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.

Gene Reports 22 (2021) 100980

Contents lists available at ScienceDirect

Gene Reports

journal homepage: www.elsevier.com/locate/genrep

Gene expression profiling of corona virus microarray datasets to identify crucial targets in COVID-19 patients

Priyanka Ramesh, Shanthi Veerappapillai, Ramanathan Karuppasamy *

Department of Biotechnology, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu, India

ARTICLE INFO

Keywords:
SARS-CoV-2
Microarray data
Hub gene identification
Cytokine storm
STRING database

ABSTRACT

The current outbreak of coronavirus disease (COVID-19) has been affecting millions of people and has caused devastating mortality worldwide. Moreover, it is to be noted that cytokine storm has become an important cause for the rising mortality. However, the efforts for the development of drugs, vaccines and treatment has also been intervened due to poor understanding of host's defense mechanism and also due to the development of cytokine storm against this viral infection. Thus, a deeper understanding of the mechanism behind the immune dysregulation and cytokine storm development might give us clues for the clinical management of the severe cases. Hence, we have implemented differential gene expression analysis together with protein-protein interaction and Gene Ontology (GO) studies with the help of Severe Acute respiratory syndrome coronavirus (SARS-CoV) data sets such as GSE1739 and GSE33267 to give us more knowledge on the host immune response for the pathogenic coronavirus which in turn reduces the mortality. A total of 79 differentially-expressed genes (DEGs) were identified in our data set using the filters such as *P*-value and log2 fold change values of less than 0.05 and 1.5 respectively. Further, network analysis and GO studies showed that differential expression of two hub genes namely ELANE and LTF which could induce higher levels of pro-inflammatory cytokines in the lungs. We are certain that differential expression of ELANE and LTF results in an excessive inflammatory reaction known as the cytokine storm and ultimately leading to death. Therefore, targeting these key drivers of cytokine storm genes appears to be the potential therapeutic targets for combating the Severe Acute respiratory syndrome coronavirus - 2 (SARS-CoV-2) infection ultimately resulting in reduced mortality. Indeed, this predictive view may open new insights for designing an immune intervention for COVID-19 in the near future resulting in the mitigation of mortality rate.

1. Introduction

SARS-CoV-2 emerged from Wuhan, China is the leading cause of COVID-19 primarily affecting the respiratory system of humans. According to the World Health Organisation (WHO) estimates, a total of 9.2 million confirmed cases and 479 thousand deaths were reported worldwide as of June 2020. Moreover, recent literature evidences has been reported that cytokine storm is the main cause of mortality in COVID-19 patients (Ruan et al., 2020). Unfortunately, no vaccine and effective drugs are available for COVID-19 treatment (Huang et al., 2020). Hence, the need for the identification of novel targets and effective therapeutics is of immense importance to overcome this pandemic situation and to reduce the rising mortality rate. Thus, the current investigation focuses on identification of key genes which plays a vital role in cytokine generation using differential gene expression analysis of microarray data.

SARS-CoV-2 belonging to β-coronavirus group expressing the symptoms in patients after 5.2 days onset of infection (Mehta et al., 2020). Patients are experiencing both respiratory and non-respiratory symptoms including kidney failure, acute heart, pneumonia, RNAemia, lung, liver injury (Kakhki et al., 2020; Chen et al., 2020) and thus regarded as an important pathogen causing respiratory diseases including acute respiratory distress syndrome (ARDS) in humans. Angiotensin receptor 2 (ACE2) was identified as the common receptor for both SARS-CoV-2 and SARS-CoV. The interaction between the host receptor ACE2 and the viral S-protein complex results in activation of immune responses as well as virus replication among patients (Junejo et al., 2020). Hence, the drastic release of proinflammatory cytokines was observed in severe COVID-19 patients which in turn hampers the antiviral response developed by the human body (Wu and Yang, 2020).

* Corresponding author.
E-mail address: kramanathan@vit.ac.in (R. Karuppasamy).

https://doi.org/10.1016/j.genrep.2020.100980
Received 18 July 2020; Received in revised form 3 October 2020; Accepted 23 November 2020
Available online 27 November 2020

Subsequently, the rapid release of cytokines attracts the inflammatory cells to the site of infection resulting in enhanced infiltration of inflammatory compounds such as monocytes and neutrophils. This infiltration causes inflammation, tissue damages, and lung organ failure (Channappanavar et al., 2019). Besides, it also resulted in severe pathological changes including alveolar damage, persistent organ dysfunction, and immunopathologic injuries (Ye et al., 2020). Moreover, the proinflammatory cytokines released into the circulatory system caused systematic cytokine storm resulting in multi-organ dysfunction (Wu and Yang, 2020; Ren et al., 2020). All these are evident that respiratory failure due to increased cytokine production is the major cause of death in COVID-19 cases (Mehta et al., 2020). To prevent deterioration of the patient's health with COVID-19, suppressing cytokine storm has become essential. Thus, targeting the genes involved in cytokine storm development mitigates the number of deaths in COVID-19.

Although considerable effort has been directed toward the development of novel therapeutics, major gap in our knowledge is the understanding of the molecular events that precipitate a cytokine storm (Lai et al., 2020) Thus, a key challenge toward assessing and, perhaps, improving the clinical outcome of the treatment of coronavirus patients is to better understand the molecular basis of the disease and its development, i.e., the key changes of gene expression patterns in coronavirus patients (Zhang et al., 2020). Numerous studies using microarray data initiated in recent years to identify hub genes that may assist in the discovery of novel therapeutics for infected patients. For instance, a study by Brahma et al. in 2018 identified hub genes that may assist in the discovery of therapeutics for patients infected with zika virus (Brahma et al., 2018). Further, hub gene identification using differential gene expression analysis of microarray data was also successful to understand the metastasis mechanism of gastric cancer and lung cancer (Yan et al., 2018; El-aarag et al., 2017). In addition, microarray data provides a platform to compare gene expression level of thousand genes simultaneously (Dwivedi, 2018). Moreover, it plays a vital role in disease diagnosis and treatment by predicting the patient's prognosis (Li et al., 2017). By focusing on these advanced strategies, the present study was designed to explore the hub genes between normal and severe COVID-19 patients using a microarray dataset retrieved from the Gene Expression Omnibus (GEO) database. However, in our study, this would not be possible because of the unavailability of microarray data set corresponds to SARS-CoV-2.

