Meta-analysis of orthogonal OMICs data from COVID-19 patients unveils prognostic markers and antiviral factors.

Abhijith Biji[1,2†], Shachee Swaraj[2†], Oyahida Khatun[2], Rohan Narayan[2], Rahila Sardar[3], Deepshikha Satish[3], Simran Mehta[4], Hima Bindhu[4], Madhumol Jeevan[4], Deepak K Saini[5], Amit Singh[2], Dinesh Gupta[3], Shashank Tripathi[2*]

[1]Undergraduate Programme, Indian Institute of Science, Bengaluru, India

[2]Microbiology & Cell biology Department, Centre for Infectious Disease Research, Indian Institute of Science, Bengaluru, India

[3]Translational Bioinformatics Group, International Centre for Genetic Engineering and Biotechnology, New Delhi, India

[4]COVID-19 Diagnostic Facility, Centre for Infectious Disease Research, Indian Institute of Science, Bengaluru, India

[5]Molecular Reproduction & Developmental Genetics, Indian Institute of Science, Bengaluru, India

*Correspondence: shashankt@iisc.ac.in

[†]These authors contributed equally to this manuscript

ABSTRACT

Coronavirus disease 2019 (COVID-19) pandemic has lasted more than a year since its first case in December 2019 and yet its social and economic burden continues to grow. While a tremendous amount of OMICs data has been generated from COVID-19 patient samples, the host antiviral response and markers of disease progression remain to be completely delineated. In this study, we have conducted a meta-analysis of published transcriptome and proteome profiles of the nasal swab and bronchioalveolar lavage fluid (BALF) samples of COVID-19 patients to identify high confidence upregulated host factors. This was followed by rank ordering, shortlisting, and validation of overexpression of a set of host factors in a nasal swab/BALF samples from a cohort of COVID-19 positive/negative, symptomatic/asymptomatic individuals. This led to the identification of host antiviral response in the upper respiratory tract and potential prognostic markers. Notably, SEPRIN B3 and Thioredoxin were identified as potential antiviral factors. In addition, several S100 family

proteins were found to be upregulated in COVID-19 specific and disease severity dependent manner. Overall, this study provides novel insights into the host antiviral mechanisms and COVID-19 disease progression.

KEYWORDS

COVID-19, SARS-CoV-2, Transcriptomics, Proteomics, Overlap Analysis, Antiviral response, Prognostic marker

INTRODUCTION

COVID-19 pandemic has been reported in 221 countries with a total of 110,526,493 cases and 2,442,986 deaths worldwide, as on February 18, 2021, 12:49 GMT (https://www.worldometers.info/coronavirus/). The causative virus SARS-CoV-2 is a member of the family *Coronaviridae* which contains a single-stranded positive-sense RNA genome that encodes ~27 proteins (1). The initial symptoms of infection are flu-like, including fever, dry cough, headache, muscle pain, and shortness of breath, which may resolve in 7-8 days in healthy individuals. The majority of infected individuals, especially young adults may remain asymptomatic, even with considerable viral load (2, 3). In severe cases, it may progress to Acute Respiratory Distress Syndrome (ARDS), multi-organ failure, and occasionally to death, especially in patients with immune-deficiency and comorbidities (4). An effective and timely IFN response is critical in resolving viral infections (5), however, SARS-COV-2 seems to have acquired multiple strategies to suppress host immune response and disrupt immune homeostasis (6). Severe cases of COVID-19 show extensive damage to lung tissues caused by a hyperactive proinflammatory response to the virus often termed as a cytokine storm. (7). The mechanisms underlying immune dysregulation brought about by the viral infection remain to be completely defined. Extensive efforts have been undertaken to understand the host response to viral infection using high throughput genomics and proteomics technologies. Several studies have explored serum diagnostic and prognostic markers by evaluating transcriptomic and proteomic changes in mild, severe and fatal cases of COVID-19 (8-11). Elevated levels of interleukins (IL6 and IL10), C-reactive proteins (CRP), MCP1, MIP1A, TNFα, procalcitonin (PCT), lactate dehydrogenase (LDH), ferritins, D-dimers and cardiac troponins in serum samples of COVID-19 patients are indicative of disease severity (8-13). Examination of host response at the primary site of infection in the upper respiratory tract, is crucial to understand viral pathogenesis. Various studies utilized the bronchoalveolar lavage and nasopharyngeal

swabs to characterize the changes in transcripts and proteins during infection to understand COVID-19 pathogenesis (14-19), which have highlighted significantly upregulated genes and biological pathways altered during infection. While proinflammatory cytokines, chemokines, enzymes in neutrophil mediated immunity and several IFN stimulated genes (ISGs) have consistently showed up in their analysis, an experimental validation and mechanistic studies are generally lacking (14-19). A detailed characterization of antiviral response in the upper respiratory tract of patients, its variation with age, sex and association with progression of disease severity remains to be accomplished.

The goal of our study was to identify genes that are upregulated during SARS-CoV-2 infection in the upper respiratory tract of patients and understand their role in viral infection and disease progression. For this, we surveyed the literature for OMICs data from COVID-19 positive patient's nasal swab and BALF samples and selected 4 transcriptomic and 3 proteomic datasets. We performed a hypergeometric distribution-based overlap analysis followed by cumulative fold change score-based prioritization to shortlist genes. This was followed by testing gene expression level in nasal swab/ BALF samples from a cohort of COVID positive, negative, symptomatic, and asymptomatic individuals, ranging from 30-60 years in age and of mixed gender. Quantitative PCR data analysis revealed upregulation of TXN, SERPINB3, S100A8, ASS1, S100A9, S100A6, S100P, DEFA3, and KRT6A in COVID-19 specific manner, whereas AGR2, LCN2, and BPIFB1 were identified to be associated with the severity of COVID-19 symptoms. Overall, our analysis reveals a specific set of host genes that are upregulated in a COVID-19 specific manner and are indicative of disease severity. In addition, we have uncovered novel potential antiviral mechanisms that can be repurposed to mitigate viral replication and pathogenesis.

MATERIALS AND METHODS:

Data collection and Processing:

Transcriptomics and protein abundance data from COVID-19 patient's naso- and oropharyngeal swab, bronchoalveolar lavage fluid (BALF), and other respiratory specimens were chosen from PubMed, BioRxiv, and MedRxiv using different combinations of keywords like "**COVID-19, SARS-CoV-2, Transcriptomics, Proteomics, BALF, swab**". Studies dealing with gene expression profiles of SARS-CoV-2 infected non-human cell lines and tissues were not

considered. The SARS-CoV-2 and COVID-19 collections in the EMBL-EBI PRIDE proteomics database (20) were retrieved and used without any modification. In the NCBI GEO database (21) the following combination of terms was used to collect relevant datasets: **((covid-19 OR SARS-COV-2) AND gse [entry type]) AND "Homo sapiens"[porgn: _txid9606]**. The retrieved datasets were then filtered by their date of publication to collect the studies published between the 1st of January 2020 and the 15th of September 2020. The filtration of datasets was carried out using two parameters, fold-change, and its significance value. Genes and proteins with a fold-change value of ≥ 1.5 and q-value ≤ 0.05 were chosen for the overlap analysis. The raw p-value was used for filtering in cases where the adjusted p-value was not provided, albeit with a more stringent cut-off of ≤ 0.01. The UniProt IDs in filtered protein abundance datasets were converted to their corresponding primary Gene Symbols using UniProt (22).

Gene set overlap analysis:

The GeneOverlap class of R package "GeneOverlap" (23) was used for testing whether two lists of genes are independent, which is represented as a contingency table, and then Fisher's exact test was used to find the statistical significance. Genes with less than 0.01 overlap p-value were selected for further analysis. The number of background genes for proteome-proteome pairwise study and the transcriptome-proteome pairwise study was 25,000, i.e., the number of protein-coding genes in Hg19. For the transcriptome-transcriptome overlap study, the number of background genes was taken to be the union of the expressed genes in both the datasets considered.

Gene Ontology, Interferome Analysis, Cellular and tissue localization analysis:

Enriched GO terms were obtained by express analysis on Metascape (24) and plotted using ggplot2 (25). The database Interferome v2.01 (26) was queried using gene symbols for identifying interferon regulated genes (IRGs) in normal samples of the respiratory system from both in vitro and in vivo experiments in humans. For cellular localization, each gene was queried on UniProt annotation (27) and Human Protein Atlas ver20.0 (28) and then manually annotated. The single-cell expression data of transcripts was also obtained from Human Protein Atlas ver20.0 (Available from http://www.proteinatlas.org/). They were further filtered to obtain cells that are associated with the immune system or respiratory tract.

