Deciphering Novel SARS CoV-2 specific Disease Pathways from RNA Sequencing Data of COVID-19 Infected A549 Cells and Potential Therapeutics using Next Generation Knowledge Discovery Platforms

Peter Natesan Pushparaj
Jeddah University: University of Jeddah
Laila Abdullah Damiati
King Abdulaziz University Faculty of Applied Medicine
Luliana Denetiu
king Abdulaziz University Faculty of Applied Medical Sciences
Sherin Bakhashab
King Abdulaziz University Faculty of Biological Sciences
Muhammad Asif
Balochistan University of Information Technology Engineering and Management Sciences, Faculty of Life Sciences
Abrar Hussain
Balochistan University of Information Technology Engineering and Management Sciences, Faculty of Life Sciences
Sagheer Ahmed
Shifa Tameer-e-Millat University, College of Pharmaceutical Sciences
Mohammed Hammid Hamdard (hamid.hamdard@ku.edu.af)
Kabul University https://orcid.org/0000-0002-5915-6574
Mahmood Rasool
King Abdulaziz University Faculty of Applied Medical Sciences

Research

Keywords: COVID-19, A549 Cells, RNA Seq, BioJupies, iPathwayGuide, Synthetic Drugs, Natural Products, Next Generation Knowledge Discovery

Posted Date: June 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-589902/v1

Abstract

Background

The coronavirus (CoV) disease identified in Wuhan, China in 2019 (COVID-19) was chiefly characterized by atypical pneumonia and severe acute respiratory syndrome (SARS) and caused by SARS CoV-2 that belongs to the family Coronaviridae. COVID-19 symptoms vary from a mild cold to more severe illnesses such as SARS, thrombosis, stroke, organ failure, and in some patients even cause mortality. Deciphering the underlying disease mechanisms is pivotal for the identification and development of COVID-19 specific drugs for effective treatment and prevent human-to-human transmission, disease complications, and mortality.

Methodology: Here, the Next Generation RNA Sequencing (RNA Seq) data using Illumina Next Seq 500 from SARS CoV-infected A549 cells and mock-treated A549 cells, were obtained from the gene expression omnibus (GEO) (GSE147507) and the Quality Control (QC) were evaluated using the CLC Genomics Workbench 20.0 (Qiagen, USA) before the RNA Seq analysis. The DEGs were imported into BioJupies to analyze to decipher COVID-19 induced biological, molecular, and cellular processes, pathways, and small molecules derived from chemical synthesis or natural sources to mimic or reverse COVID-19-specific gene signatures. Besides, we have used the iPathwayGuide (Advaita Bioinformatics USA) to identify COVID-19 specific pathways, biological, molecular, and cellular processes, and “druggable” candidates for future therapy.

Results: 141 DEGs were identified out of a total of 9665 DEGs obtained from BioJupies analysis of the RNASeq reads of the SARS CoV infected A549 cells and mock-treated A549 cells based on a p-value cut off (0.05) and a fold change cut off 1.5.

Conclusion: In conclusion, the present study unravels a novel approach of using next-generation knowledge discovery platforms to discover specific drugs for the amelioration of COVID-19 related disease pathologies.

Introduction

Coronavirus Disease 2019 (COVID-19) is caused by a type of coronavirus (CoV), severe acute respiratory syndrome (SARS) virus 2 (SARS CoV-2). COVID-19 is characterized by symptoms ranging from a mild cold to more severe ailments such as SARS, sudden stroke, gastrointestinal complications, multiple organ failure, and even cause mortality in some patients (Lamers et al., 2020; Mulay et al., 2021). Coronaviruses belong to the family of Coronaviridae and the presence of viral spike proteins in the virus gives a halo or corona-like appearance under electron microscopy (Fig. 1A). A novel coronavirus virus (nCoV) was discovered in Wuhan, China in 2019 as the cause of an outbreak of respiratory illness in humans leading to severe atypical pneumonia (Wang et al., 2020) and it is the source of the current global pandemic that hampered all walks of life (Agarwal et al., 2021)

The World Health Organization (WHO) renamed this nCoV as SARS-CoV 2 which is the causative agent for COVID-19 (Dong et al., 2020; Wang et al., 2020; Wu et al., 2020). The COVID-19 is highly transmissible and pathogenic compared to other viral infections and the exact mortality rate is yet to be determined since the pandemic is still not under control in several countries leading to unprecedented protective measures, partial or total lockdowns, travel restrictions, etc., (Mandolesi et al., 2021; Zhonghua et al., 2019). As of 4th March

2021, based on the Johns Hopkins Coronavirus Dashboard metrics, COVID-19 had already infected more than 154 million in 192 countries and territories around the world and killed about 3.2 million people. However, about 90 million people recovered from COVID-19 to date. Nevertheless, the precise mortality rate will be calculated or identified once the COVID-19 epidemic reaches a plateau. The United States of America (USA)and the WHO has declared the SARS CoV-2 outbreak a public health emergency since it is more contagious than Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) (Wu et al., 2020), and the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) (Liu et al., 2021). SARS-CoV-2 has a nucleocapsid with a positive-sense RNA genome. The host cells express SARS-CoV-2 nucleoproteins and the nucleocapsid protein (N-protein) that is the most abundant, highly immunogenic, and required for CoV RNA synthesis. The N-protein is a structural protein that attaches to the CoV RNA genome and forms a capsid around the viral RNA. Though, the spike protein (S-protein) is crucial for the attachment between the SARS CoV-2 and the Angiotensin-Converting Enzyme 2 (ACE2) surface receptors on the host cells (Fig. 1B) and thus facilitates the coronavirus to penetrate the host cells (Fig. 1C). (Ols et al., 2020).

Even though the COVID-19 vaccines are currently available as a preventive measure and many are under the research and development phase (Bezbaruah et al., 2021; Mandolesi et al., 2021), decoding the underlying pathological mechanisms is pivotal for the identification and development of COVID-19 specific drugs for the effective treatment and prevention of human-to-human transmission, COVID-19 complications, and mortality. Henceforth, in the present study, the raw RNA Seq reads (Single-End) (*FASTQ files*) in quadruplicates derived using Illumina Next Seq 500 from SARS CoV-infected A549 cells and mock-treated A549 cells, were obtained from the gene expression omnibus (GEO) (Accession Number: GSE147507) and the Quality Control (QC) were evaluated using the CLC Genomics Workbench 20.0 (Qiagen, USA) before the RNA Seq analysis. After the initial QC, the RNA Seq reads were imported into CLC Genomics Workbench 20.0 (Qiagen, USA) before the RNA Seq analysis and applied next-generation knowledge discovery platforms (NGKD) such as BioJupies (Torre et al., 2018) and iPathwayGuide (Advaita Bioinformatics, USA) to decipher the disease-specific molecular signatures and an array of small molecules derived from either synthetic or natural sources to mimic or reverse COVID-19 gene signatures. The present study outlines an innovative method of applying NGKD platforms to discover precise drugs and natural products for the potential treatment of COVID-19 related disease pathologies.

Materials And Methods

Ethical Statement

This study did not use any animal models or human subjects and was performed using the RNASeq datasets from Next Generation Sequencing experiments using A549 cells. The raw data were retrieved from the Gene Expression Omnibus (GEO) as stated in the Data Source section below. Hence, it was exempted from the Institution Review Board (IRB) approval.

NGS Data Source

The raw RNA Seq reads (Single-End) (*FASTQ format*) in quadruplicates derived using Illumina Next Seq 500 from A549 cells infected with SARS CoV-2, and mock-treated A549 cells, were obtained from the gene

expression omnibus (GEO) (*Accession No: GSE147507*) were used for analysis using high-throughput knowledge discovery platforms.