Note that genomic analysis indicates that SARS-CoV-2 is in the same beta-CoV clade as SARS-CoV. Of note, SARS-CoV-2 has been observed to share almost 80% of the genome with SARS-CoV (Lu et al., 2020). Therefore, we have characterized the gene expression profiles of SARS-CoV data and identified differentially expressed genes between normal and infected patients using the Bayesian approach in our analysis. These expression patterns were further examined by identifying molecular pathways associated with cytokine development to figure out the novel target in the treatment of coronavirus patients. Ultimately, this investigation provides significant hub genes to be targeted for the clinical management of the severe COVID-19 cases which in turn reduces the mortality rate by mitigating the cytokine storm.

2. Materials and methods

2.1. Collection of microarray data

Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) of National Centre for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) was used for extracting the raw gene expression profiling datasets of Severe Acute respiratory syndrome coronavirus (SARS-CoV). Keywords including "datasets or series", "SARS – CoV", "expression profiling by array" and "homo – sapiens" were used to search for publicly available datasets. Both cell line study and patients samples were considered for our analysis. Additionally, the following filters were implemented to retrieve the SARS - CoV datasets: 1) Study carried out using microarray human genome array platform, 2) Study should both contain control and SARS – CoV infected samples, 3) Each control and infected study have at least three samples, 4) Datasets can have replicates 5) Datasets published in journals. Implementation of these filters resulted in two datasets with GEO accession id: GSE1739 and GSE33267 containing a total of 80 samples consisting of 37 control samples and 43 SARS – CoV infected. Of note, the dataset extracted from accession id GSE1739 contained 4 control samples and 10 SARS – CoV infected samples. Literature evidence highlights that these samples were isolated from peripheral blood mononuclear cells (PBMC) of the patients and the healthy donors using Affymetrix Human HD – Focus Target array strategy. On the other hand, GSE33267 microarray data were retrieved by means of Calu – 3 cell lines using the Whole Genome microarray approach. This dataset consists of 99 samples from 3 different categories like i) control, ii) SARS – CoV infected, iii) SARS – delta ORF6 mutant infected respectively. In the present investigation, the control and SARS – CoV infected samples were considered for gene expression analysis. The platform files correspond to both studies were also obtained from the GEO database to map the gene symbol with a reference ID.

2.2. Data preprocessing and annotation

R studio is an open-source platform for R programming. Hence, it was implemented for performing differential gene expression analysis of SARS – CoV samples. The retrieved datasets were pre-processed and normalized using "affy" package of R programming (Gautier et al., 2004). This process involves background correction using Robust Multichip Average (RMA) approach, quantile normalization, and summarization of data. Subsequently, the gene expression matrix was obtained by mapping the reference ID against the official gene symbol using "annotate" package (Gentleman, 2020).

2.3. Identification of differentially expressed genes

Differentially expressed genes (DEGs) between control and infected samples were identified with the help of Student's *t*-test using limma package of R programming (Ritchie et al., 2015). In this study, multiple adjustment test was used to calculate adjusted *P*-value (adj. P.Value). Note that adj. P.Value less than 0.05 and log2 fold change ($\log_2$ FC) greater than 1.5 were fixed as thresholds for identifying DEGs. Based on the $\log_2$ FC value, the genes were classified into up-regulated and down-regulated genes. For instance, the positive value of $\log_2$ FC was regarded as up-regulated genes and negative values were regarded as down-regulated genes.

2.4. Protein-protein network construction

STRING v11 was used to develop protein-protein interaction among the differentially expressed genes (Szklarczyk et al., 2019). The combined confidence score greater than 0.9 was considered as the threshold for network construction. The network was further filtered by removing the genes with zero interactions. The generated network was imported into cytoscape v3.7.2 to identify important nodes by performing a network-based analysis of connectivity (Lopes et al., 2010). Further, the network was analyzed based on the topological parameters generated using network analyzer v4.4.6 plugin of cytoscape v3.7.2. The parameters such as betweenness centrality, closeness centrality, average shortest path length, neighborhood connectivity, stress centrality distribution, and average clustering co-efficient were analyzed. It is evident from the literature that topological parameters such as degree, betweenness centrality, closeness centrality, and average shortest path length characterizes the pattern of genes in a network system. Thus, these parameters were considered for DEG analysis as it implies biological essentiality in networks (Zhao and Liu, 2019).