Virus-Host protein-protein interaction network analysis:

The interaction data for the selected 46 genes were retrieved from publicly available interaction datasets (29). The retrieved information was then used to generate a network map. Cytoscape v3.8.0 (30) was used to construct the interaction network for virus-host protein-protein interaction. STRING database within the Cytoscape store was used to query the proteins to elucidate the interactions between the proteins significantly altered during SARS-CoV-2 infection. The resulting STRING interaction network (confidence ≥0.999 for all the proteins and confidence ≥0.90 for NAMPT; max number of interactors = 10) was merged with the virus-host PPI on Cytoscape.

Nasopharyngeal Swab Collection and RNA Isolation:

Nasopharyngeal swabs were collected around Bengaluru Urban city and brought to COVID-19 Diagnostic Facility at the Indian Institute of Science in viral transport media (VTM). RNA from patients was isolated using kits recommended and provided by the Indian Council of Medical Research. Appropriate RNA samples and VTMs were curated and selected manually based on age, sex, and severity of the disease.

qRT-PCR based measurement of gene set expression:

Equal amounts of RNA were converted into cDNA using Prime Script™ RT Reagent Kit with gDNA Eraser (Perfect Real Time) (RR047A, Takara-Bio) and then diluted with 80µl nuclease-free water. The gene expression study was conducted using PowerUp™ SYBR™ Green Master Mix (A25778, Applied Biosystems™) with 18srRNA as the internal control and appropriate primers for the genes (Supplementary Table 3).

Statistical analysis:

All statistical analyses and overlaps were performed in the R statistical environment version 4.0.3 via RStudio version 1.3.1093. All boxplots were made using the ggplot2 package in R (25) where the hinges of boxes represent the first and third quartiles. The whiskers of the boxplot extend to the value which is 1.5 times the distance between the first and third quartiles. Any data point that lies beyond it is plotted separately with a dot and is considered outliers. Each data point in the boxplot represents one of the triplicates in RT-qPCR for a particular gene in a particular patient sample. Heatmaps were generated using the R package ComplexHeatmap (31).

Ethics Statement:

This study was conducted in compliance with institutional human ethics and biosafety guidelines, (IHEC No. 13-11092020; IBSC/IISc/ST/19/2019-20), following the Indian Council of Medical Research and Department of Biotechnology recommendations.

Results:

Compilation of published transcriptomics and proteomics data from COVID-19 patient samples and overlap analysis revealed 567 upregulated genes.

Based on the selection criteria (materials and methods) four transcriptomics and three proteomics datasets were chosen and were carried down for the analysis as mentioned in the workflow (Figure 1A). All these studies identified differentially expressed genes in infected patients with healthy individuals as control (Table S1). The filtration of data was carried out to sort only significantly upregulated genes from all the datasets (Table S2). Pairwise overlap analysis was performed on the filtered genes/proteins from each study by calculating the Jaccard score and significantly overlapping genes (p-value < 0.01) between T1-T3 (14), T1-T4 (9), T1-P3 (2), T3-T4 (504), T3-P1 (10), T3-P2 (8), T3-P3 (17), T4-P1 (8). T4-P3 (15) and P1-P3 (3) were determined (Figure 1B). Similar forms of overlap analysis have been previously used to compare multiple datasets and obtain the significance of intersections (32). Union of intersections between the T-T and T-P and P-P after the overlap analysis results in 567 genes (Figure 1B). To reiterate the functional characteristics of the differentially expressed genes (DEGs), we scrutinized the biological processes and signaling pathways they are involved in. Pathway enrichment of 567 genes from the union of all intersections from overlap analysis (TT+TP+PP) shows that the genes are well enriched in a biological process like protein elongation, IFN signaling, chemotaxis of granulocytes, and inflammatory pathways (Figure 1C). Interferons (IFNs) are secreted once the virus is detected within the cell by PRRs (pathogen recognition receptors) through a multitude of pathways involving transcription factors such as interferon regulatory factors (IRF3, IRF7, and IRF9.) and NF-κB. Therefore, the possibility of these 567 genes being regulated by the IFNs was also investigated. 205 genes are type I IFN regulated, 170 genes by Type II IFNs, 327 genes are both types I and type II IFN regulated while only 16 genes were determined to be regulated by all the three classes of IFNs (Figure 1D). The 16 genes are well-renowned interferon-stimulated genes (ISGs). Some direct antiviral effector ISGs (IFITs, MX1, OAS3, and OAS1.), as well as positive regulators (STAT1) of IFN response, are among the upregulated genes among these 16 ISGs, indicating an active innate antiviral response inside the cells upon SARS-CoV-2 infection.

Ranking of genes based on cumulative score revealed differentially expressed genes regulating immune response and inflammatory signaling in COVID-19 patients.

Since proteome dictates the outcome in a cell, genes from the union of intersections that were reported at least in one of the proteomics studies were selected for further analysis. This gave a total of 46 genes that were intersecting in T-P (26), P-P (3), TT-TP (16), TP-PP (1), and TT-TP-PP (1) overlaps (Figure 2A and 2B). A cumulative score for the 46 selected significantly upregulated genes was calculated using the sum of their $\log_2$ Fold-change values and ranked (Figure 2C). The enrichment of these 46 genes in each of the datasets where the expression is reported is shown in Figure 2B. Many of these genes are directly regulated by different classes of interferons. 15 genes are regulated by IFN-I, while 8 genes by IFN-II. 20 genes are regulated by both type-I and type-II IFNs, while only 2 genes by all the three types of IFNs (Figure 2D). Most of the IFITs and other ISGs that were earlier determined in our analysis to be regulated by all the three types IFNs are no more in the list since those ISGs were only reported upregulated at transcriptome level (only in T-T overlap) and hence were lost when the genes were filtered for their upregulation at the protein level, leaving behind only MX1 and OAS3 (Figure 1C and 2D). The biological functions of the selected 46 genes were also investigated to understand their roles in COVID-19 pathophysiology. The pathways enriched were mainly related to innate immune response and defense against microbes along with inflammatory and immune signaling, neutrophil degranulation, and cellular response to TNF and interferon-gamma (Figure 2E). Since these genes are significantly upregulated during the SARS-CoV-2 infection, it is highly plausible that the hyperactive exaggerated immune response and inflammation leading to lung immunopathology maybe due to these unchecked responses (Figure 2E).

Further, to understand the pathophysiology of COVID-19, the interactions of these genes with SARS-CoV-2 proteins were inspected by analyzing the publicly available interaction data (29). Host protein-protein interactions were retrieved from the STRING database and merged with the virus-host protein-protein interactions giving a discrete picture of how the viral proteins target various cellular processes during infection. Other than NAMPT, UQCRC2, and RAB5C, it was mainly ribosomal proteins that were primary interactors to the SARS-CoV-2 proteins (Figure 2F and 2G). Ribosomal proteins, along with a eukaryotic elongation factor, were targeted by four different viral proteins (Nsp1, Nsp8, Nsp9, and N proteins of SARS-CoV-2) which suggests that cellular translation machinery is one of the

critical processes that is likely hijacked by the virus for its multiplication (Figure 2F). The interactions with other proteins (NAMPT, UQCRC2, and RAB5C) also explain an approach taken by the virus via direct interaction with host proteins to partially take over the system and utilize its resources efficiently for its benefits (Figure 2G). These interactions may be one of the potential ways of evading intracellular immune responses and disrupting the cellular processes that lead to the clearance of the virus. A number of upregulated proteins were predicted to localize in the intracellular organelles like endoplasmic reticulum, mitochondria, Golgi complex, and endosomes (Figure S1A). A thorough analysis of the selected list of 46 genes using Human Tissue Atlas revealed that they express in various types of cells of the respiratory tract and immune effector cells known to survey infection sites (Figure S1B). The relative expression levels show that genes associated with protein synthesis (ribosomal proteins and elongation factors) are highly expressed compared to any other genes and are enriched across all the tissues in the map (Figure S1B).

Validation of selected upregulated genes in a cohort of COVID-19 patients.