COVID-19 RNA Seq Data from A549 Cells – Quality Control (QC)

The raw RNA Seq reads (Single-End) in quadruplicates (*FASTQ files*) derived using Illumina Next Seq 500 from SARS CoV-infected A549 cells and mock-treated A549 cells were derived from the GEO and the Quality Control (QC) was evaluated using the CLC Genomics Workbench 20.0 (Qiagen, USA) before the RNA Seq analysis to obtain the differentially expressed genes (DEGs).

COVID-19 RNA Seq Data from A549 Cells- Differential Gene and Transcript Expression Analysis

The RNA Seq reads were imported into CLC Genomics Workbench 20.0 (Qiagen, USA) after the QC step. The RNA Seq Analysis tool in the Biomedical Genomics Analysis plugin of the CLC Genomics Workbench was used to extract all annotated transcripts using both *Homo sapiens (hg38) _Gene* (Gene track) and *Homo sapiens (hg38) _mRNA* (mRNA track) and mapped to the human reference genome (GRCh38). The Gene Expression Track (GE) was generated for SARS CoV-2 infectedA549 cells, and corresponding mock reads (Test vs Control), respectively. Furthermore, the *Differential Expression in Two Groups tool* in CLC Genomics Workbench was used to perform a statistical differential expression test for a set of Expression Tracks (*Test vs Control*). Here, a multi-factorial statistic based on a negative binomial Generalized Linear Model (GLM) is used, and the *Differential Expression in Two Groups tool* handles one factor and two groups. In this analysis, for GE, the values used are "Total Exon Reads". Differentially expressed genes *(DEGs)* were then generated for Test compared to the corresponding Control and used for further downstream analysis using next-generation knowledge discovery platforms.

BioJupies Analysis of the RNASeq Data

BioJupies with 14 different plug-ins were used to analyze DEGs generated using CLC Genomics Workbench to identify novel pathways, disease-specific gene networks, and an array of drugs and small molecules obtained from natural sources to mimic or reverse disease-specific gene signatures (Torre et al., 2018). In Biojupies, the RNASeq Datasets were user-submitted, compressed in an HDF5 data package, and uploaded to Google Cloud. Raw counts were normalized to log10-Counts Per Million (log CPM) and the differentially expressed genes were derived between the control group and the experimental group using the limma R package (Ritchie et al., 2015)). The Principal Component Analysis Function in the sklearn Python module was used to transform log CPM based on the Z-score method to generate the PCA plot and Clustergrammer (Fernandez et al., 2017)), was used to generate interactive heatmaps and the DEGs were then sent to Enrichr (Kuleshov et al., 2016). In the Volcano Plot, the DEGs were shown on the x-axis and p-values were corrected using the Benjamini-Hochberg method, transformed (–log10), and presented on the y axis (Benjamini and Yekutieli, 2001)). However, the average gene expression was shown on the x-axis in the MA plot, and the P-values were corrected using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001), transformed (–log10), and presented on the y axis. Gene ontology (GO) and Pathway enrichment analyses were performed using both the up-regulated and down-regulated genes in Enrichr. Significant GO terms and Pathways (*KEGG, WiKiPathways, and Reactome*) are computed by using a cut-off of p-value<0.1 after applying Benjamini-Hochberg correction (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001).

L1000CDS^2 and L1000FWD Queries

The L1000CDS2 analysis was performed by submitting the top 2000 DEGs to the L1000CDS2 signature search API (Duan et al., 2016). Similarly, the L1000FWD analysis was performed by submitting the top 2000 DEGs to the L1000FWD signature search API (Wang et al., 2018).

In Silico Analysis of the RNASeq Expression Data using iPathwayGuide

The Impact Analysis Method (IAM)(Draghici et al., 2007; Tarca et al.) in the iPathwayGuide was used to determine the significantly impacted gene signatures, pathways, miRNAs, and metabolites in the cell lines treated with anti-cancer drugs compared with the corresponding untreated control. The pathway score is calculated based on a p-value computed using Fisher's method. The p-value is corrected based on multiple testing corrections for false discovery rate (FDR) and Bonferroni corrections (Bonferroni, 1935; Bonferroni, 1936). The FDR has significant power but it controls only the family-wise false positives rate (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001). The pathways and the gene interactions using the DEGs were generated using the KEGG database (Kanehisa and Goto, 2000; Kanehisa et al., 2012; Kanehisa et al., 2014). For each Gene Ontology (GO) term (Ashburner et al., 2000), the number of DEGs annotated to the term when compared to the DEGs expected just by chance. iPathwayGuide uses an over-representation approach to compute the statistical significance of observing at least the given number of DEGs (Draghici, 2016; Drăghici et al., 2003). The hypergeometric distribution was used to compute the p values in iPathwayGuide analysis and corrected using FDR and Bonferroni for multiple comparisons (Draghici, 2016; Drăghici et al., 2003).

Prediction of Upstream Drugs or Natural Products using iPathwayGuide

The prediction of upstream Chemicals, Drugs, Toxicants (CDTs) is based on two types of information: i) the enrichment of DEGs from the experiment and ii) a network of interactions from the Advaita Knowledge Base (AKB v2012). The edges represent known effects that these CDTs have on various genes. A signed edge in this graph consists of a source CDT, a target gene, and a sign to indicate the type of effect: activation (+) or inhibition (-) (Drăghici et al., 2003; Draghici et al., 2007).

Upstream CDTs predicted as present (or overly abundant)

Here, the research hypothesis considers the presence of the CDT. This hypothesis is useful when investigating whether the given phenotype has been impacted by the presence of a given chemical, drug, or toxicant (Drăghici et al., 2003; Draghici et al., 2007). For each CDT u, the number of consistent DE genes downstream of u, DTA(u) is compared to the number of measured target genes expected to be both consistent and DE just by chance. iPathwayGuide uses an over-representation approach to compute the statistical significance of observing at least the given number of consistent DE genes. The p-value Ppres is computed using the hypergeometric distribution (Drăghici et al., 2003; Draghici et al., 2007). The analysis uses the standard Fisher's method to combine p-values into one test statistic (Fisher, 1925).

Upstream CDTs predicted as absent (or insufficient)

In parallel with upstream CDTs predicted as present, Pabs and Pz are used to predict upstream CDTs that are absent. This hypothesis is relevant when investigating whether the given phenotype has been impacted by the lack of a given chemical that is necessary for the well-functioning of the organism or cell. Here, the research hypothesis states that the upstream CDT is insufficient in the condition studied. For each upstream CDT u, the number of consistent DE genes downstream of u, DTI(u) is compared to the number of measured target genes expected to be both consistent and DE just by chance. Using Fisher's method as above, the analysis combines Pabs and Pz, where Pz is considered only for significant negative z-scores ($z \leq -2$) (Drăghici et al., 2003; Draghici et al., 2007).

Swiss Target Prediction of Potential Anti-COVID-19 Compounds

Isomeric Simplified Molecular Input Line Entry System (SMILES) codes of Prednisolone and Withaferin-A was used in the SwissTargetPrediction tool to find the protein targets (Bahlas et al., 2020). The ligand-based target prediction for both prednisolone and Withaferin-A was performed as described before (Bahlas et al., 2020; Kalamegam et al., 2020).

The Open Targets Platform Analysis of Anti-COVID-19 Compounds

The Open Targets Platform web tool was used to uncover the Prednisolone and Withaferin-A molecular targets associated with COVID-19 disease pathology (Bahlas et al., 2020; Carvalho-Silva et al., 2019). The Open Targets Platform uses scientific evidence to score and rank target-disease associations and aid target prioritization (Carvalho-Silva et al., 2019). The query list with about 100 eligible molecular targets of Prednisolone and Withaferin-A was utilized to discover the protein targets associated significantly ($P < 0.05$) with COVID-19.