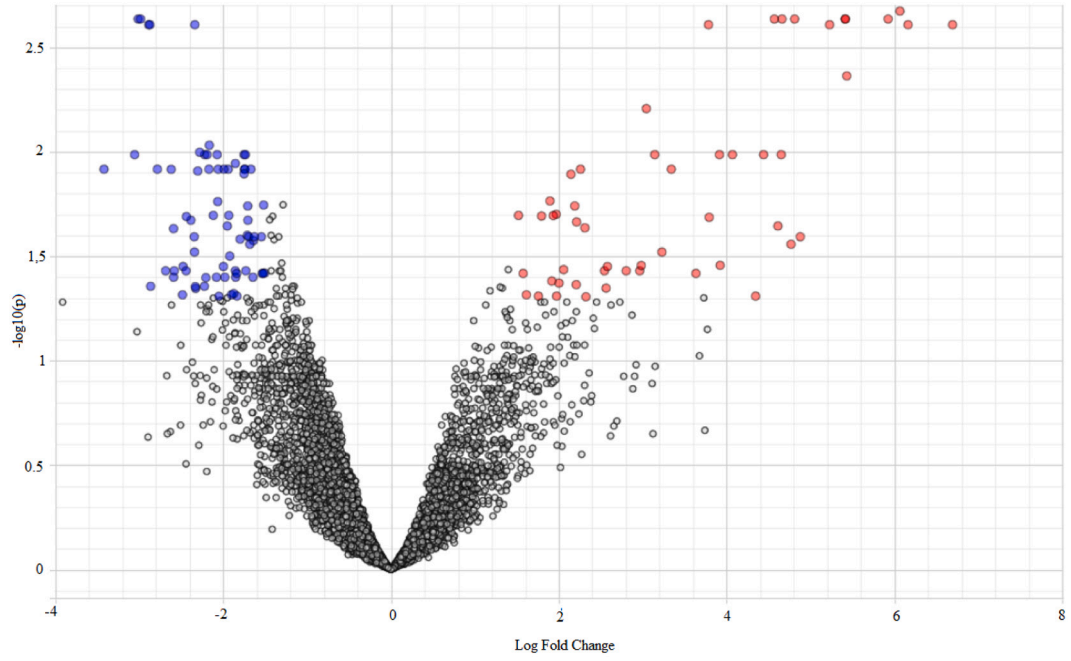
2.5. Functional enrichment analysis

Enrichment analysis of the genes was performed twice during the study. Initially, ShinyGO v0.6 was implemented not only to screen the genes based on the combined confidence score and topology parameters but also to understand the biological process of the differentially expressed genes (Ge et al., 2020). Subsequently, the GOSemSim package of R was implemented to perform Gene Ontology analysis using Wang's strategy for the most significant gene cluster in our sample (Yu et al., 2010). This analysis is of immense importance to understand the hub gene association in the development of immune response against the infection. The graphs and networks were constructed using igraph package of R and cytoscape v3.7.2 respectively (Csardi and Nepusz, 2006; Lopes et al., 2010).

3. Results and discussion

3.1. Identification of differentially expressed genes

The two different microarray datasets GSE1739 and GSE33267 were retrieved from GEO database. Totally 37 control samples and 43 SARS – CoV infected samples were considered for analysis. The preprocessing, quality check and differential gene expression analysis of the microarray dataset was performed using statistical software named R programming. On performing the quality check analysis of the raw data by mapping with the corresponding platform files, 4546 gene expression data were removed from the 27,380 genes. Further, adj. *P*-value and log2 fold change was employed to enhance the power of analyzing the gene expression data. Adj. P-value generated using student's *t*-test assisted in computing the statistical significance of gene in different samples. Moreover log2 fold change was implemented to illustrate the variation



(a)

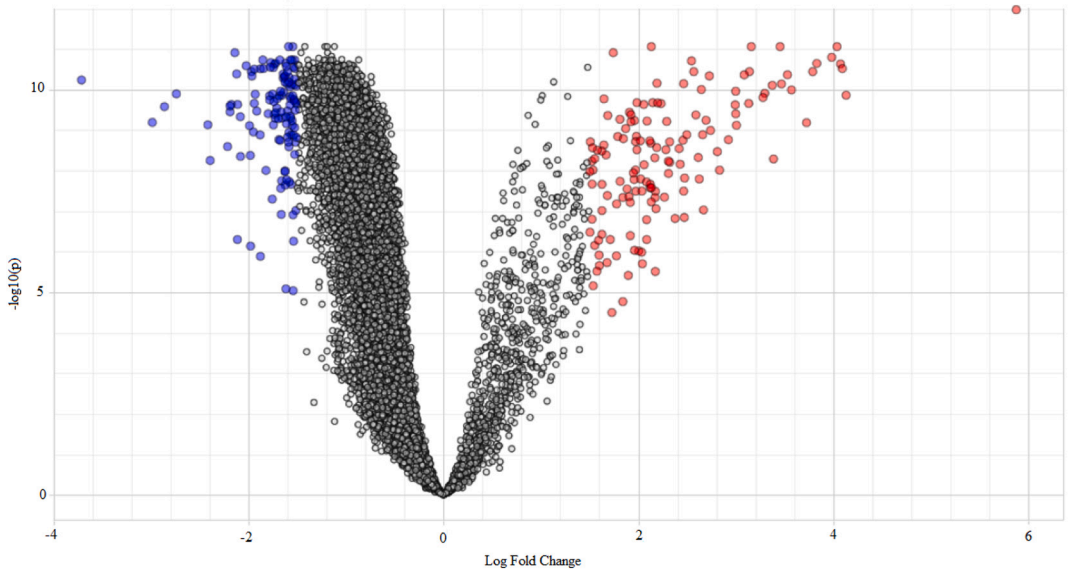


Fig. 1. Volcano plot of differentially expressed genes on analyzing (a) GSE1739 and (b) GSE33267 datasets. Red and blue circles represent up-regulated and down-regulated genes respectively. Black circles represent the genes which are not significantly expressed.

in expression level of genes (Zhao et al., 2018; Dalman et al., 2012). Hence, setting up of adj. P-value cutoff of less than 0.05 resulted in 1350 significant genes from our sample. Further data reduction accomplished with the aid of log2 fold change ($\log_2$ FC = 1.5) resulted in 180 genes that are differentially expressed between normal and infected patients. Among 180 genes, 79 DEGs (42 up-regulated and 37 down-regulated genes) were obtained from GSE1739 dataset and 101 DEGs (84 up-regulated genes and 17 down-regulated genes) were obtained from GSE33267 dataset respectively. The result of DEGs from each of our samples are illustrated using volcano plot (Fig. 1). The red, blue and black circles in Fig. 1 denotes up-regulated, down-regulated and non-significant genes identified during the analysis. The description of the DEGs is also depicted in Table 1.