For further analysis using RT-qPCR, we selected all genes with a cumulative score greater than 10. In addition to this, all genes belonging to the S100 family that came up within these 46 genes were also selected since they are known markers of inflammation. TXN (Thioredoxin) was selected because it was at the center of the Venn diagram (Figure 2A) and hence present in significant intersections between overlaps comparing Proteome-Proteome, Transcriptome-Transcriptome, and Proteome-Transcriptome. From a total of 64 patients, nasopharyngeal swabs of 16 positive-symptomatic, 16 positive-asymptomatic, 16 negative-symptomatic, and 15 negative-asymptomatic were collected and total mRNA was isolated (Table 1). The upregulation of the selected genes among 46 genes was verified by RT-qPCR on the patient samples and expressed as $\log_2$ fold-change with respect to the negative asymptomatic group (Figure S2, Figure 3A). The heatmap depicts the enrichment of the selected genes in different patient samples (Figure 3A). We determined a direct correlation between the increasing viral load in COVID-19 patients and the upregulation of selected proteins as detected by the RT-qPCR of viral envelope (E) gene and the $\log_2$ Fold-change of respective genes in the patient's sample. The data shows that the Ct value for the E gene was negatively correlated with $\log_2$ Fold-change of genes showing that viral load and disease severity are positively correlated (Figure S5). The upregulation of all the selected genes was found to be higher in positive symptomatic patients with a higher viral load than positive asymptomatic (Figure 3A and Figure S5). A comparative heatmap in Figure 3B gives an insight into the genes that are

COVID-19 specific markers, or markers for severity during SARS-CoV-2 infection, and those that are indicative of any general infection. While all the genes reported upregulated are a marker for infection (Figure 3B; NA-PS), only a few genes showed significant upregulation in a COVID-19 specific manner (Figure 3B; NS-PS). Other than S100A12 and KRT8, all other genes are significantly upregulated in symptomatic COVID-19 patients in contrast to asymptomatic COVID-19 patients and hence can be considered as prognostic markers. SERPINB3, ASS1, S100A6, and S100A9 also display a significant COVID-19-specific enrichment and hence, can be acknowledged as possible markers for prognosis of the disease (Figure 3B, NS-PS).

To understand the influence of age and sex on the COVID-19 pathology we chose to study the reported upregulation of genes in patient's samples by categorizing them based on age groups (30-40, 41-50, and 51-60) and gender (male and female) (Figure 3C, Figure 3D, Figure S3 and S4). The RT-qPCR analysis revealed that almost all the genes are induced in positive symptomatic patients. However, the S100 family of genes show upregulation in the case of positive symptomatic (at least 4-fold with respect to negative asymptomatic) and negative symptomatic (at least 1.9-fold with respect to negative asymptomatic where it is significant, Figure 4A) patients and it is independent of the age and sex of the patient indicating that they are common markers of infection (Figure 3A, 3C, and 3D). Neutrophil defensin alpha 3 (DEFA3) is interestingly upregulated in some of the positive symptomatic patients with no biases towards a specific age group or sex. DEFA3 is a small antimicrobial peptide that has been reported to display broad-spectrum antiviral activity (33-35). Hence, an upregulation during SARS-CoV-2 infection is not surprising. However, expression of DEFA3 remained undetermined in many of the patient's samples, indicating a requirement of a larger sample size to confirm its importance for prognosis.

S100 family transcripts are highly enriched in COVID-19 symptomatic patients.

An upregulation of S100 proteins is reported by many studies as an indication of viral or bacterial infections (36). Interestingly, a large proportion of selected 46 DEGs include the S100 family of proteins in our analyses (Figure 2B). S100A8, S100A9, S100A6, and S100P show a considerable enrichment in positive symptomatic patients in contrast to positive asymptomatic and negative symptomatic patient samples (Figure 4A). The extracellularly secreted S100 proteins include S100A12, S100A8, and S100A9 (Figure S1A), all of which have been shown to serve as a danger signal and in regulating immune response (37). They activate NF-kB

signalling through RAGE and TLR4 pathways stimulating the cells to produce proinflammatory cytokines at the site of infection (37). It is no surprise that their enrichment in our analyses suggests that they are significantly upregulated in COVID-19 positive patients. An age-dependent upregulation of S100A6 among COVID-19 positive patients is very evident in the age group 41-50 (Figure 4B). A similar trend is observed for its upregulation independent of sex differences. S100A6 is significantly upregulated in positive symptomatic males as well as females (Figure 4C). For patients falling in the negative category, only S100A12, S100A9, and S100P show sex-based difference and all show age-based differences in 41-50 and 51-60 age groups except for S100A6 which shows the significant lowering of gene expression in negative symptomatic patients falling in the 30-40 age group (Supplementary Figure 6A and B). The S100 family of proteins display considerable expression in macrophages and other immune-related cells in the tissue expression atlas indicating its involvement in inflammation and immune response across a broad range of pathogens (Figure S1B).

DISCUSSION

Application of high throughput OMICs technologies to understand systems level regulation of biological processes has become commonplace. A number of studies have analysed changes in global transcriptome and proteome in COVID-19 patient samples of various kinds. These studies have given overview of the cellular processes that are modulated during SARS-Co-V2 infection, however translation of this knowledge into antiviral interventions requires validation and mechanistic studies. Meta-analysis of virus-host interaction Big Data is a useful approach to narrow down on key host factors involved in viral replication and pathogenesis (32, 38). Using this approach, we have performed an integrative analysis of published transcriptomics and proteomics data from COVID-19 positive nasal swab and BALF samples, to identify host factors involved in SARS-CoV-2 pathology and disease progression. In our analysis we focussed on genes which were represented in multiple orthogonal datasets, especially in proteomics data. This was done based on assumption that changes at RNA levels must also be manifested at the protein level to bring about phenotypic changes in the infected individuals during infection. Expression of the genes shortlisted through meta-analysis was tested in a cohort of nasal swab samples collected for COVID-19 diagnosis. This included samples from COVID-19 negative and positive, and within those two, samples from symptomatic and asymptomatic individuals. This was done to ensure identification of genes which are overexpressed in COVID-19 specific manner and those which indicate disease severity.

Previous meta-analysis studies by reusing BALF/PBMC/Whole blood transcriptomic data have extensively studied the immune and metabolic pathways that contribute to/counteract SARS-CoV-2 infection (39, 40) and provided insights into therapy strategies such as the use of tocilizumab against IL-6 receptor (41). A meta-analysis study that integrates orthogonal datasets of proteomics, transcriptomics and interactomes have also narrowed down on multiple pathways that are associated with immune related functions (42).

The cumulative ranking of shortlisted genes based on their upregulation showed SERPINB3 at the top (Figure 2C). SERPINB3 belongs to the serine protease inhibitor family and it inhibits papain-like cysteine proteases such as papain, cathepsin-S, -K, and –L, thus pointing towards a possible mechanism for hindering viral entry through cathepsin inhibition or prevent viral spread by interfering with proteolytic processing of Spike protein (43, 44). Furthermore, SERPINB3 has been shown to play an important role in attenuating TNF- α mediated apoptosis and exhibits an anti-chemotactic effect for natural killer (NK) cells (45), and thus may mitigate the host inflammatory immune response. Additionally, SERPINB3 upregulation was more prominent in positive symptomatic patients when compared to negative asymptomatic or positive asymptomatic category, confirming its utility as a COVID-19-specific disease severity marker (Figure 3B). Other notable SERPIN family members like SERPINB1 (which came up during our cumulative score ranking, Figure 2C) and SERPINE1 has been shown to restrict pathogenesis of influenza, the latter by preventing influenza A glycoprotein maturation (46, 47). SERPINA1 deficiencies or mutations in populations were found to be associated with severe forms of COVID-19 and IL-6 to α1-antitrypsin (protein coded by SERPINA1) ratios positively correlated with disease severity and mortality (48, 49). Notably, several ribosomal proteins (RPs) emerged as highly upregulated proteins in the patient samples (RSP3A, RPL4, RPL5, RPL18, RPL13A, RPS4X, RPL7A, RPS9, and RPS3) as evident from the enriched GO term “peptide chain elongation”. Previous studies have shown that ribosomal proteins (RPs) are hijacked by different viruses during infection to activate IRES-mediated translation of viral proteins (50-57) . Besides an upregulation, RPs are also recorded to directly interact with viral proteins (50, 58, 59). Our inspection for possible interactions between the ribosomal proteins that came up in our list with the SARS-CoV-2 proteins revealed that nsp1, nsp8, nsp9 and nucleocapsid (N) proteins of SARS-CoV-2 directly interact with the host ribosomal proteins (Figure 3C). Considering these data, it is likely RPs in our list may be targeted by SARS-Co-V2 proteins for selective translation of viral proteins.

Several genes shortlisted through our analysis play direct or indirect role in prepping the cell for combating the infection. The GO analyses showed, pathways like “antimicrobial humoral response”, “Response to interferon-gamma", “interferon alpha-beta signalling”, “granulocyte activation”, “response to interferon-alpha” “granulocyte chemotaxis”, neutrophil degranulation” to be highly enriched. Cornification is another significantly enriched pathway which has been reported to regulate expression of proteins involved in restructuring the cellular architecture like keratin proteins during viral infection (60-65). The viruses also exploit the cytoskeletal elements for their entry, assembly and exit of viral particles, thus explaining the enrichment of pathways like “plasma membrane bounded cell projection assembly”, “cytoskeleton-dependent intracellular transport”, “actin cytoskeleton organization” and “vesicle cytoskeletal trafficking” appearing in our analysis.