Results

Here, the raw RNA Seq reads (Single-End) (*FASTQ files*) in quadruplicates derived using Illumina Next Seq 500 from SARS CoV-infectedA549 cells, and mock-treated A549 cells were obtained from the gene expression omnibus (GEO) (GSE147507) and the Quality Control (QC) were evaluated before the RNA Seq analysis using the CLC Genomics Workbench 20.0 (Qiagen, USA). The DEGs were further analyzed using BioJupies (Torre et al., 2018), and iPathwayGuide (Advaita Bioinformatics, USA) to decipher the disease-specific signatures and an array of drugs and small molecules derived from natural sources to mimic or reverse disease-specific gene signatures.

The global patterns in high-dimensional RNA seq datasets were uncovered by interactive PCA analysis **(Figure 2A).** Clustergrammer web tool was used to generate interactive heatmaps for visualization and in-depth analysis of DEGs derived from the high-dimensional RNASeq data of SARS CoV-infected A549 cells, and mock-treated A549 cells **(Figure 2A - C).** The volcano plot was generated using transformed gene fold changes using log2 and shown on the x-axis **(Figure 2D).** MA plot was based on average gene expression which was calculated using the mean of the normalized gene expression values and shown on the x-axis **(Figure 2E).**

The interactive bar chart **(Figure 3A)** shows the top small molecules identified by the L1000CDS2 query. The left panel displays the small molecules such as Calyculin A, Emetine Hydrochloride, Narciclasine, NVP-TAE684, wiskostatin, NCGC00185684-02, and Amsacrine which mimic the observed gene expression signature, while the right panel displays the small molecules such as Trichostatin A, Vorinostat, afatinib, DL-PDMP, Withaferin-A, IMD 0354, and 2-[(chloroacetyl) (4-fluorophenyl] amino-N-cyclohexyl-2 pyridine 3 which reverse it. Besides, the natural products and drugs with opposite and similar molecular signatures **(Table 1)** based on the L1000FWD tool that contains gene signatures from an array of human cell lines administered with more than 20,000 drugs and natural products. Withaferin-A, an active ingredient of the medicinal plant **(Figure 3A)**, Withania somnifera was found to reverse the COVID-19 induced molecular signatures in both L1000CDS2 and L1000FWD analyses along with other small molecule drugs.

The GO enrichment analysis for the Biological Processes, Molecular Function, and Cellular Component were generated using Enrichr **(Figure 4A).** The x-axis indicates the -log10(P-value) for each term and the significant terms enriched in each GO category were highlighted in bold. Similarly, **Figure 4B** shows the results of the pathway enrichment analysis produced using Enrichr. The x-axis indicates the -log10(P-value) for each term and the significantly enriched pathways (KEGG, Wiki Pathways, and Reactome) were highlighted in bold.

iPathwayGuide Analysis of the High Dimensional RNA Seq Data

In this experiment, **141** DEGs were identified out of a total of **9665** DEGs obtained from BioJupies analysis of the RNASeq reads of the SARS CoV infected A549 cells and mock-treated A549 cells based on a p-value cut off (0.05) and a fold change cut off **1.5**. The DEGs were analyzed in the context of pathways obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Release 96.0+/11-21, Nov 20) (Kanehisa and Goto, 2000), gene ontologies from the Gene Ontology Consortium database (2020-Oct14) (Ashburner et al., 2000), miRNAs from the iRbase (MIRBASE Version: Version22.1,10/18) and TARGETSCAN (Targetscan version: Mouse:7.2, Human:7.2) databases, network of regulatory relations from BioGRID: Biological General Repository for Interaction Datasets v4.0.189. Aug. 25th, 2020 (Szklarczyk et al., 2017), chemicals/drugs/toxicants from the Comparative Toxicogenomics Database July 2020 (Davis et al., 2019), and diseases from the KEGG database (Release 96.0+/11-21, Nov 20 (Kanehisa and Goto, 2000). iPathwayGuide analysis further showed that **34** pathways were found to be significantly impacted in the SARS CoV2 infected A549 cells compared to the mock-treated A549 cells. Also, **557** Gene Ontology (GO) terms, **224** gene upstream regulators, **451** chemical upstream regulators, and **31** diseases were found to be significantly ($P<0.05$) enriched before the correction for multiple comparisons.

The top five upstream regulators identified after Bonferroni Correction STAT2, IRF9, IFNB1, IL1B, and IRF3 were predicted as activated (Table 7). The COVID-19 infection activates key infectious disease specific immune-related signaling pathways such as Influenza A, Viral protein interaction with cytokine and cytokine receptor, Measles, Epstein-Barr virus infection, and IL-17 signaling pathway **(Table 2).** *Likewise, the* significantly enriched Gene Ontology (GO) terms such as Biological, Molecular, and Cellular Processes based on False Discovery Rate (*q value*) were identified using iPathwayGuide. The top identified Biological Processes identified were innate immune response, response to external biotic stimulus, response to other organisms, response to biotic stimulus, and defense response to other organisms, the top Molecular Functions identified include chemokine receptor binding, chemokine activity, CXCR chemokine receptor

binding, receptor-ligand activity, signaling receptor activator activity, and the top Cellular Components identified include blood microparticle, fibrinogen complex, nuclear outer membrane, extracellular space, and extracellular region for each pruning type were provided in **Table 3.**

The upstream regulator drugs obtained either based on chemical synthesis or natural sources with opposite molecular signatures were also identified based on iPathwayGuide Analysis The drugs that can significantly reverse the molecular impact of COVID-19 infection are Methyl Prednisolone, Prednisolone, Gold Sodium Thiomalate, Tofacitinib, Diclofenac, JQ1 Compound, Azathioprine, etc., **(Figure 3B).** The upstream regulator drugs and natural products with opposite molecular signatures identified using iPathwayGuide sorted based on Z score were listed in **Supplementary Table 1.**

In silico Prediction of the Molecular Targets of Prednisolone and Withaferin-A Using SwissTarget Prediction

In the present study, the SwissTargetPrediction was performed for Prednisolone and Withaferin A using the canonical SMILES code. The Open Targets Platform was applied to uncover the Withaferin-A molecular targets associated with COVID-19 disease pathology (Bahlas et al., 2020; Carvalho-Silva et al., 2019; Koscielny et al., 2017). The scientific evidence was used in the Open Targets Platform to assign a score and rank target-disease associations and help target prioritization(Carvalho-Silva et al., 2019) . Among the molecular targets of Prednisolone and Withaferin-a, 40 and 36 targets respectively were significantly associated with COVID-19 disease pathology (**Supplementary Table 02)**

Discussion

The COVID-19 is highly infectious and pathogenic compared to other viral infections and the exact mortality rate is yet to be determined since the pandemic is still not under control in several countries (Zhonghua et al., 2019)**.** Hence, deciphering the underlying pathological mechanisms is pivotal for the identification and development of COVID-19 specific drugs for the effective treatment and prevention of human-to-human transmission, COVID-19 complications, and reduce mortality. COVID-19 is usually characterized by cough, breathing problems, high body temperature, diarrhea, abdominal discomfort, and in severe cases causes' atypical pneumonia, SARS, stroke, thrombosis, multiple organ failure, and in some cases cause mortality. It was found that about 80% of the COVID-19 cases were having mild symptoms or asymptomatic, with the elderly, and individuals with other comorbid conditions were more prone to develop severe symptoms and succumb to the disease(Wang et al., 2020; Zhonghua et al., 2019).

To differentiate the COVID-19 from other influenza viruses, SARS, and MERS coronaviruses, etc. is essential in a clinical setting to design effective or efficient treatment strategies for the patients (Huang et al., 2020). Also, non-infectious diseases like idiopathic interstitial pneumonia, or cryptogenic organizing pneumonia, dermatomyositis, vasculitis, etc., must be differentially diagnosed from COVID-19 (Huang et al., 2020; Zhonghua et al., 2019).