3.2. Protein-protein network construction among the DEGs

The protein-protein network of DEGs was constructed using cytoscape v3.7.2 based on the STRING v11 database. The network consisted of 48 nodes and 187 edges. Fig. 2 demonstrates the constructed protein-protein interaction network with pink and yellow colored nodes representing up-regulated and down-regulated genes respectively. The edges in the graph denotes the strength of interaction between the two genes. These genes were functionally analyzed using ShinyGO v0.6 to understand its association with the biological processes. Fig. 3 consolidates the functions of all the selected genes. For instance, orange, green and blue color of the bar denotes the genes involvement in biological process, cellualar components and molecular functions respectively. The details of the genes and its associated function were depicted in Table 2. The resultant network was examined based on the biologically significant topological parameters reported earlier. This analysis resulted in ten genes such as neutrophil elastase (*ELANE*), myeloperoxidase (*MPO*), arginase – 1 (*ARG1*), defensin alpha – 4 (*DEFA4*), cathelicidin antimicrobial peptide (*CAMP*), matrix metallopeptidase – 9 (*MMP9*), lactotransfferin (*LTF*), neutrophil gelatinase-associated lipocalin – 2 (*LCN2*), peptidoglycan recognition protein – 1 (*PGLYRP1*) and haptoglobin (*HP*) which were regarded as highly connected nodes (Fig. 4). It is obvious to note that the screened top genes were up-regulated in their expression. In addition, protein with high degree in a biological network provides its significant association with other essential proteins. Literature evidence also support that these genes have direct correlation with the disease prognosis (Ashtiani et al., 2018). Hence, these highly connected genes are considered as hub genes in our analysis. Indeed, these hub genes were more likely to be essential to gain insight into the COVID-19 disease progression. The scores of other topological parameters including betweenness centrality, closeness acentrality and shortest average path length associated with the hub genes are shown in Table 3.

3.3. GO enrichment analysis of hub genes

GO enrichment analysis showed that the resultant hub genes were involved in different functions such as neutrophil mediated immune response, neutrophil degranulation, leukocyte degranulation, neutrophil activation, granulocytes activation, and specific granule lumen. More specifically, in biological processes, these genes are involved in neutrophil-mediated immunity, neutrophil, and granulocyte activation, neutrophil degranulation, myeloid cell activation involved in immune response and myeloid leukocyte mediated immunity. For cellular components, hub genes were enriched in secretory granule, secretory vesicle, cytoplasmic vesicle part, intracellular and extracellular vesicles. Moreover, about molecular function, the hub genes were involved in hydrolase activity, metal ion binding, and cation binding respectively. These enrichment terms provide insights into the understanding of significant roles played by hub genes in COVID-19 disease progression.

Based on the enrichment terms, DEGs were classified into three categories i) Genes involved in both cytokine production and neutrophil activation, ii) Genes involved in cytokine production alone iii) Genes involved in other immune responses. The categories i, ii, and iii contained 4, 3, and 3 genes respectively. In Fig. 5, the red colored bar represent the genes involvement in cytokine production while deep gold colored bar denotes the genes association in other molecular functions. In general, *MPO* and *ELANE* were found to be the primary granules of neutrophils whereas *LTF* and *MMP9* were found to be the secondary and tertiary granules of neutrophils respectively. These granular components combine together to form neutrophil extracellular traps (NET) which in turn is released during immune response with customized production of cytokines (Tetro, 2020). In essence, *LTF* and *ELANE* genes are found to be more significant with the higher gene expression level of about 7 and 5 folds against SARS infection (Fig. 5). In specific, recent literature evidences have also highlighted the significance of ELANE gene for causing epithelial cell injuries by enhanced proinflammatory cytokines in cystic fibrosis patients. This review also provided a strong evidence that inhibition of ELANE gene can reduce the proinflammatory cytokines production resulting in an improvement in pulmonary function (Kelly et al., 2008). Moreover, a review by Kruzel provides an evident that neutrophil mediated LTF production results in development of inflammation (Kruzel et al., 2017). In addition, *LTF* also plays an indirect role in sensing receptors for triggering the production of cytokines and chemokines such as IL-1, IL-6, IL-12, IL-15, TNF-α, INF-α, and β (Wu and Yang, 2020).

On the other hand, *MPO* and *MMP9* (4 folds) were neglected from the studies because of its lower gene expression levels (Fig. 5). Moreover, in category ii and iii, only *CAMP* and *LCN2* gene showed an increased level of gene expression. However, *CAMP* and *LCN2* have a significant role in other biological processes and immune response functions in our system. The other genes *DEFA4*, *HP*, *ARG1,* and *PGLYRP1* were found to be regulated by the cytokines during immune response against the virus (Vandenbroucke et al., 2014; Takahashi et al., 2018; Wang et al., 2001; Pesce et al., 2009). Moreover, their gene expression level is very low compared to the other investigated genes. Therefore, these genes may not be the target of choice for COVID-19 treatment. Overall, on performing differential gene expression analysis, network analysis, gene ontology studies and literature survey, it is clearly evident that ELANE and LTF gene plays significant role in generation of cytokine storm simultaneously resulting death of COVID-19 patients.

It is also worth mentioning that recent research in COVID-19 highlights that increased immune response intervenes in the treatment. For instance, the development of cytokine storm in patients resulted in acute respiratory distress syndrome (ARDS) and ultimately leads to death in many cases (Prompetchara et al., 2020; Shi et al., 2020). Moreover, 38% of 99 patients in Wuhan were reported with increased neutrophils and hence immune evasion in coronavirus treatment has become essential (Li et al., 2020; Abella et al., 2015). Importantly, the above-identified DEGs had similar processes of provoking immune responses and cytokine production against SARS-CoV. Hence targeting *ELANE* and *LTF* certainly prevents the development of cytokine storms which in turn prevents deaths of patients due to ARDS and organ failure.

Table 1
Number of significant genes identified in given datasets.

S. no	GSE ID	Number of genes	No. of up-regulated genes	No. of down-regulated genes
1	GSE1739	79	42	37
2	GSE33267	101	84	17

4. Conclusions

In the present investigation, our objective was to establish a comprehensive gene expression profile study using microarray data that corresponds to the patients with SARS-CoV in order to investigate the underlying mechanisms associated with high mortality of coronavirus. Our analysis strongly suggests that the severity of the disease seems to

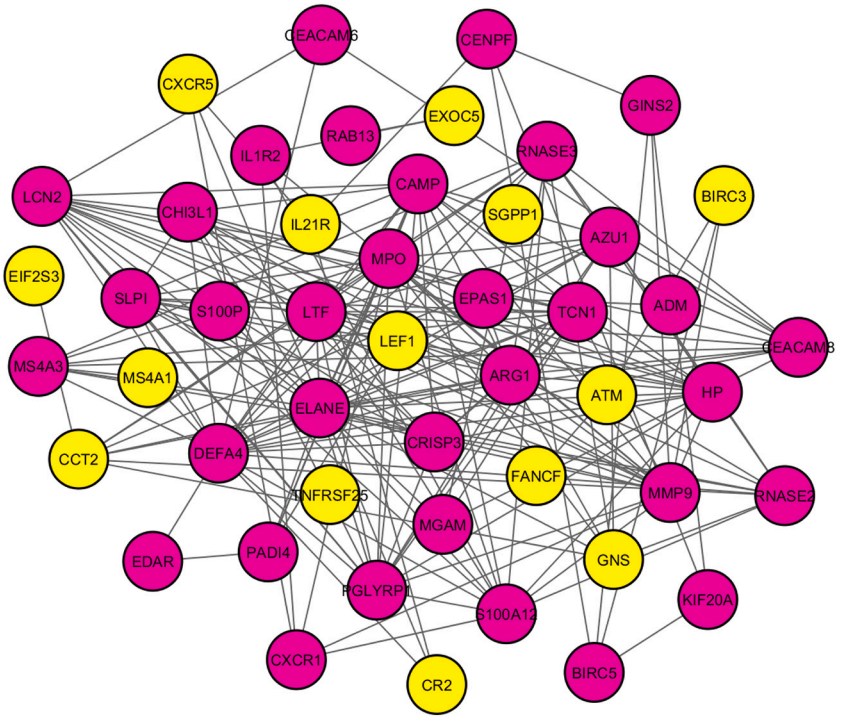


Fig. 2. Protein-protein interaction analysis after verifying using STRING database. Pink color denotes up-regulated genes and yellow color represents down-regulated genes.

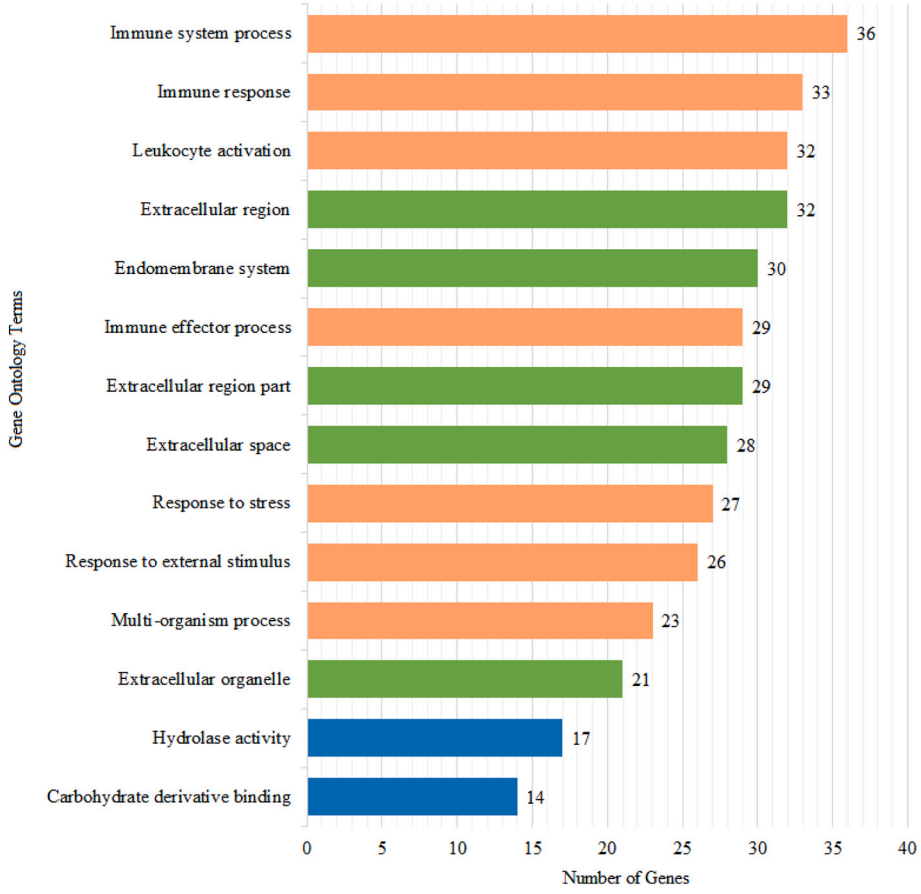


Fig. 3. Gene ontology analysis of differentially expressed genes after screening using ShinyGO database. Orange, green and blue color bar represents biological process, cellular component and molecular function of GO terms.

Table 2
Gene ontology analysis of significant genes

S. no	GO terms	No. of genes	Name of the genes
1	Immune system process	36	PGLYRP1, LEF1, CXCR5, CXCR1, CAMP, DEFA4, ELANE, LTF, CR2, ARG1, SLPI, MS4A1, S100A12, RNASE2, RNASE3, AZU1, MPO, MMP9, EPAS1, LCN2, ATM, PADI4, HP, CRISP3, IL21R, BIRC3, CEACAM6, IL1R2, CEACAM8, CHI3L1, TCN1, GNS, MS4A3, S100P, CCT2, MGAM.
2	Immune response	33	PGLYRP1, CXCR5, CXCR1, CAMP, DEFA4, ELANE, LTF, SLPI, MS4A1, S100A12, RNASE3, CR2, ARG1, LEF1, LCN2, PADI4, AZU1, CRISP3, MPO, BIRC3, CEACAM6, MMP9, IL1R2, CEACAM8, CHI3L1, TCN1, GNS, MS4A3, S100P, CCT2, RNASE2, HP, MGAM.
3	Leukocyte activation	32	PGLYRP1, LEF1, CR2, ARG1, MS4A1, ATM, CXCR5, AZU1, IL21R, MPO, LTF, CEACAM6, CRISP3, MMP9, SLPI, CEACAM8, CHI3L1, TCN1, GNS, LCN2, MS4A3, S100A12, CXCR1, S100P, CAMP, DEFA4, CCT2, RNASE2, RNASE3, ELANE, HP, MGAM.
4	Extracellular region	32	MPO, PGL,YRP1, LTF, CRISP3, MMP9, CR2, SLPI, CEACAM8, EIF2S3, CHI3L1, TCN1, GNS, RAB13, LCN2, MS4A1, S100P, CAMP, CCT2, RNASE2, AZU1, ELANE, HP, MGAM, DEFA4, CEACAM6, KIF20A, ARG1, ADM, RNASE3, IL1R2, S100A12, TNFRSF25.
5	Endomembrane system	30	CAMP, AZU1, MPO, LTF, CRISP3, CENPF, RAB13, DEFA4, ELANE, KIF20A, SGPP1, CHI3L1, MS4A3, PGLYRP1, CEACAM6, MMP9, ARG1, SLPI, CEACAM8, TCN1, GNS, LCN2, S100A12, CXCR1, S100P, CCT2, RNASE2, RNASE3, HP, MGAM.
6	Immune effector process	29	PGLYRP1, RNASE2, AZU1, ELANE, MPO, CR2, ARG1, LEF1, LTF, BIRC3, CEACAM6, CRISP3, MMP9, SLPI, CEACAM8, CHI3L1, TCN1, GNS, LCN2, MS4A3, S100A12, CXCR1, S100P, CAMP, DEFA4, CCT2, RNASE3, HP, MGAM.
7	Extracellular region part	29	MPO, PGLYRP1, LTF, CRISP3, MMP9, CR2, SLPI, CEACAM8, EIF2S3, CHI3L1, TCN1, GNS, RAB13, LCN2, MS4A1, S100P, CAMP, CCT2, RNASE2, AZU1, ELANE, HP, MGAM, DEFA4, CEACAM6, KIF20A, ARG1, ADM, RNASE3.
8	Extracellular space	28	MPO, PGLYRP1, LTF, CRISP3, MMP9, CR2, SLPI, CEACAM8, EIF2S3, CHI3L1, TCN1, GNS, RAB13, LCN2, MS4A1, S100P, CAMP, CCT2, RNASE2, AZU1, ELANE, HP, MGAM, DEFA4, CEACAM6, ARG1, ADM, RNASE3.
9	Response to stress	27	MPO, PGLYRP1, GINS2, EDAR, ATM, CAMP, DEFA4, FANCF, ELANE, HP, LTF, MMP9, EPAS1, SLPI, LCN2, S100A12, RNASE2, RNASE3, AZU1, IL1R2, CR2, ARG1, CHI3L1, ADM, PADI4, CRISP3, BIRC3.
10	Response to external stimulus	26	MPO, PGLYRP1, CXCR5, CXCR1, CAMP, DEFA4, LTF, SLPI, LEF1, LCN2, S100A12, RNASE2, RNASE3, AZU1, ELANE, IL,1R2, ARG1, ADM, ATM, MS4A1, HP, CHI3L1, RAB13, BIRC3, MMP9, CR2.
11		23	

Table 2 (*continued*)

S. no	GO terms	No. of genes	Name of the genes
	Multi-organism process		MPO, PGLYRP1, CAMP, DEFA4, FANCF, LTF, SLPI, LEF1, LCN2, S100A12, RNASE2, RNASE3, AZU1, ELANE, BIRC3, MMP9, CR2, ARG1, ADM, ATM, MS4A1, CCT2, HP.
12	Extracellular organelle	21	MPO, PGLYRP1, LTF, MMP9, CR2, SLPI, CEACAM8, EIF2S3, CHI3L1, GNS, RAB13, LCN2, MS4A1, S100P, CAMP, CCT2, RNASE2, AZU1, ELANE, HP, MGAM.
13	Hydrolase activity	17	MMP9, PGLYRP1, KIF20A, ARG1, GINS2, GNS, RAB13, PADI4, RNASE2, RNASE3, ELANE, HP, AZU1, LTF, SGPP1, EIF2S3, MGAM.
14	Carbohydrate derivative binding	14	PGLYRP1, CHI3L1, GNS, CAMP, MPO, LTF, RAB13, RNASE3, AZU1, ELANE, KIF20A, EIF2S3, ATM, CCT2.

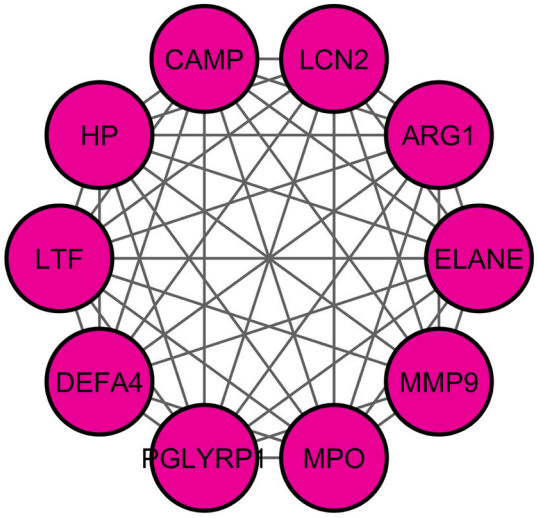


Fig. 4. Top 10 genes with higher degree of interaction.

Table 3
Identification of top 10 genes based on the features of the network.

S. no	Gene name	Degree	Betweenness centrality	Closeness centrality	Average shortest path length
1	ELANE	22	0.105408	0.516484	1.93617
2	MPO	20	0.13815	0.516484	1.93617
3	ARG1	20	0.117982	0.484536	2.06383
4	DEFA4	19	0.028527	0.447619	2.234043
5	CAMP	17	0.046265	0.474747	2.106383
6	MMP9	17	0.319459	0.51087	1.957447
7	LTF	16	0.007201	0.456311	2.191489
8	LCN2	15	0.01955	0.451923	2.212766
9	PGLYRP1	15	0.003853	0.451923	2.212766
10	HP	15	0.003853	0.451923	2.212766

change the genetic profile substantially between healthy and affected patients. In particular, the high disease mortality in some patients has been closely associated with dysregulation of the 3 hub genes such as *CAMP*, *ELANE,* and *LTF*. Note that *CAMP* serves a variety of biological roles such as immune regulation, wound healing, angiogenesis, and anticancer functions. Hence, CAMP cannot be targeted for mitigating cytokine storm. Hence, we hypothesize that only *ELANE* and *LTF* have been shown to induce higher levels of pro-inflammatory cytokines in the lungs, known as the cytokine storm phenomenon. This contributes to its

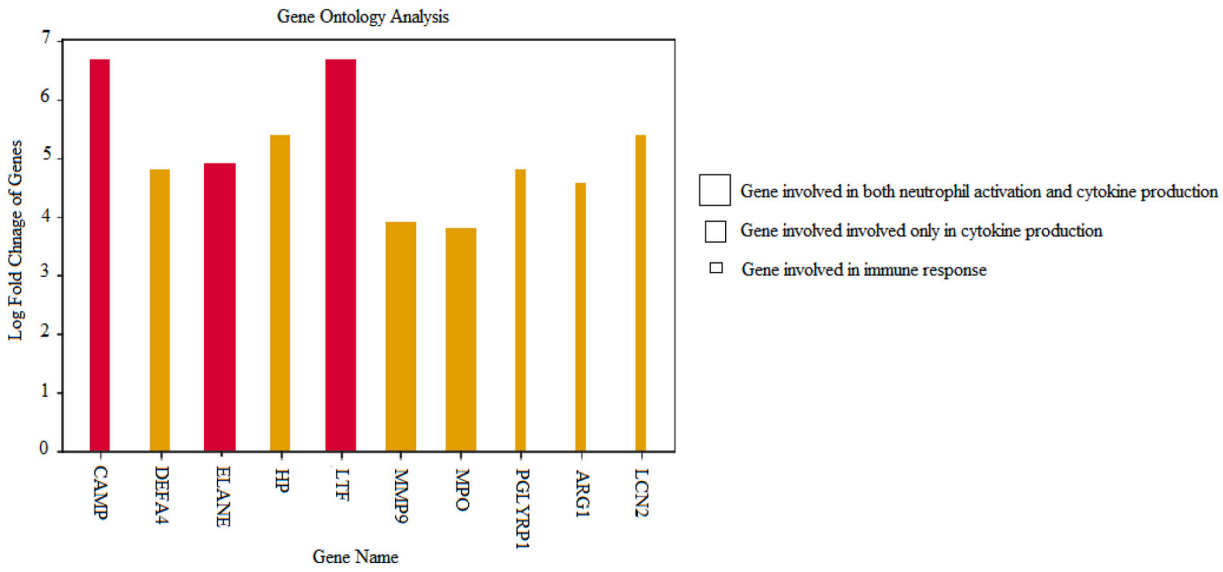


Fig. 5. Gene Ontology analysis of top 10 differentially expressed genes using DAVID database. The width of bar represents the response of genes in neutrophil activation and cytokine production. Red colored genes denotes the top three genes contributing highly in cytokine production. Deep gold color genes denotes genes involved in other functions.

high virulence and may account for the unusually high mortality rate during the outbreak. We conclude from our investigation that the inhibition of *ELANE* and *LTF* directly protects the lung by lowering the neutrophil burden and enhance host defense by protecting locally produced anti-inflammatories in severe SARS-CoV patients. Thus, timely control of the cytokine storm in its early stage through immunomodulators and cytokine antagonists, as well as the reduction of lung inflammatory cell infiltration, is the key to improve the treatment success rate of coronavirus patients.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors thank VIT for providing 'VIT SEED GRANT' for carrying out this research work.

CRediT authorship contribution statement

P.R. and V.S. performed the computational work, wrote the main manuscript, prepared tables, and figures. K.R. supervised the entire study and was involved in the analysis and interpretation of the data. All authors reviewed and approved the manuscript.

Role of funding sources

No funding was received to carry out this work.

References

Abella, V., Scotece, M., Conde, J., Gómez, R., Lois, A., Pino, J., Gómez-Reino, J.J., Lago, F., Mobasheri, A., Gualillo, O., 2015. The potential of lipocalin-2/NGAL as biomarker for inflammatory and metabolic diseases. Biomarkers 20 (8), 565–571.

Ashtiani, M., Salehzadeh-Yazdi, A., Razaghi-Moghadam, Z., Hennig, H., Wolkenhauer, O., Mirzaie, M., Jafari, M., 2018. A systematic survey of centrality measures for protein-protein interaction networks. BMC Syst. Biol. 12 (1), 80.

Brahma, R., Gurumayum, S., Naorem, L.D., Muthaiyan, M., Gopal, J., Venkatesan, A., 2018. Identification of hub genes and pathways in Zika virus infection using RNA-seq data: a network-based computational approach. Viral Immunol. 31 (4), 321–332.

Channappanavar, R., Fehr, A.R., Zheng, J., Wohlford-Lenane, C., Abrahante, J.E., Mack, M., Sompallae, R., McCray, P.B., Meyerholz, D.K., Perlman, S., 2019. IFN-I response timing relative to virus replication determines MERS coronavirus infection outcomes. J. Clin. Invest. 129 (9).

Chen, Y., Liu, Q., Guo, D., 2020. Emerging coronaviruses: genome structure, replication, and pathogenesis. J. Med. Virol. 92 (4), 418–423.

Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. InterJournal, complex systems 1695 (5), 1–9.

Dalman, M.R., Deeter, A., Nimishakavi, G., Duan, Z.H., 2012, December. Fold change and p-value cutoffs significantly alter microarray interpretations. In: BMC Bioinformatics, vol. 13, No. S2. BioMed Central, p. S11.

Dwivedi, A.K., 2018. Artificial neural network model for effective cancer classification using microarray gene expression data. Neural Comput. & Applic. 29 (12), 1545–1554.

El-aarag, S.A., Mahmoud, A., Hashem, M.H., Abd Elkader, H., Hemeida, A.E., ElHefnawi, M., 2017. In silico identification of potential key regulatory factors in smoking-induced lung cancer. BMC Med. Genet. 10 (1), 40.

Gautier, L., Cope, L., Bolstad, B.M., Irizarry, R.A., 2004. Affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20 (3), 307–315.

Ge, S.X., Jung, D., Yao, R., 2020. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics 36 (8), 2628–2629.

Gentleman, R., 2020. Annotate: Annotation for Microarrays. *R package* Version 1.66.0.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395 (10223), 497–506.

Junejo, Y., Ozaslan, M., Safdar, M., Khailany, R.A., Rehman, S., Yousaf, W. and Khan, M. A., 2020. Novel SARS-CoV-2/COVID-19: origin, pathogenesis, genes and genetic variations, immune responses and phylogenetic analysis. Gene reports, 20, pp.100752.

Kakhki, R.K., Kakhki, M.K., Neshani, A., 2020. COVID-19 target: a specific target for novel coronavirus detection. Gene Reports 20, 100740.

Kelly, E., Greene, C.M., McElvaney, N.G., 2008. Targeting neutrophil elastase in cystic fibrosis. Expert Opin. Ther. Targets 12 (2), 145–157.

Kruzel, M.L., Zimecki, M., Actor, J.K., 2017. Lactoferrin in a context of inflammation-induced pathology. Front. Immunol. 8, 1438.

Lai, C.C., Shih, T.P., Ko, W.C., Tang, H.J., Hsueh, P.R., 2020. 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 55 (3), 105924.

Li, G., Li, X., Yang, M., Xu, L., Deng, S. and Ran, L., 2017. Prediction of biomarkers of oral squamous cell carcinoma using microarray technology. Sci. Rep., 7, pp.42105.

Li, X., Geng, M., Peng, Y., Meng, L. and Lu, S., 2020. Molecular immune pathogenesis and diagnosis of COVID-19. Journal of Pharmaceutical Analysis.

Lopes, C.T., Franz, M., Kazi, F., Donaldson, S.L., Morris, Q., Bader, G.D., 2010. Cytoscape Web: an interactive web-based network browser. Bioinformatics 26 (18), 2347–2348.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395 (10224), 565–574.

Mehta, P., McAuley, D.F., Brown, M., Sanchez, E., Tattersall, R.S., Manson, J.J., HLH Across Speciality Collaboration, 2020. COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet (London, England) 395 (10229), 1033.

Pesce, J.T., Ramalingam, T.R., Mentink-Kane, M.M., Wilson, M.S., El Kasmi, K.C., Smith, A.M., Thompson, R.W., Cheever, A.W., Murray, P.J., Wynn, T.A., 2009.

Arginase-1–expressing macrophages suppress Th2 cytokine–driven inflammation and fibrosis. PLoS Pathog. 5 (4), e1000371.
Prompetchara, E., Ketloy, C., Palaga, T., 2020. Immune responses in COVID-19 and potential vaccines: lessons learned from SARS and MERS epidemic. Asian Pac. J. Allergy Immunol. 38 (1), 1–9.
Ren, L.L., Wang, Y.M., Wu, Z.Q., Xiang, Z.C., Guo, L., Xu, T., Jiang, Y.Z., Xiong, Y., Li, Y. J., Li, X.W., Li, H., 2020. Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. Chin. Med. J. 133 (9), 1015–1024.
Ritchie, M.E., Phipson, B., Wu, D.I., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43 (7), e47.
Ruan, Q., Yang, K., Wang, W., Jiang, L., Song, J., 2020. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Intensive Care Med. 46 (5), 846–848.
Shi, Y., Wang, Y., Shao, C., Huang, J., Gan, J., Huang, X., Bucci, E., Piacentini, M., Ippolito, G., Melino, G., 2020. COVID-19 Infection: The Perspectives on Immune Responses.
Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., 2019. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 47 (D1), D607–D613.
Takahashi, T., Kulkarni, N.N., Lee, E.Y., Zhang, L.J., Wong, G.C., Gallo, R.L., 2018. Cathelicidin promotes inflammation by enabling binding of self-RNA to cell surface scavenger receptors. Sci. Rep. 8 (1), 1–13.
Tetro, J.A., 2020. Is COVID-19 receiving ADE from other coronaviruses? Microbes Infect. 22 (2), 72–73.
Vandenbroucke, R.E., Vanlaere, I., Van Hauwermeiren, F., Van Wonterghem, E., Wilson, C., Libert, C., 2014. Pro-inflammatory effects of matrix metalloproteinase 7 in acute inflammation. Mucosal Immunol. 7 (3), 579–588.
Wang, Y., Kinzie, E., Berger, F.G., Lim, S.K., Baumann, H., 2001. Haptoglobin, an inflammation-inducible plasma protein. Redox Rep. 6 (6), 379–385.
Wu, D., Yang, X.O., 2020. TH17 responses in cytokine storm of COVID-19: an emerging target of JAK2 inhibitor Fedratinib. J. Microbiol. Immunol. Infect. 53 (3), 368–370.
Yan, P., He, Y., Xie, K., Kong, S., Zhao, W., 2018. In silico analyses for potential key genes associated with gastric cancer. PeerJ 6, e6092.
Ye, Q., Wang, B. and Mao, J., 2020. Cytokine storm in COVID-19 and treatment. J. Infect., 80(6), pp. 607–613.
Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., Wang, S., 2010. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics 26 (7), 976–978.
Zhang, W., Zhao, Y., Zhang, F., Wang, Q., Li, T., Liu, Z., Wang, J., Qin, Y., Zhang, X., Yan, X., Zeng, X., 2020. The use of anti-inflammatory drugs in the treatment of people with severe coronavirus disease 2019 (COVID-19): the experience of clinical immunologists from China. Clin. Immunol. 214, 108393.
Zhao, X., Liu, Z.P., 2019. Analysis of topological parameters of complex disease genes reveals the importance of location in a biomolecular network. Genes 10 (2), 143.
Zhao, B., Erwin, A., Xue, B., 2018. How many differentially expressed genes: a perspective from the comparison of genotypic and phenotypic distances. Genomics 110 (1), 67–73.