Research article

Unveiling common markers in COVID-19: ADAMTS2, PCSK9, and OLAH emerged as key differential gene expression profiles in PBMCs across diverse disease conditions

Mairembam Stelin Singh[1,2], PV Parvati Sai Arun[3,*] and Mairaj Ahmed Ansari[4,5,*]

[1] Department of Biochemistry, SCLS, Jamia Hamdard, New Delhi, India
[2] Department of Zoology, Rajiv Gandhi University, Arunachal Pradesh, India
[3] Department of Biotechnology, School of Sciences, Central Tribal University of Andhra Pradesh, Vizianagaram, Andhra Pradesh, India.
[4] Department of Biotechnology, SCLS, Jamia Hamdard, New Delhi, India
[5] Centre for Virology, SIST, Jamia Hamdard, New Delhi, India

* Correspondence: Email: arun.uoh@gmail.com; drmansari@jamiahamdard.ac.in.

Abstract: Diverse COVID-19 severity levels and a spectrum of clinical manifestations underscore the need to comprehend the underlying genetic mechanisms. Such knowledge is essential for improving disease management and therapeutic approaches. This study aims to explore and uncover pivotal genes and pathways linked to distinct COVID-19 conditions, providing insights into potential therapeutic avenues. Gene expression data from COVID-19 patients across different conditions were analyzed using differential gene expression analysis. Significant genes were subjected to pathway analysis and protein–protein interaction network analysis. Gene ontology was used to identify the functions of these genes. The genes *ADAMTS2*, *PCSK9*, and *OLAH* were upregulated across all disease conditions including SARS-CoV-2 bacterial coinfection, potentially serving as therapeutic targets. The proteins, including RPL and CEACAM, could serve as a potential therapeutic target. The deregulated genes were majorly involved in inflammation, lipid metabolism, and immune regulation. The study's findings reveal significant gene expression differences among COVID-19 disease conditions. These insights guide future research toward targeted therapies and an improved understanding of disease progression and long-term consequences.

Keywords: COVID-19; Asymptomatic COVID-19; Severe COVI-19; Gene regulation; Disease severity

1. Introduction

In late 2019, the emergence of COVID-19, driven by the severe acute respiratory syndrome coronavirus2 (SARS-CoV-2), triggered a global pandemic. Within 210 days, this virus had already claimed over 761,779 lives worldwide. SARS-CoV-2 primarily targets the upper respiratory tract, subsequently causing substantial damage to the lower respiratory tract, often resulting in severe pneumonia [1,2]. Vulnerable populations, particularly the elderly and those with preexisting comorbidities, face a heightened risk of severe health complications, marked by cytokine up-regulation and the onset of acute respiratory distress syndrome (ARDS) [3]. A substantial challenge in the management of COVID-19 lies in the stratification of disease severity andthe subsequent customization of treatment strategies [4,5]. COVID-19 encompasses a spectrum of severity,categorized into four distinct levels: asymptomatic, mild, moderate, and severe [6,7]. Each of these categories is characterized by unique clinical presentations, accompanied by differential gene regulation.Numerous research has been undertaken to identify the molecular markers closely linked to the severity of COVID-19 [8]. Among these markers, immunological indicators such as atypical immune cell counts and proinflammatory cytokine levels play a pivotal role in distinguishing between disease severities. Uniform treatment strategies, such as the use of antiviral medications, corticosteroids, or monoclonal antibodies, frequently are proved inadequate in accommodating the wide array of clinical presentations and immune reactions observed among distinct COVID-19 patients [9,10]. Throughout the COVID-19 pandemic, an extensive volume of transcriptomic data was amassed, encompassing RNA-sequencing datasets that delineate gene expression patterns in diverse COVID-19 patients [11]. Transcriptomic data analysis serves as a pivotal tool in discerning the molecular distinctionsamong patients experiencing varying degrees of COVID-19 severity, including asymptomatic, mild, moderate, and severe cases [12]. By scrutinizing the transcriptome, which encompasses the complete arrayof RNA molecules within a patient's cells, researchers can glean valuable insights into the underlying molecular mechanisms contributing to the diverse clinical outcomes observed in COVID-19 [13]. This analytical process typically entails the extraction of RNA from patient specimens, such as blood or respiratory tissues, followed by the application of high-throughput sequencing methods like RNA-seq toquantify the expression levels of thousands of genes [11–14]. Subsequently, bioinformatics tools are harnessed to pinpoint differentially expressed genes and pathwayslinked to disease severity. Comparative analysis of transcriptomic profiles across these patient groups unveils pivotal genes and regulatory pathways associated with immune responses, inflammation, viral replication, and tissue damage, thereby illuminating the molecular determinants influencing disease progression [15,16]. These revelations hold the potential to facilitate the development of more precise diagnostic tools, therapeutic approaches, and personalized treatments, ultimately advancing our comprehension of COVID-19 pathogenesis and patient outcomes.

The pathogenesis of COVID-19 represents a multifaceted and intricate interplay of various factors that contribute to the variable clinical outcomes observed in infected individuals, ranging from asymptomaticcases to severe disease presentations. To date, a comprehensive understanding of the underlying mechanisms governing the diverse disease severity spectrum remains elusive. This research endeavors to bridge this knowledge gap by focusing on the identification and characterization of pivotal genes that play a substantial role in the pathophysiological processes associated with distinct disease severities. Byelucidating the molecular determinants and their interactions, we aim to shed light on the intricate regulatory networks and pathways that are modulated by these genes. Such

insights may facilitate the development of targeted interventions and therapeutic strategies, ultimately aiding in the effectivemanagement of COVID-19 across its clinical spectrum.

2.   Materials and methods

2.1. Retrieval of datasets

In this research, we utilized the Gene Expression Omnibus (GEO) repository, a prominent public resource for the archival and dissemination of high-throughput functional genomic data derived from various techniques, including microarray and next-generation sequencing (NGS) [17]. Our study, focused on COVID-19 investigation, harnessed two specific datasets, namely GSE196822 and GSE166424, both of which were processed through the method of expression profiling by high-throughput sequencing. The GSE196822 dataset encompassed a total of 49 samples, while the GSE166424 dataset included 38 samples. All samples in these datasets were sequenced utilizing the GPL20301 Illumina HiSeq 4000 platform, specifically designed for *Homo sapiens.*

2.2. Dataset categorization for severity-associated gene analysis

The primary objective of this research is to elucidate the genetic underpinnings associated with the diverse clinical presentations observed in COVID-19, ranging from asymptomatic cases to severe manifestations. To accomplish this, a systematic categorization of the datasets has been undertaken, with a focus on disease severity stratification. This categorization facilitates the segregation of datasets into distinct groups corresponding to the varying clinical spectrums of COVID-19 severity. The detailed breakdown of dataset categorization, which delineates the criteria for severity assessment, is comprehensively documented in Table 1. Within the GSE196822 dataset, one sample from the healthy and COVID-bacterial coinfection category wasunavailable for analysis through the GEO2R platform. Consequently, these specific data points were excluded from the subsequent analytical procedures.

Table 1. Severity categorization criteria for COVID-19 datasets.

Disease severity	Datasets	
	GSE196822 (n = 49)	GSE166424 (n = 38)
Healthy	8	2
Asymptomatic COVID-19	8	30
Mild COVID-19	9	2
Moderate COVID-19	10	2
Severe COVID-19	7	2
Covid-bacterial coinfection	5	0

2.3. Differential gene expression (DGE) analysis

The raw datasets underwent preprocessing steps within the GEO2R platform, employing the R package robust multiarray (RMA). Subsequently, differential expression analysis was carried out, involving the application of cutoff criteria. Specifically, differentially expressed genes (DEGs) were extracted from the preprocessed dataset based on variables exhibiting a statistical significance with

adjusted p-values less than or equal to 0.05 and log-fold change values exceeding +1.5 for upregulated genes and falling below −1.5 for downregulated genes [18].

To visualize the significant DEGs, volcano and mean difference (MD) plots were generated, employing a default adjusted p-value threshold of 0.05 with the Benjamini and Hochberg [false discovery rate (FDR)] method. In these plots, upregulated genes are denoted in red, while downregulated genes are represented in blue. The volcano plot displays the log2-fold change against statistical significance (−log10 p-value), while the mean difference plot presents the average log2 expression. These graphical representations were generated utilizing the Limma Package.

2.4. Protein–protein interaction network (PPIN)

The top DEGs, identified as statistically significant, were subjected to a PPIN analysis using the STRING database. The STRING database is a comprehensive resource that integrates various types of protein relationships, encompassing both biological and physical interactions. A functional protein association network was constructed for the multiple DEGs *with Homo sapiens* as the reference organism [19]. The network was built with a high-confidence score threshold of 0.900 to ensure robustness and reliability. The active interactions within the network were derived from diverse sources, including text mining, experimental data, established databases, co-expression patterns, physical proximity in the genome, gene fusion events, and co-occurrence in previous studies. To further explore and visualize the significant interactions within the network, a k-means clustering approach was applied, maintaining a confidence score of 0.900, to identify major interactors.

2.5. Pathway analysis

The statistically significant DEGs were subjected to KEGG and gene ontology (GO) analysis. To maintain the study's biological relevance, we used the human species as the reference organism, and the FDR cutoff was set at 0.05, minimizing the likelihood of false positives [20].

3. Results

3.1. Data preprocessing

While the GSE196822 dataset includes samples from COVID-19 patients in India and GSE166424 includes samples from asymptomatic COVID-19 patients in Singapore, both datasets encompass a diverse range of disease severities and demographic characteristics. The inclusion of samples from different regions allows for the exploration of potential regional variations in disease presentation and molecular profiles.

By utilizing transcriptomic data from distinct geographical locations, our study has the unique opportunity to conduct cross-validation and comparative analyses. Through these analyses, we aim to identify common molecular pathways underlying COVID-19 pathogenesis, as well as unique regional signatures that may contribute to variations in disease severity and outcomes. This approach not only enhances the robustness and generalizability of our findings but also provides valuable insights into the global dynamics of COVID-19 and the molecular mechanisms driving its clinical manifestations.

Background correction and normalization are essential data preprocessing steps in microarray

analysis to ensure the accuracy and reliability of results. Background correction is performed to eliminate non-specific signals or noise in the microarray data. Microarray experiments may capture background noise, which can affect the quantification of gene expression levels. By subtracting the background signal, the accuracy of gene expression measurements will be greatly improved. Normalization is necessary to remove systematic variations between samples, such as differences in labeling efficiency, hybridization conditions, or scanner sensitivity. It ensures that the expression levels across different samples are directly comparable, making it possible to identify true biological differences. The observation of median-centered values in the boxplot (Figure 1) serves as an indicator of successful data normalization and cross-comparability. It signifies that the datasets have been effectively processed, aligning them for downstreamDGE analysis.
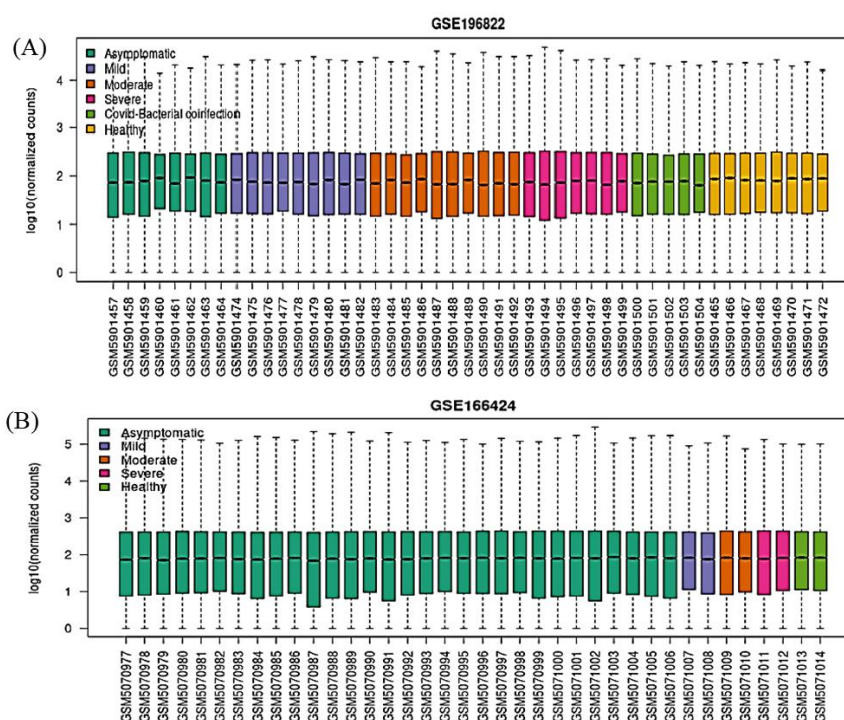

Figure 1. Normalization of the COVID-19 datasets. Note: Samples (A) and (B) in the datasets are colored according to the groups. The median-centered value ensures that all selected samples have identical value distribution. The plot represents data after the following normalization.

3.2. Differentially expressed genes

Analysis of differential gene expression is the principal application of RNA data sequencing. Genes that are variably controlled under different conditions can be identified using this strategy. The modified p-value was determined using the Benjamini-Hochberg method. For inclusion, an adjusted p-value of less than 0.05 was deemed significant. The log-fold change threshold was set at $|logFC| > 1.5$, indicating that entries with $logFC > 1.5$ were upregulated and those with $logFC < 1.5$ were downregulated. A volcano plotwas constructed utilizing the Limma package (Figure 2) to visually

depict the relationship between statistical significance (−log10 p-value) and the magnitude of change, as represented by the log2-fold change values. This graphical representation serves as an effective means to identify genes exhibiting substantial changes in expression while also highlighting their statistical significance.
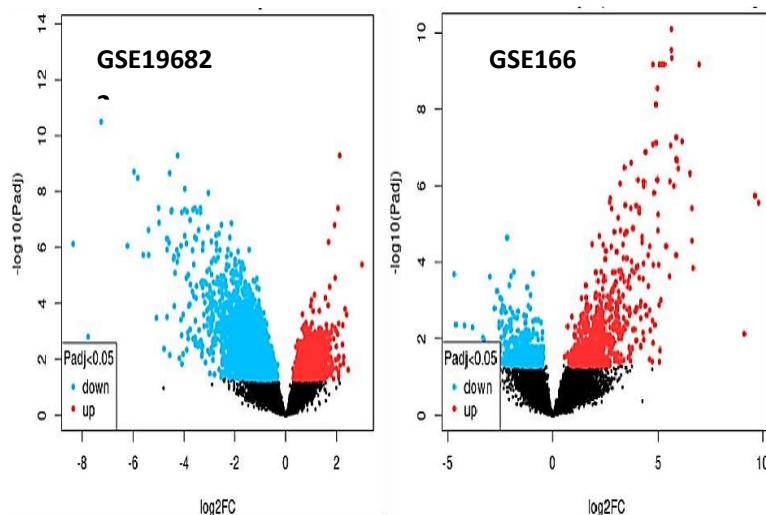

Figure 2. Differentially expressed genes in COVID-19 datasets. Note: The volcano plot is useful in representing differentially expressed genes. Highlighted genes represent the DEGs with a p-value cutoff set to 0.05, wherein the red color is indicative of upregulated genes and the blue color represents downregulated genes.

As previously outlined, the core objective of our study is to discern the genes associated with the clinical manifestations of asymptomatic, mild, moderate, and severe cases of COVID-19. To achieve this, we conducted a differential gene expression analysis, identifying the DEGs for each of the aforementioned clinical categories as follows.

3.2.1. Deregulated genes in asymptomatic COVID-19

The genes associated with various aspects of immune and inflammatory responses were found to be upregulated in asymptomatic COVID-19. *DEFA5*, *DEFA8P*, *PRTN3*, *CTSG*, and *ELANE* are defensin genes, which are known for their antimicrobial and immunomodulatory functions (Supplementary Table 1). Conversely, the downregulated genes in asymptomatic COVID-19 may indicate a dampened immune response or a specific modulation of host gene expression in these cases. Among the downregulated genes, the representatives of the TRAJ family were found to be significantly downregulated (Supplementary Table1).

3.2.2. Deregulated genes in mild COVID-19

The genes associated with neutrophils and their activation (e.g., *DEFT1P*, *CD177*, *PRTN3*, *ELANE*, and *OLAH*) were found to be associated with the clinical presentation of mild COVID-19 (Table 3). The downregulated genes in mild COVID-19 cases are less characterized (e.g., *SUSD2*,

ZNF181, *CELF2-AS1*, *CFAP97D2*, *NBEAL1*, and *LINC02067*), and their functions in the context of this disease remain unclear. In mild COVID-19 patients, we observed a commonality in gene expression patterns. Specifically, genes such as *PRTN3* and *ELANE*, which were found to be upregulated in asymptomatic cases, displayed a similar upregulation in mild cases (Supplementary Table 1 and Figure 3).
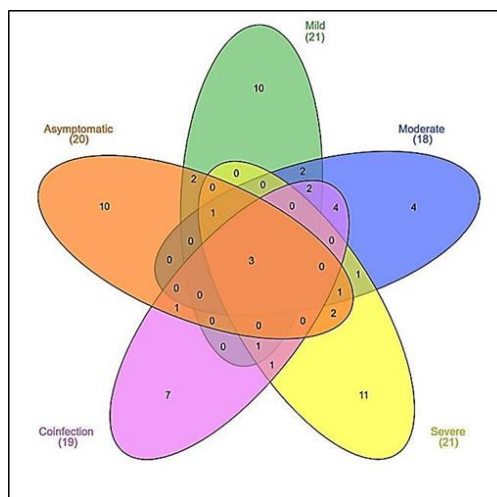

Figure 3. Overlap of gene regulation in different disease severities.

3.2.3. Deregulated genes in moderate COVID-19

In moderate COVID-19 cases, the observed upregulation of immune-related genes (*DEFT1P*, *OLAH*, *PCSK9*, *DAAM2*, *ADAMTS2*, and *DAAM2-AS1*) and the downregulation of genes associated with epigenetic regulation and cellular functions (*SLCO5A1*, *NRCAM*, *NOG*, *TET1*, *MAMDC2*, *VMO1*, and *SIRPG-AS1*) indicates a complex interplay of molecular responses to the virus (Figure 3 and Supplementary Table 2). Genes such as *TET1* and *CFAP97D2*, which exhibited downregulation in mild cases, displayed a similar downregulation in moderate cases (Figure 3).

3.2.4. Deregulated genes in severe COVID-1

Several of the genes that were upregulated display functions linked to immune and inflammatory responses. For instance, *DEFT1P*, *OLAH*, *DEFA5*, *ADAMTS2*, *DEFT1P2*, *CD177*, *PCSK9*, and *MAOA* are associated with various aspects of the immune system. The downregulated genes in severe COVID-19 include *FCER1A*, *CPA5*, *LRRN3*, *TRBJ1-2*, *STEAP1B-AS1*, *NOG*, *TRAJ49*, *TRAJ54*, *ZFP37*, and *TRBV4-1*. Many of these genes have diverse functions, with some associated with immune responses, cellular adhesion, and signaling pathways. The downregulation of *TRBJ1-2*, *TRAJ49*, and *TRAJ54*, which are related to T-cell-receptor gene segments, suggests potential immune suppression or altered T-cell responses in severe cases (Supplementary Table 2).

3.2.5. Deregulated genes in COVID-bacterial coinfection

Notably, *OLAH*, *ADAMTS2*, *DAAM2*, *DAAM2-AS1*, *MAOA*, *CD177*, *PCSK9*, *CLRN1-AS1*,

TIMP4, and *VSIG4* are amongthe upregulated genes. The downregulated genes in severe COVID-bacterial coinfection, including *CLEC4F*, *IFI27*, *VMO1*, *HLA-DPB2*, *SLCO5A1*, *HES4*, *LYPD2*, *SNORD141A*, *LINC02086*, and an additional *CLEC4F*, suggest potential alterations in immune and cellular responses (Supplementary Table 3).

3.3. Shared gene expression patterns across COVID-19 severity levels and coinfections

The data reveals intriguing patterns in gene expression across different COVID-19 severity levelsand coinfection scenarios. Notably, the genes *PRTN3* and *ELANE* show consistent upregulation in both asymptomatic and mild cases, indicating a shared immune response that might contribute to viral clearance without severe symptoms. In contrast, genes *TET1* and *CFAP97D2* are downregulated in mildand moderate cases, suggesting potential immune modulation in the transition between these severity levels (Supplementary Table 4 and Figure 3). The gene *NOG*, downregulated in both moderate and severe cases,plays a role in tissue homeostasis and cell differentiation. Its decreased expression could disrupt tissue repair mechanisms, possibly contributing to more severe clinical outcomes (Supplementary Table 4 and Figure 3). Additionally, in severe cases with bacterial coinfections, MAOA is upregulated, reflecting an alteredhost response, emphasizing the impact of coinfections on the immune response (Supplementary Table 4 and Figure 3).

LYPD2's consistent downregulation in coinfection and asymptomatic cases may signify immune suppression or altered responses to viral and bacterial pathogens (Supplementary Table 4 and Figure 3). In theasymptomatic-to-severe comparison, *DEFA5*'s upregulation implies robust defense mechanisms, while *TRAJ49* downregulation may affect immune responses by altering T-cell-receptor gene segment expression (Supplementary Table 4 and Figure 3). *DEFT1P* consistently shows upregulation in all three categories (asymptomatic, mild, and moderate), emphasizing its crucial role in mounting and maintaining immune responses (Supplementary Table 4 and Figure 3). Genes *SLCO5A1* and *SNORD141A* exhibit shared downregulation across mild, moderate, and coinfection cases, suggesting potential immune suppression (Supplementary Table 4 and Figure 3). On the other hand, *CD177* is upregulated in cases with mild, severe, and coinfections, implying a common immune response signature (Supplementary Table 4 and Figure 3).

3.4. Protein–protein interaction network (PPIN)

The PPIN analysis (Figure 4) represents the interactions between the top DEGs. These interactions reveal how different proteins encoded by these genes collaborate and communicate within a cellular context. The presence of the RPL (ribosomal protein large subunit) family as a major interaction (Figure 4) suggests a significant role of ribosomal proteins. Their prominence may indicate an upregulation ofprotein synthesis or specific cellular processes related to ribosomal functions in COVID-19. The identification of the CEACAM (carcinoembryonic antigen-related cell adhesion molecule) family as a major interaction (Figure 4) implies the importance of cell adhesion and intercellular communication.The identification of the NLR (NOD-like receptor) protein family as a major interaction (Figure 4) implies an association with the immune system and inflammation. NLR proteins are involved in the detection of pathogens and cellular stress, playing a key role in the regulation of innate immunity. Theirpresence may indicate the activation of specific immune pathways in response to COVID-19.
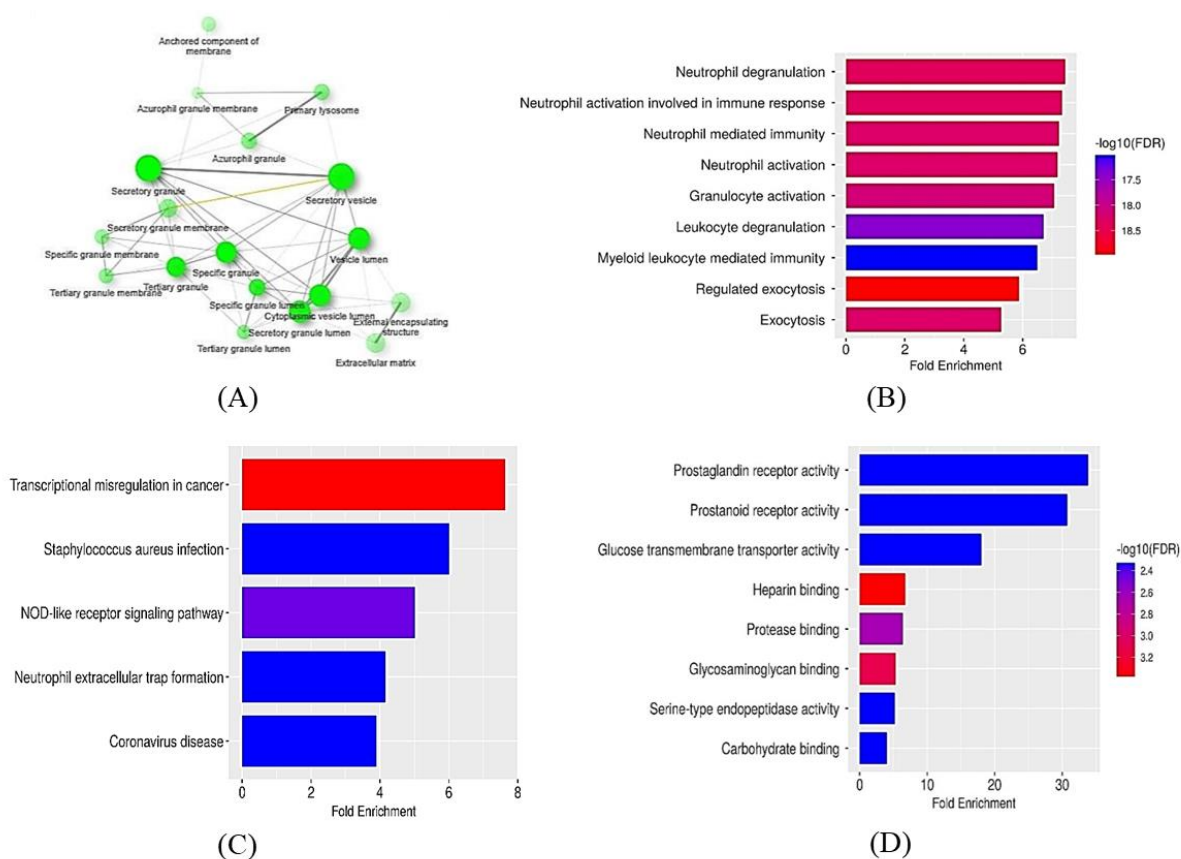
Figure 4. Protein–protein interaction network (PPIN) analysis of top differentially expressed genes(DEGs). Note: PPIN illustrates interactions among top DEGs.

3.5. Pathway analysis

In the KEGG pathway analysis, the results reveal a multifaceted landscape of pathway involvement in COVID-19. Notably, the presence of "transcriptional misregulation in cancer" suggests potentiallinks to cancer-related gene expression dysregulation. "*Staphylococcus aureus* infection" points to the condition's influence on host responses to bacterial infections. The "NOD-like receptor signalingpathway" underscores an activated innate immune response. The inclusion of "neutrophil extracellular trap (NET) formation" implies an ongoing immune response featuring neutrophil engagement (Figure 5).

The biological processes, cellular components, and molecular functions highlighted in the analysis collectively underscore the prominent role of immune responses and granule activation in the condition. This signifies an active immune response, potentially involving neutrophils and regulated exocytosis of cellular components. The significance of specific granules, lysosomes, and secretory vesicles as cellular components underscores their integral role in the studied condition, while the presence of azurophil granules and secretory vesicles further supports an active immune response (Figure 5). Additionally, the molecular functions related to receptor activity, including prostaglandin and prostanoid receptor activity, suggest the involvement of signaling pathways associated with these receptors. The identification of carbohydrate binding and glycosaminoglycan binding implies interactions with carbohydrates and proteoglycans in the molecular landscape of the condition (Figure 5).

Figure 5. Pathway analysis of the significant genes. **(A)** GO biological processes: Biological processes primarily involve neutrophil functions, granule activation, and regulated exocytosis. **(B)** GO cellular component: Cellular components indicate the involvement of granules, lysosomes, and secretory vesicles. **(C)** KEGG pathways: Pathways related to cancer, bacterial infection, immune signaling, neutrophilactivity, and coronavirus disease are significant in the analyzed DEGs. **(D)** GO molecular function: Molecular functions primarily include receptor activity, carbohydratebinding, and protease activity.

3.6. Host–protein interaction

The interactions between the SARS-CoV-2 virus and human proteins by utilizing the IntAct database. The IntAct database serves as a comprehensive repository of information related to human proteins and their interactions with genes. To discern the specific interactions between the virus and our protein of interest, we applied a rigorous filtering process. This filtering involved considering various criteria, such as the molecular interaction (MI) score, which helped us identify positive interactions, and detectionmethods employed to ensure the reliability of the interactions. Our focus was primarily on interactions that exhibited a direct and physical association between the virus and human proteins, indicative of a close and biologically significant connection.

These interactions occurred within various types of host cells, which provides valuable insights into the infection process. Specifically, the majority of host interactions were detected in embryonic kidney cells,in vitro cells (typically referring to cells cultured in a laboratory setting), carcinoma cells

(which are cancer cells), and Aethiops simian cells (Supplementary Table 5), which are commonly used in virologyresearch due to their susceptibility to viral infections. These observations underscore the relevance of these cellular environments in the context of SARS-CoV-2 infection and suggest that the virus may exhibit specific affinities for these particular cell types.

4. Discussion

The global COVID-19 pandemic has created a profound health crisis, manifesting a spectrum of disease severity, spanning from asymptomatic cases to severe conditions, presenting diverse clinical symptoms [21]. However, the intricacies underlying the progression from asymptomatic to severe stages remain poorly understood, representing a significant research gap. Addressing this knowledge deficit is imperative, as it necessitates specialized investigations to discern the pivotal genes contributing to distinct disease severities and the accompanying modulated pathways [22]. The identification of such genes and pathways holds promise for the development of targeted therapeutic interventions based on these pathways. Additionally, understanding the progression of COVID-19 from an asymptomatic to a severe stage offers critical insights for devising effective prevention and treatmentstrategies [6,7]. While prior studies have investigated the pathogenesis of COVID-19 and recognizedpotential therapeutic targets, knowledge gaps, and research priorities persist. One example is the elusive understanding of the long-term consequences of COVID-19, hindered by a scarcity of high-quality research, which remains a barrier to a comprehensive definition of conditions like long COVID and post-acute COVID.

To address the aforementioned research gap, we conducted a comprehensive analysis of DGE withinCOVID-19 datasets. This analysis enabled the identification of distinct gene profiles associated withvarying severities of COVID-19.

Age, sample processing procedures, comorbidities, and medication usage are recognized as critical determinants that can significantly influence gene expression patterns in COVID-19 patients, necessitating careful consideration in the interpretation of findings. While our study did not explicitly account for age groups or specific comorbidity profiles in the transcriptomic analysis, we ensured the inclusion of a diverse range of individuals with varying disease severities and demographic characteristics within the GSE196822 and GSE166424 datasets. Future research endeavors could delve into subgroup analyses to elucidate differential gene expression patterns among different age groups or specific comorbidity profiles, thus providing further insight into disease pathogenesis. Moreover, variations in sample processing protocols and medication usage among patients have the potential to introduce biases in transcriptomic analyses. Although detailed clinical metadata and medication histories were not available in the GEO datasets utilized in our study, we implemented stringent quality-control measures during data preprocessing to mitigate technical artifacts and ensure data reliability.

In asymptomatic COVID-19, several genes that are related to immune and inflammatory responses were found to be upregulated. *DEFA5*, *DEFA8P*, *PRTN3*, *CTSG*, and *ELANE* are defensin genes, which are known for their antimicrobial and immunomodulatory functions [23]. The upregulation of these genes may suggest an activated immune response in asymptomatic cases, potentially aiding in controlling viral replication. *ADAMTS2* is linked to extracellular matrix remodeling and could play a role in tissue repair and inflammation resolution [24]. *PCSK9* may be involved in lipid metabolism and immune regulation [25]. Among the downregulated genes, *TRAJ49*,

TRIM51EP, *UICLM*, *LYDP2*,*IGLV4-60*, *TRDJ3*, *SHISA4*, *TRDV1*, *TRAJ10*, and *TRAJ33* do not have well-documented functions in the context of COVID-19. However, their downregulation may still suggest some level of immunesuppression.

The upregulated genes in mild COVID-19, for example, *DEFT1P*, *CD177*, *PRTN3*, *ELANE*, and *OLAH*, are associated with neutrophils and their activation. These genes are critical components of the immune response, and their upregulation might indicate a heightened immune reaction in asymptomatic individuals, which could contribute to viral control without clinical symptoms [26]. *TMEM54* and *CLRN1* have roles that need further investigation in the context of COVID-19. The downregulated gene *TET1* is involved in DNA demethylation, and its downregulation might influence epigenetic regulation [27]. *SLCO5A1* is associated with organic anion transport and may have implications for drug metabolism or immune responses [28]. *KLRB1* is associated with natural killer cells and T-cell function, and its downregulation may indicate immune suppression [29]. Other downregulated genes such as *SUSD2*, *ZNF181*, *CELF2-AS1*, *CFAP97D2*, *NBEAL1*, and *LINC02067* are less characterized.

In the context of moderate COVID-19, the significant upregulation of PCSK9 is particularly noteworthy, as it is known to play a role in lipid metabolism and immune regulation. In the context of COVID-19, elevated *PCSK9* expression may indicate a potential link between lipid metabolism and the immune response in moderate cases [25]. Conversely, the downregulated genes in moderate COVID-19 include *SLCO5A1*, *SNORD141A*, *CFAP97D2*, *NRCAM*, *NOG*, *TET1*, *MAMDC2*, *VMO1*, *SIRPG-AS1*, and an additional *SLCO5A1*. Some of these genes, such as *TET1*, are associatedwith epigenetic regulation, suggesting potential alterations in DNA methylation patterns in moderatecases [27]. Others, like *SLCO5A1* and *NRCAM*, have roles in cell adhesion and transport processes,the downregulation of which could impact various cellular functions during infection [28].

The upregulated genes in severe COVID-19—*DEFT1P*, *OLAH*, *DEFA5*, *ADAMTS2*, *DEFT1P2*, *CD177*, *PCSK9*, and *MAOA*—are associated with various aspects of the immune system. *DEFA5*, *DEFT1P*, and *CD177* are involved in neutrophil activation and host defense mechanisms [26]. Upregulation of these genes may indicate a heightened immune response in severe cases to combat the viral infection. *MAOA* is linked to the monoamine pathway and might be involved in regulating inflammatory responses [30]. The downregulation of *TRBJ1-2*, *TRAJ49*, and *TRAJ54*, which are related to T-cell-receptor gene segments, suggests potential immune suppression or altered T-cell responses in severe cases [31].

Lastly, in Covid-bacterial coinfection, *OLAH* and *ADAMTS2* are genes associated with extracellularmatrix remodeling that might be involved in tissue repair processes. *CD177*, *PCSK9*, and *MAOA* areknown for their roles in immune regulation and lipid metabolism. Their upregulation may indicate animmune response and potential involvement in the host's defense against bacterial coinfection. *DAAM2* and *DAAM2-AS1* have functions that require further investigation in the context ofcoinfection. *TIMP4* and *VSIG4* are involved in tissue remodeling and immune regulation, respectively, and their upregulation suggests a potential immune response against the coinfection [32]. *CLRN1-AS1*'s function in the context of coinfection remains to be elucidated. *IFI27* is an interferon-induced gene, and its downregulation may indicate a modulation of the interferon response in coinfection.

In the present study, the genes *ADAMTS2*, *PCSK9*, and *OLAH* were found to be significantly upregulated across all disease severities. These findings align with past research wherein it is reported that ADAMTS2, a metalloprotease enzyme, is involved in the proinflammatory response to COVID-

19 [24]. A study found that *PCSK9*, a gene that encodes for a protein involved in lipid metabolism, is upregulated in severe COVID-19 patients [33]. *OLAH*, a gene that encodes for an enzyme involved in lipid metabolism, was upregulated in severe COVID patients [33].

The proteins RPL, CEACAM, and NLR were identified as major interactors in the PPIN [34]. A study found that CEACAM5 and CEACAM6 are highly expressed in developing neutrophils/neutrophil progenitors in COVID-19 [35]. The study also found that CEACAM5 is an important surface-attachment factor that facilitates the entry of the Middle East respiratory syndrome coronavirus (MERS-CoV). Disrupting the interaction between CEACAM5 and MERS-CoV spiked with anti-CEACAM5 antibody, recombinant CEACAM5 protein, or small interfering RNA (siRNA) knockdown of CEACAM5 significantly inhibited the entry of MERS-CoV [36]. The neutrophil/lymphocyte ratio (NLR) has been suggested as a good predictive marker of disease severity and mortality in COVID-19 patients [37].

The findings from pathway analysis suggest that significant DEGs are involved in immune responses,particularly neutrophil-related processes, and may play a role in response to infections, both bacterial and viral. The presence of pathways related to cancer and coronavirus disease implies potential connections between the pathologies. These findings can guide further research intothe molecular mechanisms underlying the studied condition and provide a foundation for understanding its pathogenesis and potential therapeutic targets.

ML algorithms have emerged as powerful tools for analyzing large datasets and uncovering complex relationships between patient characteristics and disease outcomes. In the context of COVID-19, MLholds promise for prognostic evaluation, biomarker identification, and treatment optimization [38]. By leveraging diverse data sources such as clinical records, imaging data, and genomic information, ML algorithms can assist in predicting disease severity, identifying relevant biomarkers, and personalizing treatment strategies. However, it is crucial to address certain challenges associated with the adoption of ML in clinical practice [39]. One major hurdle is the interpretability of ML models, often referred to as the "black box" problem, which limits clinicians' ability to understand and trust the predictions generated by these algorithms. Additionally, there is a need for explainable AI modelsthat provide insights into the decision-making process of ML algorithms, enhancing their utility in clinical settings. Furthermore, efforts should be directed toward improving clinicians' understandingof ML models to facilitate their integration into routine practice [40]. Despite these challenges, the potential of ML in revolutionizing COVID-19 research and clinical care is substantial, and further exploration and refinement of ML-based approaches are warranted to unlock their full potential in improving patient outcomes.

5. Conclusions

The comprehensive analysis of differential gene expression in varying severities of COVID-19 has shed light on the intricate molecular mechanisms underlying the disease's progression. We observed distinct gene profiles associated with asymptomatic, mild, moderate, and severe COVID-19, as well as coinfections. Notably, genes like ADAMTS2, PCSK9, and OLAH were consistently upregulated across all disease severities, aligning with existing research on their roles in inflammation, lipid metabolism, and immune regulation. These findings emphasize the significance of immune responses, particularly neutrophil-related processes, in the context of COVID-19 severity. Understanding the gene expression patterns associated with different disease severities not only

contributes to our comprehension of COVID-19 pathogenesis but also holds promise for the development of targeted therapeutic interventions.

The conclusions drawn from our study provide a foundational understanding of the molecular mechanisms underlying COVID-19 severity, offering valuable insights into potential avenues for future research and clinical applications. Moving forward, our findings could significantly impact the development of personalized medicine approaches for the treatment of severe COVID-19. Specifically, the identification of distinct gene expression profiles associated with varying disease severities presents an opportunity to tailor treatment strategies based on individual patient characteristics. By leveraging transcriptomic data and integrating it with clinical parameters, future research endeavors could refine prognostic models to predict disease progression and treatment response with greater accuracy. This personalized approach holds immense potential to optimize clinical decision-making, enabling healthcare professionals to tailor interventions based on the specific molecular signatures observed in patients. Strategies such as immunomodulation, anti-inflammatory therapy, or antiviral treatments could be tailored to target the specific molecular mechanisms identified in different severities of the disease, potentially improving clinical outcomes for patients. Additionally, the integration of multi-omics data, including genomic, transcriptomic, and proteomic data, could provide a comprehensive understanding of the complex interplay between genetic factors and disease severity, paving the way for more precise and effective treatment approaches. However, it is important to acknowledge the limitations of our present study. We must address the need for larger datasets and more extensive clinical validation to strengthen the generalizability of our findings.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

All the authors declare that there is no conflict of interest.

References

1. Yang H, Lan Y, Yao X, et al. (2020) The chest CT features of coronavirus disease 2019 (COVID-19) in China: A meta-analysis of 19 retrospective studies. *Virol J* 17: 159. https://doi.org10.1186/s12985-020-01432-9

2. Muralidar S, Ambi SV, Sekaran S, et al. (2020) The emergence of COVID-19 as a global pandemic: Understanding the epidemiology, immune response and potential therapeutic targets of SARS-CoV-2. *Biochimie* 179: 85–100. https://doi.org 10.1016/j.biochi.2020.09.018

3. Chiumello D, Modafferi L, Fratti I (2022) Risk factors and mortality in elderly ARDS COVID-19 compared to patients without COVID-19. *J Clin Med* 11: 5180. https://doi.org/10.3390/jcm11175180

4. Malik YS, Kumar N, Sircar S, et al. (2020) Coronavirus disease pandemic (COVID-19): Challenges and a global perspective. *Pathogens* 9: 519. https://doi.org/10.3390/pathogens9070519

5. Cunningham AC, Goh HP, Koh D (2020) Treatment of COVID-19: Old tricks for new challenges. *Crit Care* 24: 91. https://doi.org/10.1186/s13054-020-2818-6

6. Kim GU, Kim MJ, Ra SH, et al. (2020) Clinical characteristics of asymptomatic and symptomatic patients with mild COVID-19. *Clin Microbiol infect* 26: 948. https://doi.org/10.1016/j.cmi.2020.04.040

7. Gandhi RT, Lynch JB, Del Rio C (2020) Mild or moderate Covid-19. *N Engl J Med* 383: 1757-1766. https://doi.org/10.1056/NEJMcp2009249

8. Sayampanathan AA, Heng CS, Pin PH, et al. (2021) Infectivity of asymptomatic versus symptomatic COVID-19. *Lancet* 397: 93-94. https://doi.org/10.1016/S0140-6736(20)32651-9

9. Yin SW, Zhou Z, Wang JL, et al. (2021) Viral loads, lymphocyte subsets and cytokines in asymptomatic, mildly and critical symptomatic patients with SARS-CoV-2 infection: A retrospective study. *Virol Jl* 18: 126. https://doi.org/10.1186/s12985-021-01597-x

10. Iwamura APD, Tavares da Silva MR, Hümmelgen AL, et al. (2021) Immunity and inflammatory biomarkers in COVID-19: A systematic review. *Rev Med Virol* 31: e2199. https://doi.org/10.1002/rmv.2199

11. Sadanandam A, Bopp T, Dixit S, et al. (2020) A blood transcriptome-based analysis of disease progression, immune regulation, and symptoms in coronavirus-infected patients. *Cell Death Discov* 6: 141. https://doi.org/10.1038/s41420-020-00376-x

12. Chakraborty C, Sharma AR, Bhattacharya M, et al. (2021) Understanding gene expression and transcriptome profiling of COVID-19: An initiative towards the mapping of protective immunity genes against SARS-CoV-2 infection. *Front Immunol* 12: 724936. https://doi.org/10.3389/fimmu.2021.724936

13. Gardinassi LG, Souza COS, Sales-Campos H, et al. (2020) Immune and metabolic signatures of COVID-19 revealed by transcriptomics data reuse. *Front Immunol* 11: 1636. https://doi.org/10.3389/fimmu.2020.01636

14. Kwan PKW, Cross GB, Naftalin CM, et al. (2021) A blood RNA transcriptome signature for COVID-19. *BMC Med Genomics* 14: 155. https://doi.org/10.1186/s12920-021-01006-w

15. Li G, Ruan S, Zhao X, et al. (2021) Transcriptomic signatures and repurposing drugs for COVID-19 patients: Findings of bioinformatics analyses. *Comput Struct Biotec J* 19: 1–15. https://doi.org/10.1016/j.csbj.2020.11.056

16. El-Aarag SA, Mahmoud A, ElHefnawi M (2022) Identifying potential novel insights forCOVID-19 pathogenesis and therapeutics using an integrated bioinformatics analysis of host transcriptome. 194: 770–780. https://doi.org/10.1016/j.ijbiomac.2021.11.124

17. Barrett T, Wilhite SE, Ledoux P, et al. (2012) NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res* 41: D991–D995. https://doi.org/10.1093/nar/gks1193

18. Venugopal BS, Dinesh S, Sharma S, et al. (2023) Exploring the therapeutic potential of green tea phytocompounds for pancreatic cancer: An mRNA differential gene expression and pathway analysis study. *Biomedicine* 43: 1231–1238. https://orcid.org/0009-0006-2673-1858

19. Vora D, Kapadia H, Dinesh S, et al. (2023) *Gymnema sylvestre* as a potential therapeutic agent for PCOS: insights from mRNA differential gene expression and molecular docking analysis. *Futur J Pharm Sci* 9: 76. https://doi.org/10.1186/s43094-023-00529-6

20. Kapadia H, Dinesh S, Sharma S, et al. (2023) Integrating network pharmacology and molecular docking for the identification of key genes and therapeutic targets of Nigella sativa in multiple sclerosis treatment. *Biomedicine* 43: 936–944. https://doi.org/10.51248/.v43i3.2867

21. Bhowmik R, Nath R, Sharma S, et al. (2022) High-throughput screening and dynamic studies of selected compounds against SARS-CoV-2. *Int J Appl Pharm* 14: 251–260.

22. Alabdali AYM, Chinnappan S, Abd Razik BM, et al. (2021) Impact of COVID-19 on multiple body organ failure: A review. 54–59.

23. Abdeen S, Bdeir K, Abu-Fanne R, et al. (2021) Alpha-defensins: Risk factor for thrombosis in COVID-19 infection. *Br J Haematol* 194: 44–52. https://doi.org/10.1111/bjh.17503

24. de Seabra Rodrigues Dias IR, Cao Z, Kwok HF (2022) Adamalysins in COVID-19–Potential mechanisms behind exacerbating the disease. *Biomed Pharm* 150: 112970. https://doi.org/10.1016/j.biopha.2022.112970

25. Barkas F, Milionis H, Anastasiou G, et al. (2021) Statins and PCSK9 inhibitors: What is their role in coronavirus disease 2019? *Med Hypotheses* 146: 110452. https://doi.org/10.1016/j.mehy.2020.110452

26. Wang J, Li Q, Yin Y, et al. (2020) Excessive neutrophils and neutrophil extracellular traps in COVID-19. *Front Immunol* 11: 2063. https://doi.org/10.3389/fimmu.2020.02063

27. Kouidou S, Malousi A, Andreou AZ (2020) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection: Triggering a lethal fight to keep control of the ten-eleven translocase (TET)-associated DNA demethylation? *Pathogens* 9: 1006. https://doi.org/10.3390/pathogens9121006

28. Couto N, Al-Majdoub ZM, Gibson S, et al. (2020) Quantitative proteomics of clinically relevant drug-metabolizing enzymes and drug transporters and their intercorrelations in the human small intestine. *Drug Metab Dispos* 48: 245–254. https://doi.org/10.1124/dmd.119.089656

29. Hachim M, Hachim I, Hannawi S (2022) Killer cell lectin-like receptor subfamily B member 1 (KLRB1), lymphocyte function-associated antigen 3 (CD58), and Niban1 (FAM129A) are unique markers for developed (central and effector) memory T cells in rheumatoid arthritis. *Saudi Med J* 43: 646.

30. Mpekoulis G, Kalliampakou KI, Milona RS, et al. (2022) Significance of catecholamine biosynthetic/metabolic pathway in SARS-CoV-2 infection and COVID-19 severity. *Cells* 12: 12. https://doi.org/10.3390/cells12010012

31. Hou X, Wang G, Fan W, et al. (2021) T-cell receptor repertoires as potential diagnostic markers for patients with COVID-19. *Int J Infect Dis* 113: 308–317. https://doi.org/10.1016/j.ijid.2021.10.033

32. Liang X, Li P, Jiang J, et al. (2023) Transcriptomics unveils immune metabolic disruption and a novel biomarker of mortality in patients with HBV-related acute-on-chronic liver failure. *JHEP Rep* 5: 100848. https://doi.org/10.1016/j.jhepr.2023.100848

33. Rombauts A, Bódalo Torruella M, Abelenda-Alonso G, et al. (2023) Dynamics of gene expression profiling and identification of high-risk patients for severe COVID-19. *Biomedicines* 11: 1348. https://doi.org/10.3390/biomedicines11051348

34. Rofeal M, Abd El-Malek F (2020) Ribosomal proteins as a possible tool for blocking SARS-CoV 2 virus replication for a potential prospective treatment. *Med Hypotheses* 143: 109904. https://doi.org/10.1016/j.mehy.2020.109904

35. Huang R, Meng T, Zha Q, et al. (2021) The predicting roles of carcinoembryonic antigen and its underlying mechanism in the progression of coronavirus disease 2019. *Crit Care* 25: 234. https://doi.org/10.1186/s13054-021-03661-y

36. Chan CM, Chu H, Wang Y, et al. (2016) Carcinoembryonic antigen-related cell adhesion molecule 5 is an important surface attachment factor that facilitates entry of Middle East respiratory syndrome coronavirus. *J Virol* 90: 9114–9127. https://doi.org/10.1128/jvi.01133-16

37. Parthasarathi A, Padukudru S, Arunachal S, et al. (2022) The role of neutrophil-to-lymphocyte ratio in risk stratification and prognostication of COVID-19: A systematic review and meta-analysis. *Vaccines* 10: 1233. https://doi.org/10.3390/vaccines10081233

38. Yagin FH, Cicek İB, Alkhateeb A, et al. (2023) Explainable artificial intelligence model for identifying COVID-19 gene biomarkers. *Comput Biol Med* 154: 106619. https://doi.org/10.1016/j.compbiomed.2023.106619

39. Bouchareb Y, Khaniabadi PM, Al Kindi F, et al. (2021) Artificial intelligence-driven assessment of radiological images for COVID-19. *Comput Biol Med* 136: 104665.

40. Došilović FK, Brčić M, Hlupić N (2018) Explainable artificial intelligence: A survey In: *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. https://doi.org/10.23919/MIPRO.2018.8400040