From analysis of potential direct interactions between the viral and host proteins during infection based on published interactome data, UQCRC2 (Cytochrome b-c1 complex subunit 2) was found to be interacting with 5 different SARS-CoV-2 proteins, orf7a, nsp7, nsp2, orf9c, and Membrane (M) (Figure 2G). We reasoned this unique finding as a mechanism of induction of oxidative mitochondrial dysfunction that affects ATP production (66, 67). UQCRC2 directly interacts with critical components of respiratory chain complexes II and IV that have been reported to be upregulated during viral infections, affecting ATP production in the infected cells that is utilized by the virus for its replication (66). Nicotinamide phosphoribosyl transferase (NAMPT), which catalyses a rate-limiting step in de novo NAD biosynthesis, was another cellular factor in our list , with potential interactions with SARS-Co-V2 M, Orf8 and Orf9c proteins (Fig 2G) (68-70). A recent finding on SARS-CoV-2 reports that the virus causes a downregulation of the de novo NAD biosynthetic pathway (71). This suggests that NAMPT, being an enzyme critical in de novo pathway, maybe a direct target for modulation which is evident in our analyses.

A subset of upregulated proteins is secretory including S100A12, S100A8, S100A9, SERPINB3, DEFA3, LCN2 and TXN. LCN2, which came up in our study was previously shown to be an important biomarker for viral infection (72, 73), and was also reported to be upregulated in transcriptomic and proteomic studies in COVID-19 patients (74, 75). It has been shown that virus replication and inflammation goes down when thioredoxin reductase, a reducing agent of TXN is inhibited using auranofin, pointing towards the importance of redox

environment during SARS-CoV2 pathogenesis (76), akin to many inflammatory responses which are governed by the redox status of the cell (77). We observed upregulation of members of the S100 family of genes like S100A8, S100A9, S100A6, and S100P in severe COVID-19 specific manner. It was reported that increased S100A8/A9 (calprotectin) levels in serum of COVID-19 patients has been correlated with severe forms of the disease (78, 79). These were also observed with increased incidence of neutrophil extracellular trap formation and clinical thrombosis in patients (80). Our findings also align with transcriptomic studies on lung tissue of fatal COVID-19 cases which report an upregulation in S100A12, S100A8, S100A9, and S100P in patients (81). Taken together S100 family of genes can be indicative of COVID-19 infection and disease progression and can be considered as prognostic markers.

Our comprehensive analyses of transcriptome and proteome data along with experimental validation of upregulated genes in COVID-19 cohort lead us to a set of host factors that have direct implications in the SARS-CoV-2 infection and associated diseases. The age and sex-based differences in gene expression that we observed highlight the varied individual responses. Many previous studies have also noted the biases in COVID-19 outcomes among age and sex groups (82-85), drawing attention to their robustness of immune responses and overall physiology. Overall, our data suggests that S100A6, S100A9, S100A8, SERPINB3, TXN, and ASS1 exhibit upregulation with the disease severity and progression in COVID-19 patients. However, an exhaustive investigation to dissect their direct roles and correlation with infection is required to certainly declare them as COVID-19 prognostic markers.

Strikingly, we observed several cellular pathways and signalling cascades being directly targeted by SARS-CoV-2 through the understanding of transcriptome and proteome changes in bronchoalveolar lavage fluids (BALF) and nasopharyngeal swab samples from COVID-19 positive symptomatic and asymptomatic patients. Our study discusses the ungrateful aspects of COVID-19 infection and host response at the site of infection. In any infection, knowledge of affected cellular signalling pathways opens the road to drug repurposing and novel therapeutic strategies. Drugs that target host factors rather than viral proteins help surpass the problem of high mutation rates of viruses that could reduce the efficacy of the latter. Studying enriched pathways and genes also provide insights into unique antiviral immune responses by the host or immune evasion strategies deployed by the virus. The findings of this study explore both these aspects and also look into the strategies shared by SARS-CoV-2 compared to other previously known viral infections. It reaffirms the need for integrated approaches for

investigating host-pathogen interactions as opposed to isolated pathways in a beautifully chaotic cell.

FUNDING: This study was supported by funds from the DBT-IISc partnership program (DBT (IED/4/2020-MED/DBT)) and the Infosys Young Investigator award (YI/2019/1106) in the ST lab. ST is an Intermediate fellow of Wellcome trust-DBT India Alliance (IA/I/18/1/503613). AS is a Senior fellow of Wellcome trust-DBT India Alliance (IA/S/16/2/50270). AB is supported by KVPY (Kishore Vaigyanik Protsahan Yojana) fellowship from DST, India. SS is supported by PMRF (Prime Minister's Research Fellowship) from the Ministry of Education, India.

ACKNOWLEDGMENTS: We thank Venitha Bernard for technical help with data analysis. We thank Prof. Umesh Varshney and Prof. K.N.Balaji for their administrative support.

REFERENCES

1. Coronaviridae Study Group of the International Committee on Taxonomy of V. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020;5(4):536-44.
2. Hasanoglu I, Korukluoglu G, Asilturk D, Cosgun Y, Kalem AK, Altas AB, et al. Higher viral loads in asymptomatic COVID-19 patients might be the invisible part of the iceberg. Infection. 2021;49(1):117-26.
3. Li Y, Shi J, Xia J, Duan J, Chen L, Yu X, et al. Asymptomatic and Symptomatic Patients With Non-severe Coronavirus Disease (COVID-19) Have Similar Clinical Features and Virological Courses: A Retrospective Single Center Study. Front Microbiol. 2020;11:1570.
4. Naqvi AAT, Fatima K, Mohammad T, Fatima U, Singh IK, Singh A, et al. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. Biochim Biophys Acta Mol Basis Dis. 2020;1866(10):165878.
5. Samuel CE. Antiviral actions of interferons. Clin Microbiol Rev. 2001;14(4):778-809, table of contents.
6. de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. Nat Rev Microbiol. 2016;14(8):523-34.
7. Casadevall A, Pirofski LA. The damage-response framework of microbial pathogenesis. Nat Rev Microbiol. 2003;1(1):17-24.
8. Henry BM, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. Clin Chem Lab Med. 2020;58(7):1021-8.
9. Velavan TP, Meyer CG. Mild versus severe COVID-19: Laboratory markers. Int J Infect Dis. 2020;95:304-7.

10. Pan F, Yang L, Li Y, Liang B, Li L, Ye T, et al. Factors associated with death outcome in patients with severe coronavirus disease-19 (COVID-19): a case-control study. Int J Med Sci. 2020;17(9):1281-92.
11. Liu F, Li L, Xu M, Wu J, Luo D, Zhu Y, et al. Prognostic value of interleukin-6, C-reactive protein, and procalcitonin in patients with COVID-19. J Clin Virol. 2020;127:104370.
12. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395(10223):497-506.
13. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020;395(10223):507-13.
14. Zhou Z, Ren L, Zhang L, Zhong J, Xiao Y, Jia Z, et al. Heightened Innate Immune Responses in the Respiratory Tract of COVID-19 Patients. Cell Host Microbe. 2020;27(6):883-90 e2.
15. Grant RA, Morales-Nebreda L, Markov NS, Swaminathan S, Querrey M, Guzman ER, et al. Circuits between infected macrophages and T cells in SARS-CoV-2 pneumonia. Nature. 2021.
16. Xiong Y, Liu Y, Cao L, Wang D, Guo M, Jiang A, et al. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. Emerg Microbes Infect. 2020;9(1):761-70.
17. Rivera B, Leyva A, Portela MM, Moratorio G, Moreno P, Duran R, et al. Quantitative proteomic dataset from oro- and naso-pharyngeal swabs used for COVID-19 diagnosis: Detection of viral proteins and host's biological processes altered by the infection. Data Brief. 2020;32:106121.
18. Akgun E, Tuzuner MB, Sahin B, Kilercik M, Kulah C, Cakiroglu HN, et al. Proteins associated with neutrophil degranulation are upregulated in nasopharyngeal swabs from SARS-CoV-2 patients. PLoS One. 2020;15(10):e0240012.
19. Maras JS, Sharma S, Bhat A, aggrawal R, Gupta E, Sarin SK. Multi-Omics integration analysis of respiratory specimen characterizes baseline molecular determinants associated with COVID-19 diagnosis. 2020:2020.07.06.20147082.
20. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. 2019;47(D1):D442-D50.
21. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013;41(Database issue):D991-5.
22. Huang H, McGarvey PB, Suzek BE, Mazumder R, Zhang J, Chen Y, et al. A comprehensive protein-centric ID mapping service for molecular data integration. Bioinformatics. 2011;27(8):1190-1.
23. L. S, ISoMaM. S. GeneOverlap: Test and visualize gene overlaps. R package version 1.26.0 ed2020.
24. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523.
25. Wickham H. ggplot2: Elegant Graphics for Data Analysis: Springer-Verlag New York; 2016.
26. Rusinova I, Forster S, Yu S, Kannan A, Masse M, Cumming H, et al. Interferome v2.0: an updated database of annotated interferon-regulated genes. Nucleic Acids Res. 2013;41(Database issue):D1040-6.
27. UniProt C. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47(D1):D506-D15.
28. Thul PJ, Akesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, et al. A subcellular map of the human proteome. Science. 2017;356(6340).
29. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. 2020;583(7816):459-68.
30. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.