The COVID-19 infection of the A549 cells activated the upstream genes such as STAT2, IRF9, IFNB, IL1B, IRF3, etc. The biological processes such as type I interferon signaling pathway, defense response to the virus, negative regulation of viral genome replication, and the interferon-gamma-mediated signaling pathway were differentially regulated. The molecular functions such as chemokine activity, CXCR chemokine receptor

binding, 2'-5'-oligoadenylate synthetase activity, double-stranded RNA binding, and protein ADP-ribosylase activity were enriched in the COVID infected cells. Cytokines are the hormones of the immune system and are important for both innate and adaptive host responses, cell growth and differentiation, repair and development, cellular homeostasis, and cell death (Harakeh et al., 2020). Cytokines are glycoproteins released upon any external stimuli and bind with specific cell surface receptors on the plasma membrane of target cells to elicit their responses (Pushparaj, 2019)

The cytokine/chemokine storm observed in the moderate to severe cases of COVID-19 is caused by a significant increase in the levels of an array of circulating cytokines and chemokines such as IL-6, IL-8, TNFα, CXCL-10, etc., and contribute to the poor prognosis (Vaninov, 2020) In general, the viruses develop mechanisms to avoid detection and the subsequent destruction inside the host by repurposing and copying cytokine and cytokine receptor genes.(Pushparaj, 2019; Pushparaj et al., 2007; Sowmya et al., 2011). Similarly, the COVID-19 induced cytokines and cytokine receptors, chemokines, and other specific cytokine receptors and binding proteins to destabilize and modify the host cytokine responses and immune networks (Pushparaj et al., 2007). Here, the COVID-19 induced chemokines and cytokines may either augment or prevent cytokine signaling and may significantly change or attenuate various arms of host immunity. Besides, cellular processes such as blood microparticle fibrinogen complex were activated in the COVID infected A549 cells. The increase in the cellular processes such as blood microparticle as observed in the present study was further confirmed by a recent study showing the increased circulating blood microparticles and activated platelets in COVID-19 patients (Zahran et al., 2021).

COVID-19 pandemic is currently managed by vaccines, convalescent plasma, monoclonal antibodies, anti-viral medications such as remdesivir, and taking preventive procedures such as wearing masks, hand hygiene measures, social distancing, etc. (Sewell et al., 2020). In our present study, Withaferin-A was predicted to oppose the molecular signatures triggered by COVID-19. The open targets platform analysis predicted that 36 targets were found to play a role in the COVID-19 disease pathology. Withaferin-A is one of the constituents of the medicinal plant, Withania somnifera (*Indian Ginseng or Ashwagandha*). The active ingredients mainly include withanolides, saponins, alkaloids, and steroidal lactones. W. somnifera is used in herbal formulations in traditional medicine possess antioxidant, anti-anxiety, anti-inflammatory, anti-bacterial, and aphrodisiac properties, etc (Sood et al., 2018). Ashwagandha has neuroprotective, cardioprotective, immunomodulating, and anti-cancer properties (Singh et al., 2021). Also, a recent in silico screening study recognized that ashwagandha has natural compounds against COVID-19 (Srivastava et al., 2020).

Similarly, the traditional Chinese herbal formulation, JinFuKang is composed of 12 medicinal plants, and each dosage included 10 mL (Cassileth et al., 2009; Jiao et al., 2015). JinFuKang has anti-cancer properties and an array of medicinal benefits (Que et al., 2021). Since the antiviral remdesivir is very minimal in reducing mortality (Wilt et al., 2021), the use of corticosteroids increases the possibility of secondary infections (Gopalaswamy and Subbian, 2021), and monoclonal antibody therapies are either expensive or difficult to procure, it is better to investigate the potential compounds identified in our study for the potential treatment and amelioration of COVID-19 and related pathologies.

Conclusion

In conclusion, the present study unravels an innovative approach of using next-generation high throughput RNA sequencing technologies coupled with NGKD platforms to decipher specific drugs obtained either synthetically or from natural products for the betterment of COVID-19. Further studies are required to validate the drugs such as Prednisolone, Methylprednisolone, Diclofenac, JQ1 compound, etc. and the natural products such as Withaferin-A and JinFuKang in COVID-19 infection model systems such as primary human alveolar epithelial cells and human small intestinal organoids (hSIOs) (Mulay et al., 2021)to infer the mechanism(s) of action before pre-clinical trials and clinical trials for the potential treatment of COVID-19 related disease pathologies.

Abbreviations

CDTs Chemicals, Drugs, Toxicants

CoV Coronavirus Virus

DEGs Differentially expressed genes

FDR False Discovery Rate

GEO Gene Expression Omnibus

GO Counts Per Million (log CPM) ontology

IAM Impact Analysis Method

NGKD applied next-generation knowledge discovery

QC Quality Control

Declarations

Acknowledgments

The research work was funded by the DSR project number DF-509-142-1441 .Therefore, the authors gratefully acknowledge technical and financial support from the King Abdulaziz University, DSR, Jeddah. Saudi Arabia.

Disclosure statement

All authors declare that they have no conflict of interest.

Authors Contribution

PNP, LAD, LD, SB, and MR designed the experiments. PNP and MR conducted the experiments. PNP, LAD, LD, SB, and MR analyzed the data. PNP and MR AM HA wrote the manuscript. PNP and MR finally revised the manuscript. All authors contributed to the editing of the manuscript and the scientific discussions.

References

Agarwal, A., Leisegang, K., Panner Selvam, M.K., Durairajanayagam, D., Barbarosie, C., Finelli, R., Sengupta, P., Dutta, S., Majzoub, A., Pushparaj, P.N., 2021. An online educational model in andrology for student training in the art of scientific writing in the COVID-19 pandemic. Andrologia 53, e13961.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., 2000. Gene ontology: tool for the unification of biology. Nature genetics 25, 25-29.

Bahlas, S., Damiati, L.A., Al-Hazmi, A.S., Pushparaj, P.N., 2020. Decoding the Role of Sphingosine-1-Phosphate in Asthma and Other Respiratory System Diseases Using Next Generation Knowledge Discovery Platforms Coupled With Luminex Multiple Analyte Profiling Technology. Frontiers in cell and developmental biology 8, 444.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 57, 289-300.

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Annals of statistics, 1165-1188.

Bezbaruah, R., Borah, P., Kakoti, B.B., Al-Shar, N.A., Chandrasekaran, B., 2021. Developmental landscape of potential vaccine candidates based on viral vector for prophylaxis of COVID-19. Frontiers in Molecular Biosciences 8.

Carvalho-Silva, D., Pierleoni, A., Pignatelli, M., Ong, C., Fumis, L., Karamanis, N., Carmona, M., Faulconbridge, A., Hercules, A., McAuley, E., 2019. Open Targets Platform: new developments and updates two years on. Nucleic acids research 47, D1056-D1065.

Cassileth, B.R., Rizvi, N., Deng, G., Yeung, K.S., Vickers, A., Guillen, S., Woo, D., Coleton, M., Kris, M.G., 2009. Safety and pharmacokinetic trial of docetaxel plus an Astragalus-based herbal formula for non-small cell lung cancer patients. Cancer chemotherapy and pharmacology 65, 67-71.

Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. The Lancet infectious diseases 20, 533-534.

Draghici, S., 2016. Statistics and data analysis for microarrays using R and bioconductor. CRC Press.

Drăghici, S., Khatri, P., Martins, R.P., Ostermeier, G.C., Krawetz, S.A., 2003. Global functional profiling of gene expression. Genomics 81, 98-104.

Draghici, S., Khatri, P., Tarca, A.L., Amin, K., Done, A., Voichita, C., Georgescu, C., Romero, R., 2007. A systems biology approach for pathway level analysis. Genome research 17, 1537-1545.

