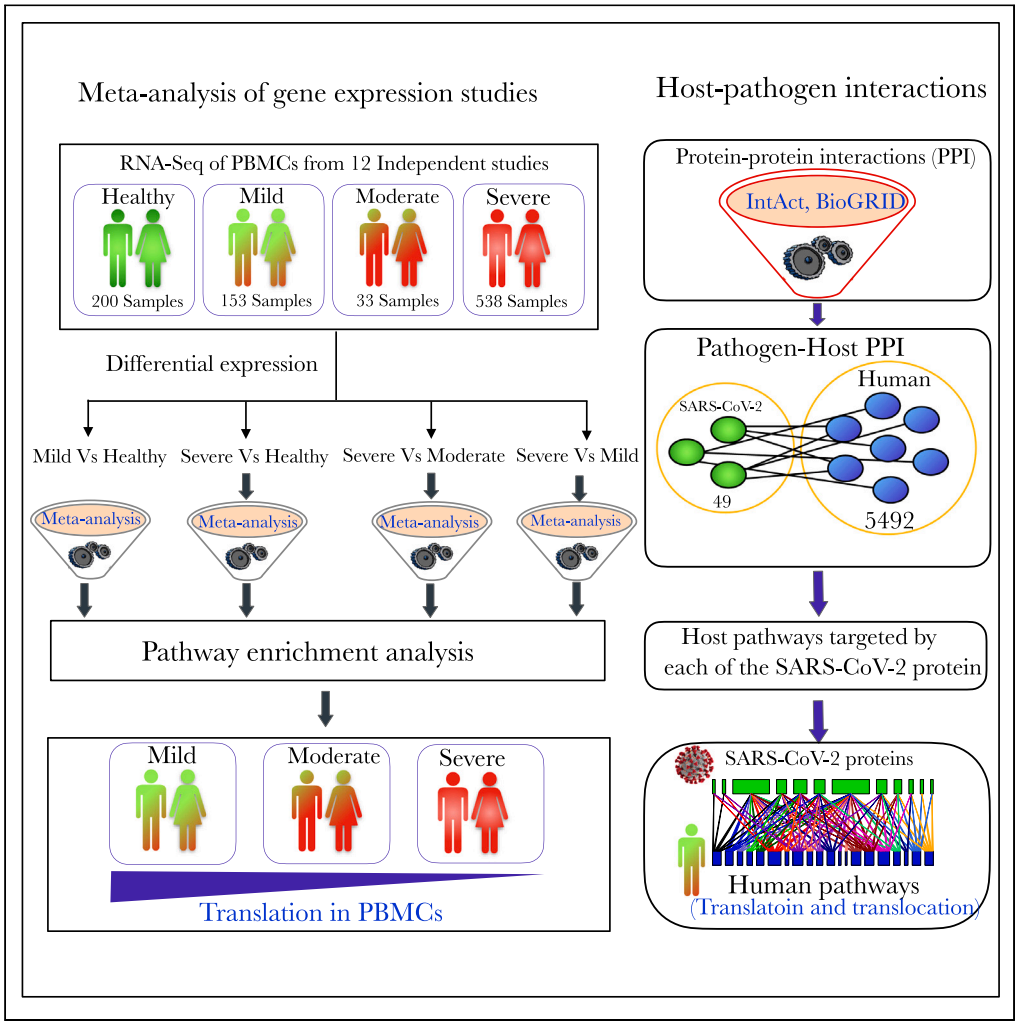
iScience

CellPress
OPEN ACCESS

Article

Systems-wide view of host-pathogen interactions across COVID-19 severities using integrated omics analysis



Mairembam Stelin Singh, Anand Pyati, R. Devika Rubi, Rajasekaran Subramanian, Vijaykumar Yogesh Muley, Mairaj Ahmed Ansari, Sailu Yellaboina

bio.sailu@gmail.com

Highlights

Multi-omics insights into virus-host interplay within PBMCs in Covid-19

SARS-CoV-2 hijacks host protein translation and translocation pathways

Host translation is upregulated in mild and downregulated in severe disease conditions

Singh et al., iScience *27*, 109087
March 15, 2024
https://doi.org/10.1016/j.isci.2024.109087

iScience

CellPress
OPEN ACCESS

Article

Systems-wide view of host-pathogen interactions across COVID-19 severities using integrated omics analysis

Mairembam Stelin Singh,[1,2] Anand Pyati,[3] R. Devika Rubi,[4] Rajasekaran Subramanian,[4] Vijaykumar Yogesh Muley,[5] Mairaj Ahmed Ansari,[6,7] and Sailu Yellaboina[3,8,*]

SUMMARY

The mechanisms explaining the variability in COVID-19 clinical manifestations (mild, moderate, and severe) are not fully understood. To identify key gene expression markers linked to disease severity, we employed an integrated approach, combining host-pathogen protein-protein interaction data and viral-induced host gene expression data. We analyzed an RNA-seq dataset from peripheral blood mononuclear cells across 12 projects representing the spectrum of disease severity. We identified genes showing differential expression across mild, moderate, and severe conditions. Enrichment analysis of the pathways in host proteins targeted by each of the SARS-CoV-2 proteins revealed a strong association with processes related to ribosomal biogenesis, translation, and translocation. Interestingly, most of these pathways and associated cellular machinery, including ribosomal biogenesis, ribosomal proteins, and translation, were upregulated in mild conditions but downregulated in severe cases. This suggests that COVID-19 exhibits a paradoxical host response, boosting host/viral translation in mild cases but slowing it in severe cases.

INTRODUCTION

The global COVID-19 pandemic, driven by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is an unparalleled global health crisis, resulting in an estimated 14.83 million reported deaths worldwide.[1] The virus first enters the body through the respiratory tract and can infect cells in a wide variety of tissues, including the lungs, heart, kidneys, liver, gastrointestinal tract, central nervous system, and blood.[2–5] COVID-19 disease severity can vary widely among individuals.[6] Mild cases typically present flu-like symptoms such as fever, cough, and fatigue, and patients usually recover without the need for hospitalization. Moderate cases involve more pronounced symptoms, including shortness of breath and pneumonia, and may require medical intervention. Severe cases of COVID-19 can exhibit acute respiratory distress syndrome (ARDS) and multi-organ failure, requiring intensive care and mechanical ventilation.

Protein-protein interactions between viral and host proteins are essential for various steps in the viral life cycle, including viral entry into host cells, viral protein synthesis, assembly of new viral particles, and release of mature viruses.[7] Therefore, identifying interaction networks between viruses and human proteins is crucial for understanding the molecular mechanisms by which viruses hijack and manipulate host cell processes under the different conditions of COVID-19 disease severities. High-throughput proteomics methods such as affinity purification and mass-spectrophotometry have identified several interactions of SARS-CoV-2 proteins with human proteins which are involved in a variety of viral processes, including replication, transcription, and translation.[8–11]

The detection of long nucleocapsid sequences of SARS-CoV-2 in peripheral blood mononuclear cells (PBMCs)[12] and the production of virus progeny *in vitro*[13] suggest that PBMCs may harbor SARS-CoV-2 and could serve as a potential source for multi-organ spread in the human body, potentially elucidating certain pathological observations in organs like the brain and heart. In addition, PBMCs play a pivotal role as versatile immunological and pathological responders. Therefore, analyzing PBMC RNA-seq data plays a pivotal role in unraveling the intricate web of viral host-pathogen interactions, especially in the context of viral infections like SARS-CoV-2, while providing valuable insights into how the virus manipulates host gene expression, immune responses, and overall disease progression.

A large amount of PBMC RNA-seq data have been generated across different COVID-19 disease severities, populations, and study groups. Studies have shown that PBMCs infected with SARS-CoV-2 exhibit upregulation of pathways involved in inflammatory immune

[1]Department of Biochemistry, SCLS, Jamia Hamdard, New Delhi, India
[2]Department of Zoology, Rajiv Gandhi University, Itanagar, Arunachal Pradesh, India
[3]All India Institute of Medical Sciences, Bibinagar, Hyderabad, Telangana 508126, India
[4]Department of Computer Science and Engineering, Keshav Memorial Institute of Technology, Hyderabad, Telangana State, India
[5]Independent Researcher, Hingoli, Maharashtra 431513, India
[6]Department of Biotechnology, SCLS, Jamia Hamdard, New Delhi, India
[7]Centre for Virology, SIST, Jamia Hamdard, New Delhi, India
[8]Lead contact
*Correspondence: bio.sailu@gmail.com
https://doi.org/10.1016/j.isci.2024.109087

responses, cell death, and blood coagulation in severe COVID-19.[14–18] Conducting a meta-analysis of these data offers a powerful way to identify reliable gene expression patterns associated with mild, moderate, and severe COVID-19, enhancing statistical power across different populations and study groups. Furthermore, integrating these meta-analytic transcriptome data with protein-protein interaction data could help us investigate how gene expression changes modulate protein-protein interactions, thereby elucidating the role of host-pathogen protein interactions in different disease severities.