526 31. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in
527 multidimensional genomic data. Bioinformatics. 2016;32(18):2847-9.
528 32. Bushman FD, Malani N, Fernandes J, D'Orso I, Cagney G, Diamond TL, et al. Host cell factors in
529 HIV replication: meta-analysis of genome-wide studies. PLoS Pathog. 2009;5(5):e1000437.
530 33. Klotman ME, Chang TL. Defensins in innate antiviral immunity. Nat Rev Immunol.
531 2006;6(6):447-56.
532 34. Quinones-Mateu ME, Lederman MM, Feng Z, Chakraborty B, Weber J, Rangel HR, et al. Human
533 epithelial beta-defensins 2 and 3 inhibit HIV-1 replication. AIDS. 2003;17(16):F39-48.
534 35. Proud D, Sanders SP, Wiehler S. Human rhinovirus infection induces airway epithelial cell
535 production of human beta-defensin 2 both in vitro and in vivo. J Immunol. 2004;172(7):4637-45.
536 36. Donato R, Cannon BR, Sorci G, Riuzzi F, Hsu K, Weber DJ, et al. Functions of S100 proteins. Curr
537 Mol Med. 2013;13(1):24-57.
538 37. Xia C, Braunstein Z, Toomey AC, Zhong J, Rao X. S100 Proteins As an Important Regulator of
539 Macrophage Inflammation. Front Immunol. 2017;8:1908.
540 38. Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein DA, et al. Meta- and
541 Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4 in Virus Budding. Cell Host
542 Microbe. 2015;18(6):723-35.
543 39. Sheerin D, Abhimanyu, Wang X, Johnson WE, Coussens A. Systematic evaluation of
544 transcriptomic disease risk and diagnostic biomarker overlap between COVID-19 and tuberculosis: a
545 patient-level meta-analysis. medRxiv. 2020.
546 40. Gardinassi LG, Souza COS, Sales-Campos H, Fonseca SG. Immune and Metabolic Signatures of
547 COVID-19 Revealed by Transcriptomics Data Reuse. Front Immunol. 2020;11:1636.
548 41. Coomes EA, Haghbayan H. Interleukin-6 in Covid-19: A systematic review and meta-analysis.
549 Rev Med Virol. 2020;30(6):1-9.
550 42. Parkinson N, Rodgers N, Head Fourman M, Wang B, Zechner M, Swets MC, et al. Dynamic
551 data-driven meta-analysis for prioritisation of host genes implicated in COVID-19. Sci Rep.
552 2020;10(1):22303.
553 43. Schick C, Pemberton PA, Shi GP, Kamachi Y, Cataltepe S, Bartuski AJ, et al. Cross-class inhibition
554 of the cysteine proteinases cathepsins K, L, and S by the serpin squamous cell carcinoma antigen 1: a
555 kinetic analysis. Biochemistry. 1998;37(15):5258-66.
556 44. Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, et al. Characterization of spike glycoprotein of SARS-CoV-2
557 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun. 2020;11(1):1620.
558 45. Vidalino L, Doria A, Quarta S, Zen M, Gatta A, Pontisso P. SERPINB3, apoptosis and
559 autoimmunity. Autoimmun Rev. 2009;9(2):108-12.
560 46. Dittmann M, Hoffmann HH, Scull MA, Gilmore RH, Bell KL, Ciancanelli M, et al. A serpin shapes
561 the extracellular environment to prevent influenza A virus maturation. Cell. 2015;160(4):631-43.
562 47. Gong D, Farley K, White M, Hartshorn KL, Benarafa C, Remold-O'Donnell E. Critical role of
563 serpinB1 in regulating inflammatory responses in pulmonary influenza infection. J Infect Dis.
564 2011;204(4):592-600.
565 48. McElvaney OJ, McEvoy NL, McElvaney OF, Carroll TP, Murphy MP, Dunlea DM, et al.
566 Characterization of the Inflammatory Response to Severe COVID-19 Illness. Am J Respir Crit Care Med.
567 2020;202(6):812-21.
568 49. Yang C, Chapman KR, Wong A, Liu M. alpha1-Antitrypsin deficiency and the risk of COVID-19:
569 an urgent call to action. Lancet Respir Med. 2021.
570 50. Dong HJ, Zhang R, Kuang Y, Wang XJ. Selective regulation in ribosome biogenesis and protein
571 production for efficient viral translation. Arch Microbiol. 2020.
572 51. Huang JY, Su WC, Jeng KS, Chang TH, Lai MM. Attenuation of 40S ribosomal subunit abundance
573 differentially affects host and HCV translation and suppresses HCV replication. PLoS Pathog.
574 2012;8(6):e1002766.
575 52. Landry DM, Hertz MI, Thompson SR. RPS25 is essential for translation initiation by the
576 Dicistroviridae and hepatitis C viral IRESs. Genes Dev. 2009;23(23):2753-64.

53. Fukushi S, Okada M, Stahl J, Kageyama T, Hoshino FB, Katayama K. Ribosomal protein S5 interacts with the internal ribosomal entry site of hepatitis C virus. J Biol Chem. 2001;276(24):20824-6.
54. Bhat P, Shwetha S, Sharma DK, Joseph AP, Srinivasan N, Das S. The beta hairpin structure within ribosomal protein S5 mediates interplay between domains II and IV and regulates HCV IRES function. Nucleic Acids Res. 2015;43(5):2888-901.
55. Majzoub K, Hafirassou ML, Meignin C, Goto A, Marzi S, Fedorova A, et al. RACK1 controls IRES-mediated translation of viruses. Cell. 2014;159(5):1086-95.
56. Shi Z, Fujii K, Kovary KM, Genuth NR, Rost HL, Teruel MN, et al. Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide. Mol Cell. 2017;67(1):71-83 e7.
57. Wood J, Frederickson RM, Fields S, Patel AH. Hepatitis C virus 3'X region interacts with human ribosomal proteins. J Virol. 2001;75(3):1348-58.
58. Chen Y, Lu Z, Zhang L, Gao L, Wang N, Gao X, et al. Ribosomal protein L4 interacts with viral protein VP3 and regulates the replication of infectious bursal disease virus. Virus Res. 2016;211:73-8.
59. Liu QH, Ma FF, Guan GK, Wang XF, Li C, Huang J. White spot syndrome virus VP51 interact with ribosomal protein L7 of Litopenaeus vannamei. Fish Shellfish Immunol. 2015;44(1):382-8.
60. Domachowske JB, Bonville CA, Rosenberg HF. Cytokeratin 17 is expressed in cells infected with respiratory syncytial virus via NF-kappaB activation and is associated with the formation of cytopathic syncytia. J Infect Dis. 2000;182(4):1022-8.
61. Hamaguchi M, Nishikawa K, Toyoda T, Yoshida T, Hanaichi T, Nagai Y. Transcriptive complex of Newcastle disease virus. II. Structural and functional assembly associated with the cytoskeletal framework. Virology. 1985;147(2):295-308.
62. De BP, Lesoon A, Banerjee AK. Human parainfluenza virus type 3 transcription in vitro: role of cellular actin in mRNA synthesis. J Virol. 1991;65(6):3268-75.
63. Huang YT, Romito RR, De BP, Banerjee AK. Characterization of the in vitro system for the synthesis of mRNA from human respiratory syncytial virus. Virology. 1993;193(2):862-7.
64. Arcangeletti MC, Pinardi F, Missorini S, De Conto F, Conti G, Portincasa P, et al. Modification of cytoskeleton and prosome networks in relation to protein synthesis in influenza A virus-infected LLC-MK2 cells. Virus Res. 1997;51(1):19-34.
65. Staufenbiel M, Epple P, Deppert W. Progressive reorganization of the host cell cytoskeleton during adenovirus infection. J Virol. 1986;60(3):1186-91.
66. Fu X, Jiang X, Chen X, Zhu L, Zhang G. The Differential Expression of Mitochondrial Function-Associated Proteins and Antioxidant Enzymes during Bovine Herpesvirus 1 Infection: A Potential Mechanism for Virus Infection-Induced Oxidative Mitochondrial Dysfunction. Mediators Inflamm. 2019;2019:7072917.
67. Derakhshan M, Willcocks MM, Salako MA, Kass GEN, Carter MJ. Human herpesvirus 1 protein US3 induces an inhibition of mitochondrial electron transport. J Gen Virol. 2006;87(Pt 8):2155-9.
68. Belenky P, Bogan KL, Brenner C. NAD+ metabolism in health and disease. Trends Biochem Sci. 2007;32(1):12-9.
69. Wu C, Liu Y, Yang Y, Zhang P, Zhong W, Wang Y, et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. Acta Pharm Sin B. 2020;10(5):766-88.
70. Garten A, Schuster S, Penke M, Gorski T, de Giorgis T, Kiess W. Physiological and pathophysiological roles of NAMPT and NAD metabolism. Nat Rev Endocrinol. 2015;11(9):535-46.
71. Heer CD, Sanderson DJ, Voth LS, Alhammad YMO, Schmidt MS, Trammell SAJ, et al. Coronavirus and PARP expression dysregulate the NAD Metabolome: a potentially actionable component of innate immunity. bioRxiv. 2020.
72. Sawatzky J, Soo J, Conroy AL, Bhargava R, Namasopo S, Opoka RO, et al. Biomarkers of Systemic Inflammation in Ugandan Infants and Children Hospitalized With Respiratory Syncytial Virus Infection. Pediatr Infect Dis J. 2019;38(8):854-9.