Fernandez, N.F., Gundersen, G.W., Rahman, A., Grimes, M.L., Rikova, K., Hornbeck, P., Ma'ayan, A., 2017. Clustergrammer, a web-based heatmap visualization and analysis tool for high-dimensional biological data. Scientific data 4, 1-12.

Fisher, R.A., 1925. Theory of statistical estimation, Mathematical Proceedings of the Cambridge Philosophical Society. Cambridge University Press, pp. 700-725.

Gopalaswamy, R., Subbian, S., 2021. Corticosteroids for COVID-19 Therapy: Potential Implications on Tuberculosis. International Journal of Molecular Sciences 22, 3773.

Harakeh, S., Kalamegam, G., Pushparaj, P.N., Al-Hejin, A., Alfadul, S.M., Al Amri, T., Barnawi, S., Al Sadoun, H., Mirza, A.A., Azhar, E., 2020. Chemokines and their association with body mass index among healthy Saudis. Saudi journal of biological sciences 27, 6-11.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The lancet 395, 497-506.

Jiao, L., Wang, Y., Xu, L., You, M., 2015. Lung cancer prevention and therapy using the JinFuKang herbal mixture. Current Pharmacology Reports 1, 346-353.

Kalamegam, G., Alfakeeh, S.M., Bahmaid, A.O., AlHuwait, E.A., Gari, M.A., Abbas, M.M., Ahmed, F., Abu-Elmagd, M., Pushparaj, P.N., 2020. In vitro evaluation of the anti-inflammatory effects of thymoquinone in osteoarthritis and in silico analysis of inter-related pathways in age-related degenerative diseases. Frontiers in cell and developmental biology 8.

Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 28, 27-30.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M., 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic acids research 40, D109-D114.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2014. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic acids research 42, D199-D205.

Koscielny, G., An, P., Carvalho-Silva, D., Cham, J.A., Fumis, L., Gasparyan, R., Hasan, S., Karamanis, N., Maguire, M., Papa, E., 2017. Open Targets: a platform for therapeutic target identification and validation. Nucleic acids research 45, D985-D994.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic acids research 44, W90-W97.

Lamers, M.M., Beumer, J., van der Vaart, J., Knoops, K., Puschhof, J., Breugem, T.I., Ravelli, R.B., van Schayck, J.P., Mykytyn, A.Z., Duimel, H.Q., 2020. SARS-CoV-2 productively infects human gut enterocytes. Science 369, 50-54.

Liu, Q., Zhao, X., Ma, J., Mu, Y., Wang, Y., Yang, S., Wu, Y., Wu, F., Zhou, Y., 2021. Selenium (Se) plays a key role in the biological effects of some viruses: Implications for COVID-19. Environmental research 196, 110984.

Mandolesi, M., Sheward, D.J., Hanke, L., Ma, J., Pushparaj, P., Vidakovics, L.P., Kim, C., Adori, M., Lenart, K., Loré, K., 2021. SARS-CoV-2 protein subunit vaccination of mice and rhesus macaques elicits potent and durable neutralizing antibody responses. Cell Reports Medicine, 100252.

Mulay, A., Konda, B., Garcia Jr, G., Yao, C., Beil, S., Villalba, J.M., Koziol, C., Sen, C., Purkayastha, A., Kolls, J.K., 2021. SARS-CoV-2 infection of primary human lung epithelium for COVID-19 modeling and drug discovery. Cell reports, 109055.

Ols, S., Yang, L., Thompson, E.A., Pushparaj, P., Tran, K., Liang, F., Lin, A., Eriksson, B., Hedestam, G.B.K., Wyatt, R.T., 2020. Route of vaccine administration alters antigen trafficking but not innate or adaptive immunity. Cell reports 30, 3964-3971. e3967.

Pushparaj, P.N., 2019. Multiple analyte profiling (xMAP) technology coupled with functional bioinformatics strategies: potential applications in protein biomarker profiling in autoimmune inflammatory diseases, Essentials of Bioinformatics, Volume II. Springer, pp. 151-165.

Pushparaj, P.N., Manikandan, J., Melendez, A.J., 2007. The Immune Response in Patients with SARS: differential gene expression profiling. Nature Precedings, 1-1.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research 43, e47-e47.

Sood, A., Mehrotra, A., Dhawan, D.K., Sandhir, R., 2018. Indian Ginseng (Withania somnifera) supplementation ameliorates oxidative stress and mitochondrial dysfunctions in experimental model of stroke. Metabolic brain disease 33, 1261-1274.

Sowmya, G., Shamini, G., Anita, S., Sakharkar, M., Mathura, V., Rodriguez, H., Levine, A.J., Singer, E., Commins, D., Somboonwit, C., 2011. HIV-1 envelope accessible surface and polarity: clade, blood, and brain. Bioinformation 6, 48.

Srivastava, A., Siddiqui, S., Ahmad, R., Mehrotra, S., Ahmad, B., Srivastava, A., 2020. Exploring nature's bounty: identification of Withania somnifera as a promising source of therapeutic agents against COVID-19 by virtual screening and in silico evaluation. Journal of Biomolecular Structure and Dynamics, 1-51.

Tarca, A.L., Draghici, S., Khatri, P., Hassan, S.S., Mittal, P., Kim, J.-s., Kim, C.J., Kusanovic, J.P., Romero, R., A Novel Signaling Pathway Impact Analysis (SPIA).

Torre, D., Lachmann, A., Ma'ayan, A., 2018. BioJupies: automated generation of interactive notebooks for RNA-Seq data analysis in the cloud. Cell systems 7, 556-561. e553.

Vaninov, N., 2020. In the eye of the COVID-19 cytokine storm. Nature Reviews Immunology 20, 277-277.

Wang, C., Horby, P.W., Hayden, F.G., Gao, G.F., 2020. A novel coronavirus outbreak of global health concern. The lancet 395, 470-473.

Wang, Z., Lachmann, A., Keenan, A.B., Ma'ayan, A., 2018. L1000FWD: fireworks visualization of drug-induced transcriptomic signatures. Bioinformatics 34, 2150-2152.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., 2020. A new coronavirus associated with human respiratory disease in China. Nature 579, 265-269.

Zahran, A.M., El-Badawy, O., Ali, W.A., Mahran, Z.G., Mahran, E.E.M., Rayan, A., 2021. Circulating microparticles and activated platelets as novel prognostic biomarkers in COVID-19; relation to cancer. PloS one 16, e0246806.

Zhonghua, L., Xing, B., Za, Z., Liuxingbingxue, Z., 2019. Novel coronavirus pneumonia emergency response epidemiology team. The epidemiological characteristics of an outbreak of, 145-151.

Tables

Table 1. Natural Products and Drugs with Opposite Molecular Signatures based on L1000FWD web-based Tool.

Signature ID	Drugs or Natural Products	Similarity Score	p-value	q-value	Z-score	combined score
CPC019_VCAP_24H:BRD-K50234570-001-06-6:10	EMF-bca1-16	-0.0569	2.03e-10	8.67e-07	1.67	-16.15
ERG005_VCAP_6H:BRD-K88378636-001-02-8:20	withaferin-a	-0.0544	1.54e-09	2.74e-06	1.65	-14.57
CPC006_HCC515_24H:BRD-A28105619-001-01-3:10	cucurbitacin-i	-0.0531	2.51e-09	3.98e-06	1.81	-15.59
CPC006_HCC515_6H:BRD-K16406336-311-01-2:10	methylene-blue	-0.0544	6.11e-09	8.44e-06	1.77	-14.52
CPC016_MCF7_24H:BRD-K08547377-003-03-2:10	irinotecan	-0.0506	1.88e-08	2.12e-05	1.70	-13.16
CPC001_VCAP_24H:BRD-K12516989-001-01-9:10	zaprinast	-0.0442	6.06e-08	5.64e-05	1.93	-13.91
CPC016_NPC_24H:BRD-A22783572-065-01-3:10	vinblastine	-0.0493	1.28e-07	1.08e-04	1.69	-11.63
CPC004_PC3_6H:BRD-A69815203-001-05-0:10	cyclosporin-a	-0.0455	1.90e-07	1.51e-04	1.84	-12.37
CPC008_PC3_6H:BRD-K66037923-001-04-4:10	BRD-K66037923	-0.0480	1.96e-07	1.53e-04	1.76	-11.79
MUC.CP003_MCF7_24H:BRD-K02407574-001-04-8:0.3704	parbendazole	-0.0468	2.45e-07	1.79e-04	1.63	-10.75

Table 2. Natural Products and Drugs with Similar Molecular Signatures based on L1000FWD web-based Tool.

