



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

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



A network-based systems biology approach for identification of shared Gene signatures between male and female in COVID-19 datasets

Md Shahjaman^{a,*}, Md Rezanur Rahman^b, Md Rabiul Auwul^a

^a Department of Statistics, Begum Rokeya University, Rangpur, 5400, Bangladesh

^b Department of Biotechnology and Genetic Engineering, Islamic University, Kushtia, Bangladesh

ARTICLE INFO

Keywords:

Coronavirus
SARS-CoV-2
COVID-19
Sex-specific biomarkers
Robust voom
Hub-proteins

ABSTRACT

The novel coronavirus (SARS-CoV-2) has expanded rapidly worldwide. Now it has covered more than 150 countries worldwide. It is referred to as COVID-19. SARS-CoV-2 mainly affects the respiratory systems of humans that can lead up to serious illness or even death in the presence of different comorbidities. However, most COVID-19 infected people show mild to moderate symptoms, and no medication is suggested. Still, drugs of other diseases have been used to treat COVID-19. Nevertheless, the absence of vaccines and proper drugs against the COVID-19 virus has increased the mortality rate. Albeit sex is a risk factor for COVID-19, none of the studies considered this risk factor for identifying biomarkers from the RNASeq count dataset. Men are more likely to undertake severe symptoms with different comorbidities and show greater mortality compared with women. From this standpoint, we aim to identify shared gene signatures between males and females from the human COVID-19 RNASeq count dataset of peripheral blood cells using a robust voom approach. We identified 1341 overlapping DEGs between male and female datasets. The gene ontology (GO) annotation and pathway enrichment analysis revealed that DEGs are involved in various BP categories such as nucleosome assembly, DNA conformation change, DNA packaging, and different KEGG pathways such as cell cycle, ECM-receptor interaction, progesterone-mediated oocyte maturation, etc. Ten hub-proteins (UBC, KIAA0101, APP, CDK1, SUMO2, SP1, FN1, CDK2, E2F1, and TP53) were unveiled using PPI network analysis. The top three miRNAs (mir-17-5p, mir-20a-5p, mir-93-5p) and TFs (PPARG, E2F1 and KLF5) were uncovered. In conclusion, the top ten significant drugs (roscovitine, curcumin, simvastatin, fulvestrant, troglitazone, alvocidib, L-alanine, tamoxifen, serine, and doxorubicin) were retrieved using drug repurposing analysis of overlapping DEGs, which might be therapeutic agents of COVID-19.

1. Introduction

Coronaviruses (CoVs) belong to the Coronaviridae family, one of eight families whose members infect humans and vertebrates. CoVs made up of single-stranded RNA. The upper respiratory tract is the main region of humans infected by the CoVs [1–3]. However, some other regions, such as the gastrointestinal, hepatic, and central nervous systems of humans, can also be infected by the CoVs. In 2002–2003 severe acute respiratory syndromes associated with coronavirus (SARS-CoV-1) was emerged in China and spread to the other four countries. SARS-CoV-1 infected around 8000 cases with a case-fatality ratio (CFR) of 11% [4]. The Middle East respiratory syndrome-associated coronavirus (MERS-CoV) is another type of CoVs emerged in 2012. The number of reported deaths was 858 out of 2494 infected cases by MERS-CoV

with a higher CFR, 34% [5]. The novel coronavirus-2019 (COVID-19), also known as SARS-CoV-2, has been declared as a pandemic by the world health organization (WHO) [6,7]. It has given global challenges and threats to the whole human with an enormous loss of lives worldwide [8]. It first appeared in Wuhan, China, in December 2019 [4]. There are several variants of SARS-CoV-2 such as B.1.1.7 (alpha), B.1.351 (beta), P.1 (gamma), B.1.427 (epsilon) and B.1.617.2 (delta) [9, 10]. The alpha variant was the first outbreak in the United Kingdom (UK) in November 2020. The beta variant was first detected in South Africa in October 2020. Gamma variant, also known as Brazilian variant, was first detected in January 2021. The delta variant of SARS-CoV-2 was detected in late 2020 in India. Alpha variant and delta variant are both more transmissible than the original virus identified in China. Most people infected with the COVID-19 show mild to moderate respiratory

* Corresponding author.

E-mail address: shahjaman_brur@yahoo.com (M. Shahjaman).

<https://doi.org/10.1016/j.imu.2021.100702>

Received 15 April 2021; Received in revised form 12 August 2021; Accepted 13 August 2021