In this study, we developed a computational pipeline to integrate gene expression and protein-protein interaction data to identify the potential cellular processes and pathways targeted by pathogens. The pipeline is used to identify SARS-CoV-2-human protein interactions. First, we performed a meta-analysis and combined the meta-analytic transcriptome data with SARS-CoV-2-human protein-protein interaction data to identify key marker genes associated with mild, moderate, and severe COVID-19. We analyzed RNA-seq data from 12 independent studies containing 1,960 PBMC samples with varying disease severity conditions. We observed that the virus primarily targets host proteins involved in translation, ribosomal biogenesis, and membrane trafficking. Furthermore, the expression of translation and ribosomal biogenesis related genes was downregulated in severe COVID-19 cases which may selectively cease host protein synthesis and affect host cellular functions.

RESULTS

Differentially expressed genes spanning Covid-19 disease severities

We obtained RNA-seq gene expression datasets from NCBI Gene Expression Omnibus (GEO).[19] We only included human studies related to PBMC gene expression in response to SARS-CoV-2 infection (Tables S1–S4). The tables present a list of NCBI Sequence Read Archive (SRA) project IDs, various disease severity conditions, list of comparisons for differential expression, and the number of samples available for each disease severity category. In total, we compiled RNA-seq data from 12 independent studies.

We conducted comparisons between different conditions, including mild vs. healthy, severe vs. healthy, severe vs. mild, and severe vs. moderate, to identify differentially expressed genes. The p values of the differentially expressed genes for each comparison across the various projects were combined using our previously published meta-analytic method.[20,21] For the meta-analysis of mild vs. healthy conditions, we included a total of 146 mild cases and 68 healthy cases. In the meta-analysis of severe vs. healthy conditions, there were 513 severe cases and 227 healthy cases. For the meta-analysis of severe vs. mild conditions, the dataset comprised 114 severe cases and 267 mild cases. Lastly, the meta-analysis of severe vs. moderate conditions included 71 severe cases and 52 moderate cases.

Table S5 shows the P-values and Z-scores of differentially expressed genes across the different comparisons. Figures 1A–1D show the number of genes that are upregulated in each comparison, as well as the overlap of these genes with the human genes whose products interact with SARS-CoV-2 and with those involved in immune defense. Most of the genes whose products interact with SARS-CoV-2 are upregulated in mild conditions, while some are also downregulated in severe conditions.

Figure 1E shows the enrichment of various Reactome pathways in differentially expressed genes across the comparisons and whose products also interact with various SARS-CoV-2 proteins. The pathways involved in the response to starvation, translation process, RNA processing, signaling by robo receptors, and regulation of expression of slits and robos are downregulated in severe disease conditions compared to healthy, mild, and moderate disease conditions. Furthermore, pathways associated with the innate immune system, hemostasis, and platelet activation and signaling show significant upregulation across disease severities, especially in severe cases.

Figure 1F shows the enrichment of various Reactome pathways in host genes that are highly differentially expressed across the comparisons and whose products are also involved in immune defense. The pathways involved in neutrophil degradation, Toll-like receptor signaling, IL-1 signaling, programmed cell death, and apoptosis are highly upregulated across the conditions.

Higher order functions of differentially expressed genes across different disease severities

Enrichment analysis was performed using our published method[21] for differentially expressed genes across different disease severities. This analysis included pathways, gene ontology categories, and other functional modules from the Molecular Signature Database (MSigDB).[22] To address redundancy, gene sets with a large number of overlapping genes were pooled and analyzed as pathway modules. Table S6 presents the enrichment scores for these modules across the different comparisons, with higher scores indicating greater relevance to the observed gene expression changes.

The genes involved in ribosomal biogenesis, translation, non-coding RNA processing, and mRNA processing are downregulated in severe disease conditions compared to healthy, mild, and moderate disease conditions (Figures 2A, 2C, and 2D). The genes involved in the electron transport chain/oxidative phosphorylation and mitochondrial components are also downregulated in severe disease conditions (Figures 2B and 2D). In addition, the genes involved in vesicle-mediated transport, the Golgi complex, and the endoplasmic reticulum are upregulated across all three disease conditions, but are slightly more upregulated in mild conditions than in severe conditions (Figure 2D).

Most SARS-CoV-2 proteins interact with host proteins involved in translation and translocation

We have combined the published SARS-CoV-2 protein interactions with human proteins from IntAct and BioGRID databases which are two major public repositories for protein-protein interactions. There were a total of 35,101 interactions between 49 SARS-CoV-2 and 5,492 host proteins. To identify the higher order functions of host-proteins, we have carried out hypergeometric enrichment analysis of genesets from mSigDB. We have developed a modified version of traditional hypergeometric analysis by carrying enrichment analysis of pathways in target

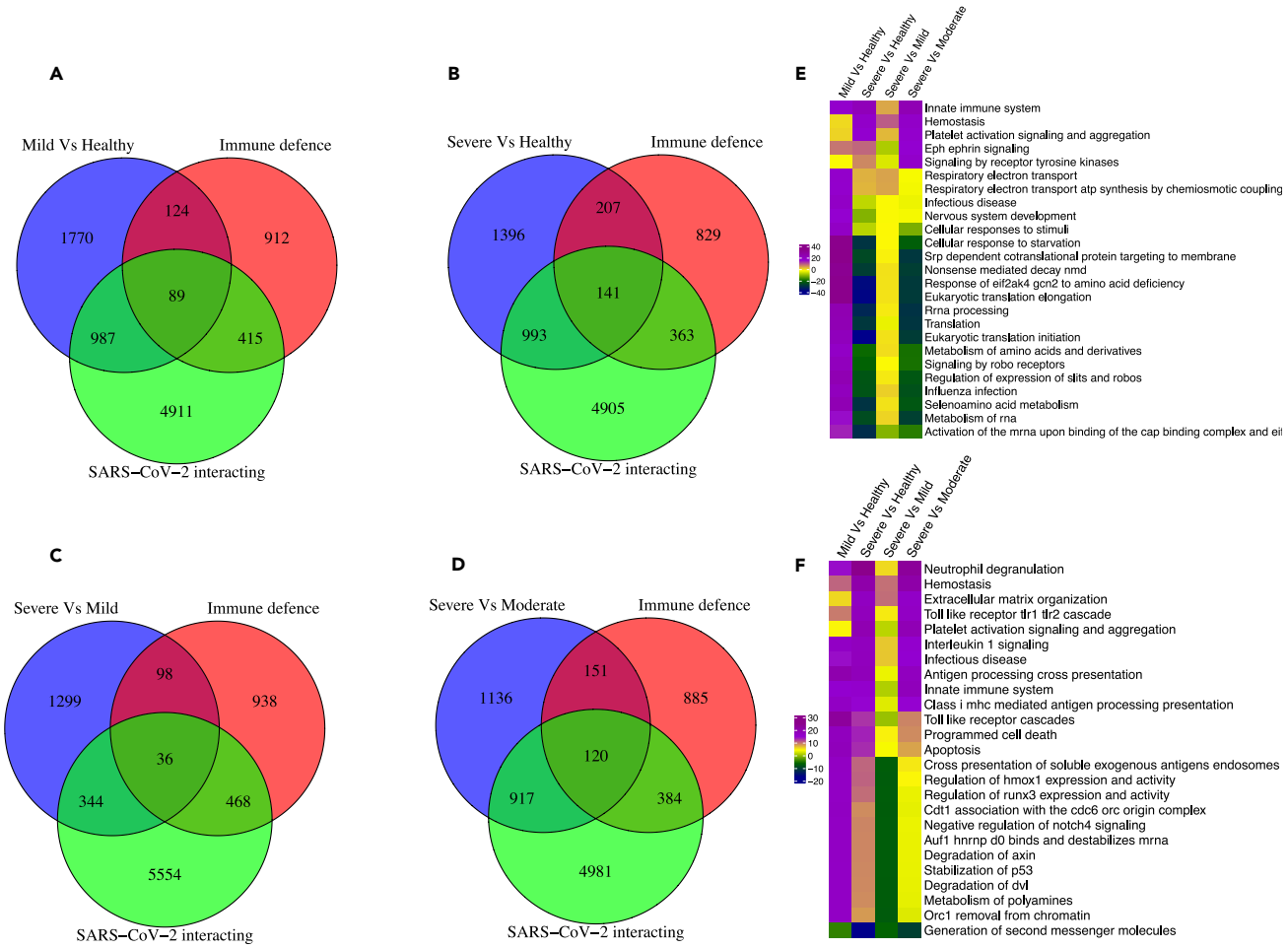


Figure 1. Differentially expressed genes

Venn diagrams showing the overlap between differentially expressed genes in various comparisons, genes associated with immune defense, and genes whose protein products interact with SARS-CoV-2 proteins. The comparisons are as follows: (A) Mild vs. Healthy; (B) Severe vs. Healthy; (C) Severe vs. Mild; (D) Severe vs. Moderate. Heatmaps showing pathway enrichment in highly upregulated and highly downregulated genes across distinct conditions; (E) Magenta signifies strong enrichment of pathways in the genes that are concurrently upregulated across conditions and interact with SARS-CoV-2 proteins, while dark blue signifies strong enrichment of pathways among genes that are concurrently downregulated across conditions and interact with SARS-CoV-2 proteins; (F) Magenta signifies strong enrichment of pathways among genes that are concurrently upregulated across conditions and involved in immune defense, whereas while dark blue signifies strong enrichment of pathway among genes that are concurrently downregulated across conditions and involved in immune defense. The value range for the color bar of the heatmaps represents the Z-scores or normalized statistics of enrichment P-value of pathways. These analyses provide insights into gene expression relationships, potential interactions with SARS-CoV-2, and immune defense involvement across diverse conditions.

gene lists as well as in remaining human genes after exclusion of target gene lists. The final two column enrichment scores for each of the pathways is summed to get the final list of enrichment scores (See STAR Methods section). Table S7 shows the enrichment scores for the Reactome pathways in target interaction gene-list of each of the SARS-CoV-2 proteins. Majority of the viral proteins are found to be interacting with host proteins involved translation initiation and elongation, ribosomal biogenesis, ribosomal proteins, protein localization, vesicle mediated transport, membrane trafficking (Figures 3A and 3B). In addition, proteins of endoplasmic reticulum and Golgi complex which play a major role in protein sorting and trafficking are targeted by viral proteins (Figures 3C and 3D). Viral proteins ORF6 and N predominantly target host proteins involved in translation related processes (Figures 3A–3C) and other viral proteins mainly ORF7a, NSP4, NSP6 and M target host proteins involved in intracellular transport, protein localization, endoplasmic reticulum and Golgi complex associated processes (Figures 3C and 3D).

Expression of human genes targeted by SARS-CoV-2 proteins in different disease severities

We conducted a rank-based enrichment analysis of human gene sets whose products interact with SARS-CoV-2 proteins.[23] Table S8 presents the enriched pathways in human genes targeted by each of the SARS-CoV-2 proteins across different conditions. In contrast, Figure 4A

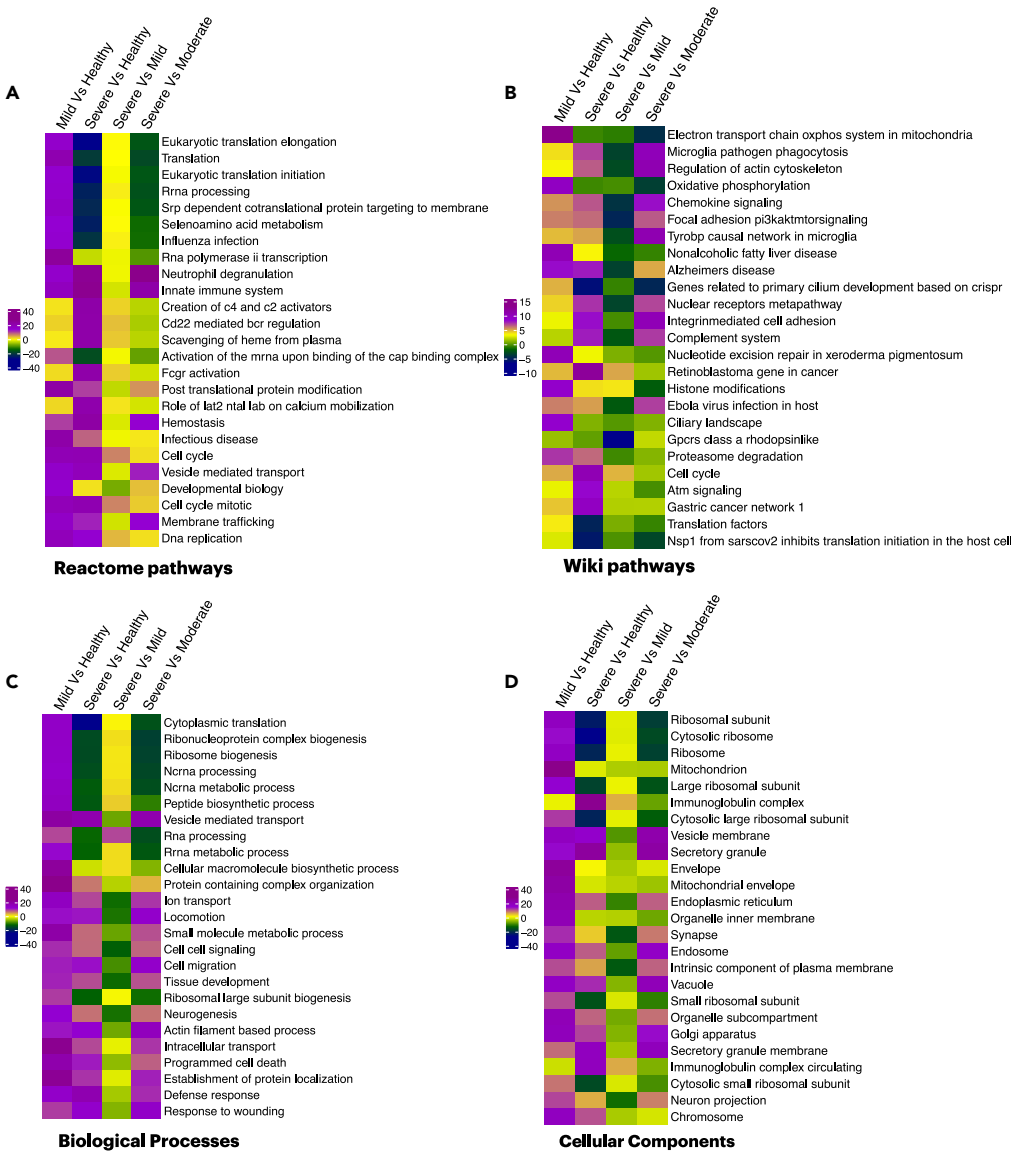


Figure 2. Higher-order functions of differentially expressed genes

Heatmaps showing the higher-order functions of differentially expressed genes across the different comparisons. Magenta color signifies strong upregulation, while dark blue color signifies strong downregulation.

(A) Reactome pathways; (B) Wiki pathways; (C) Biological Process; (D) Cellular Component. The value range for the color bar of the heatmaps represents the Z-scores or normalized statistics of enrichment P-value of pathways.

displays the heatmap representing the enrichment score of human gene sets targeted by each viral protein across disease severities. Most human proteins interacting with SARS-CoV-2 are upregulated in mild conditions and downregulated in severe disease conditions. Specifically, the interacting partners of the N protein primarily consist of ribosomal proteins (Figure 4A), which are upregulated in mild and moderate conditions compared to severe disease conditions. Enrichment of Reactome pathways in human protein targets of SARS-CoV-2 shows (Figure 4B) that the majority of them are involved in metabolism of RNA, translation, cellular response to starvation, nonsense mediated decay (targets of ORF6, N, NSP4, NSP6, M, ORF7a, E, ORF3A, ORF7B, NSP13, ORF1AB, NSP5, NSP14), post-translational modifications, vesicle mediated transport and membrane trafficking (NSP4, NSP6, M, ORF7a, E, ORF3A, ORF7B, S and SARS2). Some of them are also involved in cell cycle (ORF6, NSP13, ORF1AB, NSP7, NSP16 and NSP12). The protein encoded by ORF6 interacts with the largest number of human proteins involved in all of the aforementioned cellular processes such as RNA processing, translation, translocation and cell cycle. Figure S1 shows enrichment of biological process gene ontology terms, where N, NSP1 and ORF1AB target unique human proteins related to processing of various types of RNAs (mRNA, rRNA and ncRNA) and ribosomal biogenesis. Figure S2 shows enrichment of various cellular

CellPress
OPEN ACCESS

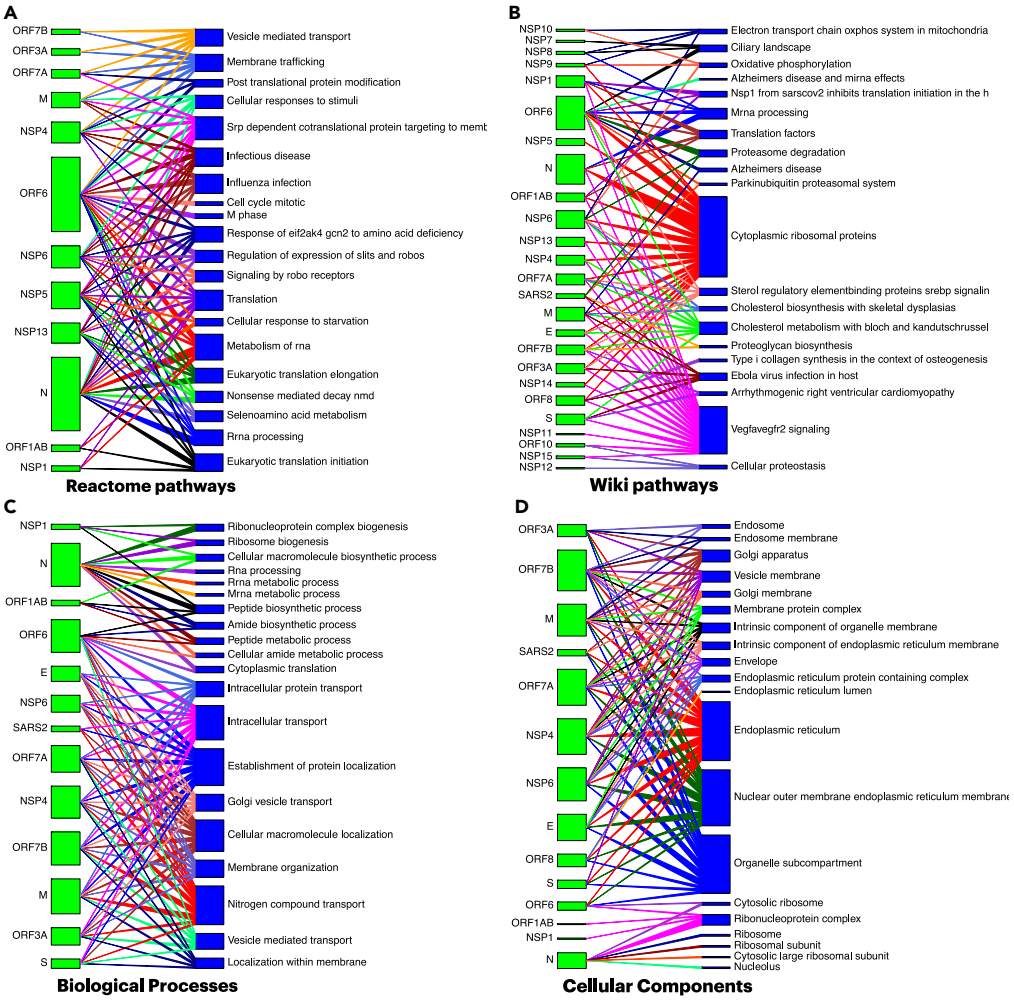


Figure 3. Bipartite graphs illustrating SARS-CoV-2 protein interactions with human pathways or processes

In the figure, bipartite graphs depict the intricate interactions between distinct SARS-CoV-2 proteins (shown in green/left boxes) and human pathways (displayed in blue/right boxes). In this representation, each node is symbolized by a box, with the height of the box correlates with the number of interactions associated with the corresponding protein or pathway. The thickness of the edge represents the strength of the interaction i.e., hypergeometric enrichment score of pathways within human proteins known to interact with specific SARS-CoV-2 proteins. Edges stemming from each pathway or gene ontology (GO) term are visually differentiated by distinct colors to make the interactions less confusing. The graphs are divided into categories to capture the annotations of genes using diverse gene-sets: (A) Reactome pathways; (B) Wiki pathways; (C) Biological Process; (D) Cellular Component.

components, where N has unique host target proteins related to ribosomal subunits and nucleolus. ORF7B, M, NSP4, NSP6, ORF7A, E, ORF3A, S, ORF6, ORF8 and SARS2 proteins products primarily interact with endoplasmic reticulum and Golgi complex proteins.

DISCUSSION

The results presented in this work provide valuable insights into the host's molecular responses to SARS-CoV-2 infection across varying disease severities. These findings deepen our understanding of the dynamic changes within host cells when exposed to the virus and their potential implications for COVID-19 pathogenesis.

Pathway analysis of host-proteins targeted by SARS-CoV-2 proteins provides a comprehensive view of how the virus interacts with and manipulates various cellular processes. These findings shed light on the remarkably precise targeting patterns of viral proteins and their intricate interactions with indispensable human cellular processes. Most viral proteins interact with host proteins involved in translation initiation and elongation, ribosomal biogenesis, ribosomal proteins, nonsense-mediated decay, vesicle-mediated transport, membrane trafficking and protein localization. This implies the virus's adept manipulation of these cellular processes to redirect cellular resources to facilitate viral protein synthesis and replication, consistent with prior research on viral hijacking of host machinery.[24] It is surprising that the majority of the pathways and cellular components i.e., ribosomal biogenesis, ribosomal proteins and translation interacting with SARS-CoV-2, are upregulated in mild conditions and downregulated in severe conditions. This suggests that the differential regulation of these pathways may contribute to

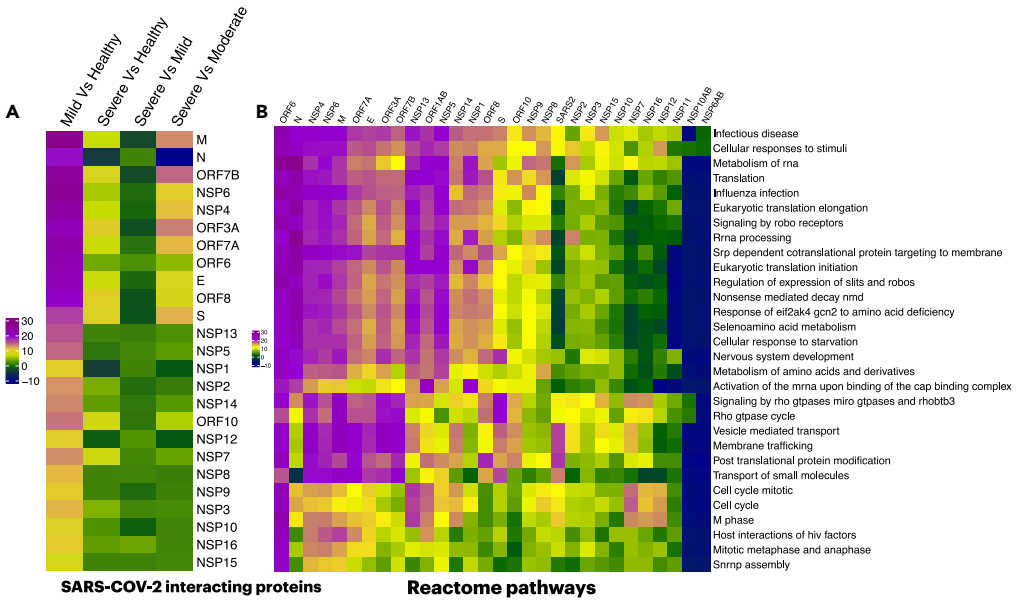


Figure 4. SARS-CoV-2 targeted human proteins and their functions

Heatmaps showing insights into the enrichment patterns of human genes whose protein products interact with SARS-CoV-2 proteins, along with their higher-order functions.

(A) Heatmap (left) displays the rank-based enrichment scores of human protein targets for each of the SARS-CoV-2 proteins across diverse comparisons. Magenta represents strong upregulation, while dark blue signifies pronounced downregulation; (B) Heatmap (right) portrays the hypergeometric-based enrichment of various Reactome pathways within the gene set targeted by each SARS-CoV-2 protein. This analysis sheds light on the relationships between SARS-CoV-2 proteins and human genes, revealing their potential roles in pathway regulation and biological functions. The value range for the color bar of the heatmaps represents the Z-scores or normalized statistics of enrichment P-value of pathways.

differences in disease severity, potentially impairing essential cellular functions such as protein synthesis, and host's ability to mount an effective antiviral defense and causing broader health consequences.

One interesting observation is that viral proteins ORF6 and N primarily target processes involved in translation, while other viral proteins such as ORF7a, NSP4, NSP6, and M primarily target proteins involved in intracellular transport, protein localization, and proteins of the endoplasmic reticulum and Golgi complex. It reveals the virus's diverse strategies for hijacking host cellular machinery to facilitate its replication and spread. It also demonstrates the virus's ability to manipulate key cellular pathways to promote its replication and survival within the host. Particular interest is the protein encoded by ORF6, which targets multiple cellular processes, including RNA processing, translation, translocation, and cell cycle regulation. This suggests that ORF6 plays a central role in hijacking host cellular machinery for viral replication and propagation. The unique targets of viral proteins N, NSP1, and ORF1AB related to the processing of various types of RNAs and ribosomal biogenesis emphasizes the direct impact of these viral proteins on RNA processing and ribosome function, potentially contributing to the dysregulation of protein synthesis in host cells. Furthermore, the protein N has unique targets related to ribosomal subunits and the nucleolus, indicating its involvement in hijacking ribosomal function and protein synthesis. On the other hand, several viral proteins primarily target proteins of the endoplasmic reticulum and Golgi complex, which are critical for viral protein synthesis and the assembly of new virions, and their subsequent transport and release in extracellular milieu.

While genes whose protein products interact with SARS-CoV-2 and are involved in translation are downregulated in severe disease conditions, those whose protein products interact with SARS-CoV-2 and play a role in the endoplasmic reticulum, Golgi complex, Golgi vesicle transport, vesicle-mediated transport, Golgi membrane, vesicle membrane, and intracellular transport are upregulated across all disease conditions, including mild, moderate, and severe, albeit slightly more upregulated in mild conditions. These results shed light on the distinct responses of genes linked to translation and intracellular transport processes in the presence of the virus. Intriguingly, a key observation emerges: genes associated with translation are downregulated in severe disease conditions, while genes engaged in processes associated with the endoplasmic reticulum, Golgi complex, Golgi vesicle transport, vesicle-mediated transport, Golgi membrane, vesicle membrane, and intracellular transport are consistently upregulated across all disease conditions including mild, moderate, and severe with a slightly more pronounced upregulation in mild conditions. The endoplasmic reticulum and Golgi apparatus play central roles in protein processing, sorting, and trafficking within the cell. The upregulation of these processes might signify the host's strategy to ensure efficient production and transportation of antiviral proteins and immune molecules. Additionally, it may also help in the transport of viral proteins for viral assembly.

The pathways, neutrophil degranulation, Toll-like receptor signaling, IL-1 signaling, programmed cell death, apoptosis, and Class I MHC mediated antigen presentation and processing are upregulated across the disease conditions. These pathways play pivotal roles in the immune response to viral infections. Upregulation of the neutrophil degranulation pathway indicates an active recruitment and activation of

neutrophils in response to the virus. This is expected as neutrophils are known to be involved in combating viral infections and SARS-CoV-2 infection has been shown to be associated with increased neutrophil count and neutrophil degranulation in nasopharyngeal and lung tissues.[25–27] Increased neutrophil-to-lymphocyte ratio has shown to be an independent risk factor for mortality.[28] Activation of Toll-like receptor signaling pathways triggers the production of pro-inflammatory cytokines and chemokines, which help recruit immune cells and enhance the immune response.[29] IL-1 can induce the production of other pro-inflammatory cytokines, recruit immune cells, and promote inflammation.[30,31] The upregulation of programmed cell death pathways, particularly apoptosis, suggests that the host is actively targeting infected cells for destruction. Class I MHC mediated antigen presentation and processing is crucial for presenting viral antigens to cytotoxic T cells, triggering an immune response against infected cells. The consistent upregulation of these pathways across different disease conditions indicates that the host's immune system is actively engaged in responding to the SARS-CoV-2 virus, irrespective of disease severity. However, the sustained upregulation of these pathways, particularly in severe cases, may also contribute to the hyperinflammatory response observed in severe COVID-19 cases, potentially leading to tissue damage and adverse outcomes.

The downregulation of electron transport chain (ETC) and mitochondrial component genes in severe COVID-19 highlights potential mitochondrial dysfunction and energy metabolism disruption in immune cells. Mitochondria's role in ATP production, essential for immune cell function, suggests compromised antiviral responses in severe cases. Dysfunctional mitochondria can lead to reactive oxygen species accumulation, exacerbating cell death, inflammation and subsequent tissue damage.[32] This downregulation could also impact the intensity of apoptosis and autophagy — the key regulators of cell survival and immune responses. Investigating the virus's direct or immune response-mediated effect on mitochondria is crucial.

In response to SARS-CoV-2 infection, there is a significant upregulation of pathways associated with the innate immune system, hemostasis and platelet activation signaling across all disease conditions, particularly in severe cases. The innate immune system, when activated, releases various molecules such as cytokines and chemokines, which are proinflammatory and recruit immune cells respectively to the infection site and assist in combating the virus.[33] Several studies have reported an association between progression to severe COVID-19 and dysregulated secretion of proinflammatory cytokines.[34] Hemostasis, the process of blood clotting, is another pathway that is upregulated across the different disease conditions of disease severity. This process is vital in preventing excessive bleeding, and its disruption can lead to severe bleeding complications. Prior research has established that individuals afflicted with COVID-19 exhibit a range of hemostasis irregularities, notably coagulation dysfunction, which is a significant contributor to mortality.[35,36] Hemostasis parameters, such as prothrombin time, have been shown to be good prognostic indicators for patients with a poor outcome.[37] Additionally, upregulation of platelet activation signaling shows its possible role in underlying increased thrombotic events as previously reported.[38] Platelets release molecules like thromboxane A2, which acts as a potent vasoconstrictor to stop bleeding. These findings align with clinical observations of elevated inflammation and coagulation abnormalities in severe COVID-19 cases, often resulting in tissue damage and multi-organ failure.[39] Therefore, dissecting the contribution of platelets to COVID-19 critical illness is key to understanding SARS-CoV-2 infection pathogenesis and identify novel therapeutic strategies.

Beyond the direct impact of SARS-CoV-2, several interconnected factors intricately modulate the immune response, ultimately shaping the PBMC transcriptome and consequently impacting COVID-19 disease severity. Studies suggest that variations in the relative proportions of specific blood cell subsets, such as lymphocytes, monocytes, and neutrophils, contribute to distinct patterns of gene expression within PBMCs. These variations in gene expression are further influenced by individual factors such as gender, age, and body mass index (BMI).[40]

Studies have identified distinct gene expression profiles in PBMCs of males and females infected with SARS-CoV-2.[41] These differences involve genes related to immune function, inflammation, and cellular response to viral infection. Higher testosterone levels in males may be associated with a more severe inflammatory response and poorer outcomes, while estrogen in females may exert immunomodulatory effects leading to favorable outcomes.[42,43] Additionally, the two X chromosomes in females offer an advantage compared to the single X chromosome in males. These X chromosomes encode immune regulatory genes, potentially leading to lower viral load levels and less inflammation.[44] Aging has been shown to reprogram the immune cell landscape, leading to a decline in naive and memory T cells and an increase in exhausted and regulatory T cells, with accompanying changes in B cells, monocytes, and dendritic cells.[45] These alterations contribute to a weakened immune response, explaining the increased susceptibility of older individuals to severe COVID-19.[45,46] Notably, COVID-19 exacerbates age-induced immune cell polarization and promotes inflammation and cellular senescence.

Pre-existing medical conditions, particularly those associated with the metabolic syndrome, significantly worsen COVID-19 outcomes. Each factor within the metabolic syndrome (obesity, hypertension, and diabetes) independently impacts distinct immune populations, leading to a cumulative effect that further weakens the immune system when multiple conditions are present.[47] Additionally, pre-existing respiratory conditions like asthma and chronic obstructive pulmonary disease (COPD) are associated with particularly severe COVID-19 outcomes. These chronic conditions can alter gene expression patterns and disrupt the immune response to SARS-CoV-2 infection, increasing susceptibility to lung damage and other complications.[48]

Individual genetic makeup plays a key role in determining susceptibility and response to viral infections, including COVID-19. Differences in genes related to immune response, inflammation, and viral susceptibility significantly impact disease severity. Studies have identified specific genetic variants associated with increased risk, highlighting the importance of genetic predisposition.[49,50]

In conclusion, this study deepens our understanding of the host's molecular responses to SARS-CoV-2 infection across various disease severities. The analysis of pathways targeted by SARS-CoV-2 proteins provides a comprehensive view of how the virus interacts with and manipulates cellular processes. The virus demonstrates a remarkable ability to manipulate functions involved in translation, ribosome function, RNA processing, intracellular transport, and membrane dynamics. The downregulation of key cellular pathways involved in translation and

ribosomal biogenesis in severe conditions suggests a potential link between disrupted protein synthesis and disease severity. The distinct regulation of these pathways suggests their involvement in influencing disease outcomes by potentially impairing essential cellular functions and antiviral defense mechanisms. Understanding these pathway dysregulations is pivotal for advancing our knowledge of COVID-19 pathogenesis and guiding therapeutic strategies. They offer potential targets for interventions aimed at disrupting the virus's ability to hijack host cellular machinery and manipulate host responses.

Limitations of the study and future directions

It is important to note that the findings of this study are based on transcriptomic data, protein-protein interactions, and pathway enrichment analysis. These methods provide a valuable overview of the molecular changes that occur in response to SARS-CoV-2 infection, but they cannot provide definitive information about the underlying mechanisms. It is essential to establish a clear causal relationship between viral infection of PBMCs and differential translation regulation in mild and severe conditions. Furthermore, investigating whether similar mechanisms exist in other viral diseases such as influenza would be highly valuable.

Given that our analysis relies on transcriptome data, which does not directly reflect the proteome levels associated with the condition, we provide the first lines of evidence for several future experimental studies to establish a definitive link between translation and COVID-19 disease severity. Notably, these experiments should focus on immune cells with high translational activity, such as B cells.

First, western blot analysis should be performed on patient samples (PBMCs) from varying COVID-19 severity groups to compare the expression levels of key proteins involved in translation and ribosomal biogenesis with those of healthy controls. Second, *in vitro* translation assays should be conducted on cell cultures or patient samples from different severity groups to directly measure the impact on protein synthesis rates. Additionally, *in vitro* cell culture models (B cells) should be established and infected with SARS-CoV-2 to monitor changes in translation and ribosomal pathways, thereby mimicking disease conditions. Furthermore, ribosome profiling,[51] also known as Ribo-seq (ribosome sequencing), should be employed to determine which genes are being translated and how much protein is being produced from each gene in both host cells and the pathogen across the different conditions of COVID-19 disease severity. Utilizing ribosome profiling can identify specific genes or pathways affected in severe cases, enhancing our understanding of the connection between dysregulated protein synthesis and disease severity.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- METHOD DETAILS
  - Gene expression data collection
  - RNA-Seq data analysis
  - Differential expression analysis
  - Combining p-values of differential expression
  - Enrichment analysis of genesets/pathways
  - Data visualization

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109087.

ACKNOWLEDGMENTS

We acknowledge the BRAHM: High-Performance Computational facility of the Indian Biological Data Centre, Regional Centre for Biotechnology, Faridabad, INDIA (https://ibdc.rcb.res.in/; DBT Grant no. BT/TCB/IBDC/2019) to carry out the NGS data analysis and pathway enrichment analysis. We also extend our gratitude to the Department of Biotechnology for their past funding support. This work was supported by the intramural research program of AIIMS Bibinagar.

AUTHOR CONTRIBUTIONS

S.Y.: conceived of the study, data visualization and interpretation of the results. S.Y. and M.S.S.: data analysis and curation. S.Y.: Manuscript preparation. A.P., R.D., R.S., M.S.S., V.Y.M., and M.A.: Contributed to the interpretation of the results, reviewed and edited the manuscript. All authors read and approved the final manuscript.

CellPress
OPEN ACCESS

DECLARATION OF INTERESTS

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

Received: September 11, 2023
Revised: November 7, 2023
Accepted: January 29, 2024
Published: February 2, 2024

REFERENCES

1. Msemburi, W., Karlinsky, A., Knutson, V., Aleshin-Guendel, S., Chatterji, S., and Wakefield, J. (2023). The WHO estimates of excess mortality associated with the COVID-19 pandemic. Nature *613*, 130–137.
2. Dhakal, B.P., Sweitzer, N.K., Indik, J.H., Acharya, D., and William, P. (2020). SARS-CoV-2 infection and cardiovascular disease: COVID-19 heart. Heart Lung Circ. *29*, 973–987.
3. Damiani, S., Fiorentino, M., De Palma, A., Foschini, M.P., Lazzarotto, T., Gabrielli, L., Viale, P.L., Attard, L., Riefolo, M., and D'Errico, A. (2021). Pathological post-mortem findings in lungs infected with SARS-CoV-2. J. Pathol. *253*, 31–40.
4. Lei, H.Y., Ding, Y.H., Nie, K., Dong, Y.M., Xu, J.H., Yang, M.L., Liu, M.Q., Wei, L., Nasser, M.I., Xu, L.Y., et al. (2021). Potential effects of SARS-CoV-2 on the gastrointestinal tract and liver. Biomed. Pharmacother. *133*, 111064.
5. Moustafa, A., Khalel, R.S., and Aziz, R.K. (2021). Traces of SARS-CoV-2 RNA in peripheral blood cells of patients with COVID-19. OMICS A J. Integr. Biol. *25*, 475–483.
6. Gandhi, R.T., Lynch, J.B., and Del Rio, C. (2020). Mild or moderate Covid-19. N. Engl. J. Med. *383*, 1757–1766.
7. Durmuş Tekir, S.D., and Ülgen, K.Ö. (2013). Systems biology of pathogen-host interaction: networks of protein-protein interaction within pathogens and pathogen-human interactions in the post-genomic era. Biotechnol. J. *8*, 85–96.
8. Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature *583*, 459–468.
9. Jiang, Y., Tong, K., Yao, R., Zhou, Y., Lin, H., Du, L., Jin, Y., Cao, L., Tan, J., Zhang, X.D., et al. (2021). Genome-wide analysis of protein–protein interactions and involvement of viral proteins in SARS-CoV-2 replication. Cell Biosci. *11*, 140.
10. Zhou, Y., Liu, Y., Gupta, S., Paramo, M.I., Hou, Y., Mao, C., Luo, Y., Judd, J., Wierbowski, S., Bertolotti, M., et al. (2023). A comprehensive SARS-CoV-2–human protein–protein interactome reveals COVID-19 pathobiology and potential host therapeutic targets. Nat. Biotechnol. *41*, 128–139.
11. Chen, Z., Wang, C., Feng, X., Nie, L., Tang, M., Zhang, H., Xiong, Y., Swisher, S.K., Srivastava, M., and Chen, J. (2021). Interactomes of SARS-CoV-2 and human coronaviruses reveal host factors potentially affecting pathogenesis. EMBO J. *40*, e107776.
12. Pagano, N., Laurent-Rolle, M., Hsu, J.C., Vogels, C.B., Grubaugh, N.D., and Manuelidis, L. (2020). Long SARS-CoV-2 Nucleocapsid Sequences in Blood Monocytes Collected Soon after Hospital Admission.
13. Pontelli, M.C., Castro, I.A., Martins, R.B., Veras, F.P., Serra, L.L., Nascimento, D.C., Cardoso, R.S., Rosales, R., Lima, T.M., Souza, J.P., et al. (2020). Infection of human lymphomononuclear cells by SARS-CoV-2. Preprint at bioRxiv. https://doi.org/10.1101/2020.07.28.225912.
14. Arunachalam, P.S., Wimmers, F., Mok, C.K.P., Perera, R.A.P.M., Scott, M., Hagan, T., Sigal, N., Feng, Y., Bristow, L., Tak-Yin Tsang, O., et al. (2020). Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. Science *369*, 1210–1220.
15. Kwan, P.K.W., Cross, G.B., Naftalin, C.M., Ahidjo, B.A., Mok, C.K., Fanusi, F., Permata Sari, I., Chia, S.C., Kumar, S.K., Alagha, R., et al. (2021). A blood RNA transcriptome signature for COVID-19. BMC Med. Genom. *14*, 155–158.
16. Välikangas, T., Junttila, S., Rytkönen, K.T., Kukkonen-Macchi, A., Suomi, T., and Elo, L.L. (2022). COVID-19-specific transcriptomic signature detectable in blood across multiple cohorts. Front. Genet. *13*, 929887.
17. Agrawal, P., Sambaturu, N., Olgun, G., and Hannenhalli, S. (2022). A path-based analysis of infected cell line and COVID-19 patient transcriptome reveals novel potential targets and drugs against SARS-CoV-2. Front. Immunol. *13*, 918817.
18. Huang, J., Wang, Y., Zha, Y., Zeng, X., Li, W., and Zhou, M. (2022). Transcriptome analysis reveals hub genes regulating autophagy in patients with severe COVID-19. Front. Genet. *13*, 908826.
19. Clough, E., and Barrett, T. (2016). The gene expression omnibus database. Methods Mol. Biol. *1418*, 93–110.
20. Cinghu, S., Yellaboina, S., Freudenberg, J.M., Ghosh, S., Zheng, X., Oldfield, A.J., Lackford, B.L., Zaykin, D.V., Hu, G., and Jothi, R. (2014). Integrative framework for identification of key cell identity genes uncovers determinants of ES cell identity and homeostasis. Proc. Natl. Acad. Sci. USA *111*, E1581–E1590.
21. Parvati Sai Arun, P.V., Miryala, S.K., Rana, A., Kurukuti, S., Akhter, Y., and Yellaboina, S. (2018). System-wide coordinates of higher order functions in the host-pathogen environment upon Mycobacterium tuberculosis infection. Sci. Rep. *8*, 5079.
22. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The molecular signatures database hallmark gene set collection. Cell Syst. *1*, 417–425.
23. Arun, P.V.P.S., Miryala, S.K., Chattopadhyay, S., Thiyyagura, K., Bawa, P., Bhattacharjee, M., and Yellaboina, S. (2016). Identification and functional analysis of essential, conserved, housekeeping and duplicated genes. FEBS Lett. *590*, 1428–1437.
24. Banerjee, A.K., Blanco, M.R., Bruce, E.A., Honson, D.D., Chen, L.M., Chow, A., Bhat, P., Ollikainen, N., Quinodoz, S.A., Loney, C., et al. (2020). SARS-CoV-2 disrupts splicing, translation, and protein trafficking to suppress host defenses. Cell *183*, 1325–1339.e21.
25. Wang, J., Li, Q., Yin, Y., Zhang, Y., Cao, Y., Lin, X., Huang, L., Hoffmann, D., Lu, M., and Qiu, Y. (2020). Excessive neutrophils and neutrophil extracellular traps in COVID-19. Front. Immunol. *11*, 2063.
26. Bankar, R., Suvarna, K., Ghantasala, S., Banerjee, A., Biswas, D., Choudhury, M., Palanivel, V., Salkar, A., Verma, A., Singh, A., et al. (2021). Proteomic investigation reveals dominant alterations of neutrophil degranulation and mRNA translation pathways in patients with COVID-19. iScience *24*, 102135.
27. Akgun, E., Tuzuner, M.B., Sahin, B., Kilercik, M., Kulah, C., Cakiroglu, H.N., Serteser, M., Unsal, I., and Baykal, A.T. (2020). Proteins associated with neutrophil degranulation are upregulated in nasopharyngeal swabs from SARS-CoV-2 patients. PLoS One *15*, e0240012.
28. Liu, Y., Du, X., Chen, J., Jin, Y., Peng, L., Wang, H.H.X., Luo, M., Chen, L., and Zhao, Y. (2020). Neutrophil-to-lymphocyte ratio as an independent risk factor for mortality in hospitalized patients with COVID-19. J. Infect. *81*, e6–e12.
29. Barton, G.M., and Medzhitov, R. (2003). Toll-like receptor signaling pathways. Science *300*, 1524–1525.
30. Dinarello, C.A. (2018). Overview of the IL-1 family in innate inflammation and acquired immunity. Immunol. Rev. *281*, 8–27.
31. Shakoory, B., Carcillo, J.A., Chatham, W.W., Amdur, R.L., Zhao, H., Dinarello, C.A., Cron, R.Q., and Opal, S.M. (2016). Interleukin-1 receptor blockade is associated with reduced mortality in sepsis patients with features of the macrophage activation syndrome: Re-analysis of a prior Phase III trial. Crit. Care Med. *44*, 275–281.
32. Chen, T.H., Chang, C.J., and Hung, P.H. (2023). Possible Pathogenesis and Prevention of Long COVID: SARS-CoV-2-Induced Mitochondrial Disorder. Int. J. Mol. Sci. *24*, 8034.
33. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet (North Am. Ed.) *395*, 497–506.
34. Blanco-Melo, D., Nilsson-Payant, B.E., Liu, W.C., Uhl, S., Hoagland, D., Møller, R., Jordan, T.X., Oishi, K., Panis, M., Sachs, D.,

et al. (2020). Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell *181*, 1036–1045.e9.
35. Arachchillage, D.R.J., and Laffan, M. (2020). Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. J. Thromb. Haemostasis *18*, 1233–1234.
36. Connors, J.M., and Levy, J.H. (2020). COVID-19 and its implications for thrombosis and anticoagulation. Blood. J. Am. Soc. Hematol. *135*, 2033–2040.
37. Wang, L., He, W.B., Yu, X.M., Hu, D.L., and Jiang, H. (2020). Prolonged prothrombin time at admission predicts poor clinical outcome in COVID-19 patients. World J. Clin. Cases *8*, 4370–4379.
38. Khider, L., Gendron, N., Goudot, G., Chocron, R., Hauw-Berlemont, C., Cheng, C., Rivet, N., Pere, H., Roffe, A., Clerc, S., et al. (2020). Curative anticoagulation prevents endothelial lesion in COVID-19 patients. J. Thromb. Haemostasis *18*, 2391–2399.
39. Vinayagam, S., and Sattu, K. (2020). SARS-CoV-2 and coagulation disorders in different organs. Life Sci. *260*, 118431.
40. Eady, J.J., Wortley, G.M., Wormstone, Y.M., Hughes, J.C., Astley, S.B., Foxall, R.J., Doleman, J.F., and Elliott, R.M. (2005). Variation in gene expression profiles of peripheral blood mononuclear cells from healthy volunteers. Physiol. Genom. *22*, 402–411.
41. Klein, S.L., and Flanagan, K.L. (2016). Sex differences in immune responses. Nat. Rev. Immunol. *16*, 626–638.
42. Zheng, S., Zou, Q., Zhang, D., Yu, F., Bao, J., Lou, B., Xie, G., Lin, S., Wang, R., Chen, W., et al. (2022). Serum level of testosterone predicts disease severity of male COVID-19 patients and is related to T-cell immune modulation by transcriptome analysis. Clin. Chim. Acta *524*, 132–138.
43. Ramírez-de-Arellano, A., Gutiérrez-Franco, J., Sierra-Diaz, E., and Pereira-Suárez, A.L. (2021). The role of estradiol in the immune response against COVID-19. Hormones (Basel) *20*, 657–667.
44. Zheng, Y., Liu, X., Le, W., Xie, L., Li, H., Wen, W., Wang, S., Ma, S., Huang, Z., Ye, J., et al. (2020). A human circulating immune cell landscape in aging and COVID-19. Protein Cell *11*, 740–770.
45. Lo Tartaro, D., Neroni, A., Paolini, A., Borella, R., Mattioli, M., Fidanza, L., Quong, A., Petes, C., Awong, G., Douglas, S., et al. (2022). Molecular and cellular immune features of aged patients with severe COVID-19 pneumonia. Commun. Biol. *5*, 590.
46. Lewis, S.A., Sureshchandra, S., Zulu, M.Z., Doratt, B., Jankeel, A., Ibraim, I.C., Pinski, A.N., Rhoades, N.S., Curtis, M., Jiang, X., et al. (2021). Differential dynamics of peripheral immune responses to acute SARS-CoV-2 infection in older adults. Nat. Aging *1*, 1038–1052.
47. Kreutmair, S., Kauffmann, M., Unger, S., Ingelfinger, F., Núñez, N.G., Alberti, C., De Feo, D., Krishnarajah, S., Friebel, E., Ulutekin, C., et al. (2022). Preexisting comorbidities shape the immune response associated with severe COVID-19. J. Allergy Clin. Immunol. *150*, 312–324.
48. Konwar, C., Asiimwe, R., Inkster, A.M., Merrill, S.M., Negri, G.L., Aristizabal, M.J., Rider, C.F., MacIsaac, J.L., Carlsten, C., and Kobor, M.S. (2021). Risk-focused differences in molecular processes implicated in SARS-CoV-2 infection: corollaries in DNA methylation and gene expression. Epigenet. Chromatin *14*, 54.
49. Taylor, K., Das, S., Pearson, M., Kozubek, J., Pawlowski, M., Jensen, C.E., and Gardner, S. (2020). Analysis of genetic host response risk factors in severe COVID-19 patients. Preprint at medRxiv. https://doi.org/10.1101/2020.06.17.20134015v2.
50. Edahiro, R., Shirai, Y., Takeshima, Y., Sakakibara, S., Yamaguchi, Y., Murakami, T., Morita, T., Kato, Y., Liu, Y.C., Motooka, D., et al. (2023). Single-cell analyses and host genetics highlight the role of innate immune cells in COVID-19 severity. Nat. Genet. *55*, 753–767.
51. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. Sci. Technol. Humanit. *324*, 218–223.
52. Overmyer, K.A., Shishkova, E., Miller, I.J., Balnis, J., Bernstein, M.N., Peters-Clarke, T. M., Meyer, J.G., Quan, Q., Muehlbauer, L.K., Trujillo, E.A., et al. (2021). Large-scale multi-omic analysis of COVID-19 severity. Cell Syst. *12*, 23–40.
53. Rother, N., Yanginlar, C., Lindeboom, R.G., Bekkering, S., van Leent, Buijsers, B., Jonkman, I., de Graaf, M., Baltissen, M., Lamers, L.A., et al. (2020). Hydroxychloroquine inhibits trained immunity–implications for COVID-19. Preprint at medRxiv. https://doi.org/10.1101/2020.06.08.20122143.
54. Ryan, F.J., Hope, C.M., Masavuli, M.G., Lynn, M.A., Mekonnen, Z.A., Yeow, A.E.L., Garcia-Valtanen, P., Al-Delfi, Z., Gummow, J., Ferguson, C., et al. (2022). Long-term perturbation of the peripheral immune system months after SARS-CoV-2 infection. BMC Med. *20*, 1–23.
55. Lévy, Y., Wiedemann, A., Hejblum, B.P., Durand, M., Lefebvre, C., Surénaud, M., Lacabaratz, C., Perreau, M., Foucat, E., Déchenaud, M., et al. (2021). CD177, a specific marker of neutrophil activation, is associated with coronavirus disease 2019 severity and death. iScience, 24102711.
56. Knabl, L., Lee, H.K., Wieser, M., Mur, A., Zabernigg, A., Knabl, L., Sr., Rauch, S., Bock, M., Schumacher, J., Kaiser, N., et al. (2022). BNT162b2 vaccination enhances interferon-JAK-STAT-regulated antiviral programs in COVID-19 patients infected with the SARS-CoV-2 Beta variant. Commun. Med. *2*, 17.
57. Dhindsa, S., Zhang, N., McPhaul, M. J., Wu, Z., Ghoshal, A. K., Erlich, E.C., Mani, K., Randolph, G.J., Edwards, J.R., Mudd, P.A., and Diwan, A. (2021). Association of circulating sex hormones with inflammation and disease severity in patients with COVID-19. JAMA Netw. Open *4*, e2111398.
58. Zhang, J., Lin, D., Li, K., Ding, X., Li, L., Liu, Y., Liu, D., Lin, J., Teng, X., Li, Y., et al. (2021). Transcriptome analysis of peripheral blood mononuclear cells reveals distinct immune response in asymptomatic and re-detectable positive COVID-19 patients. Front. Immunol. *12*, 716075.
59. Giroux, N.S., Ding, S., McClain, M.T., Burke, T.W., Petzold, E., Chung, H.A., Rivera, G.O., Wang, E., Xi, R., Bose, S., et al. (2022). Differential chromatin accessibility in peripheral blood mononuclear cells underlies COVID-19 disease severity prior to seroconversion. Sci. Rep. *12*, 11714.
60. Jergović, M., Watanabe, M., Bhat, R., Coplen, C.P., Sonar, S.A., Wong, R., Castaneda, Y., Davidson, L., Kala, M., Wilson, R.C., et al. (2023). T-cell cellular stress and reticulocyte signatures, but not loss of naïve T lymphocytes, characterize severe COVID-19 in older adults. GeroScience *45*, 1–16.
61. Sirén, J., Välimäki, N., and Mäkinen, V. (2014). HISAT2-fast and sensitive alignment against general human population. IEEE/ACM Trans. Comput. Biol. Bioinforma. *11*, 375–388.
62. Liao, Y., Smyth, G. K., and Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. *47*, e47.
63. Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta *2*, e107.
64. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. Gigascience *10*, giab008.
65. Tange, O. (2018). GNU parallel 2018. Lulu.com.
66. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics *31*, 2032–2034.
67. Love, M., Anders, S., and Huber, W. (2014). Differential analysis of count data–the DESeq2 package. Genome Biol. *15*, 10–1186.
68. Parvati Sai Arun, Miryala, S.K., Rana, A., Kurukuti, S., Akhter, Y., and Yellaboina, S. (2018). System-wide coordinates of higher order functions in host-pathogen environment upon Mycobacterium tuberculosis infection. Sci. Rep. *8*, 5079.
69. Leinonen, R., Sugawara, H., and Shumway, M.; International Nucleotide Sequence Database Collaboration (2011). The sequence read archive. Nucleic Acids Res. *39*, D19–D21.
70. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890.
71. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915.
72. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.
73. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.
74. Kinsella, R.J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., et al. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database *2011*, bar030.
75. Stowell, S., and Stowell, S. (2014). Probability distributions. Using R for statistics, 87–98.
76. Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal, Complex Syst. *1695*, 1–9.
77. Que, X., Checconi, F., Petrini, F., and Gunnels, J.A. (2015). Scalable community detection with the louvain algorithm. In 2015 IEEE International Parallel and Distributed Processing Symposium (IEEE), pp. 28–37.

78. Hermjakob, H., Montecchi-Palazzi, L., Lewington, C., Mudali, S., Kerrien, S., Orchard, S., Vingron, M., Roechert, B., Roepstorff, P., Valencia, A., et al. (2004). IntAct: an open source molecular interaction database. Nucleic Acids Res. *32*, D452–D455.
79. Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. Nucleic Acids Res. *34*, D535–D539.
80. Kachitvichyanukul, V., and Schmeiser, B. (1985). Computer generation of hypergeometric random variates. J. Stat. Comput. Simulat. *22*, 127–145.
81. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B *57*, 289–300.
82. Ihaka, R., and Gentleman, R. (1996). R: a language for data analysis and graphics. J. Comput. Graph Stat. *5*, 299–314.
83. Dormann, C.F., Gruber, B., and Fründ, J. (2008). Introducing the bipartite package: analysing ecological networks. interaction *1*, 8–11.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
PRJNA639275	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA639275	Arunachalam et al., 2020[14]
PRJNA660067	https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA660067	Overmyer KA et al., 2021[52]
PRJNA662985	https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA662985	Zheng HY et al., 2020[44]
PRJNA670179	https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA670179	Rother N et al., 2020[53]
PRJNA717662	https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA717662	Ryan FJ et al., 2022[54]
PRJNA718349	https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA718349	Lévy Y et al., 2019[55]
PRJNA730810	https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA730810	Knabl L et al., 2022[56]
PRJNA735729	https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA735729	Dhindsa S et al., 2021[57]
PRJNA744408	https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA744408	Zhang J et al., 2021[58]
PRJNA849921	https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA849921	Giroux NS et al., 2022[59]
PRJNA876121	https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA876121	Jergović M et al., 2021[60]
Software and algorithms		
HISAT2	https://daehwankimlab.github.io/hisat2/	Sirén et al., 2014[61]
SUBREAD	https://subread.sourceforge.net/	Liao et al., 2019[62]
fastp	https://github.com/OpenGene/fastp	Chen et al., 2023[63]
Samtools	https://htslib.org/	Danecek et al., 2021[64]
parallel	https://www.gnu.org/software/parallel/	Tange et al., 2018[65]
Sambamba	https://lomereiter.github.io/sambamba/	Tarasov et al., 2015[66]
DESeq2	https://bioconductor.org/	Love et al., 2014[67]
Pathway Enrichment analysis	https://pubmed.ncbi.nlm.nih.gov/29567998/	Arun et al., 2018[68]

RESOURCE AVAILABILITY

All datasets used for this study are available in public domains. The information has been provided in the key resources table.

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Sailu Yellaboina (bio.sailu@gmail.com).

Material availability

This study did not generate new unique reagents.

Data and code availability

- All datasets used for this study are available in public domains. Any additional information required to reanalyze the data reported in this work is available from the lead contact upon reasonable request.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Gene expression data collection

We obtained RNA-Seq gene expression data sets from the Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) and ArrayExpress (ebi.ac.uk/biostudies/arrayexpress/) repositories. In order to eliminate redundancy in the gene expression data, we included datasets from the GEO only if they were not already present in the ArrayExpress. We have included only human studies related to PBMC gene expression in response to SARS-CoV-2 infection. NGS data from the SRA database was downloaded using a fastq-dump of NCBI

SRA toolkit[69] in FASTQ format. The gene expression data includes PBMC from healthy individuals, as well as individuals with mild, moderate and severe Covid-19 disease conditions, as detailed in Tables S1–S4. Generally, mild cases exhibit no signs of pneumonia on imaging and mild clinical symptoms. Moderate cases are characterized by symptoms such as fever, respiratory issues, and radiological evidence of pneumonia. Severe cases exhibit symptoms such as dyspnea, low blood oxygen saturation, extensive lung infiltrates, respiratory failure and septic shock.

RNA-Seq data analysis

The RNA-Seq data was preprocessed to remove low-quality reads and adapter sequences using the FASTP software.[70] HISAT2 tool[71] was used to create the index of the reference genome and performed the alignment of RNA-Seq reads against the reference genome version GRch38 from the ENSEMBL database (https://asia.ensembl.org/). The resulting alignment file was processed to remove PCR duplicates, sort the alignments, and index the resulting BAM file using SAMtools.[72] Finally the number of reads mapped to each gene were quantified using the featureCounts tool.[73] To ensure comparability of gene expression data with pathway annotations, we mapped the Ensembl gene ids to NCBI Entrez gene symbols. We obtained the mapping information from NCBI gene information and Ensembl biomart.[74]

Differential expression analysis

We used the 'estimateSizeFactors' function of DeSeq2[67] using the 'geometric mean method' to calculate size factors for each sample, which represent the normalization factors required to adjust the raw read counts for differences in library size and sequencing depth. These size factors are then used to normalize the count data, so that the expression levels between samples become comparable and unbiased. Finally we performed variance-stabilizing transformation (VST) in DESeq2 to stabilize the variance across the mean count values, making the data more suitable for differential expression analysis.

Combining p-values of differential expression

We combined the p-values of differential expression for each gene across disease severity groups using our previously described method.[20] First, we transformed the initial p-values of each gene obtained from the differential expression analysis of the gene expression dataset into z-scores using the inverse cumulative distribution function (CDF).[75] Then, we took the weighted sum of the z-scores and divided it by the square root of the sum of the weights to generate the average z-score. Here, the weights are equivalent to the mean of the number of replicated samples in the two conditions being compared. Therefore, the differential expression P-value coming from a higher number of replicates will have more weight. Finally, the average z-score was converted to a p-value using the CDF.

Enrichment analysis of genesets/pathways

Annotated functions of humans, i.e., pathways, motifs, gene ontology (GO) and immunological signatures were collected from Molecular Signature Database (MSigDB).[22] Gene set enrichment analysis in a ranked list of differential expressed genes was calculated similarly to the method published previously.[76] Differential expression P-values of each gene were converted into normal statistics Z-score using inverse cumulative distribution function (qnorm) in R.[77] Mean Z-score of a pathway was calculated by dividing the square of the number of genes in a corresponding pathway. Gene-set enrichment score (Mean Z-score) was calibrated against the background distribution, by using randomly sampled n (number of genes in a pathway) scores and calculating mean Z-score. This process was repeated over 20000 times. The mean and standard deviation of the sampling distribution thus obtained was used for correction of the original score.

Since there is an extensive overlap between the genes of pathway databases such as reactome, biocarta, wiki , KEGG, PID and gene ontologies we pooled the genesets of different categories to form a module of gene-sets. Overlap between the gene-sets was calculated using jaccard coefficient and the module of gene-sets were detected using the Louvain algorithm which was implemented in 'igraph' R package.[76] Finally the enrichment score of geneset modules was calculated by taking the median of individual gensets enrichment scores.

We prepared a database of experimentally verified protein-protein interactions between SARS-CoV-2 and human proteins by combining the data from IntAct and Biogrid.[78,79] Enrichment scores (P-value) of pathways in each set of human proteins that interact with various SARS-CoV-2 proteins were calculated using a hypergeometric distribution test using the 'phyper' function in R.[80] P-values were adjusted for the number of pathways / multiple comparisons using Benjamini-Hochberg correction.[81] The final P-values of each gene were converted into normal statistics Z-score using inverse cumulative distribution function (qnorm) in R.[77] Similarly we calculated the enrichment score (Z-score) for pathways in the proteins other than the interacting target proteins of SARS-CoV-2 protein. Final enrichment score of the pathway is obtained by taking the absolute maximum of both the Z-scores.

Data visualization

All the analysis and visualization was carried using R statistical software.[82] Heatmaps were drawn with the R package 'ComplexHeatmap' whereas bipartite graphs were drawn with plot 'plotweb' function in R package 'bipartite'.[83] We have used a 'rescale' function in R package 'scales' to normalize the matrix before generating heatmaps.