Signature ID	Drugs or Natural Products	similarity score	p-value	q-value	Z-score	combined score
CPC013_SKB_24H:BRD-K61175124-001-01-0:10	BRD-K61175124	0.0556	2.63e-13	1.12e-08	-1.83	23.09
CPC016_SKB_24H:BRD-A06352508-001-02-9:10	SB-218078	0.0544	1.17e-12	1.67e-08	-1.87	22.28
CPC006_HT29_24H:BRD-A67788537-001-01-7:120	salermide	0.0493	1.78e-12	1.91e-08	-1.85	21.75
CPC002_PC3_6H:BRD-A22684332-003-03-1:10	procaterol	0.0582	3.07e-12	2.63e-08	-1.64	18.87
CPC007_HT29_24H:BRD-A09719808-001-02-3:10	BRD-A09719808	0.0506	8.82e-12	6.29e-08	-1.81	20.02
CPC019_VCAP_6H:BRD-K23282736-001-01-1:10	BRD-K23282736	0.0594	1.12e-11	6.83e-08	-1.78	19.53
CPC007_HT29_6H:BRD-A69470004-019-04-0:10	BRD-A69470004	0.0556	5.64e-11	3.02e-07	-1.69	17.30
CPC013_SKB_24H:BRD-K74623475-001-02-7:10	BRD-K74623475	0.0480	1.67e-10	7.94e-07	-1.86	18.18
CPC006_A673_6H:BRD-K84924563-001-01-2:40	BRD-K84924563	0.0531	3.62e-10	1.29e-06	-1.68	15.85
CPC013_SKB_24H:BRD-K16541732-001-01-3:10	BRD-K16541732	0.0493	7.61e-10	1.92e-06	-1.81	16.48

Table 3. Top pathways and their associated p-values are stated in the table.

Pathway name	Pathway Id	p-value	p-value (FDR)	p-value (Bonferroni)
Influenza A	05164	**7.635e-7**	**6.389e-5**	**9.849e-5**
Viral protein interaction with cytokine and cytokine receptor	04061	**1.533e-6**	**6.389e-5**	**1.978e-4**
Measles	05162	**1.665e-6**	**6.389e-5**	**2.148e-4**
Epstein-Barr virus infection	05169	**2.019e-6**	**6.389e-5**	**2.605e-4**
IL-17 signaling pathway	04657	**2.476e-6**	**6.389e-5**	**3.194e-4**

* the p-value corresponding to the pathway was computed using only over-representation analysis.

Table 4. Top identified biological processes. Only the top scoring biological process for each pruning type is described below the table.

Pruning Type: None				Pruning Type: High-specificity		Pruning Type: Smallest Common Denominator	
GO Term	p-value	p-value (FDR)	p-value (Bonferroni)	GO Term	p-value	GO Term	p-value
innate immune response	1.000e-24	4.427e-22	4.427e-22	type I interferon signaling pathway	2.380e-12	type I interferon signaling pathway	9.961e-14
response to external biotic stimulus	1.000e-24	6.549e-22	1.965e-21	defense response to virus	8.301e-12	defense response to virus	4.012e-13
response to other organism	1.000e-24	6.549e-22	1.965e-21	negative regulation of viral genome replication	2.490e-8	interferon-gamma-mediated signaling pathway	1.383e-10
response to biotic stimulus	3.000e-24	2.075e-21	8.301e-21	interferon-gamma-mediated signaling pathway	8.301e-6	negative regulation of viral genome replication	1.868e-8
defense response to other organism	4.900e-23	2.712e-20	1.356e-19	innate immune response			

Table 5. Top identified molecular functions. Only the top scoring molecular function for each pruning type is described below the table.

Pruning Type: None				Pruning Type: High-specificity		Pruning Type: Smallest Common Denominator	
GO Term	**p-value**	**p-value (FDR)**	**p-value (Bonferroni)**	**GO Term**	**p-value**	**GO Term**	**p-value**
chemokine receptor binding	**1.300e-9**	**3.897e-7**	**5.629e-7**	chemokine activity	**7.794e-7**	chemokine receptor binding	**5.629e-7**
chemokine activity	**1.800e-9**	**3.897e-7**	**7.794e-7**	CXCR chemokine receptor binding	**4.546e-5**	2'-5'-oligoadenylate synthetase activity	**0.040**
CXCR chemokine receptor binding	**1.300e-8**	**1.876e-6**	**5.629e-6**	2'-5'-oligoadenylate synthetase activity	**0.030**	double-stranded RNA binding	**0.040**
receptor ligand activity	**1.700e-6**	**1.840e-4**	**7.361e-4**	double-stranded RNA binding	**0.030**	protein ADP-ribosylase activity	**0.086**
signaling receptor activator activity	**2.400e-6**	**2.078e-4**	**0.001**	protein ADP-ribosylase activity	**0.068**	ISG15 transferase activity	**0.109**

Table 6. Top identified cellular components. Only the top scoring cellular component for each pruning type is described below the table.

Pruning Type: None				Pruning Type: High-specificity		Pruning Type: Smallest Common Denominator	
GO Term	**p-value**	**p-value (FDR)**	**p-value (Bonferroni)**	**GO Term**	**p-value**	**GO Term**	**p-value**
blood microparticle	**9.800e-8**	**3.763e-5**	**3.763e-5**	blood microparticle	**3.763e-5**	blood microparticle	**3.763e-5**
fibrinogen complex	**0.002**	0.358	0.768	fibrinogen complex	**0.358**	fibrinogen complex	**0.358**
nuclear outer membrane	**0.003**	0.358	1.000	nuclear outer membrane	**0.358**	nuclear outer membrane	**0.358**
extracellular space	**0.006**	0.425	1.000	specific granule lumen	**0.553**	extracellular region	**0.553**
extracellular region	**0.008**	0.425	1.000	costamere	**0.553**	specific granule lumen	**0.553**

Table 7. Top upstream regulators after Bonferroni Correction is given in the table.