627 73. Bogorodskaya M, Fitch KV, Burdo TH, Maehler P, Easly RM, Murray GR, et al. Serum Lipocalin
628 2 (Neutrophil Gelatinase-Associated Lipocalin) in Relation to Biomarkers of Inflammation and Cardiac
629 Stretch During Activation of the Renin-Angiotensin-Aldosterone System in Human Immunodeficiency
630 Virus. J Infect Dis. 2019;220(9):1420-4.
631 74. Li G, Wang J, He X, Zhang L, Ran Q, Xiong A, et al. An integrative analysis identifying
632 transcriptional features and key genes involved in COVID-19. Epigenomics. 2020;12(22):1969-81.
633 75. Zeng HL, Chen D, Yan J, Yang Q, Han QQ, Li SS, et al. Proteomic characteristics of
634 bronchoalveolar lavage fluid in critical COVID-19 patients. FEBS J. 2020.
635 76. Rothan HA, Stone S, Natekar J, Kumari P, Arora K, Kumar M. The FDA-approved gold drug
636 auranofin inhibits novel coronavirus (SARS-COV-2) replication and attenuates inflammation in human
637 cells. Virology. 2020;547:7-11.
638 77. Griffiths HR, Gao D, Pararasa C. Redox regulation in metabolic programming and
639 inflammation. Redox Biol. 2017;12:50-7.
640 78. Chen L, Long X, Xu Q, Tan J, Wang G, Cao Y, et al. Elevated serum levels of S100A8/A9 and
641 HMGB1 at hospital admission are correlated with inferior clinical outcomes in COVID-19 patients. Cell
642 Mol Immunol. 2020;17(9):992-4.
643 79. Shi H, Zuo Y, Yalavarthi S, Gockman K, Zuo M, Madison JA, et al. Neutrophil calprotectin
644 identifies severe pulmonary disease in COVID-19. J Leukoc Biol. 2020.
645 80. Zuo Y, Zuo M, Yalavarthi S, Gockman K, Madison JA, Shi H, et al. Neutrophil extracellular traps
646 and thrombosis in COVID-19. J Thromb Thrombolysis. 2020.
647 81. Wu M, Chen Y, Xia H, Wang C, Tan CY, Cai X, et al. Transcriptional and proteomic insights into
648 the host response in fatal COVID-19 cases. Proc Natl Acad Sci U S A. 2020;117(45):28336-43.
649 82. Takahashi T, Ellingson MK, Wong P, Israelow B, Lucas C, Klein J, et al. Sex differences in
650 immune responses that underlie COVID-19 disease outcomes. Nature. 2020;588(7837):315-20.
651 83. Haitao T, Vermunt JV, Abeykoon J, Ghamrawi R, Gunaratne M, Jayachandran M, et al. COVID-
652 19 and Sex Differences: Mechanisms and Biomarkers. Mayo Clin Proc. 2020;95(10):2189-203.
653 84. Lieberman NAP, Peddu V, Xie H, Shrestha L, Huang ML, Mears MC, et al. In vivo antiviral host
654 transcriptional response to SARS-CoV-2 by viral load, sex, and age. PLoS Biol. 2020;18(9):e3000849.
655 85. O'Driscoll M, Ribeiro Dos Santos G, Wang L, Cummings DAT, Azman AS, Paireau J, et al. Age-
656 specific mortality and immunity patterns of SARS-CoV-2. Nature. 2021;590(7844):140-5.

657

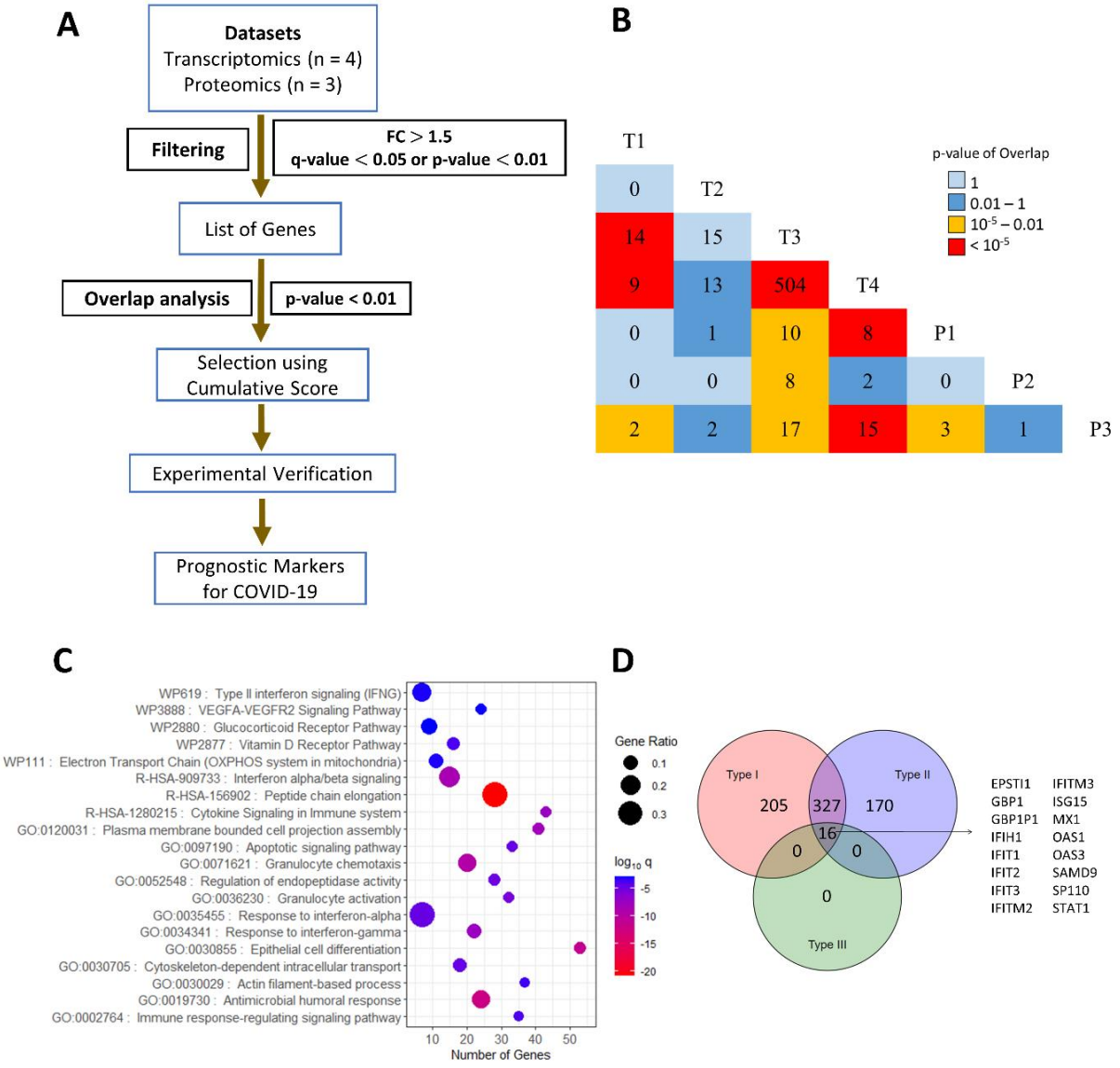


Figure 1: Meta-analysis pipeline for gene prioritisation and associate pathway analysis. **A)** Three proteomics and four transcriptomics datasets were chosen to obtain biomarkers for COVID-19 in humans. Genes/proteins that came up in these studies with a fold change greater than 1.5 and a q-value less than 0.05 (p-value less than 0.01 was taken in cases where q value is not provided) were subjected to pairwise overlap analysis. Genes that fall under significant intersections and represented in at least one proteomic dataset were sorted using cumulative scores to be experimentally verified. **B)** Triangular heatmap showing pairwise overlaps between transcriptomic and proteomic datasets. The number within each box denotes the number of genes that showed up between the corresponding intersections. The colour of a box denotes the significance of overlap. **C)** Gene ontology of all genes (567) in the significant intersections obtained during the overlap analysis plotted with the number of genes in each

term on the X-axis, proportion of genes enriched compared to the total number of genes in each term as the size of dots and the colour representing $\log_{10}$ p-adj value (q-value) of enrichment. **D)** Venn diagram showing the number of genes that are induced by Type I, II, or III interferons. The analysis was performed on Interferome v2.01 using the union of significant intersections (567)

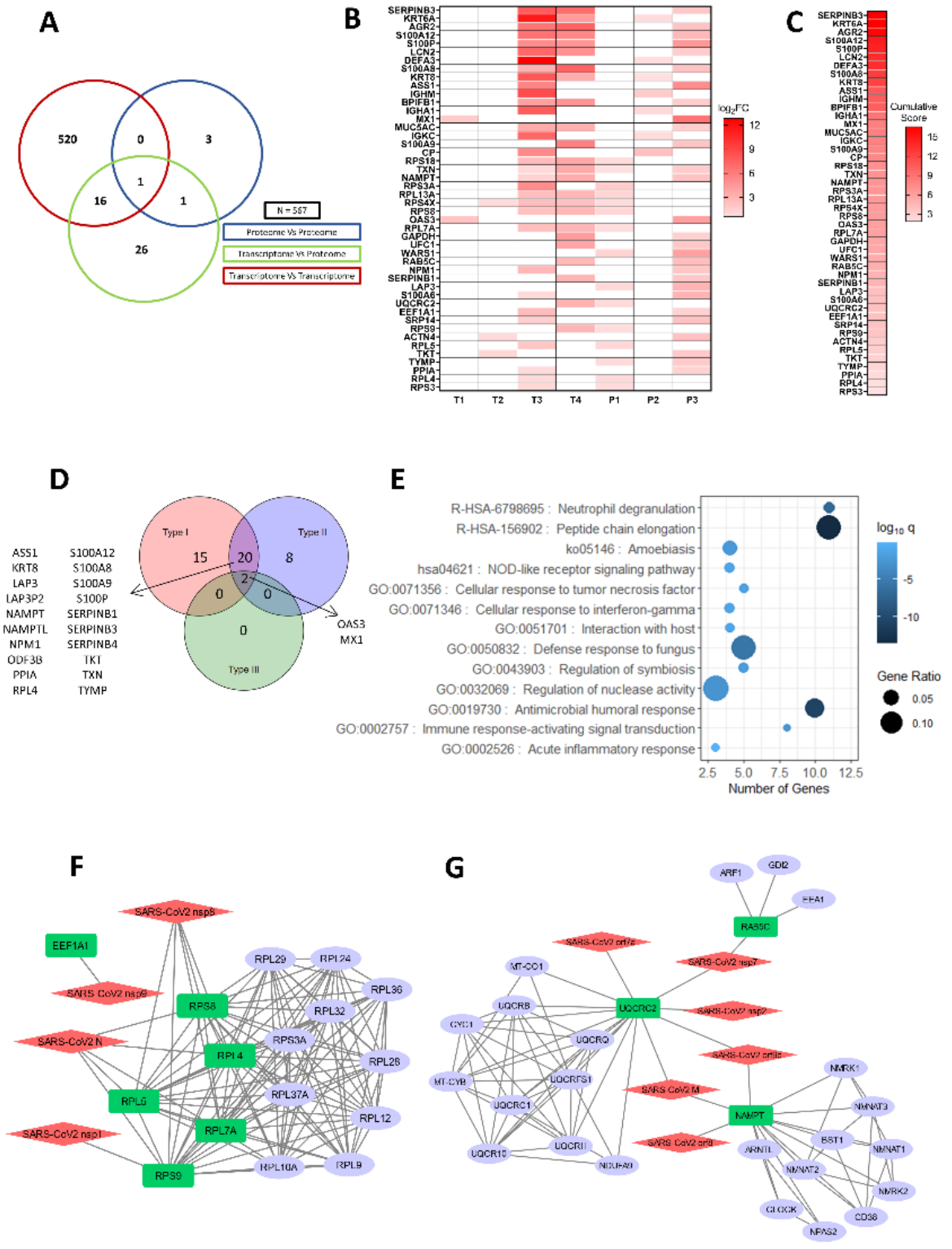
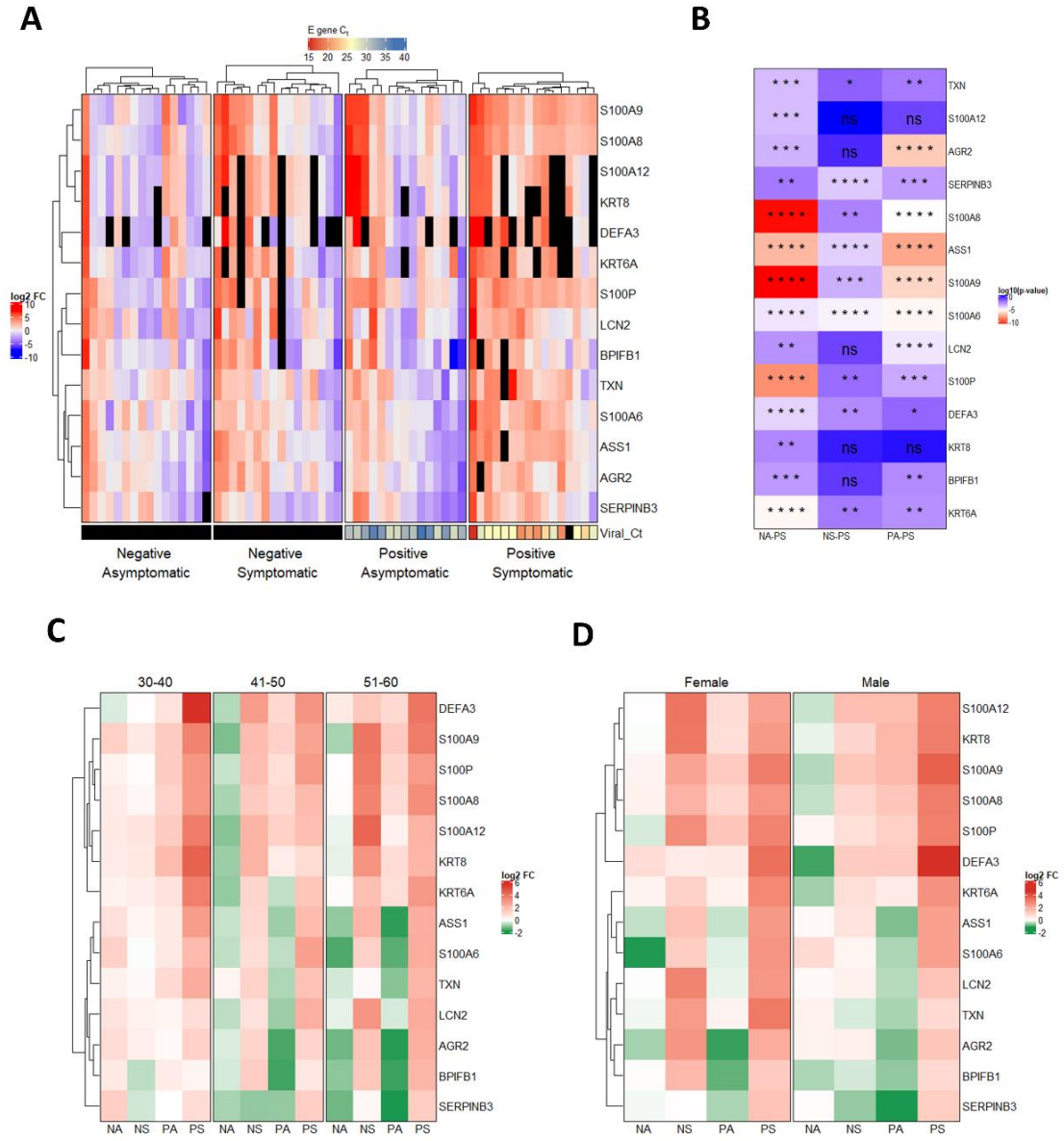


Figure 2: Cumulative score ranking, pathway and interactome analysis of selected host factors. **A)** Venn diagram of genes obtained from significant intersections among proteomic or transcriptomic datasets after pairwise overlap analysis. **B)** Genes in the Venn diagram that were found in at least one proteomic dataset with their log_2FC values in the datasets where they are present. Boxes coloured in white denote that the gene is not present in the particular dataset.

C) Genes arranged in descending order of cumulative scores obtained as a sum of log_2FC values in the datasets where they are present. **D)** Venn diagram showing the number of interferon-induced genes performed using Interferome v2.01 for 46 selected genes. **E)** Gene ontology of 46 genes plotted with the number of genes in each term on the X-axis, the proportion of genes enriched compared to the total number of genes in each term as the size of dots and the colour representing log_{10} p-adj value (q-value) of enrichment. **F, G)** Virus-host protein-protein interactions among SARS-CoV2 proteins and significant genes in the overlap analysis that shows up in at least one proteomic dataset modelled using Cytoscape v3.8.0. A STRING interactome for the primary interactors of SARS-CoV-2 proteins was merged (confidence ≥0.999 for all the proteins and confidence ≥0.90 for NAMPT; max number of interactors = 10). Red: SARS-CoV-2 proteins, Green: Host proteins (primary interactor), blue: STRING interactors (other cellular proteins interacting with the primary interactors).



693

694 **Figure 3: RT-qPCR validated expression profile of selected genes in different categories**
695 **of COVID-19 cohort A)** RT-qPCR was performed on RNA isolated from COVID-19 patients
696 for 14 genes and average $\log_2$ Fold-change values (with respect to Negative Asymptomatic
697 group) of PCR triplicates are shown in a heatmap. Each column represents a patient and
698 clustering was performed for columns and within row slices. The bottom annotation shows the
699 C_t value for the viral gene encoding Envelope (E) protein with a corresponding legend on the
700 top. Black boxes denote 'value unknown'. **B)** Differences between groups for each gene were
701 computed using the Mann-Whitney rank-sum test (paired = FALSE) without averaging PCR
702 replicates. The $\log_{10}$ (p-value) of comparisons is shown in the heatmap. The comparisons are
703 Negative asymptomatic vs Positive symptomatic (NA-PS), Negative symptomatic vs Positive
704 symptomatic (NS-PS), Positive asymptomatic vs Positive symptomatic (PA-PS). *P < 0.05;

705 **P < 0.01; ***P < 0.001; ****P < 0.0001; ns – not significant. **C)** log_2 Fold-change values
706 are grouped based on age groups 30-40, 41-50, and 51-60. Each row represents the average of
707 log_2 Fold-change values for patients falling into the particular age group and respective disease
708 status. **D)** log_2 Fold-change values are grouped according to sex. Each row represents the
709 average of log_2 Fold-change values for patients falling into the particular sex and respective
710 disease status.

711

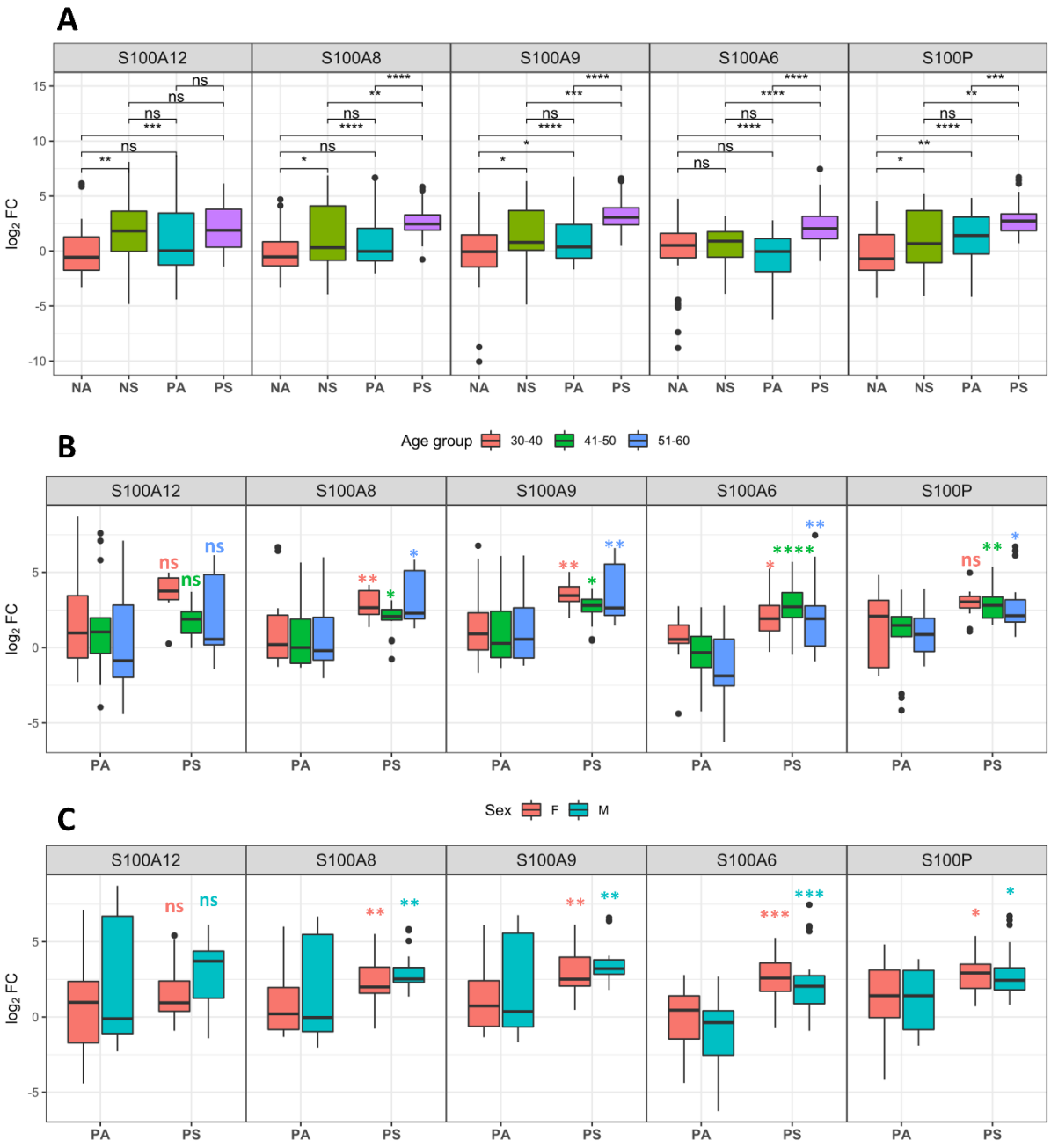


Figure 4: Expression profile of the S100 family of genes in different categories of COVID-19 cohort. **A)** log_2 Fold-change values are plotted as box plots for the S100 family of genes for patients who are Negative asymptomatic (NA), Negative symptomatic (NS), Positive asymptomatic (PA), and Positive symptomatic (PS) without averaging PCR replicates. **B)** Age-wise and **C)** Sex-wise differences in gene expression among Positive asymptomatic (PA) and Positive symptomatic (PS) patients were plotted with the significance of the comparison between groups shown on the corresponding box. Differences between groups for each gene

720 were computed using the Mann-Whitney rank-sum test (paired = FALSE). *P < 0.05; **P <
721 0.01; ***P < 0.001; ****P < 0.0001; ns – not significant.

722

Patient Status	Number of patients	Average age	Number of males	Number of females	Number in the age group 30-40	Number in the age group 41-50	Number in the age group 51-60
Negative Asymptomatic	16	43.9	8	8	5	6	5
Negative Symptomatic	16	41.7	12	4	9	4	3
Positive Asymptomatic	15	44.3	8	8	6	6	4
Positive Symptomatic	16	45	8	8	5	5	6

Table 1: Summary of individual and different categories in the COVID-19 cohort used for RT-qPCR based validation analysis. All samples were collected from Bangalore Urban area for diagnostic purposes.