to losing thousands of human lives, COVID-19 causes massive damages in the global economy. When numerous coronaviruses were studied, only seven are known that affects human health and severe diseases have happened for three of them, including severe acute respiratory syndrome coronavirus (SARS-CoV), middle east respiratory syndrome coronavirus (MERS-CoV) and, the current pandemic, SARS-CoV-2 virus [2]. For SARS-CoV and MERS-CoV, two serious global epidemics happened in 2003 and 2012 [3], respectively, but did not declare them as a pandemic. However, SARS-CoV-2 is a single-stranded RNA virus that showed 89.1% nucleotide similarity and spread more easily than others. COVID-19 patients with a number of pre-existing medical conditions (e.g., diabetes, heart disease, cancer) are more likely to suffer severe COVID-19 and poor therapeutic outcomes compared to normal infected people. Indeed, this virus affects multiple organs severely in the human body. Regarding cancer patients, a study was conducted over 55,000 confirmed COVID-19 cases in China where the death rate was 7.6% that indicated five times higher death risk than COVID-19 patients without comorbidities (1.4%) [4].Due to the relative weakness of patients for COVID-19, the question has been risen about the effects of various cancers and associated comorbidities. There is no adequate evidence about direct interaction among COVID-19 and various cancers. The frailty and cancer therapeutics are not easily modifiable where the interactions happened due to the cellular pathways of cancers and SARS-CoV-2 that could be focused by therapeutic intervention.

Numerous works of COVID-19 and cancer gene expressions happened to investigate and identify altered pathways that could serve as resources for studying COVID-19 and its cancer comorbidities. Also, it causes the changes of many potentially shared molecular factors that could interact with cancers. However, many existing and clinical databases cannot be utilized due to the lack of available bioinformatics pipelines. Therefore, we implemented a methodology that investigated possible comorbidity interactions of COVID-19 with a number of cancers relating to breast, lung, colon, kidney, liver, prostate, bladder and thyroid by examining the gene expression profiling. This analysis has been used to combine gene expressions, gene ontology and molecular instances by manipulating gene set enrichment analysis (GSEA) and semantic similarity, respectively. Therefore, various significant genes, GO terms and pathways were determined as the proximities and identified a potential interacting biological process (BP) for each disease.

## Materials and methods

Bioinformatics and integrative procedures [5] were used to investigate the relations among COVID-19 and various cancers that are described as follows:

### Data collection

The experimental datasets were obtained from the Gene Expression Omnibus (GEO) database, National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/geo/) and Array Express Database of European Bioinformatics Institute (EBI) (https://www.ebi.ac.uk/arrayexpress/). There were two query results found for COVID-19. Four main principles were used to identify appropriate microarray transcriptomics datasets, which are given as follows:

1. **Redundancy:** Several datasets are generated using similar conditions or explored with various methods. In these circumstances, no equivalent samples are not required more than one time.
2. **Typology:** Datasets required more accurate structural form such as sequential data.
3. **Relevance:** Datasets must be linked to specific pathology that gives certain importance about the biological relationships. Several samples does not contain its own pathology, hence they are imperfect for further analysis.
4. **Species:** Datasets must be gathered from clinical sources and not derived from non-human species.

### Gene Set Enrichment Analysis

GSEA is a functional process where a group of genes, their enhanced expressions and the effects of case versus control tissues are identified using statistical approaches. Also, they are recognized by genes and protein set that associates with particular disease phenotypes based on similar biological functionalities, chromosomal location and regulation [6]. However, the transcriptomic and proteomic data are investigated in this condition. Further, DNA microarray or next-generation sequencing (NGS)

**Table 1.** Selected COVID-19 datasets

| Disease name | Dataset | Tissue/source | Control | Case |
|---|---|---|---|---|
| COVID-19 | GSE147507 [7] | Human lung | 3 | 3 |
| | PBMC-COVID-19 [8] | Peripheral blood | 3 | 3 |

**Table 2.** Selected cancer datasets

| Disease name | Dataset | Tissue/source | Control | Case |
|---|---|---|---|---|
| Breast cancer (BC) | GSE98528 [9] | Invasive lobular carcinoma | 9 | 39 |
| | GSE107300 [10] | Lung metastatic subline | 6 | 6 |
| | GSE110332 [11] | Breast cancer cell SUM159 | 3 | 3 |
| | GSE124646 [12] | Breast biopsy | 10 | 90 |
| | GSE125989 [13] | Breast biopsy | 16 | 16 |
| Colon cancer (CC) | GSE78051 [14] | Colorectal cancer cells | 3 | 3 |
| | GSE92921 [15] | Colon tissue biopsy | 33 | 26 |
| | GSE94154 [16] | Colorectal adenocarcinoma cells | 3 | 3 |
| | GSE110425 [17] | Colon cancer cell | 6 | 6 |
| | GSE115716 [18] | pN1-LS174T cells | 9 | 18 |
| Kidney cancer (KC) | GSE105261 [19] | Kidney biopsy | 9 | 35 |
| | GSE117890 [20] | Kidney tissue | 6 | 5 |
| Liver cancer (LC) | GSE63067 [21] | Liver tissue | 7 | 11 |
| | GSE102079 [22] | Liver tissue | 14 | 243 |
| Bladder & prostate cancer (BPC) | GSE118123 [23] | Prostate cancer cell | 3 | 3 |
| | GSE122306 [24] | Bladder cancer cell | 6 | 6 |
| Thyroid cancer (TC) | GSE3678 [25] | Thyroid | 7 | 7 |
| | GSE65144 [26] | Thyroid | 13 | 12 |
| | GSE85457 | Thyroid | 3 | 4 |

data is explored by comparing genes from two cells or tissues and scrutinizing gene expressions depending on several states. These gene sets are interrelated with the phenotypic differences under the list of up- and down-regulated genes In this study, we gathered two COVID-19 and various cancer samples from GEO and EBI repository. The brief description of these datasets is given as follows (see Tables 1 and 2):

## Pathway

Molecular pathways are perturbed in diseased conditions and identification of them enriched by the DEGs provides critical signaling pathways and drug targets. We utilized KEGG database [27] to identify COVID-19 pathways overlapped with different cancers enriched by the DEGs.

## Ontology

GO is a conceptual model where biological information can be explored as a compatible and widespread structure. It represents genes and their related attributes across all species. The main purpose of GO is to represent, maintain, develop and annotate gene and gene products in details. Three GO domains are considered such as BP, molecular function and cellular component. However, pathological processes, experimental conditions and temporal information are not captured properly in this process. Alternatively, disease ontology (DO) denotes an open-source model that represents expansive information about inherited, developmental and acquired human diseases [28]. In this study, DO terms were extracted for the corresponding diseases such as COVID-19 DO ID: 00080600, breast cancer DO ID: 1612, colon cancer DO ID: 219, obesity DO ID: 9970, liver cancer DO ID: 3571, kidney cancer DO ID: 263, thyroid gland cancer DO ID: 1781, urinary bladder cancer DO ID: 11054 and prostate cancer DO ID: 10283. These DO IDs were retrieved from https://disease-ontology.org/. But, the result of SARS-CoV-2 is not available, hence we used DO ID of SARS coronavirus to compare DO with others.

## Semantic similarity

Semantic similarity is a function that measures the proximity between two terms annotating to the biological entities on a given ontology. Numerous methods are employed to organize common ancestor terms in view of the annotation statistics. In this work, the relations included more significant terms among genes, GO and DO than particular evaluations. The Wang method fits in this purpose because graph-based method constructs the topology and inherits by the selected ontology.

A directed acyclic graph is defined as $DAG_K = (K, T_K, E_K)$, where GO term $K$, the set of ancestor terms $T_K$ and edges $E_K$ (semantic relations) are related to $DAG_K$. The semantic value $S_K$ is manipulated as

$$\begin{cases} S_K(K) = 1 & t = K \\ S_K(t) = \max\{w_e * S_K(t') \,|t' \in \ children\ of(t)\} & t \neq K \end{cases}, \quad (1)$$

where $t$ and $t'$ specifies the generic and child term individually. According to the relation, the semantic contribution $w_e$ is assigned as 0 and 1 between $t$ and $t'$ and global semantic value for $K$ is calculated in Eq. 2.

$$SV(K) = \sum_{t \in T_K} S_K(t) \quad (2)$$

If $DAG_K = (K, T_K, E_K)$ and $DAG_L = (L, T_L, E_L)$ are measured using two terms $K$ and $L$, then the semantic similarity is given in Eq. 3.

$$\text{sim}(K, L) = \frac{\sum_{t \in T_K \cap T_L} [S_K(t) + S_L(t)]}{SV(K) + SV(L)} \quad (3)$$

Given are two term sets $X_1 = x_{11}, x_{12}, \ldots, x_{1m}$ and $X_2 = x_{21}, x_{22}, \ldots, x_{2n}$, where $m$ and $n$ are denoted as the length of the first and second set, respectively. The best-match average (BMA) method [29] generates the semantic similarity between the two

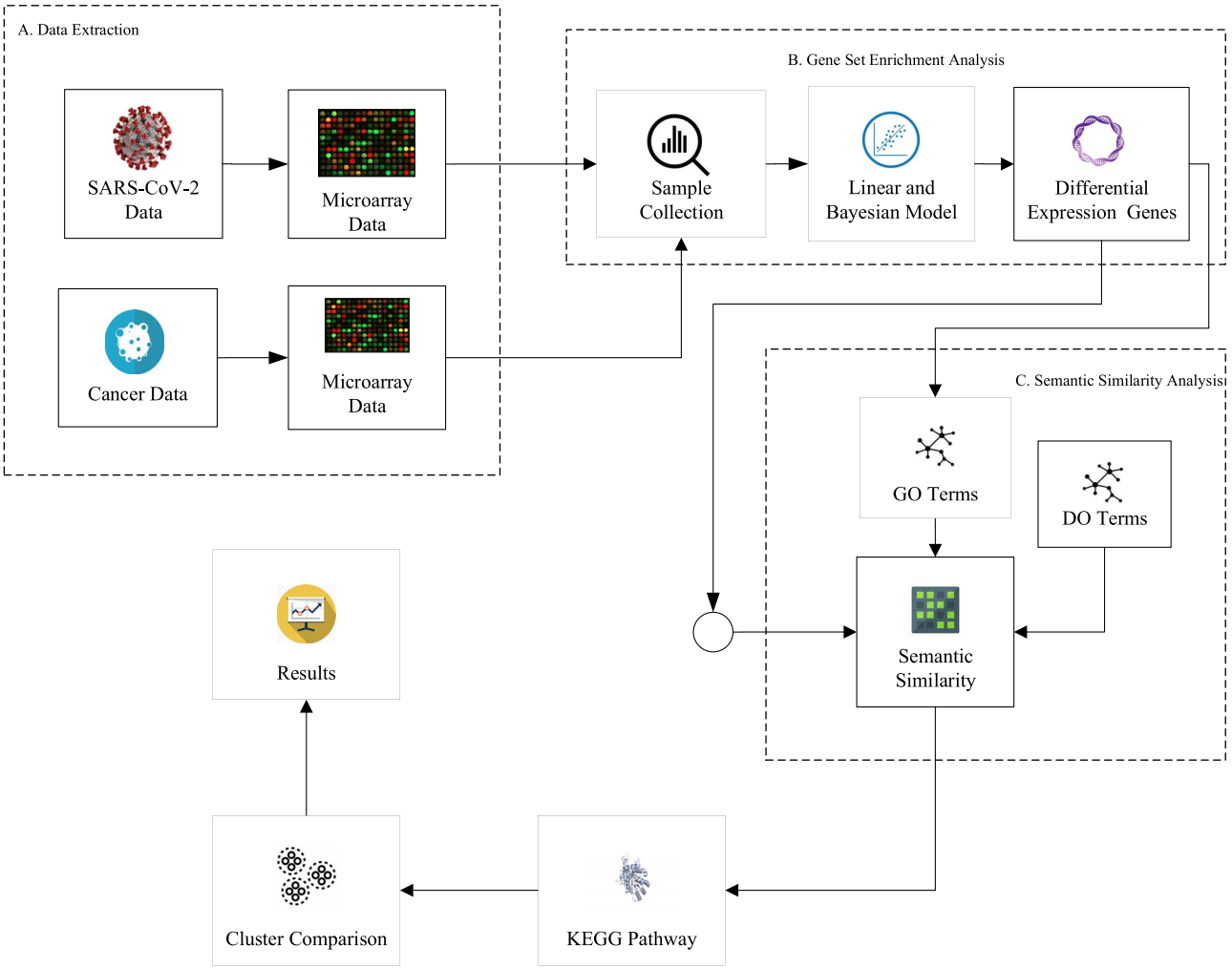


Figure 1. Working pipeline.

sets (see Eq. 4):

$$\mathrm{sim}_{BMA}(X_1, X_2) = \frac{\sum_{i=1}^{m} \max_{1\le j\le m} \mathrm{sim}(x_{1i}, x_{2j}) + \sum_{j=1}^{n} \max_{1\le i\le l} \mathrm{sim}(x_{1i}, x_{2j})}{m+n}, \quad (4)$$

with i,j indices on $K_1$,$L_1$ terms.

## Designing of pipeline

Figure 1 shows the steps of the pipeline:

1. In the data extraction process, the selected COVID-19 and cancer datasets were downloaded and explored matrix information. The normalization was performed to convert them into expression classes. Subsequently, DEGs were identified as a linear and Bayesian method by comparing the expression of healthy controls or treated COVID-19 patients.
2. These samples were manually gathered to conduct this work. Then, we reviewed, selected and classified GEO samples (GSM) very meticulously rather than the automatic selection process.
3. Differential expression can be used for identifying significant genes altered in a particular condition. To identify DEGs, a linear and Bayesian method was applied [27]. We considered three statistical criteria, namely P-value, adjusted P-value (False Discovery Rate) and absolute logFC values to screen statistically significant DEGs.
4. In the GO term test, the class called topGOdata was created, which picked GO terms and genes to implement filtering function. The mapping had been engaged for annotation where Fisher's exact test was used to explore the relationship between GO terms and genes.
5. After the mapping of the semantic similarity, the performance among all the selected pathologies were compared by means of genes, GO terms, DO terms for discrimination of the intimacy among the designated datasets.
6. Cluster comparison was used to fetch significant pathologies and enrichment test was dependent on DEGs and KEGG pathways for COVID-19 and cancers.
7. Finally, the output of this process provided a statistical summary, genes-GO terms, GO graph topology, gene semantic similarity matrix (and dendrogram), GO semantic similarity matrix (and dendrogram), DO semantic similarity matrix (and dendrogram), KEGG enrichment graph and the list of the common pathways pathologies [30]. In addition, the list

**Table 3.** Statistical summary of COVID-19 datasets used in this study. Columns 3, 4, 5 and 6 represent the number of unfiltered genes, the number of significant DEG with threshold for *P*-value, adjusted *P*-value and logFC, respectively. Columns 7 and 8 show the number of raw GO terms and significant GO terms with Fisher test, respectively.

| Dataset | Source tissue | Raw genes | *P*-value | Adjusted *P*-value | logFC | GO terms | Fisher test |
|---|---|---|---|---|---|---|---|
| GSE147507 | Human lung | 108 | 108 | 108 | 108 | 1117 | 446 |
| PBMC-COVID-19 | Peripheral blood mononuclear cells | 1745 | 1745 | 1745 | 1745 | 4386 | 1745 |

**Table 4.** Statistical summaries of cancer comorbidities datasets. Columns 3, 4, 5 and 6 specify the number of unfiltered genes, the number of significant DEG with the threshold for *P*-value, adjusted *P*-value and logFC, respectively. Columns 7 and 8 show the number of raw GO terms and significant GO terms by Fisher test, respectively. Data set legend: BC-GSE98528, GSE107300, GSE110332, GSE124646 and GSE125989; CC-GSE78051, GSE92921, GSE94154, GSE110425, GSE115716; KC-GSE105261 and GSE117890; LC-GSE63067 and GSE102079; BPC-GSE118123 and GSE122306; and TC-GSE3678, GSE65144 and GSE85457

| Dataset | Source tissue (case control) | Raw genes | *P*-value | Adjusted *P*-value | LogFC | GO terms/ raw GSEA | Fisher test |
|---|---|---|---|---|---|---|---|
| GSE118123 | Prostate cancer cell | 54675 | 4255 | 6 | 2 | 172 | 56 |
| GSE122306 | Bladder cancer cell | 54675 | 3086 | 6 | 17 | 133 | 70 |
| GSE98528 | Invasive lobular carcinoma | 46446 | 3522 | 0 | 50 | 127 | 63 |
| GSE107300 | Lung metastatic subline | 47302 | 8500 | 3326 | 8816 | 100 | 55 |
| GSE110332 | Breast cancer cell SUM159 | 22277 | 3340 | 594 | 58 | 170 | 70 |
| GSE124646 | Breast biopsy | 22283 | 5878 | 3183 | 1005 | 134 | 52 |
| GSE125989 | Breast biopsy | 22277 | 1778 | 104 | 2097 | 122 | 31 |
| GSE78051 | Colon cancer cell | 47323 | 3642 | 77 | 0 | 68 | 38 |
| GSE92921 | Colon tissue biopsy | 54675 | 8359 | 1558 | 525 | 150 | 90 |
| GSE94154 | Colorectal adenocar-cinoma cells | 54675 | 12141 | 4360 | 328 | 103 | 72 |
| GSE110425 | Colon cancer cell | 47323 | 1743 | 0 | 1 | 133 | 60 |
| GSE115716 | pN1-LS174T cells | 47323 | 6438 | 360 | 171 | 164 | 27 |
| GSE105261 | Kidney biopsy | 48107 | 8090 | 2359 | 816 | 159 | 54 |
| GSE117890 | Kidney tissue | 47309 | 5905 | 56 | 13925 | 212 | 142 |
| GSE63067 | Liver tissue | 54676 | 4448 | 0 | 227 | 248 | 168 |
| GSE102079 | Liver tissue | 54613 | 16539 | 8578 | 1263 | 179 | 74 |
| GSE3678 | Thyroid biopsy | 54675 | 6290 | 1615 | 1215 | 227 | 121 |
| GSE65144 | Thyroid biopsy | 54675 | 16368 | 10748 | 10911 | 151 | 74 |
| GSE85457 | Thyroid biopsy | 54613 | 7777 | 297 | 16055 | 329 | 189 |

**Table 5.** Summary of results along with the pipeline steps for the selected pathologies. The features from left to right are denoted as selected disease, source, number of data sets, number of selected data sets and number of upregulated and downregulated DEGs

| Disease | Origin/tissue | Dataset | Selected dataset | DEG up | DEG down |
|---|---|---|---|---|---|
| COVID-19 | Lung and blood | 2 | 2 | 1239 | 614 |
| Breast cancer | Breast | 260 | 5 | 5603 | 6443 |
| Colon cancer | Colon | 120 | 5 | 372 | 674 |
| Kidney cancer | Kidney | 80 | 2 | 7145 | 7613 |
| Liver cancer | Liver | 80 | 2 | 788 | 718 |
| Thyroid cancer | Thyroid | 240 | 3 | 15004 | 13187 |
| Bladder & prostate | Bladder & prostate | 80 | 2 | 11 | 8 |

of DEGs constructed gene networks corresponding to the information related to pathways/pathologies.

Then, we represented this work using two R scripts that are available at https://github.com/shahriariit/COVID-Cancer-Comorbidities. To build this bioinformatics pipeline, we used several Bioconductor packages [31] such as: "GEOquery" [32] for downloading GEO data and transformation of expression set class; "LIMMA" [33] for microarray data analysis, linear models and identifying DEGs on microarray data; "genefilter" [34] for keeping basic tasks of filtering genes; "topGO" for verifying GO terms and topology of DAG; "GOSemSim" [35] for the semantic similarity assessment among the diseases; "DOSE" [36] for the semantic similarity assessment among DO terms; and "clusterProfiler" [37] for the enrichment analysis with KEGG pathways.

## Results

### Statistical analysis of transcriptomic data

To identify common dysregulated DEGs between COVID-19 and cancers, we comprehensively analyzed the available transcriptomics datasets. The statistical summary of the COVID-19 and their cancer comorbidities have been presented in Tables 3

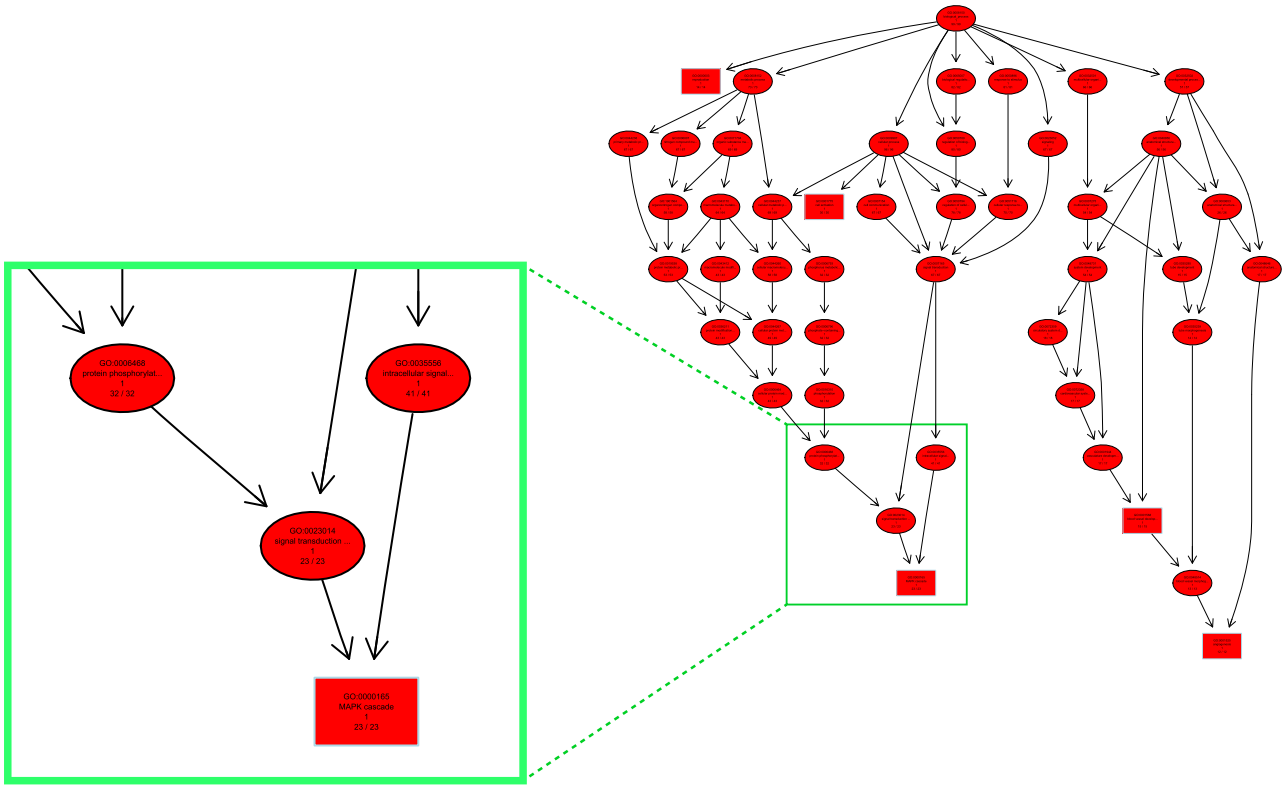


Figure 2. Example of the GO graph with GSEA on GSE147507 data set. The rectangles represent the top five GO terms after the test. The red and orange colors indicate the most significant GO terms.

and 4, respectively. The 4th and 5th column of Table 4 provides the number of DEGs that retains the statistical threshold of *P*-value $< 0.05$. We extracted up- and down-regulated genes based *P*-value and absolute $Log_2$ fold change ($logFC \geq 1$) for GSEA analysis where *logFC* denotes the direction of gene expression. In the 6th column, the number of significant DEGs is presented that gives the specific *logFC* threshold and used them for GO mapping. In the 7th column, the number of annotated GO terms of DEGs are provided. Later, Fisher's exact test was employed to extract statistically significant terms based on gene counting. The classical enrichment analysis was performed to evaluate the over-representation of these terms within the DEGs group. We summarize GO terms in the last column of Tables 3 and 4. For example, GO graph on GSE147507 represents the hierarchy and zoom on significant GO terms at Figure 2.

## KEGG pathway

To clarify the significance of the DEGs from transcriptomic datasets, we have performed gene ontologies and pathway analysis. The pathway-based analysis represents how complex diseases associates with other underlying molecular mechanisms [27]. Moreover, the following framework is provided on the BP involved in each COVID-19 study.

- GSE-147507: reproduction, MAPK cascade, angiogenesis, blood vessel development, cell activation
- PBMC-COVID-19: multicellular organismal process, developmental process, anatomical structure development, multicellular organism development, system development

## GO enrichment and construction of GO terms tree

We compared the DEGs identified from genome-wide transcriptomic datasets of COVID-19 and selected cancers and identified several common dysregulated genes (MARCO, VCAN, ACTB, LGALS1, HMOX1, TIMP1, OAS2, GAPDH, MSH3, FN1, NPC2, JUND, CHI3L1, GPNMB, SYTL2, CASP1, S100A8, MYO10, IGFBP3, APCDD1, COL6A3, FABP5, PRDX3, CLEC1B, DDIT4, CXCL10 and CXCL8) that are found common between COVID-19 and cancer (see Figure 3). To provide insights into the functional interactions of the identified genes, a protein–protein interaction network is created around the common DEGs using GeneMania web-utility considering co-expression, physical interaction, pathway, co-localization, generic interaction, predicted and shared protein domains.