Upstream Regulator (u)	DTA(u)	DT(u)	p-value	p-value (FDR)	p-value (Bonferroni)
STAT2	11	11	**1.655e-14**	**6.848e-12**	**8.092e-12**
IRF9	10	10	**2.801e-14**	**6.848e-12**	**1.370e-11**
IFNB1	6	7	**1.526e-6**	**2.488e-4**	**7.464e-4**
IL1B	7	8	**1.188e-4**	**0.014**	0.058
IRF3	3	3	**1.464e-4**	**0.014**	0.072

Table 8. COVID-19 Associated Targets Regulated by Prednisolone and Withaferin-A

Disease	Drug or Natural Product	Number of associated targets	Therapeutic Area	All targets
COVID-19	Prednisolone	40	infectious disease	DPP4 JAK1 NR3C1 JAK2 AR PTGS1 CHRNA4 IL6 PDE10A SLC5A2 FLT3 MAPK14 TYK2 OPRM1 KIT PPARG KDR ABL1 NR3C2 ESR2 CNR1 ADORA3 MPEG1 PGR ADAM17 CD38 MTOR MPO EGFR SLC6A3 MAPK1 ALK NOS2 SLC5A1 BRD4 MAPK3 ADK LCK RORA SHBG
COVID-19	Withaferin-A	36	infectious disease	NR3C1 PTGS2 AR HMGCR PTGS1 GSK3B F10 PDE4D GSK3A IMPDH1 PDE3A PDE3B PDE10A MAPK14 JAK3 NR3C2 IKBKB ADORA2A PGR REN PARP1 ERBB2 CCR1 MAPK1 ALK HDAC3 PRKCB BRAF IL6ST CXCR3 MAPK8 IARS1 BRD4 BCL2L1 MAPK3 MDM2

Figures

Figure 1A

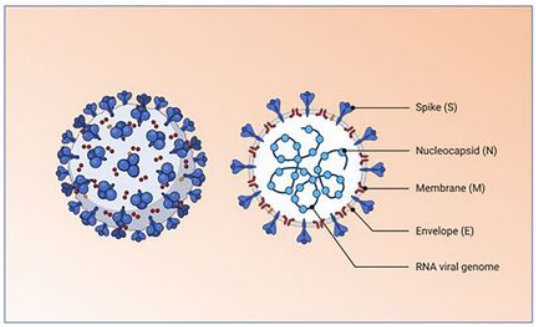

Figure 1B

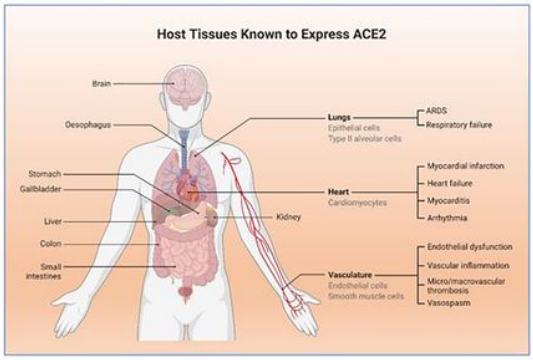

Figure 1C

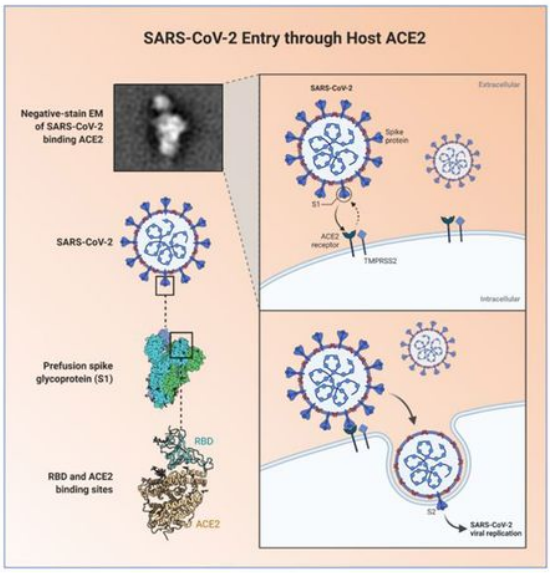

Figure 1

Structure of SARS-CoV 2 (A) The structure depicted based on electron microscopic observations of coronavirus showing the surface protein particles S, N, M, and E and the provides a corona like shape (B) The host tissues expressing ACE2 receptors (C) The mechanism of entry of SARS CoV2 into the host cells (This figure was created using the graphic tools offered by BioRender.com with an academic license)

Figure 2A

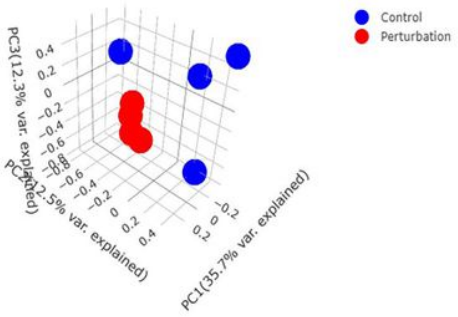

Figure 2B

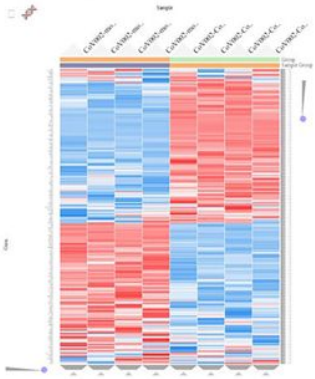

Figure 2C

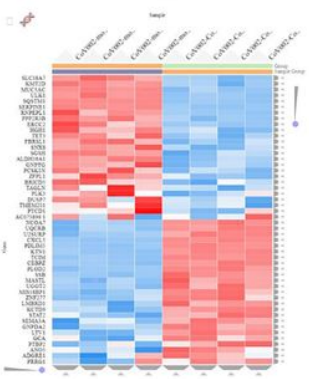

Figure 2D

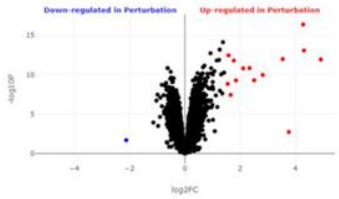

Figure 2E

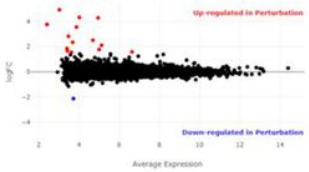

Figure 2

(A) Principal Component Analysis (PCA) was applied to identify global patterns in high-dimensional RNASeq datasets. (B) Interactive heatmaps were generated top 250 DEGs and (C) the top 50 DEGs using Clustergrammer web tool for visualizing and analyzing high-dimensional RNASeq data (D) Volcano plot was generated using transformed gene fold changes using log2 and displayed on the x axis (E) MA plot was based on average gene expression which was calculated using mean of the normalized gene expression values and displayed on the x axis.

Figure 3A

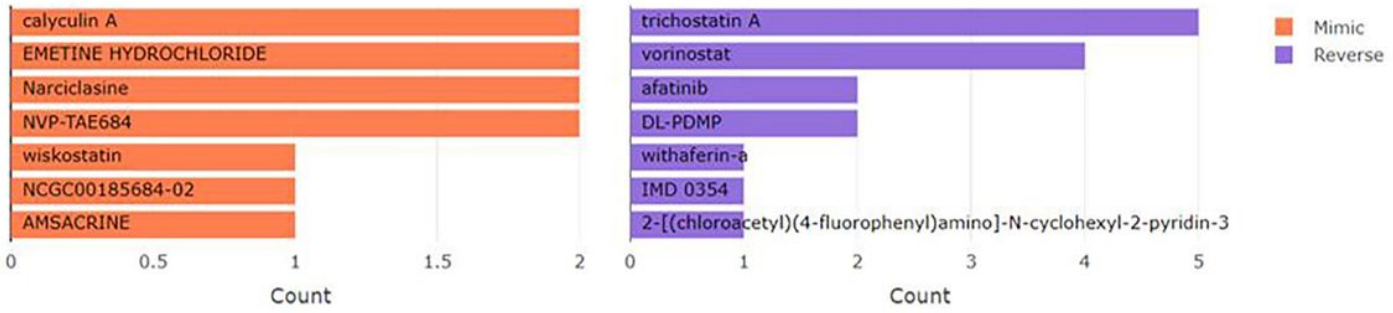

Figure 3B

Figure 3

(A) The figure contains an interactive bar chart displaying the top small molecules identified by the L1000CDS2 query. The left panel displays the small molecules which mimic the observed gene expression signature, while the right panel displays the small molecules which reverse it. (B) Bar graphs show the synthetic drugs and natural compounds with similar (mimic) and opposite (reverse) molecular signatures based on iPathwayGuide Analysis.