Available online 18 August 2021

2352-9148/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

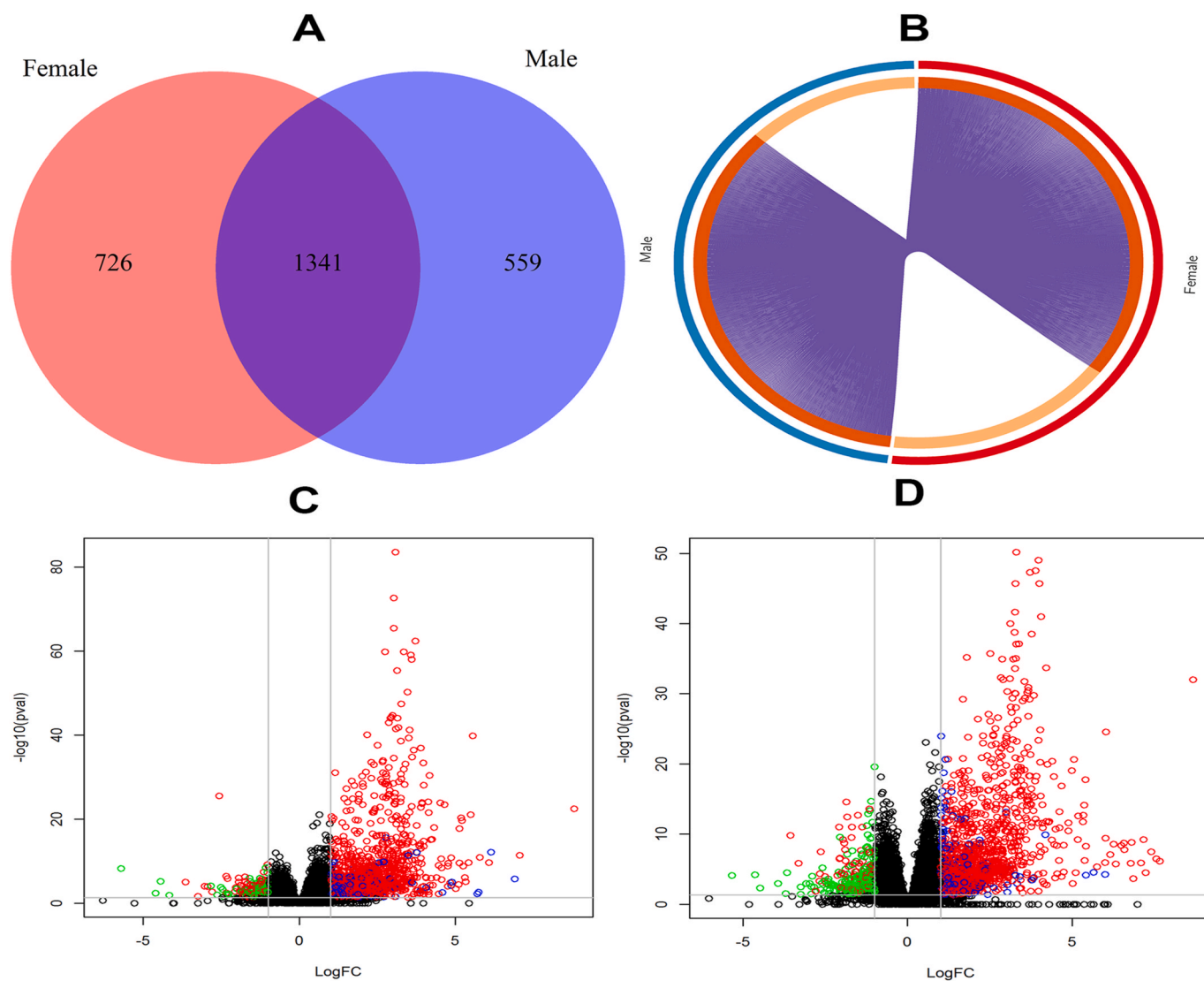


Fig. 1. Differentially expressed genes identification profiles. (A) Venn diagram of DEGs identified by robust voom approach from male and female dataset of COVID-19, (B) circos plot at gene level of overlapping 1341 DEGs between male and female, (C) volcano plot of COVID-19 male dataset, (D) volcano plot of COVID-19 female dataset.

illness (like cold, fever, and cough), and no special treatment is required [11–13]. Older adults and people with different comorbid diseases such as cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to experience serious respiratory illness and they may need hospitalization, intensive care, or a ventilator to support them breath, or people even die [14,15]. Therefore, candidate drugs and vaccines of COVID-19 are urgently needed. The traditional de novo drug discovery procedure is expensive and requires a long time. Drug repurposing is another way to explore the candidate drugs amid the existing drugs using the bioinformatics and integrative systems biology approach, which could shorten the time and expense compared to the traditional procedure. However, the identification of biomarkers is so important for further downstream analysis like drug discovery. Also, it is very challenging because there is a large number of a gene relative to the small number of samples. In general, biomarkers are correlated and sometimes show the same activities in the complex regulatory networks and pathways. So it is necessary to understand the underlying mechanism and functions of biomarkers [16–18].

Demographic variables such as age and sex are the main risk factors of COVID-19 disease. Men are more likely to undergo severe symptoms with different comorbidities and show more significant mortality than

women [19,20]. One of the main reasons is that men are more likely to participate in smoking and drinking alcohol [21]. Other reasons are chromosomal factors (sex-specific hormones and steroids) and gender-specific factors (behaviors and social activities). Recently, Blanco-Melo et al. revealed transcriptional signatures and pathways of SARS-CoV-2 by identifying differentially expressed genes (DEGs) from the RNA sequencing (RNA-Seq) data [22]. Previous studies also examined DEGs and molecular gene ontology and pathway analysis using lung epithelial cells [8]. The premeditated HIV, Ebola, and malaria drugs have been tested to prevent COVID-19 [23,24]. However, the absence of vaccines and proper drugs against the COVID-19 has increased the mortality rate worldwide. Therefore, common biomarkers between males and females may play an important role in discovering drugs against the COVID-19. No specific studies were performed to identify biomarkers from gene expression levels by considering the sex differences using a robust approach. Hence, in this paper, we aim to identify shared gene signatures between males and females from the human RNAseq dataset in blood. To conduct gene ontology (GO) and pathway enrichment analysis, the overlapped differentially expressed genes (DEGs) or biomarkers between males and females were employed to conduct gene ontology (GO) and pathway enrichment analysis. These

Table 1

Gene Ontology (GO) enrichment analysis using 1341 overlapping DEGs between male and female dataset.

GO of Biological Process (BP)	No. of Gene	Adjust.p-value
nucleosome assembly	35	2.49E-13
DNA conformation change	50	4.58E-13
DNA packaging	41	4.58E-13
mitotic nuclear division	45	8.59E-13
chromatin assembly	35	3.30E-12
protein-DNA complex assembly	40	2.14E-11
nucleosome organization	35	2.70E-11
chromosome segregation	46	2.84E-11
chromatin assembly or disassembly	36	4.23E-11
sister chromatid segregation	33	1.34E-10
GO of Cellular Component (CC)	No. of Gene	Adjust.p-value
DNA packaging complex	37	2.95E-19
nucleosome	35	1.41E-18
chromosomal region	59	1.02E-16
condensed chromosome		1.39E-15
kinetochore	32	4.28E-13
condensed chromosome kinetochore	28	4.28E-13
chromosome		6.14E-15
protein-DNA complex	38	5.41E-13
condensed chromosome	38	3.90E-12
collagen-containing extracellular matrix	52	7.17E-10
GO of Molecular Function (MF)	No. of Gene	Adjust.p-value
protein heterodimerization activity	56	3.07E-06
heparin binding	21	0.002536
sulfur compound binding	27	0.002536
extracellular matrix structural constituent	21	0.005097
glycosaminoglycan binding	24	0.005741
icosanoid receptor activity	5	0.015201
cGMP binding	5	0.015201
peptidase regulator activity	22	0.015201
extracellular matrix binding	9	0.015201
amyloid-beta binding	11	0.043064

Table 2

Top ten KEGG pathways using 1341 overlapping DEGs between male and female dataset.

KEGG pathway	No. of Gene	Adjust. p-value
Cell cycle	29	1.58E-07
ECM-receptor interaction	21	6.57E-06
Progesterone-mediated oocyte maturation	20	7.16E-05
Oocyte meiosis	23	9.88E-05
Systemic lupus erythematosus	22	0.000135
Alcoholism	25	0.00017
Dilated cardiomyopathy	16	0.006848
Focal adhesion	27	0.007077
Platelet activation	17	0.033341
Arrhythmogenic right ventricular cardiomyopathy	12	0.049718
Regulation of lipolysis in adipocytes	10	0.049718

analyses revealed that the mutual DEGs are involved in various BP categories such as nucleosome assembly, DNA conformation change, DNA packaging, and different KEGG pathways such as cell cycle, ECM-receptor interaction, progesterone-mediated oocyte maturation, etc. Finally, the ten hub genes (UBC, KIAA0101, APP, CDK1, SUMO2, SP1, FN1, CDK2, E2F1, and TP53) were revealed from protein-protein interaction (PPI) network analysis and underwent in the online databases to explore the candidate drugs of COVID-19.

2. Materials and methods

2.1. Data acquisition and identification of differentially expressed genes from peripheral blood cells of SARS-CoV-2

The RNA-Seq dataset of COVID-19 was retrieved from Gene Expression Omnibus (GEO) [25] with the accession number GSE152418

under the platform GPL24676 [26]. This dataset comprises 34 samples. Among them, 17 samples came from peripheral blood cells with SARS-CoV-2 infected patients, and 17 samples came from peripheral blood cells of healthy control people. There is an outlier sample in peripheral blood cells with SARS-CoV-2; therefore, we have discarded this sample from this dataset [27]. Among the 16 samples of SARS-CoV-2 infected patients, 7 came from males, and 9 came from female patients. For distinct identification of DEGs between males and females, firstly, we divided the whole dataset into two independent datasets. One dataset consists of 7 male SARS-CoV-2 infected patients, and the other datasets consist of 9 female SARS-CoV-2 infected patients, where the number of healthy control people was the same in both datasets (17). However, information on variants of SARS-CoV-2 infected patients is unavailable in the original publication that provided this dataset [26]. For robust identification of DEGs from both male and female datasets, we employed a robust voom approach [28]. The DEGs were identified using adjusted p-value <0.05 and absolute log2 fold change (FC) ≥ 1 [29]. The overlapped DEGs between male and female dataset was used for further downstream analyses.

2.2. Gene ontology and pathway enrichment analysis

To decode identified DEGs' biological functions and pathways, we used Database for Annotation, Visualization and Integrated Discovery (DAVID) and Metascape [30,31]. Different gene ontology (GO) categories such as biological process (BP), cellular component (CC), molecular functions (MF), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were mined using this database. Adjusted p-value <0.05 of the hypergeometric test was used to declare significant categories. We also considered a minimum number of 2 genes in each category as the cut-off.

2.3. Protein-protein interaction analysis and hub-protein identification

Protein-protein interaction (PPI) has been carried out using the online tool NetworkAnalyst [32]. STRING was used to construct the PPI. The hub-proteins were extracted with a high confidence level (900) of PPI and based on degrees >30. The hub-protein were then visualized using the GeneMANIA web-server [33].

2.4. DEGs-miRNA network analysis to identify the potential micro RNAs

The DEGs-miRNA interaction analysis was performed using the miRTarBase database and visualize via NetworkAnalyst [32]. This database collected fifty thousand miRNA-target interactions. To determine the significant core-miRNAs, we considered the degree of interaction >30.

2.5. DEGs-transcription factor regulatory network analysis

The DEGs-TFs interaction analysis was conducted using the JASPAR database and visualize via NetworkAnalyst. JASPAR is a collection of large curated and non-redundant DNA binding TFs respiratory [34]. We retrieved hub-TFs with a cutoff value of degree >40.

2.6. Hub proteins specific drug repositioning

The Drug Signatures Database (DSigDB) [35] and DrugMatrix [36] were used via Enrichr [37] for the identification of potential drug candidates using hub proteins. A collection of 22,527 gene sets comprises 17,389 distinct compounds covering 19,531 genes contain in DSigDB.

2.7. Sensitivity and specificity analysis of the hub proteins

Sensitivity and specificity analysis has been performed by different classification methods using the R package *MLSeq* and *caret*. This

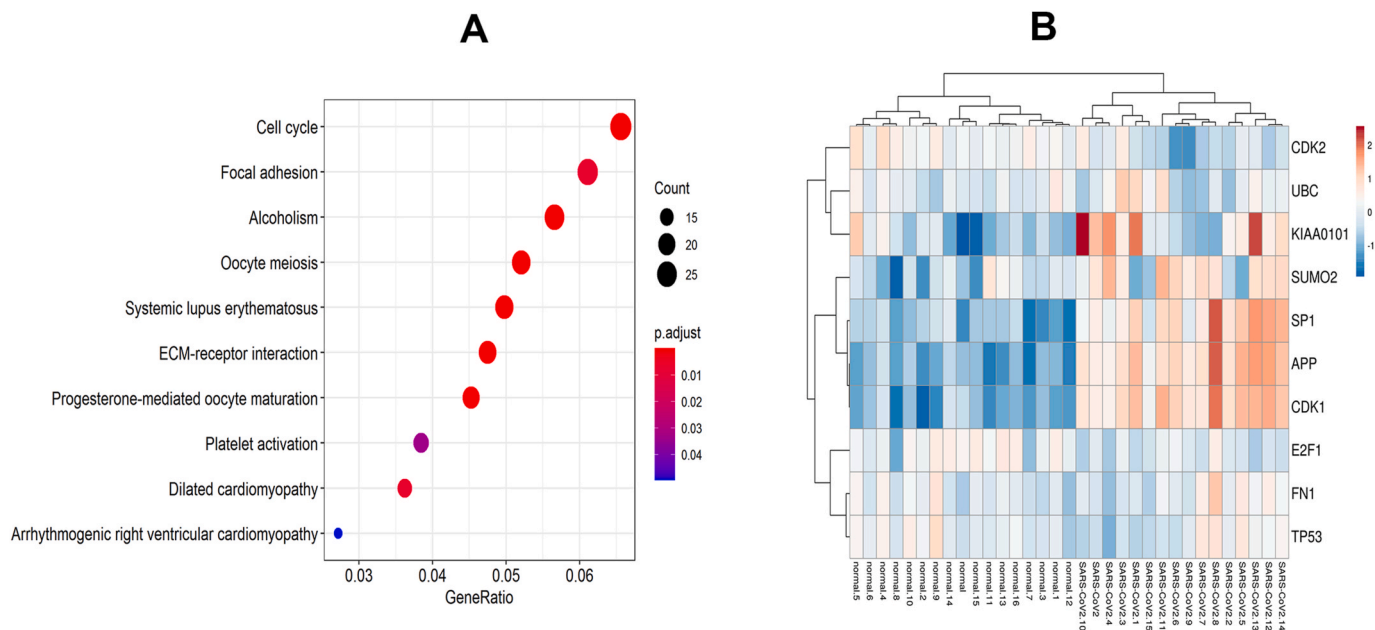


Fig. 2. Heatmap and pathway enrichment analysis of DEGs. (A) KEGG pathway enrichment analysis, (B) heatmap of 10 hub-proteins.

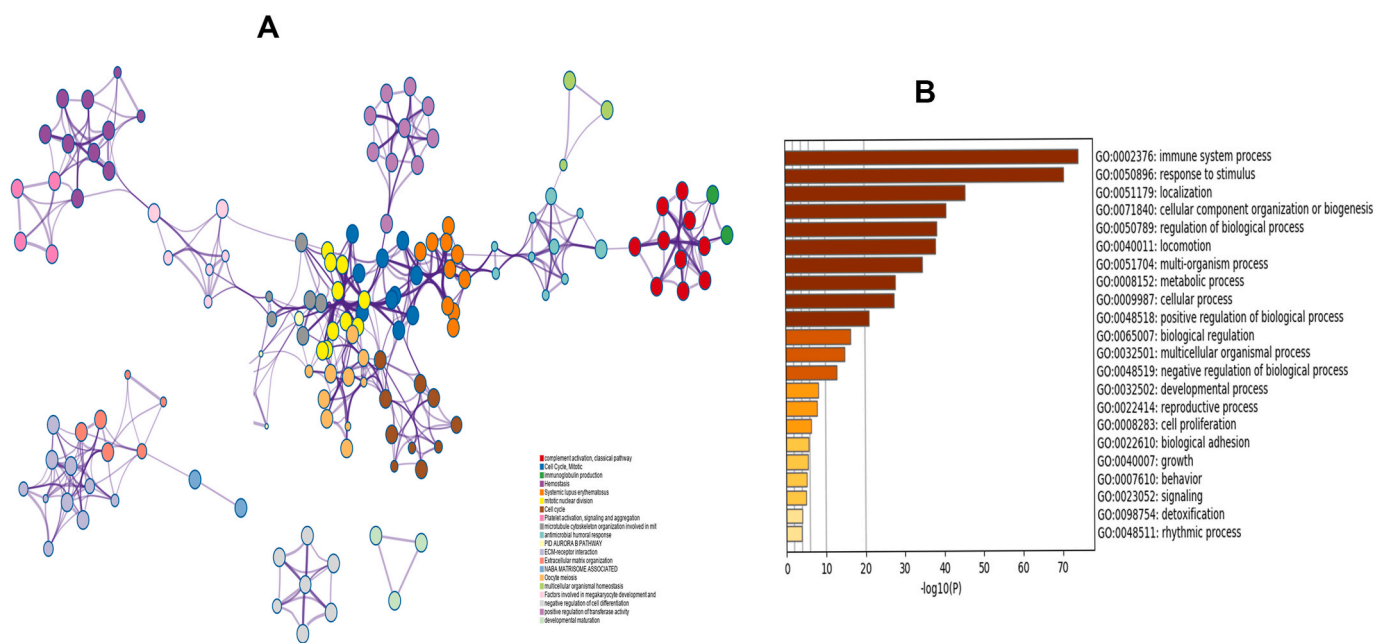


Fig. 3. Gene ontology analysis of DEGs. (A) network of enriched terms colored by cluster identity, (B) barplot of enriched terms using overlapping DEGs colored by p-value.

analysis was accomplished to verify the sample classification performance of the identified hub proteins based on two independent datasets.

3. Results

Identification of differentially expressed genes from peripheral blood cells of SARS-CoV-2.

We applied a robust voom approach to identify the DEGs from the male and female SARS-CoV-2 dataset. A total of 2067 and 1900 DEGs were identified from the male and female datasets, respectively. There are 1341 overlapping mutual DEGs identified between males and females, depicted in a Venn diagram of Fig. 1A. These overlapping DEGs were then used for further downstream analyses. The overlapping DEGs

between male and female has been shown in a circus at the gene level (Fig. 1B). The DEGs between SAR-CoV-2 infected male and female patients versus healthy control have been shown in volcano plots of Fig. 1C and D, respectively. In these figures, the green and red colors represent the down-regulated and up-regulated DEGs.

3.1. Functional annotation and pathway enrichment analysis

Genes like to interact with each other, and they do not work alone. Sometimes most of the genes show similar biological functions and pathways. To interpret the biological mechanism of 1341 overlapping DEGs, we performed GO and KEGG pathway enrichment analysis. GO analysis revealed that BPs are mainly enriched in nucleosome assembly,

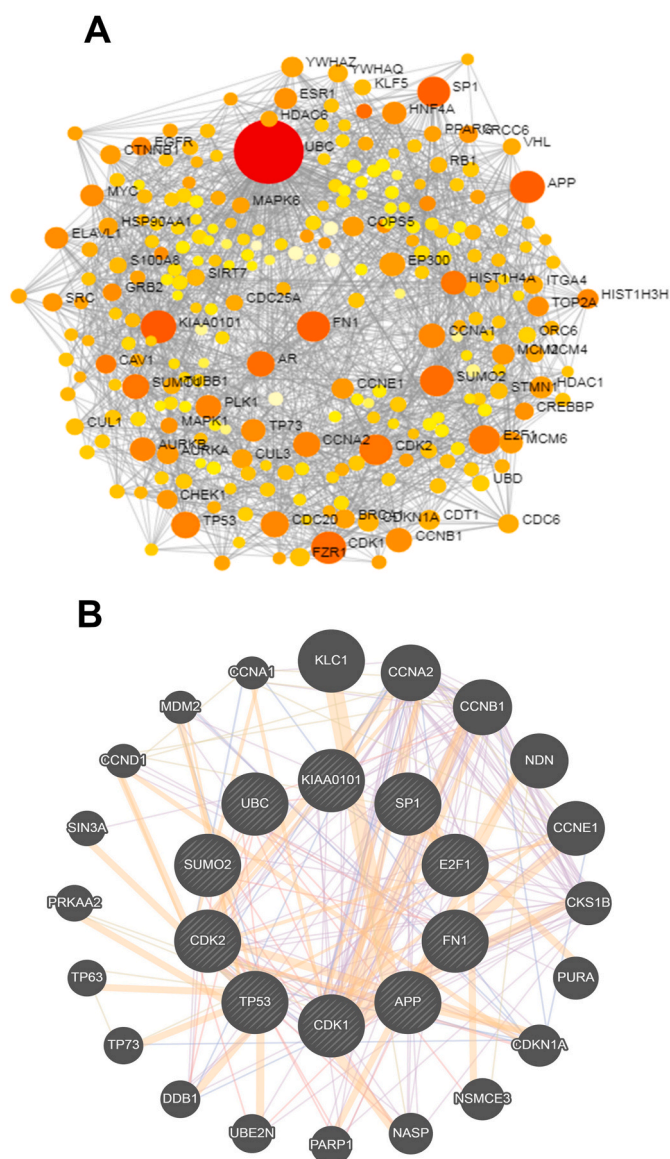


Fig. 4. PPI network analysis of overlapping DEGs. (A) PPI network of 1341 common DEGs identified between male and female, (B) hub-proteins network.

DNA conformation change, DNA packaging, etc. The top three significantly enriched CCs are DNA packaging complex, nucleosome, chromosomal region. The MFs are enriched in protein heterodimerization activity, heparin-binding, sulfur compound binding. The top-ranked significant GO categories were summarized in Table .1. From KEGG pathway enrichment analysis, we uncovered various pathways such as cell cycle, ECM-receptor interaction, progesterone-mediated oocyte maturation, alcoholism, and so on that are statistically significantly using a hypergeometric test with adjusted p-value <0.05. The top ten KEGG pathways have been summarized in Table .2 and plotted against enrichment ratio using a dot diagram in Fig. 2A. In addition, from Metascape, we explored that overlapping DEGs were significantly enriched in the immune system process, response to stimulus, metabolic process, developmental process ($P < 0.05$, Fig. 3A and Fig. 3B).

3.2. Determination of hub-proteins using protein-protein interaction analysis

Ten hub-proteins (UBC, KIAA0101, APP, CDK1, SUMO2, SP1, FN1, CDK2, E2F1 and TP53) were discovered using protein-protein

Table 3
Top ten drug candidates identified based on the hub-proteins.

Drug name	Mechanism of Action	FDA Status	Treatment
Roscovitine	Kinase Inhibitors	Investigational	Breast cancer, lung cancer, leukemia
Curcumin	Tyrosinase inhibitor	Approved	Colorectal Cancer, Pancreatic Cancer, liver
Simvastatin	Cholesterol lowering agent	Approved	Hyperlipidemia, diabetes mellitus, chronic kidney disease
Fulvestrant	Synthetic estrogen receptor antagonist	Approved	Metastatic breast cancer
Troglitazone	Antidiabetic and hepatotoxic agent	Approved	Type II diabetes mellitus
Alvocidib	Pan-ckd inhibitor, Kinase Inhibitors	Experimental	Esophageal cancer, leukemia, lung cancer, liver cancer
L-alanine	Glycine receptor agonist	Investigational	Metabolism of sugars and fatty acid, muscle growth, immune system
Tamoxifen	Antineoplastic nonsteroidal selective estrogen receptor modulator (SERM)	Approved	Breast cancer
Serine	Weak endogenous glycine receptor agonist	Approved	Muscle growth, immune system
Doxorubicin	Topo II nhibitor, immunosuppressive antineoplastic antibiotic activity	Approved	Leukemia, neuroblastoma, breast cancer, ovarian cancer

interaction (PPI) analysis. The PPI network has been shown in Fig. 4A. The hub-proteins with higher degrees of interaction determined using the topological analysis of PPI have also been displayed in Fig. 4B. The logarithmic values of RNASeq count expression of hub-proteins were shown in a heatmap plot in Fig. 2B.

3.3. Identification of potential miRNAs from DEGs-miRNA network

MicroRNAs (miRNAs) are non-coding RNAs which regulate the gene expression by controlling their target messenger mRNAs (mRNAs) for translational repression and degradation. Using DEGs-miRNA network analysis we extracted ten potential miRNAs (hsa-mir-17-5p, hsa-mir-20a-5p, hsa-mir-93-5p, hsa-mir-6499-3p, hsa-mir-92a-3p, hsa-mir-16-5p, hsa-mir-24-3p, hsa-mir-193b-3p, hsa-mir-192-5p, hsa-mir-98-5p). The DEGs-miRNA network has been shown in Fig. S1.

3.4. Identification of potential transcription factors

Transcription factors (TFs) are the proteins that regulate the transcription of genes from DNA to mRNA by binding DNA sequences. Therefore, the gene-TFs regulatory network has been carried out in Fig. S2 to revealed key TFs. The core TFs were found from this analysis are PPARG, E2F1, KLF5, FOXC1, and GATA2.

3.5. Identification of candidate drugs based on hub proteins

The top ten significant candidate drug agents identified by Enrichr are roscovitine, curcumin, simvastatin, fulvestrant, troglitazone, alvocidib, L-alanine, tamoxifen, serine, and 2-Butanone. They were summarized in Table .3. The other significant drug agents were also retrieved from this database such as water, tamoxifen, doxorubicin, roscovitine, ns-398, vinblastine, aflodac, resveratrol, rapamycin, mechlorethamine. The top 20 drug candidates have been presented in

Table 4
Performance evaluation of 10 hub-genes using six classifiers.

Methods	ACC	LACC	UACC	Sensitivity	Specificity	PPV	NPV
SVM	0.833	0.587	0.946	0.917	0.75	0.831	0.813
NBLDA	0.982	0.794	0.998	0.998	0.992	0.997	0.996
PLDA	0.868	0.635	0.951	0.847	0.889	0.914	0.871
PLDA2	0.903	0.669	0.973	0.861	0.944	0.951	0.873
voomDLDA	0.981	0.735	0.983	0.917	0.986	0.972	0.941
voomNSC	0.951	0.725	0.993	0.903	0.999	0.999	0.923

ACC = accuracy, LACC = lower limit of ACC, UACC = upper limit of ACC, PPV = positive predictive value, NPV = negative predictive value.

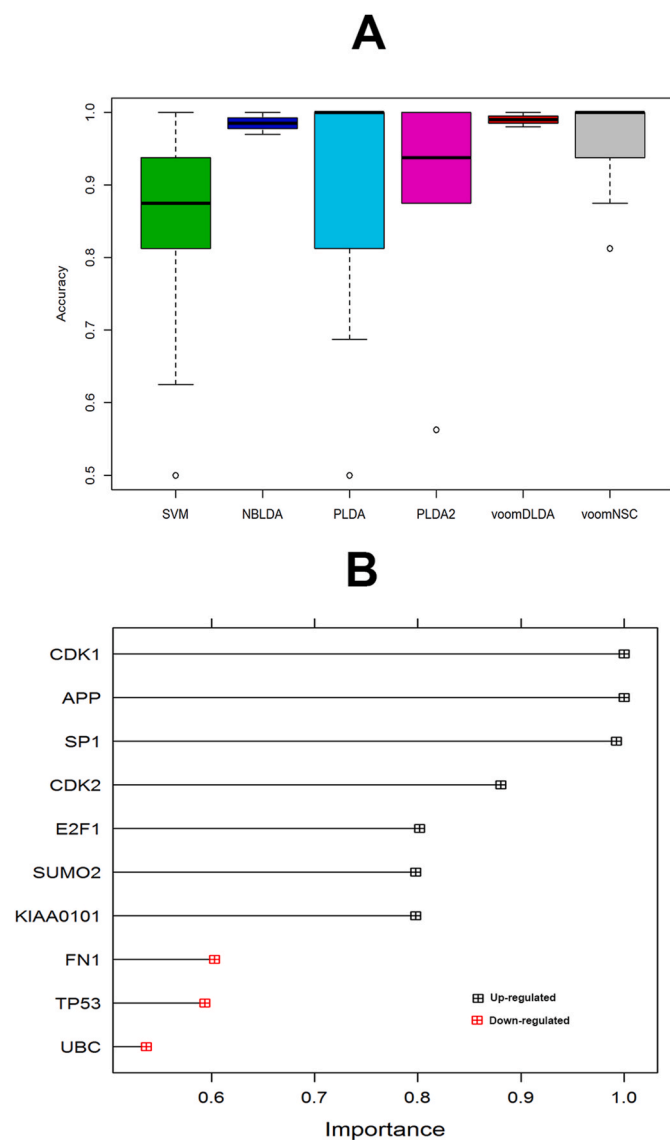


Fig. 5. Performance evaluation of 10 hub-gene using boxplot of Accuracies. (A) boxplot of accuracies of six classifiers, (B) Ranking of 10 hub-genes according to their importance using SVM.

Table S1.

3.6. Sensitivity and specificity analysis of the hub proteins

To investigate the sample discriminative performance of the identified ten hub-genes we computed various performance measures such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy (ACC) by the six classifiers (SVM,

PLDA, PLDA2, NBLDA, voomDLDA, voomNSC). To execute this task, we randomly divided this dataset (GSE152418) into a training dataset and a test dataset. The training dataset consists of 9 control and 8 COVID-19 samples. The rest of the samples belong to the test dataset. After that, ten hub-genes were selected from both datasets to construct the reduced training and test datasets. We performed 5-fold cross-validation to train the six classifiers, and the above performance measures were recorded. The average values of these performance indices were presented in Table 4. The accuracy values in this table indicate that the discrete distribution-based methods such as NBLDA, PLDA2, voomDLDA, and voomNSC performed well than SVM. The boxplot of the accuracies (Fig. 5A) also demonstrates the same results as Table 4. The proposed ten hub-genes were ranked according to their importance using SVM (Fig. 5B). We also showed count gene expression values of ten hub-genes between control, male and female sample in Fig. 6. This figure depicts that three hub-genes (CDK2, E2F1, and TP53) were down-regulated and seven hub-genes (UBC, KIAA0101, APP, CDK1, SUMO2, SP1, FN1) were up-regulated.

4. Discussion

Though sex-specific biomarker identification is crucial for developing drugs or therapies from the RNASeq count gene expression level of COVID-19 disease, none of the studies considered it yet. Men are more likely to become a serious condition in the presence of different comorbidities than women, and the mortality rate of COVID-19 is larger in men than women. Therefore, biomarkers may also be different between males and females of COVID-19. From this point of view, in this study, we aimed to identify the common DEGs between males and females of COVID-19. We identified 1341 overlapping DEGs between males and females. These DEGs were then undergone for GO and KEGG pathway analysis using DAVID and Metascape to explore the biological mechanisms of DEGs. From GO analysis we revealed that the top three categories BP (nucleosome assembly, DNA conformation change, DNA packaging), CC (DNA packaging complex, nucleosome, chromosomal region), and MF (protein heterodimerization activity, heparin-binding, sulfur compound binding) were significantly enriched. Using metaScape, we discovered that DEGs are involved in different cancer pathways, immune system, response to stimulus, metabolism process, and so on. DAVID also divulged some important KEGG pathways such as cell cycle, ECM-receptor interaction, progesterone-mediated cyte maturation etc. using the overlapping DEGs. Furthermore, we conducted PPI to identify hub-proteins, DEGs-miRNA to identify potential miRNAs, gene-TFs to determine core TFs. Ten hub-proteins (UBC, KIAA0101, APP, CDK1, SUMO2, SP1, FN1, CDK2, E2F1, and TP53) were identified using PPI topological network analysis. UBC is also known as Ubiquitin C is a protein-coding gene. Phlyctenulosis and cystic fibrosis diseases were found to associate with UBC [38]. Disease-associated with KIAA0101 (protein-coding gene) is thyroid carcinoma and heart conduction [39]. The GO annotation related to amyloid-beta precursor protein (APP) is app-related and Alzheimer's disease are associated with this gene [40]. CDK1 (Cyclin-dependent kinase 1) is the protein-coding gene. The diseases related to genes involve retinoblastoma, breast cancer, and glioblastoma multiforme. Pathways associated with these genes include

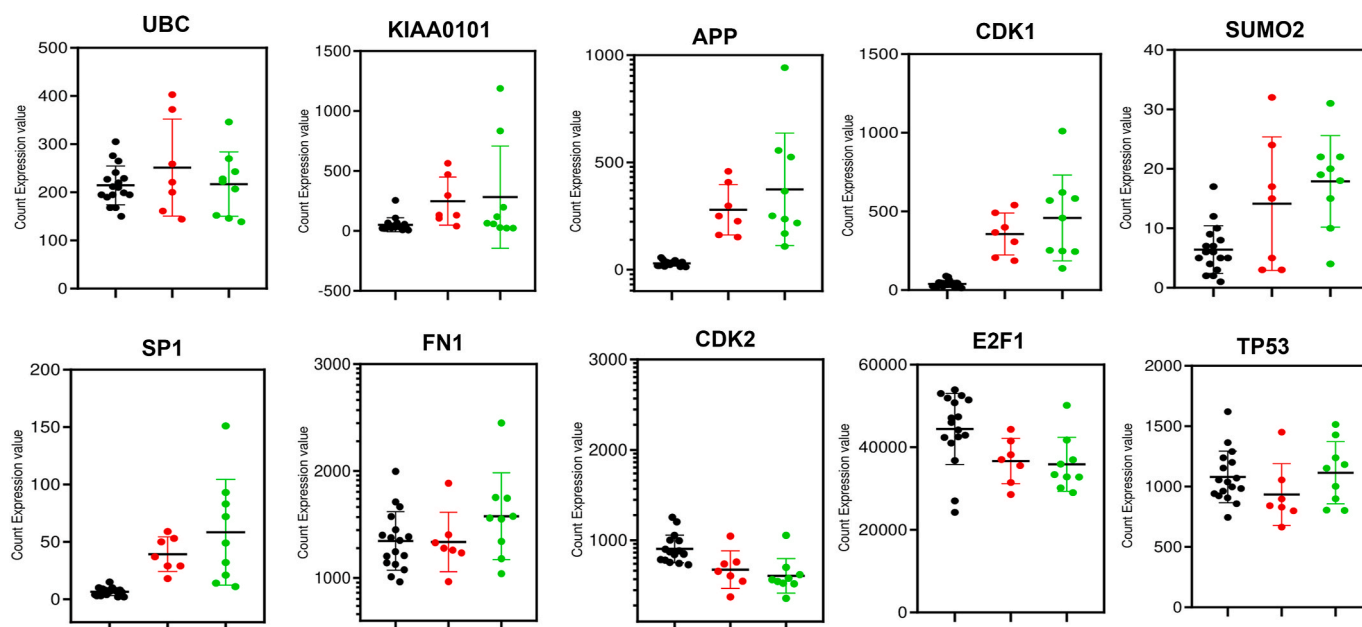


Fig. 6. Gene expression pattern of 10 hub-gene in the control, male and female datasets.

the ATM pathway and cell cycle [41]. SUMO2 is a protein-coding gene, and it is related to Gordon holmes syndrome [42]. The GO annotation related to SP1 includes DNA-binding transcription factor activity, and Huntington's disease was associated with this gene [43]. Heparin-binding and protease binding are found from GO annotation of the FN1 gene [44]. CDK2 (Cyclin-dependent kinase 2) is also a protein-coding gene. Diseases associated with CDK2 include breast Cancer and glioblastoma multiforme [45]. Linked diseases of E2F transcription factor 1 (E2F1) are retinoblastoma and glioblastoma multiforme. Pathways of this gene include regulation of activated PAK-2p34 by proteasome-mediated degradation and E2F transcription factor network [46]. The GO annotation of the TP53 gene includes DNA-binding transcription factor activity and protein heterodimerization activity [47]. The top three miRNAs (mir-17-5p, mir-20a-5p, mir-93-5p) and TFs (PPARG, E2F1 and KLF5) were uncovered. Finally, using drug repurposing analysis top ten significant drugs (roscovitine, curcumin, simvastatin, fulvestrant, troglitazone, alvocidib, L-alanine, tamoxifen, serine, and doxorubicin) were retrieved for therapeutic targets of COVID-19. Most of the drugs are FDA-approved and have been used to treat different types of cancer diseases.

5. Conclusions

The novel coronavirus (SARS-CoV-2) has expanded rapidly in today's world. SARS-CoV-2 mainly affects the respiratory systems of humans that can lead up to severe illness or even death with comorbidities. Still, drugs of other diseases have been used to treat the COVID-19. Nevertheless, in the absence of vaccines and proper drugs against COVID-19 has increased the mortality rate. Furthermore, COVID-19 infected men are more likely to experience severe illness than women. Hence, sex might be a major risk factor of COVID-19, and biomarkers related to the sex might be useful for discovering drugs against the COVID-19. Therefore, this paper attempts to identify the biomarkers by considering the sex effects using a robust voom approach. A total of 1341 overlapping DEGs were identified between males and females datasets. Using these DEGs' PPI analysis, we explored ten hub-proteins (UBC, KIAA0101, APP, CDK1, SUMO2, SP1, FN1, CDK2, E2F1, and TP53) that are involved in some important and interesting pathways of cancer-related disease. In sum, using these hub-proteins ten significant candidate drugs (roscovitine, curcumin, simvastatin, fulvestrant,

troglitazone, alvocidib, L-alanine, tamoxifen, serine, and doxorubicin) were retrieved that might be therapeutic targets of COVID-19 disease. Our future work will be covered with a network-based gene co-expression analysis based on COVID-19 datasets.

Author contributions

MSJ conceived and designed the study; MSJ analyzed data; MSJ wrote the draft manuscript; MRR and MRA reviewed and edited the manuscript.

Availability of data and material

Gene expression profiling data with accession GSE152418 is publicly available at Gene Expression Omnibus database (<https://www.ncbi.nlm.nih.gov/geo/>).

Funding

This research received no external funding.

Ethical approval

This clinical review does not require ethical approval as it is not a human or animal research.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgments

Authors would like to thank Dr. Mohsina Ahasan, Assistant Professor, Department of English, Begum Rokeya University, Rangpur, Bangladesh to proofread our manuscript for English corrections.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.imu.2021.100702>.

References

- [1] Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* 2019;17:181–92. <https://doi.org/10.1038/s41579-018-0118-9>.
- [2] Weiss SR, Navas-Martin S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol Mol Biol Rev* 2005;69:635–64. <https://doi.org/10.1128/MMBR.69.4.635-664.2005>.
- [3] Forni D, Cagliani R, Clerici M, Sironi M. Molecular evolution of human coronavirus genomes. *Trends Microbiol* 2017;25:35–48.
- [4] Chan-Yeung M, Xu RH. Sars. *epidemiology. Respirology* 2003;8(Suppl (s1)):S9–14. <https://doi.org/10.1046/j.1440-1843.2003.00518.x>.
- [5] Middle East respiratory syndrome coronavirus MERS-CoV. WHO; November 2019. retrieved 20 July 2020.
- [6] Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. *Acta bio-medica Atenei Parm* 2020;91:157–60.
- [7] Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB. Pharmacologic treatments for coronavirus disease 2019 (COVID-19): a review. *J Am Med Assoc* 2020;323:1824–36.
- [8] Islam, T.; Rahman, M.R.; Aydin, B.; Agga, K.Y.; Shahjaman, M. Integrative transcriptomics analysis of lung epithelial cells and identification of repurposable drug candidates for COVID-19. *Eur J Pharmacol* 2020, v-887.
- [9] Koyama T, Platt D, Parida L. Variant analysis of SARS-CoV-2 genomes. *Bull World Health Organ* 2020;98(7):495–504. <https://doi.org/10.2471/BLT.20.253591>.
- [10] Mahase E. Covid-19: what have we learnt about the new variant in the UK? *BMJ* 2020;371:m4944. <https://doi.org/10.1136/bmj.m4944>.
- [11] Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol* 2020;92:418–23.
- [12] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395:497–506.
- [13] Loo KY, Letchumanan V, Ser HL. COVID-19: Insights into potential vaccines. *Microorganisms* 2021;9(3):605. <https://doi.org/10.3390/microorganisms9030605>.
- [14] Goyal R, Jialal I. Diabetes mellitus type 2. In: *Treasure Island (FL)*; 2020.
- [15] Nowakowska M, Zghebi SS, Ashcroft DM, Buchan I, Chew-Graham C, Holt T, Mallen C, Van Marwijk H, Peek N, Perera-Salazar R. The comorbidity burden of type 2 diabetes mellitus: patterns, clusters and predictions from a large English primary care cohort. *BMC Med* 2019;17:145.
- [16] Shahjaman M, Rahman MR, Islam SMS, Mollah MNH. A robust approach for identification of cancer biomarkers and candidate drugs. *Medicina (Kaunas)* 2019; 55. <https://doi.org/10.3390/medicina55060269>.
- [17] Rahman MR, Islam T, Shahjaman M, Zaman T, Faruquee HM, Jamal MAHM. Discovering biomarkers and pathways shared by Alzheimer's disease and Ischemic Stroke to identify novel therapeutic targets. *Medicina (Kaunas)* 2019;55. <https://doi.org/10.3390/medicina55050191>.
- [18] Jin JM, Bai P, He W, Wu F, Liu XF, Han DM, Liu S, Yang JK. Gender differences in patients with COVID-19: Focus on severity and mortality. *Frontiers in public health* 2020;8:152. <https://doi.org/10.3389/fpubh.2020.00152>.
- [19] Xu K, Chen Y, Yuan J, Yi P, Ding C, Wu W. Factors associated with prolonged viral RNA shedding in patients with coronavirus disease 2019 (covid-19). *Clin Infect Dis* 2020;71(15):799–806. <https://doi.org/10.1093/cid/ciaa351>.
- [20] Zheng S, Fan J, Yu F, Feng B, Lou B, Zou Q. Viral load dynamics and disease severity in patients infected with sars-cov-2 in Zhejiang Province, China, January–March 2020: retrospective cohort study. *BMJ* 2020;369:m1443.
- [21] Coronaviruses Bwire GM. Why men are more vulnerable to covid-19 than women? [published online ahead of print, 2020 Jun 4]. *SN Compr Clin Med* 2020:1–3. <https://doi.org/10.1007/s42399-020-00341-w>.
- [22] Blanco-Melo D, Nilsson-Payant B, Liu WC, Moeller R, Panis M, Sachs D, Albrecht R. SARS-CoV-2 launches a unique transcriptional signature from in vitro, ex vivo, and in vivo systems. *bioRxiv*; 2020.
- [23] Pradhan A, Olsson PE. Sex differences in severity and mortality from COVID-19: are males more vulnerable? *Biol Sex Differ* 2020;11:53. <https://doi.org/10.1186/s13293-020-00330-7>.
- [24] Ge HP, Wang XF, Yuan XN, Xiao G, Wang CZ, Deng TC. The epidemiology and clinical information about covid-19. *Eur J Clin Microbiol* 2020;39(6):1011–9. <https://doi.org/10.1007/s10096-020-03874-z>.
- [25] Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 2013;41:991–5. <https://doi.org/10.1093/nar/gks1193>. PMID:23193258.
- [26] Arunachalam PS, Wimmers F, Mok CKP, Perera RAPM, Scottl M, Haganl T. Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* 2020;369:1210–20. <https://doi.org/10.1126/science.abc6261>.
- [27] Auwul MR, Rahman MR, Gov E, Shahjaman M, Moni MA. Bioinformatics and machine learning approach identifies potential drug targets and pathways in COVID-19. *Briefings in Bioinformatics*. 2021bbab120. <https://doi.org/10.1093/bib/bbab120>.
- [28] Shahjaman M, Mollah MMH, Rahman MR, Islam SM, Mollah MNH. Robust identification of differentially expressed genes from RNA-seq data. *Genomics* 2020; 112:2000–10. <https://doi.org/10.1016/j.ygeno.2019.11.012>.
- [29] Rahman MR, Islam T, Shahjaman M, Islam MR, Lombardo SD, Bramanti P, et al. Discovering common pathogenetic processes between COVID-19 and diabetes mellitus by differential gene expression pattern analysis. *Briefings in Bioinformatics* 2021:bbab262. <https://doi.org/10.1093/bib/bbab262>.
- [30] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nat Protoc* 2009;4(1):44–57.
- [31] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019;10(1):1523. <https://doi.org/10.1038/s41467-019-09234-6>. PMID: 30944313; PMCID: PMC6447622.
- [32] Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res* 2019;47:W234–41.
- [33] Montojo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, Morris Q, Bader GD. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics* 2010;26(22):2927–8.
- [34] Fornes O, Castro-Mondragon JA, Khan A, Jaspas 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2019. <https://doi.org/10.1093/nar/gkz1001>.
- [35] Yoo M, Shin J, Kim J, Ryall KA, Lee K, Lee S, Jeon M, Kang J, Tan AC. DSigDB: drug signatures database for gene set analysis. *Bioinformatics* 2015;31(18):3069–71. <https://doi.org/10.1093/bioinformatics/btv313>.
- [36] Ganter B, Snyder RD, Halbert DN, Lee MD. Toxicogenomics in drug discovery and development: mechanistic analysis of compound/class-dependent effects using the DrugMatrix database. *Pharmacogenomics* 2006;7(7):1025–44. <https://doi.org/10.2217/14622416.7.7.1025>. PMID:17054413.
- [37] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016;8(W1):44. <https://doi.org/10.1093/nar/gkw377>. W90-7.
- [38] Ralat LA, Kalas V, Zheng Z, Goldman RD, Sosnick TR, Tang WJ. Ubiquitin is a novel substrate for human insulin-degrading enzyme. *J Mol Biol* 2011;25(3):454–66. <https://doi.org/10.1016/j.jmb.2010.12.026>. 406.
- [39] Povlsen LK, Beli P, Wagner SA, Poulsen SL, Sylvestersen KB, Poulsen JW, Nielsen ML, Bekker-Jensen S, Mailand N, Choudhary C. Systems-wide analysis of ubiquitylation dynamics reveals a key role for PAF15 ubiquitylation in DNA-damage bypass. *Nat Cell Biol* 2012;14(10):1089–98. <https://doi.org/10.1038/ncb2579>.
- [40] Bodmer S, Podlisy MB, Selkoe DJ, Heid I, Fontana A. Transforming growth factor-beta bound to soluble derivatives of the beta amyloid precursor protein of Alzheimer's disease. *Biochem Biophys Res Commun* 1990;171(2):890–7. [https://doi.org/10.1016/0006-291x\(90\)91229-1](https://doi.org/10.1016/0006-291x(90)91229-1).
- [41] Medema RH, Macurek L. Checkpoint recovery in cells: how a molecular understanding can help in the fight against cancer. *F1000 Biol Rep*. 2011;3:10. <https://doi.org/10.3410/B3-10>.
- [42] 35 Tatham MH, Geoffroy MC, Shen L, Plechanovova A, Hattersley N, Jaffray EG, Palvimo JJ, Hay RT. RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat Cell Biol* 2008;10(5):538–46. <https://doi.org/10.1038/ncb1716>.
- [43] Bonfiglioglio D, Gabriele S, Aquila S, Qi H, Belmonte M, Catalano S, Andò S. Peroxisome proliferator-activated receptor gamma activates fas ligand gene promoter inducing apoptosis in human breast cancer cells. *Breast Cancer Res Treat* 2009;113(3):423–34. <https://doi.org/10.1007/s10549-008-9944-1>.
- [44] Sage J, Leblanc-Noblesse E, Nizard C, Sasaki T, Schnebert S, Perrier E, Kurfurst R, Brömme D, Lalmanach G, Lecaillon F. Cleavage of nidogen-1 by cathepsin S impairs its binding to basement membrane partners. *PLoS One* 2012;7(8):e43494. <https://doi.org/10.1371/journal.pone.0043494>.
- [45] Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat Rev Canc* 2009;9(3):153–66. <https://doi.org/10.1038/nrc2602>.
- [46] Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* 2009;10(4):252–63. <https://doi.org/10.1038/nrg2538>.
- [47] Wei X, Xu H, Kufe D. Human MUC1 oncoprotein regulates p53-responsive gene transcription in the genotoxic stress response. *Canc Cell* 2005;7(2):167–78. <https://doi.org/10.1016/j.ccr.2005.01.008>.