Similarly, the common genes of COVID-19 (lung and blood tissues) and individual cancers are shown in Tables 6 and 7, respectively, which are obtained from the comparison between COVID-19 and cancer comorbidities.

## Semantic similarity analysis of the KEGG pathways

We performed the semantic similarity of the pathways enriched by the DEGs in order to prioritize and evaluate their proximity. Figure 4 shows the semantic similarity matrix for DEGs of the selected pathologies. The COVID-19 (PBMC) is highly connected to BC5_GSE124646, BC4_GSE110332 and TC3_GSE65144 when the values of semantic similarity are above 0.7. While COVID-19 (lung data) is highly associated with LC1_GSE63067 at the same semantic similarity values. When

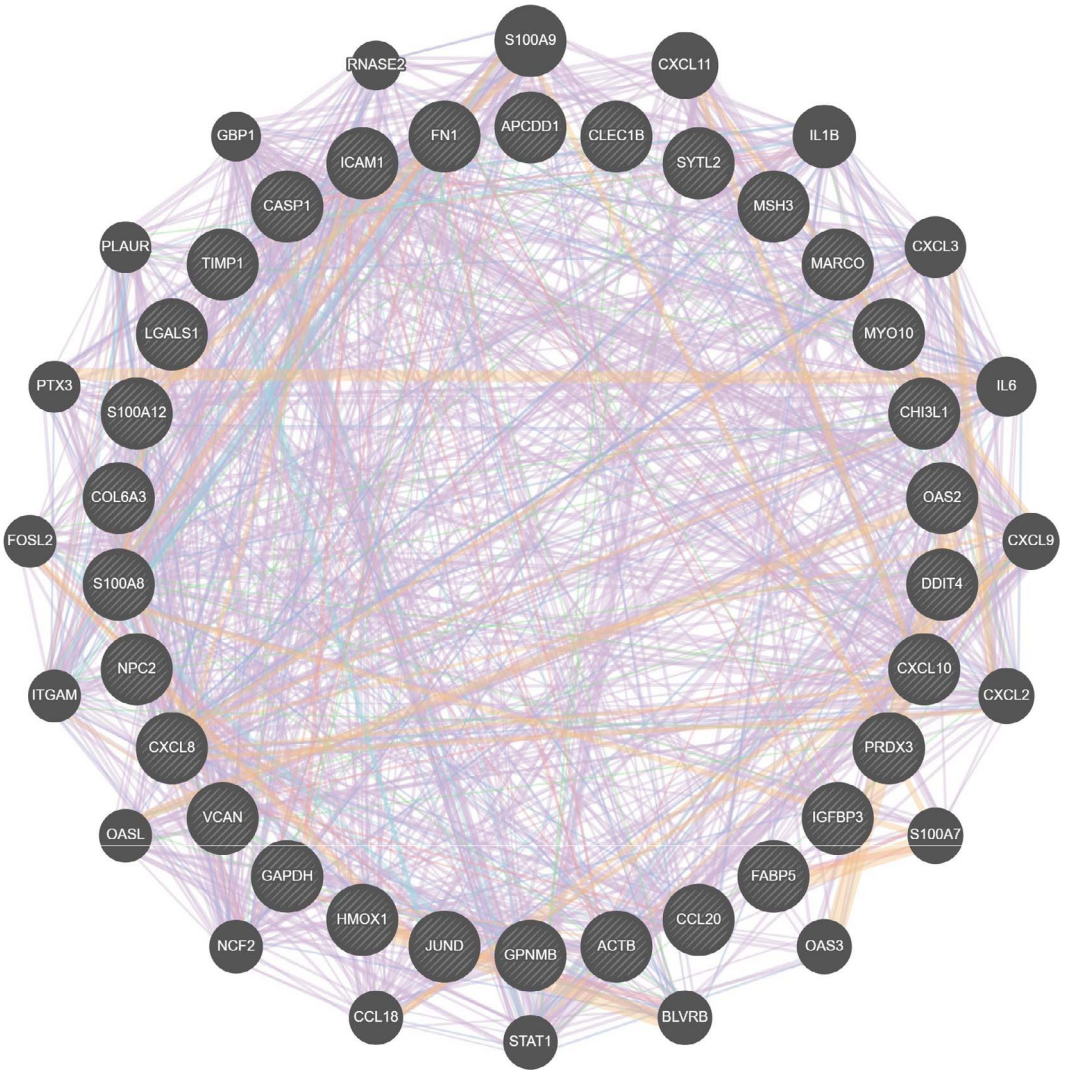


Figure 3. Network on common differential expressed genes between COVID-19 and its cancer comorbidities.

we consider semantic similarity value above 0.6 and less 0.7, then COVID-19 (lung) and COVID-19 (PBMC) are associated with several cancers like LC2_GSE102079, KD2_GSE105261, CC4_GSE94154, BC6_GSE125989, BC2_GSE98528, TC2_GSE3678 and CC3-_GSE92921 individually. At 0.5 semantic similarity score, COVID-19 (PBMC) is related to LC1_GSE63067. This matrix showed that TC1_GSE3468, BC3_GSE107300, BC2_GSE98528 and BP1_GSE1181123 provided low semantic similarity value with other cancers.

Figure 5 represents the semantic similarity matrix of GO terms. Over the value of 0.7 and less than 0.8, all datasets are found well-clustered among themselves except COVID-19 (lung) and LC1_GSE63067. When the semantic similarity value was 0.8 and less than 0.9, LC2_GSE102079, TC1_GSE3467, TC3_GSE65144 and KD2_GSE105261 are also well-clustered with several cancer pathologies. When the semantic similarity value was 0.9 or above 0.9, TC1_GSE3468, TC2_GSE3678, CD1_GSE893333, BC4_GSE110332 and KD2-_GSE105261 are represented well-clustered.

Figure 6 shows DO terms for SARS-CoV where COVID-19, breast cancer, kidney cancer, liver cancer and thyroid cancer are related with 0.09 threshold. Again, colon cancer contains 0.07 similarity value that is less connected than others. Instead, DO terms for SARS-CoV-2 are not available in the DO repository where it shows blank values in the generated graph. Hence, we used terms of SARS-CoV in this work.

However, Figures 7 and 8 show KEGG pathway association with selected datasets. This analysis is useful to understand how complex diseases may be related to each other through their underlying molecular mechanisms [27]. It represents the relationships between KEGG pathways of COVID-19 and associated cancer data sets. These pathways enriched by DEGs are shown in the dot plot where each row represents them associated with COVID-19 and various cancers. The domination of genes is determined by the dimension of the circles in the pathway and the range of the circles is computed the statistical validation for P-value = 0.05.

Common recurring pathways between COVID-19 (lung) and others pathologies are found including viral protein interaction with cytokine and cytokine receptor, Toll-like receptor signaling pathway, Influenza A, prion diseases, cytokine–cytokine receptor interaction, Rheumatoid arthritis, IL-17 signaling pathway, TNF signaling pathway and NOD-like receptor signaling pathway, among others.

**Table 6.** Common GO terms among COVID-19 (lung) and cancers

| GSE ID | GO ID | GO term |
|---|---|---|
| Common GO terms between COVID-19 and breast cancer | | |
| BC1_GSE95165 | GO:0002376 | Immune system process |
| BC4_GSE110332 | GO:0002376 | Immune system process |
| BC5_GSE124646 | GO:0002376 | Immune system process |
| BC6_GSE125989 | GO:0002376 | Immune system process |
| BC7_GSE135427 | GO:0002376 | Immune system process |
| | GO:0002520 | Immune system development |
| BC8_GSE89333 | GO:0001568 | Blood vessel development |
| | GO:0001775 | Cell activation |
| | GO:0001932 | Regulation of protein phosphorylation |
| | GO:0001934 | Positive regulation of protein phosphorylation |
| | GO:0001944 | Vasculature development |
| | GO:0002376 | Immune system process |
| Common GO terms between COVID-19 and colon cancer | | |
| CC3_GSE92921 | GO:0001775 | Cell activation |
| | GO:0001816 | Cytokine production |
| | GO:0001817 | Regulation of cytokine production |
| | GO:0002376 | Immune system process |
| | GO:0002682 | Regulation of immune system process |
| | GO:0002684 | Positive regulation of immune system process |
| Common GO terms between COVID-19 and kidney cancer | | |
| KD2_GSE105261 | GO:0001775 | Cell activation |
| | GO:0002252 | Immune effector process |
| | GO:0002253 | Activation of immune response |
| | GO:0002376 | Immune system process |
| | GO:0002682 | Regulation of immune system process |
| | GO:0002684 | Positive regulation of immune system process |
| KD2_GSE105261 | GO:0001775 | Cell activation |
| | GO:0002376 | Immune system process |
| | GO:0006955 | Immune response |
| | GO:0007166 | Cell surface receptor signaling pathway |
| | GO:0009605 | Response to external stimulus |
| | GO:0010033 | Response to organic substance |
| KD3_GSE117890 | GO:0001775 | Cell activation |
| | GO:0002376 | Immune system process |
| | GO:0002682 | Regulation of immune system process |
| Common GO terms between COVID-19 and lung cancer | | |
| LC1_GSE63067 | GO:0001775 | Cell activation |
| | GO:0001816 | Cytokine production |
| | GO:0001932 | Regulation of protein phosphorylation |
| | GO:0001934 | Positive regulation of protein phosphorylation |
| | GO:0002252 | Immune effector process |
| | GO:0002263 | Cell activation involved in immune response |
| | GO:0002274 | Myeloid leukocyte activation |
| | GO:0002275 | Myeloid cell activation involved in immune response |
| | GO:0002366 | Leukocyte activation involved in immune response |
| | GO:0002376 | Immune system process |
| | GO:0002443 | Leukocyte mediated immunity |
| | GO:0002444 | Myeloid leukocyte mediated immunity |
| | GO:0002446 | Neutrophil mediated immunity |
| | GO:0002682 | Regulation of immune system process |
| | GO:0002684 | Positive regulation of immune system process |
| LC2_GSE102079 | GO:0002252 | Immune effector process |
| | GO:0002376 | Immune system process |
| | GO:0002682 | Regulation of immune system process |
| Common GO terms between COVID-19 and thyroid cancer | | |
| TC2_GSE3678 | GO:0001775 | Cell activation |
| | GO:0002376 | Immune system process |
| | GO:0006955 | Immune response |
| | GO:0007166 | Cell surface receptor signaling pathway |
| TC2_GSE3678 | GO:0000165 | MAPK cascade |
| | GO:0001568 | Blood vessel development |
| | GO:0001775 | Cell activation |
| | GO:0001932 | Regulation of protein phosphorylation |
| | GO:0001944 | Vasculature development |
| | GO:0002274 | Myeloid leukocyte activation |
| | GO:0002376 | Immune system process |
| TC3_GSE65144 | GO:0001775 | Cell activation |
| | GO:0002376 | Immune system process |
| | GO:0002682 | Regulation of immune system process |
| TC4_GSE85457 | GO:0001568 | Blood vessel development |
| | GO:0001944 | Vasculature development |
| | GO:0002376 | Immune system process |
| | GO:0002682 | Regulation of immune system process |

**Table 7.** Common GO term among COVID-19 (PBMC) and cancers

| GSE ID | GO ID | GO term | GSE ID | GO ID | GO term |
|---|---|---|---|---|---|
| Common GO terms between COVID-19 and breast cancer | | | Common GO terms between COVID-19 and colon cancer | | |
| BC1_GSE95165 | GO:0002376 | Immune system process | CC2_GSE79462 | GO:0007275 | Multicellular organism development |
| | GO:0007275 | Multicellular organism development | | GO:0009653 | Anatomical structure morphogenesis |
| | GO:0032501 | Multicellular organismal process | | GO:0030154 | Cell differentiation |
| | GO:0032502 | Developmental process | | GO:0032501 | Multicellular organismal process |
| BC2_GSE98528 | GO:1901564 | Organonitrogen compound metabolic process | | GO:0032502 | Developmental process |
| | GO:0042221 | Response to chemical | | GO:0042221 | Response to chemical |
| | GO:0030154 | Cell differentiation | | GO:0048513 | Animal organ development |
| | GO:0048869 | Cellular developmental process | CC3_GSE92921 | GO:0001775 | Cell activation |
| | GO:0032502 | Developmental process | | GO:0002376 | Immune system process |
| BC3_GSE107300 | GO:0010033 | Response to organic substance | | GO:0006955 | Immune response |
| | GO:0032501 | Multicellular organismal process | | GO:0007166 | Cell surface receptor signaling pathway |
| | GO:0042221 | Response to chemical | | GO:0007275 | Multicellular organism development |
| BC4_GSE110332 | GO:0002376 | Immune system process | | GO:0009605 | Response to external stimulus |
| | GO:0006955 | Immune response | | GO:0010033 | Response to organic substance |
| | GO:0007166 | Cell surface receptor signaling pathway | CC4_GSE94154 | GO:0002376 | Immune system process |
| | GO:0007275 | Multicellular organism development | | GO:0006955 | Immune response |
| | GO:0009605 | Response to external stimulus | | GO:0007166 | Cell surface receptor signaling pathway |
| | GO:0010033 | Response to organic substance | | GO:0009605 | Response to external stimulus |
| BC5_GSE124646 | GO:0002376 | Immune system process | | GO:0010033 | Response to organic substance |
| | GO:0006955 | Immune response | | GO:0032501 | Multicellular organismal process |
| | GO:0007166 | Cell surface receptor signaling pathway | CC5_GSE110425 | GO:0010033 | Response to organic substance |
| | GO:0007275 | Multicellular organism development | | GO:0042221 | Response to chemical |
| | GO:0009653 | Anatomical structure morphogenesis | CC6_GSE115200 | GO:0007275 | Multicellular organism development |
| | GO:0010033 | Response to organic substance | | GO:0030154 | Cell differentiation |
| | GO:0030154 | Cell differentiation | | GO:0032501 | Multicellular organismal process |
| | GO:0032501 | Multicellular organismal process | | GO:0032502 | Developmental process |
| BC6_GSE125989 | GO:0002376 | Immune system process | | GO:0042221 | Response to chemical |
| | GO:0006928 | Movement of cell or subcellular component | CC7_GSE115716 | GO:0050896 | Response to stimulus |
| | GO:0007166 | Cell surface receptor signaling pathway | | GO:0051716 | Cellular response to stimulus |
| | GO:0007275 | Multicellular organism development | Common GO terms between COVID-19 and kidney cancer | | |
| | GO:0007399 | Nervous system development | KD1_GSE51571 | GO:0006928 | Movement of cell or subcellular component |
| BC7_GSE135427 | GO:0002376 | Immune system process | | GO:0007275 | Multicellular organism development |
| | GO:0007275 | Multicellular organism development | | GO:0009653 | Anatomical structure morphogenesis |
| | GO:0009653 | Anatomical structure morphogenesis | KD3_GSE117890 | GO:0001775 | Cell activation |
| BC8_GSE89333 | GO:0001775 | Cell activation | | GO:0002376 | Immune system process |
| | GO:0002376 | Immune system process | | GO:0006955 | Immune response |
| | GO:0006955 | Immune response | | GO:0007166 | Cell surface receptor signaling pathway |
| | GO:0007166 | Cell surface receptor signaling pathway | | GO:0007275 | Multicellular organism development |

*(Continued)*

**Table 7.** Continued

| GSE ID | GO ID | GO term |
|---|---|---|
| Common GO terms between COVID-19 and breast cancer | | |
| BP1_GSE118123 | GO:0009605 | Response to external stimulus |
| | GO:0010033 | Response to organic substance |
| | GO:0042221 | Response to chemical |
| BP2_GSE122306 | GO:0007275 | Multicellular organism development |
| | GO:0009605 | Response to external stimulus |
| | GO:0010033 | Response to organic substance |
| Common GO terms between COVID-19 and lung cancer | | |
| LC1_GSE63067 | GO:0001775 | Cell activation |
| | GO:0002376 | Immune system process |
| LC2_GSE102079 | GO:0002376 | Immune system process |
| | GO:0006955 | Immune response |
| | GO:0007275 | Multicellular organism development |
| | GO:0009605 | Response to external stimulus |
| Common GO terms between COVID-19 and thyroid cancer | | |
| TC1_GSE3467 | GO:0032501 | Multicellular organismal process |
| TC3_GSE65144 | GO:0001775 | Cell activation |
| | GO:0002376 | Immune system process |
| | GO:0006928 | Movement of cell or subcellular component |
| | GO:0006955 | Immune response |
| | GO:0007166 | Cell surface receptor signaling pathway |
| | GO:0007275 | Multicellular organism development |
| | GO:0010033 | Response to organic substance |
| TC4_GSE85457 | GO:0002376 | Immune system process |
| | GO:0006928 | Movement of cell or subcellular component |
| | GO:0006955 | Immune response |

## Discussion

Bioinformatics is a very important and fast-growing field that can investigate the cause and interaction of various diseases in the medical sciences. The main purpose of this work is to explore the association between COVID-19 and its cancer comorbidities to understand the complexities of cancer patients if they are infected by the SARS-CoV-2. The entire research process relies on the different methods and techniques used for knowledge extraction in bioinformatics. Therefore, we examined the most recent COVID-19 and numerous cancers transcriptomic data in the publicly accessible repositories. In this integrated bioinformatics framework, numerous packages were implemented from the Bioconductor repository using R. GSEA is used to study COVID-19 in terms of the pathways and different ontologies such as GO and DO terms. We also began this test from the set of DEGs and defined GSEA taking into account the most relevant GO terms. In order to show the proximity between different diseases according to chosen ontologies, the usefulness of semantics similarity was again used. Furthermore, GSM documents were noted manually and samples were divided into control and case instead of automatic selection of GEO samples. Then, we created models using manually curated datasets instead of the automatic selection with GEO samples. In order to show the proximity between different diseases according to chosen ontologies, we used semantic similarity approach again. Then, all results containing genes, GO and DO terms were compared to evaluate semantic similarity. There is still no effective method to define the functional similarities based on gene annotation information from dissimilar data sources. Hence, GO terms are effective to address the consistent explanations about genes in different data sources. Instead, DO provides an open source ontology for the incorporation of biomedical data in human disease. It produces a consistent description of gene products with disease perspectives for supporting functional genomics. Several metrics like P-value and logFC thresholds are used in this work. For the P-value of 0.05 and absolute logFC of 1, the variances among sets of DEGs and GO terms are extracted. Consequently, we determined KEGG pathway graph that showed the connectivity of COVID-19 and other diseases. Our analysis identified a number of common dysregulated genes between COVID-19 and cancers. Among the identified common genes, MARCO and OAS2 were identified as dysregulated in breast cancer as consistent with previous report [38, 39]. Previous studies suggested OAS2 as prognostic markers of breast cancer [39]. Another gene, VCAN was identified as a new prognostic gene in gastric cancer [40]. The critical role of ACTB was also found in lung cancer [41]. Overexpression of LGALS1 gene was reported in oral cancer and has been detected as key players for various tumor including prostate, thyroid, bladder and ovarian cancer [42]. Higher expression of HMOX1 gene was revealed in cancer corroborating our findings [43]. TIMP1 was established as anti-apoptotic roles in colon cancer and suggested that it might be critical for cell proliferation, invasion and metastasis of colon cancer [44]. Again, the rest of the identified genes has represented key roles in the development and progression of cancer as consistent with previous findings. In order to shed light on biological pathways commonly altered in COVID-19 and cancer, we identified several pathogenetic processes and molecular pathways that may potentially clarify the potential mechanisms of COVID-19 in cancer patients. Our study highlighted immune system processes and cytokine-mediated inflammations as key BPs of COVID-19 and cancer. The chronic

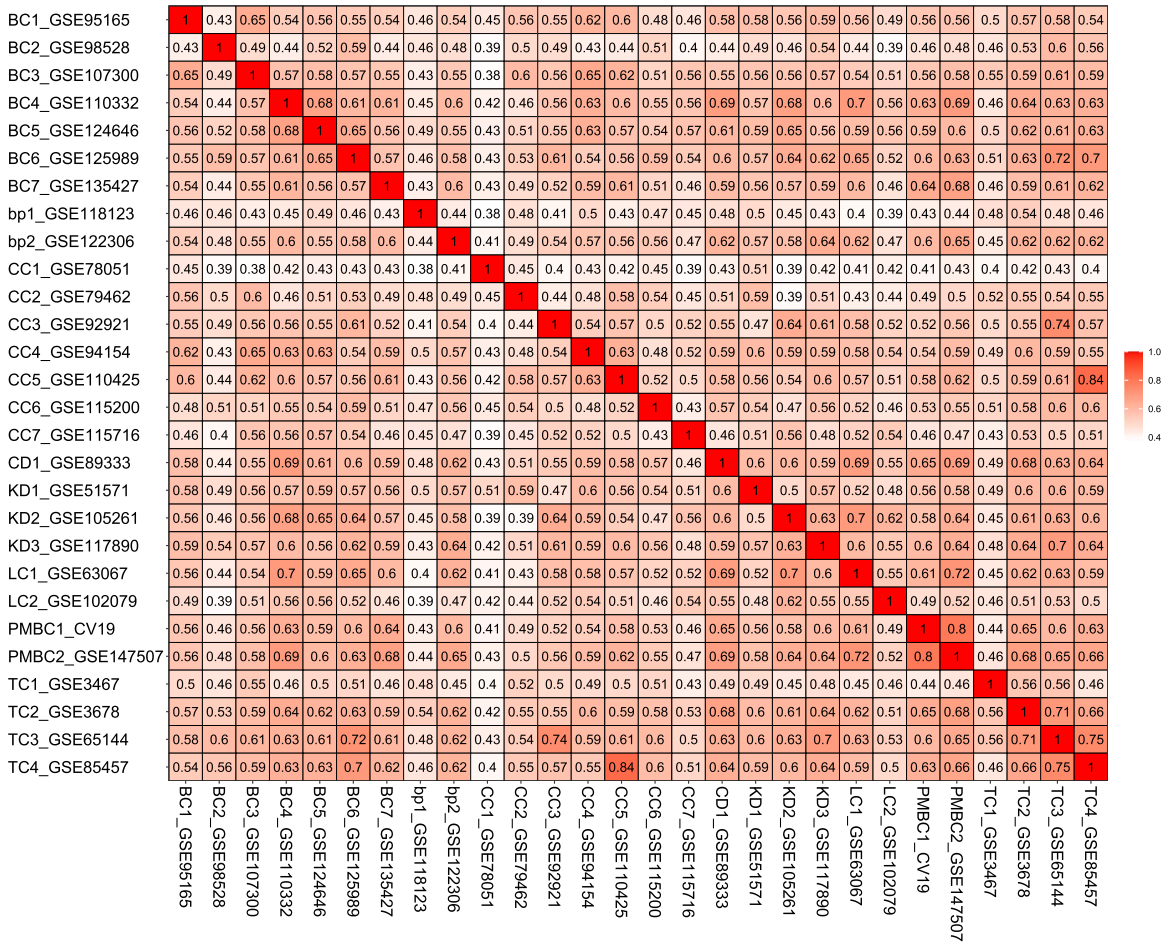


Figure 4. Semantic similarity matrix for differential expressed genes. The two-letter suffix before the GSE codes referred to the following: BC, breast cancer; CC, colon cancer; PMVC1 and PMVC2, COVID-19; KD, kidney cancer; LC, liver cancer; TC, thyroid cancer; and BP, bladder prostate cancer. The number after the two letters indicates the logFC threshold.

inflammation has been recognized as a causative factor for the progression of cancer [45]. Immune systems, cytokines overproduction and cytokine-mediated signaling provided key features in lung inflammation in response to COVID-19 infections [46], which are consistent with our findings. This study identified wnt signaling pathways, IL-17 signaling pathway, TNF signaling pathways as key signaling pathways associated with COVID-19 and cancers which is consistent with previous reports that identified these altered pathways in COVID-19 [46]. Specifically, the "cytokine storm" seen in COVID-19 patients is the result of severe immune response by the host that deteriorate the conditions of the patients [46]. In line with this evidence, we may suggest the dysregulated immune systems play a critical role in COVID-19 patients with cancers. Several previous studies employed whole genome transcriptomic data, identified gene signatures and elucidated immunopathological features and potential marker focused on COVID-19 [46–52], which were consistent with our findings; However, molecular associations between COVID-19 and different cancers have not been found yet. For the first time, we elucidated molecular cell pathways shared between COVID-19 and cancer individually.

It demonstrated the likelihood of reusing the data available from the analytical perspective. For further research, various works related to comorbidities and transcriptomics have been published. However, owing to legal or ethical concerns they are not open to the media at all times. In this study, we represented the datasets with more cell types and resources that investigated robust results than single cells and resources. Several challenges were considered while developing this pipeline. Firstly, it was not only concerned about control versus patients but also scrutinized genetic variants to show the risk of this disease and its variants. Secondly, the standard of data is not similar in all cases. For instance, it took a lot of effort to prepare GEO series data. Therefore the microarray data quality (e.g., the arrayQualityMetrics package) was retained and ideal for semi-automated analysis. However, this approach provides an automated way of gathering, comparing and evaluating microarray data. In this study, we have implemented a comprehensive bioinformatics pipeline where several common pathogenetic processes are detected and shared by COVID-19 and cancer that may aid the clinicians and bench scientists to further dismantle the complex interconnections of the patients. The pipeline can also be used to investigate COVID-19 and other comorbidities, which is freely accessible for clinical researchers to use.

## Conclusion

We have developed an R pipeline that incorporates bioinformatics methods to identify the infectome, diseasome and comorbidities relationship among the infections and diseases. In this

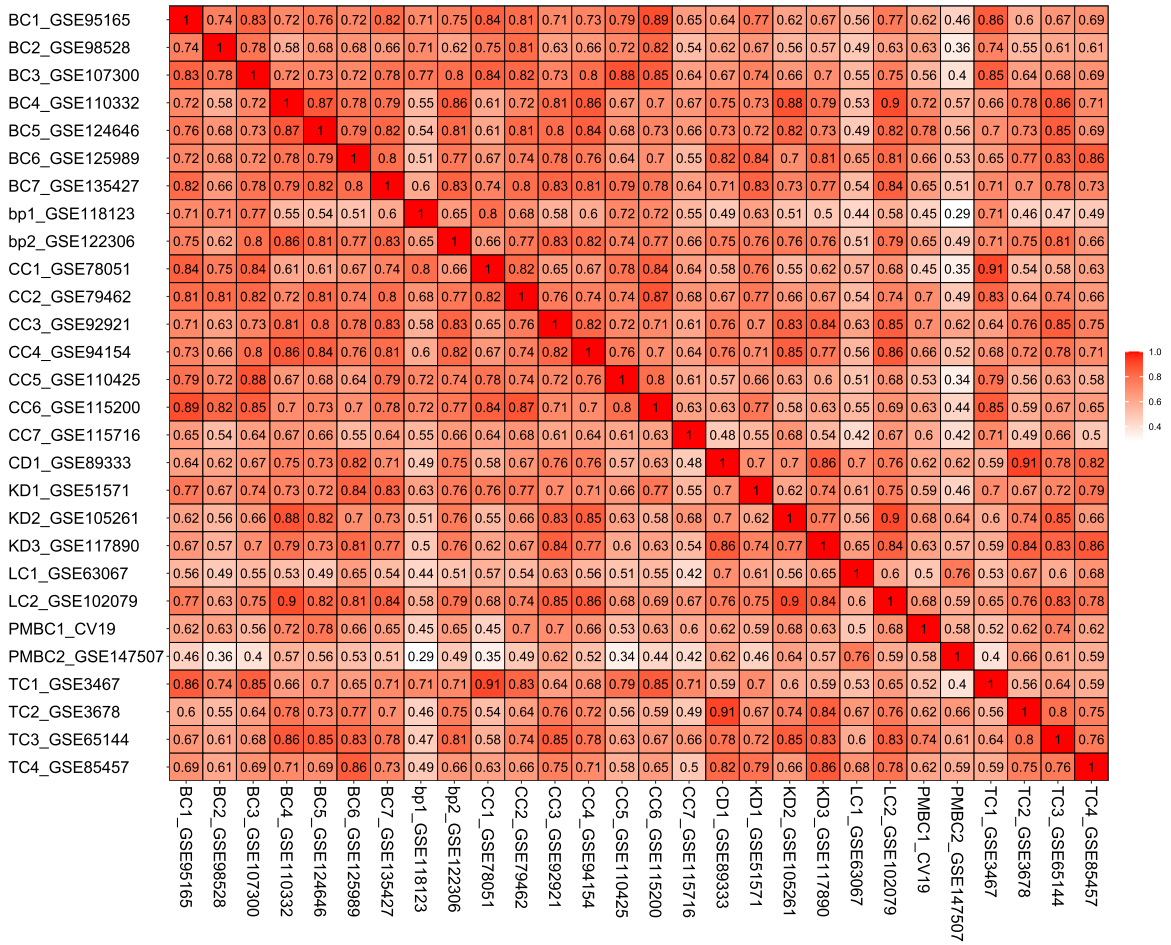


Figure 5. Semantic similarity matrix of GO terms. The number after each pair of entries represent the logFC threshold.

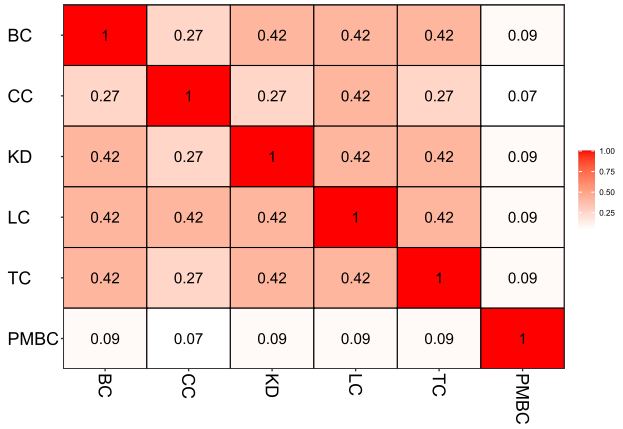


Figure 6. Semantic similarity matrix for DO terms (SARS-CoV)

study, a large set of transcriptomic datasets of COVID-19 and different cancers have been utilized and identified molecular associations between them using our developed pipeline. Our analysis showed common dysregulated genes shared between COVID-19 and cancers. We detected immune systems processes as major dysregulated pathways in COVID-19 and common cancers. Such study is also helpful in evidence-based guidelines on COVID-19 in patients with cancer as our suggested pipeline combines an integrated structure for discovering COVID-19 molecular pathways and various pathologies. Our pipeline can also be used for infectome, diseasome and comorbidities analysis of other diseases by using a large set of transcriptomic data. We are unable to test this technique with further records because of the lack of COVID-19 data, which will be available for the research on COVID-19 by the scientist. We now suggest to incorporate more genome-wide transcriptomic data once it will be available to get more comprehensive understanding of the COVID-19 in cancer comorbid patients. Our pipeline can be an enormous opportunity for clinicians and scientists to provide new insights into COVID-19 pathways in cancer patients despite constraints on the availability of more transcriptomic data.

## Key Points

- This work developed a bioinformatics pipeline and has been applied to detect infectome, diseasome and comorbidities between COVID-19 and cancer diseases.
- Bioinformatics analysis of COVID-19 and its malignant comorbidities are required to evaluate their roles for clinical and further implications of COVID-19.
- Several approaches such as gene set enrichment analysis and semantic similarity are used to investigate


Figure 7. KEGG pathway enrichment analysis for COVID-19 lung tissues.

COVID-19 and its malignant comorbidities in this work.

- Numerous transcriptomic datasets are explored common genes, gene ontology, DO and pathways.

## References


1. S. K. Bandyopadhyay, S. Dutta Machine learning approach for confirmation of covid-19 cases: positive, negative, death and release. medRxiv,2020.
2. Wang Y, Sun J, Zhu A, *et al*. Current understanding of middle east respiratory syndrome coronavirus infection in human and animal models. *J Thorac Dis* 2018; **10**:S2260.
3. A. T. A. Qarawi, S. J. Ng, A. Gad, *et al*.. *Awareness and preparedness of hospital staff against novel coronavirus (COVID-2019): a global survey-study protocol*. 2020.
4. People most at risk of dying from virus, https://www.thechronicle.com.au/news/people-most-at-risk-of-dying-from-virus/3965445/, 2020. (19 May 2020, date last accessed).
5. Del Prete E, Facchiano A, Liò P. Bioinformatics methodologies for coeliac disease and its comorbidities. *Brief Bioinform* 2020; **21**:355–67.
6. Subramanian A, Tamayo P, Mootha VK, *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* 2005; **102**:15545–50.
7. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, *et al*. Imbalanced host response to sars-cov-2 drives development of covid-19. *Cell* 2020.
8. Xiong Y, Liu Y, Cao L, *et al*. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in covid-19 patients. *Emerg Microbes Infect* 2020; **9**:761–70.
9. Reed AEM, Lal S, Kutasovic JR, *et al*. LobSig is a multigene predictor of outcome in invasive lobular carcinoma. *NPJ Breast Cancer* 2019; **5**:1–11.
10. Sethuraman A, Brown M, Krutilina R, *et al*. BHLHE40 confers a pro-survival and pro-metastatic phenotype to breast cancer cells by modulating hbegf secretion. *Breast Cancer Res* 2018; **20**:117.
11. Kannan A, Philley JV, Hertweck KL, *et al*. Cancer testis antigen promotes triple negative breast cancer metastasis and is


Figure 8. KEGG pathway enrichment analysis for COVID-19 blood tissues.

traceable in the circulating extracellular vesicles. *Sci Rep* 2019; **9**:1–12.

12. Sinn BV, Fu C, Lau R, *et al.* SET ER/PR: a robust 18-gene predictor for sensitivity to endocrine therapy for metastatic breast cancer. *NPJ Breast Cancer* 2019; **5**:1–8.
13. Iwamoto T, Niikura N, Ogiya R, *et al.* Distinct gene expression profiles between primary breast cancers and brain metastases from pair-matched samples. *Sci Rep* 2019; **9**: 1–8.
14. S.-H. Kim, Y.-Y. Park, S.-N. Cho, *et al.* Krüppel-like factor 12 promotes colorectal cancer growth through early growth response protein 1. *PLoS One* 2016; 11.
15. K. Gotoh, E. Shinto, Y. Yoshida, *et al.* Prognostic model of stage ii/iii colon cancer constructed using gene expression subtypes and KRAS mutation status. J Clin Exp Oncol 2018;2.
16. N. V. Ciudad CJ, Oleaga C Gene expression after 24h cocoa extract incubation in HT29 cells. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=;GSE94154, 2018. (19 May 2020, date last accessed).
17. Jordheim LP, Chettab K, Cros-Perrial E, *et al.* Unexpected growth-promoting effect of oxaliplatin in excision repair cross-complementation group 1 transfected human colon cancer cells. *Pharmacology* 2018; **102**:161–8.
18. V. Freihen, K. Rönsch, J. Mastroianni, *et al.* SNAIL1 employs $\beta$-catenin-LEF1 complexes to control colorectal cancer cell invasion and proliferation. *Int J Cancer* **146** 2020; 2229–42.
19. Nam H-Y, Chandrashekar DS, Kundu A, *et al.* Integrative epigenetic and gene expression analysis of renal tumor progression to metastasis. *Mol Cancer Res* 2019; **17**:84–96.
20. Lucarelli G, Rutigliano M, Sallustio F, *et al.* Integrated multiomics characterization reveals a distinctive metabolic signature and the role of NDUFA4L2 in promoting angiogenesis, chemoresistance, and mitochondrial dysfunction in clear cell renal cell carcinoma. *Aging (Albany NY)* 2018; **10**:3957.
21. I. Frades, E. Andreasson, J. M. Mato, *et al.* Integrative genomic signatures of hepatocellular carcinoma derived from nonalcoholic fatty liver disease. PLoS One2015;10.
22. Chiyonobu N, Shimada S, Akiyama Y, *et al.* Fatty acid binding protein 4 (FABP4) overexpression in intratumoral hepatic stellate cells within hepatocellular carcinoma with metabolic risk factors. *Am J Pathol* 2018; **188**:1213–24.

23. Elliott B, Millena AC, Matyunina L, *et al*. Essential role of jund in cell proliferation is mediated via myc signaling in prostate cancer cells. *Cancer Lett* 2019; **448**:155–67.
24. di Martino E, Alder O, Hurst CD, *et al*. ETV5 links the FGFR3 and hippo signalling pathways in bladder cancer. *Sci Rep* 2019; **9**:1–12.
25. I. Reyes PTC versus paired normal thyroid tissue, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=;GSE3678. (19 May 2020, date last accessed).
26. Von Roemeling CA, Marlow LA, Pinkerton AB, *et al*. Aberrant lipid metabolism in anaplastic thyroid carcinoma reveals stearoyl coa desaturase 1 as a novel therapeutic target. *J Clin Endocrinol Metabol* 2015; **100**:E697–709.
27. Rahman MH, Peng S, Hu X, *et al*. Bioinformatics methodologies to identify interactions between type 2 diabetes and neurological comorbidities. *IEEE Access* 2019; **7**: 183948–70.
28. Schriml LM, Arze C, Nadendla S, *et al*. Disease ontology: a backbone for disease semantic integration. *Nucleic Acids Res* 2012; **40**:D940–6.
29. C. Pesquita, D. Faria, H. Bastos, *et al*. Metrics for GO based protein semantic similarity: a systematic evaluation. In *BMC Bioinformatics*, vol. 9. Springer, 2008, p. S4.
30. Kanehisa M, Sato Y, Kawashima M, *et al*. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2016; **44**:D457–62.
31. Huber W, Carey VJ, Gentleman R, *et al*. Orchestrating high-throughput genomic analysis with bioconductor. *Nat Methods* 2015; **12**:115.
32. Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and bioconductor. *Bioinformatics* 2007; **23**:1846–7.
33. Ritchie ME, Phipson B, Wu D, *et al*. limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res* 2015; **43**:e47–7.
34. R. Gentleman, V. Carey, W. Huber, F. Hahne. Genefilter: methods for filtering genes from high-throughput experiments. R package version 1, 2015.
35. Yu G, Li F, Qin Y, *et al*. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* 2010; **26**:976–8.
36. Yu G, Wang L-G, Yan G-R, *et al*. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* 2015; **31**:608–9.
37. Yu G, Wang L-G, Han Y, *et al*. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012; **16**:284–7.
38. Georgoudaki A-M, Prokopec KE, Boura VF, *et al*. Reprogramming tumor-associated macrophages by antibody targeting inhibits cancer progression and metastasis. *Cell Rep* 2016; **15**:2000–11.
39. Zhang Y, Yu C. Prognostic characterization of oas1/oas2/oas3/oasl in breast cancer. *BMC Cancer* 2020; **20**:1–12.
40. Binang HB, Wang Y-s, Tewara MA, *et al*. C. Wang. *Vcan–a novel prognostic marker for gastric cancer*, 2019.
41. Walter RFH, Werner R, Vollbrecht C, *et al*. Actb, cdkn1b, gapdh, grb2, rhoa and sdcbp were identified as reference genes in neuroendocrine lung cancer via the ncounter technology. *PLoS One* 2016; **11**:e0165181.
42. Li J-M, Tseng C-W, Lin C-C, *et al*. Upregulation of lgals1 is associated with oral cancer metastasis. *Ther Adv Med Oncol* 2018; **10**:1758835918794622.
43. Bekeschus S, Freund E, Wende K, *et al*. Hmox1 upregulation is a mutual marker in human tumor cells exposed to physical plasma-derived oxidants. *Antioxidants* 2018; **7**:151.
44. Song G, Xu S, Zhang H, *et al*. Timp1 is a prognostic marker for the progression and metastasis of colon cancer through fak-pi3k/akt and mapk pathway. *J Exp Clin Cancer Res* 2016; **35**:1–12.
45. Gonzalez H, Hagerling C, Werb Z. Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes Dev* 2018; **32**:1267–84.
46. Islam T, Rahman MR, Aydin B, *et al*. Integrative transcriptomics analysis of lung epithelial cells and identification of repurposable drug candidates for covid-19. *Eur J Pharmacol* 2020; **173594**.
47. Y. Xiong, Y. Liu, L. Cao, *et al*. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in covid-19 patients, *Emerg Microbes Infect* 2020; **9** 761–70. doi: 10.1080/22221751.2020.1747363. pMID: 32228226.
48. Gardinassi LG, Souza COS, Sales-Campos H, *et al*. Immune and metabolic signatures of covid-19 revealed by transcriptomics data reuse. *Front Immunol* 2020; **11**:1636. doi: 10.3389/fimmu.2020.01636.
49. P. S. Arunachalam, F. Wimmers, C. K. P. Mok, *et al*. Systems biological assessment of immunity to mild versus severe covid-19 infection in humans. *Science* 2020; **369** 1210–1220. doi:10.1126/science.abc6261. arXiv: https://science.sciencemag.org/content/369/6508/1210.full.pdf.
50. Nain, Z., Rana, H. K., Liò, P., *et al*. Pathogenetic profiling of COVID-19 and SARS-like viruses. *Brief Bioinform* 2020. Aug 11. doi: 10.1093/bib/bbaa173.
51. T. A. Taz, K. Ahmed, B. K. Paul, *et al*. Network-based identification genetic effect of SARS-CoV-2 infections to Idiopathic pulmonary fibrosis (IPF) patients. *Brief Bioinform* 2020. https://doi.org/10.1093/bib/bbaa235. doi: 10.1093/bib/bbaa235.
52. M. A. Moni, J. M. W. Quinn, N. Sinmaz, M. A. Summers Gene expression profiling of SARS-CoV-2 infections reveal distinct primary lung cell and systemic immune infection responses that identify pathways relevant in COVID-19 disease. *Brief Bioinform* 2020. doi: 10.1093/bib/bbaa376.