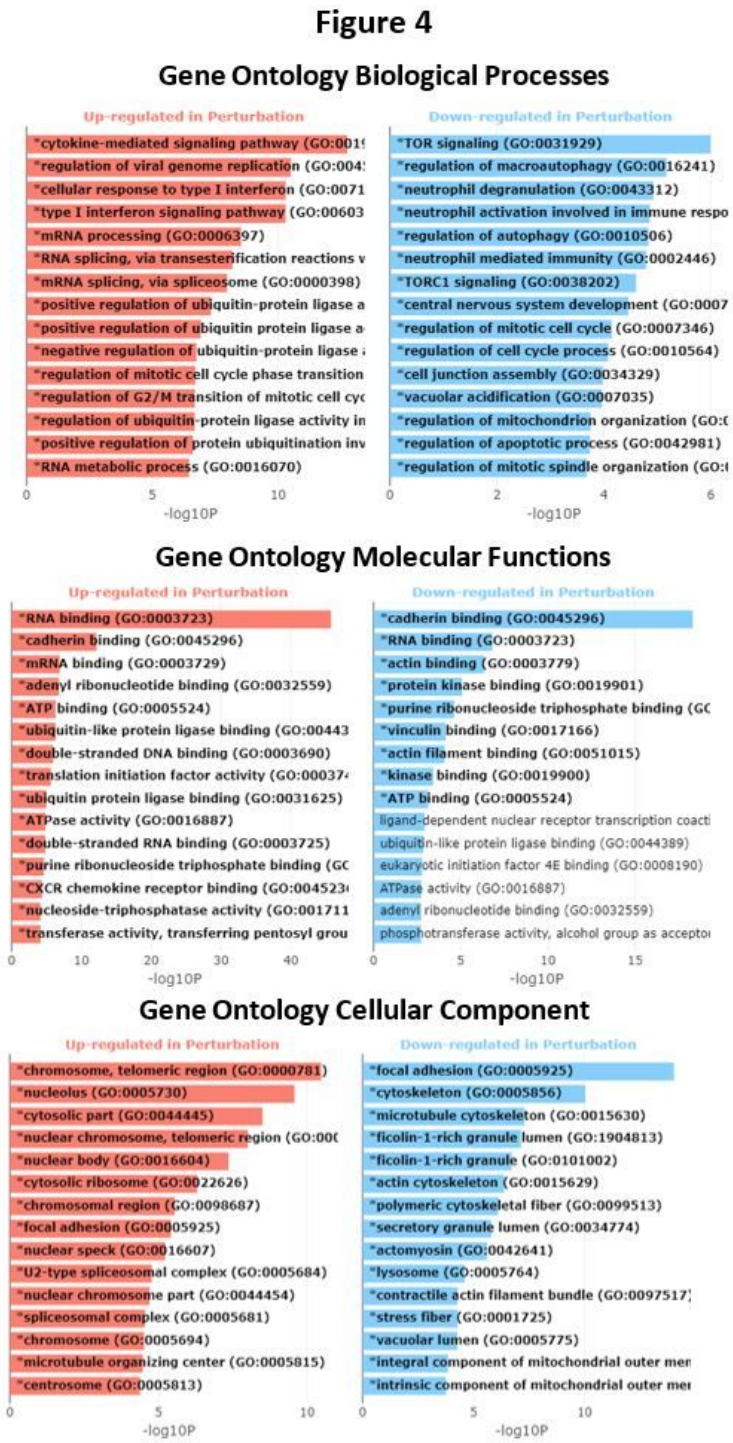

Figure 4

A. Gene Ontology Enrichment Analysis. The figure contains interactive bar charts displaying the results of the Gene Ontology enrichment analysis generated using Enrichr. The x axis indicates the -log10(P-value) for each term. Significant terms are highlighted in bold.

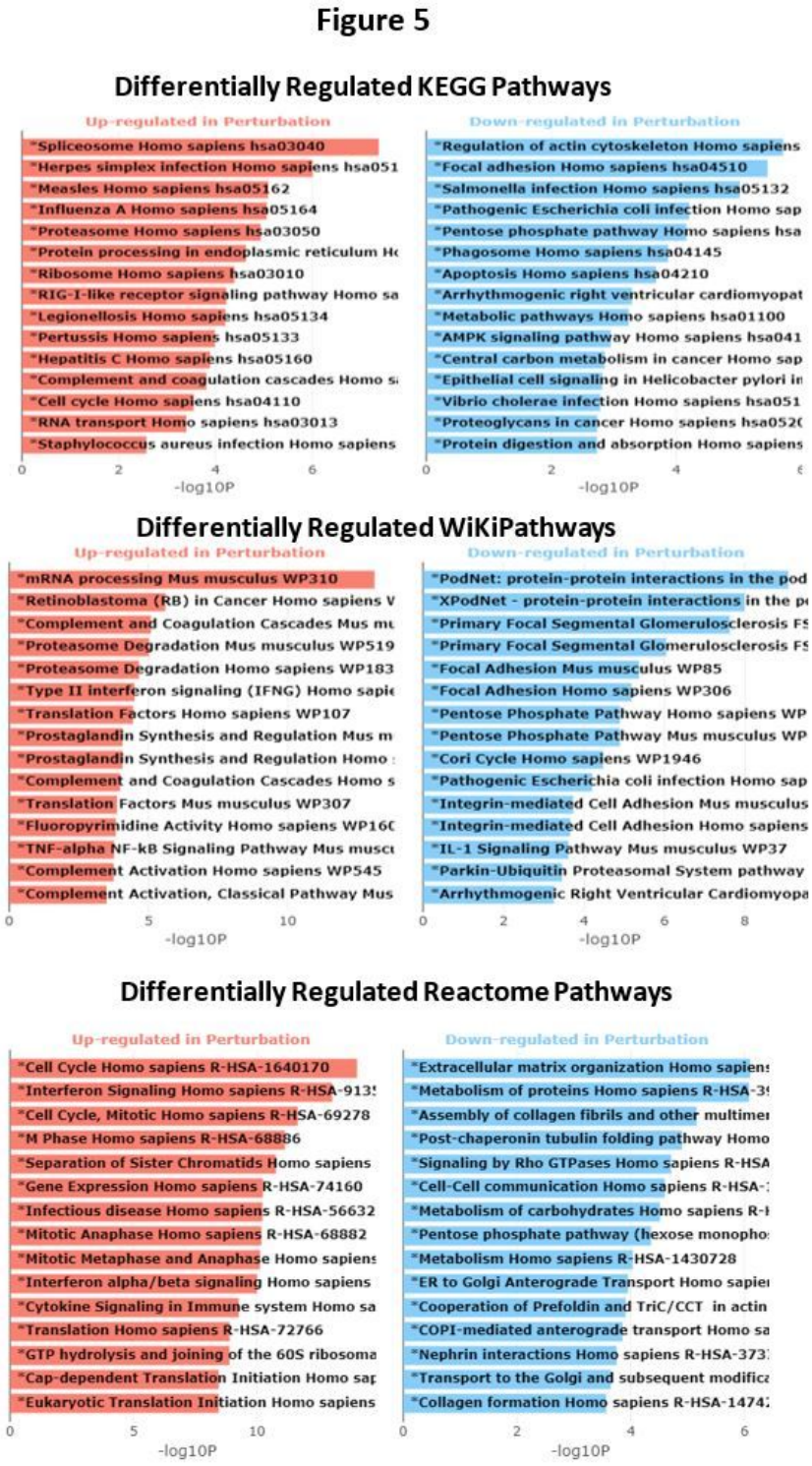

Figure 5

B. Pathway Enrichment Analysis. The figure contains interactive bar charts displaying the results of the pathway enrichment analysis generated using Enrichr. The x axis indicates the -log10(P-value) for each term. Significant terms are highlighted in bold.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AnnswerthesubmistinQur.docx