



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

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



Research paper

Analyzing host-viral interactome of SARS-CoV-2 for identifying vulnerable host proteins during COVID-19 pathogenesis

Jayanta Kumar Das^a, Swarup Roy^{b,*}, Pietro Hiram Guzzi^{c,*}

^a Department of Pediatrics, School of Medicine, Johns Hopkins University, MD, USA

^b Network Reconstruction & Analysis (NetRA) Lab, Department of Computer Applications, Sikkim University, Gangtok, India

^c Department of Medical and Surgical Sciences, Magna Graecia University, Catanzaro, Italy



ARTICLE INFO

Keywords:

SARS-CoV-2

COVID-19

Protein-protein interaction

Centrality

Pathways

Disease

ABSTRACT

The development of therapeutic targets for COVID-19 relies on understanding the molecular mechanism of pathogenesis. Identifying genes or proteins involved in the infection mechanism is the key to shedding light on the complex molecular mechanisms. The combined effort of many laboratories distributed throughout the world has produced protein and genetic interactions. We integrated available results and obtained a host protein-protein interaction network composed of 1432 human proteins. Next, we performed network centrality analysis to identify critical proteins in the derived network. Finally, we performed a functional enrichment analysis of central proteins. We observed that the identified proteins are primarily associated with several crucial pathways, including cellular process, signaling transduction, neurodegenerative diseases. We focused on the proteins that are involved in human respiratory tract diseases. We highlighted many potential therapeutic targets, including *RBX1*, *HSPA5*, *ITCH*, *RAB7A*, *RAB5A*, *RAB8A*, *PSMCS*, *CAPZB*, *CANX*, *IGF2R*, and *HSPA1A*, which are central and also associated with multiple diseases.

1. Introduction

The world is experiencing an unprecedented pandemic due to a massive outbreak of Severe Acute Respiratory Syndrome Corona Virus 2 (SARS-CoV-2) infected viral disease, COVID-19. SARS-CoV-2 is a large enveloped coronavirus (family-*Coronaviridae*, subfamily-*Coronavirinae*) with non-segmented, single-stranded, and positive-sense RNA genomes (Wrapp et al., 2020), transmits rapidly through human to human contacts. Although SARS-CoV-2 is similar to other known coronaviruses, i. e., SARS-CoV or MERS-CoV (Perlman and Netland, 2009; De Groot et al., 2013; Weber et al., 2021), however it shows high rate of infection (Liu et al., 2020; Surveillances, 2020; Milano and Cannataro, 2020). Hence, there is a need to understand the disease pathogenesis of SARS-CoV-2 to develop effective therapies and vaccines.

The SARS-CoV-2 virus causes damage in multiple organs as the disease progresses from an asymptomatic phase to a life-threatening disease (Servick, 2020; Cho et al., 2013). Therefore, accurate molecular diagnosis of COVID-19 disease is essential by collecting the proper respiratory tract specimen (Whetton et al., 2020; Ortuso et al., 2021). In this context, the integrated analysis (Antonelli et al., 2019) of various data-sets, including clinical and imaging data, may explain, and

hopefully predict, the longitudinal effects of SARS-CoV-2 infection (Tang et al., 2020; Kumar Das et al., 2021). In particular, many independent projects, throughout the world have studied in genomics and proteomics levels (Kumar Das et al., 2021), and then they integrated these outcomes with clinical case studies. These works have produced data about the effect of the infection at a molecular scale, evidencing genes and proteins' role, such as the interactions among viral and human proteins. Interactions between a host and its pathogen are primarily driven by interactions among the host proteins and pathogen proteins, referred to as host-pathogen protein-protein interaction (PPI) network. The SARS-CoV-2 virus-host interactome have been studied computationally, focusing various virulence factors that are influencing SARS-CoV-2 pathogenesis and interacting mechanism (Guzzi et al., 2020; Hoffmann et al., 2021; Messina et al., 2020; Li et al., 2021). Further, many recent works also used host-viral protein-protein interaction network as an input to elucidate potential drug targets or repurposed drug molecules (Beck et al., 2020; Das et al., 2021; Gordon et al., 2020; Zhou et al., 2020). Host-pathogen protein interactions provide important insights into the molecular mechanisms of pathogenicity (Memišević et al., 2015) and for understanding virulence factors influencing SARS-CoV-2 pathogenesis (Li et al., 2021; Thiel et al., 2003).

* Corresponding authors.

E-mail addresses: sroy01@cus.ac.in (S. Roy), hguzzi@unicz.it (P.H. Guzzi).

<https://doi.org/10.1016/j.meegid.2021.104921>

Received 7 February 2021; Received in revised form 4 May 2021; Accepted 7 May 2021

Available online 15 May 2021

1567-1348/© 2021 Elsevier B.V. All rights reserved.

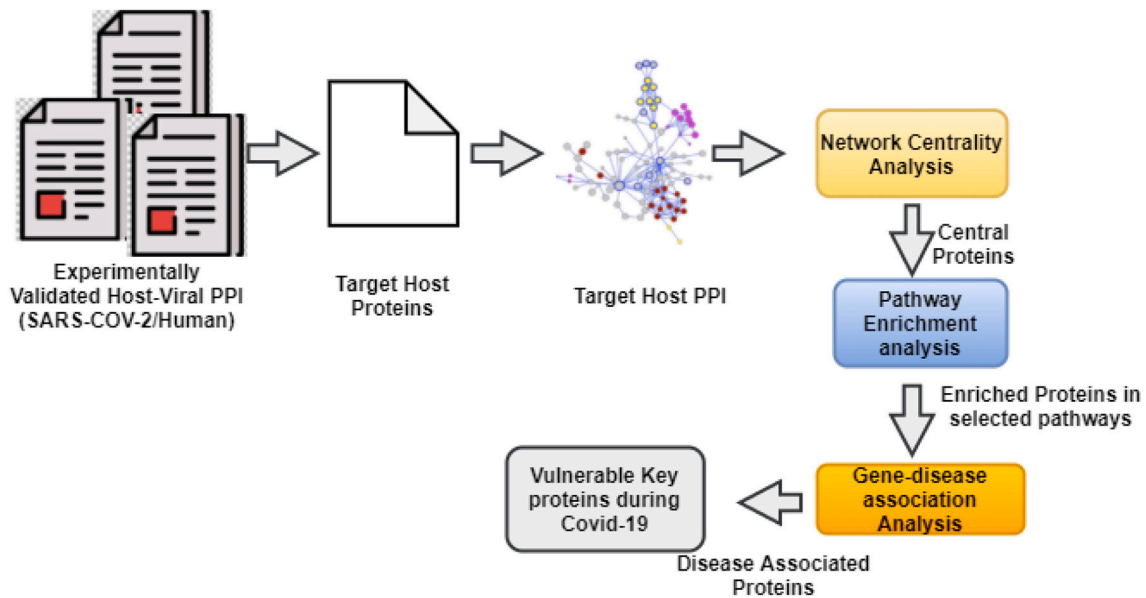


Fig. 1. The complete work-flow design of the current study.

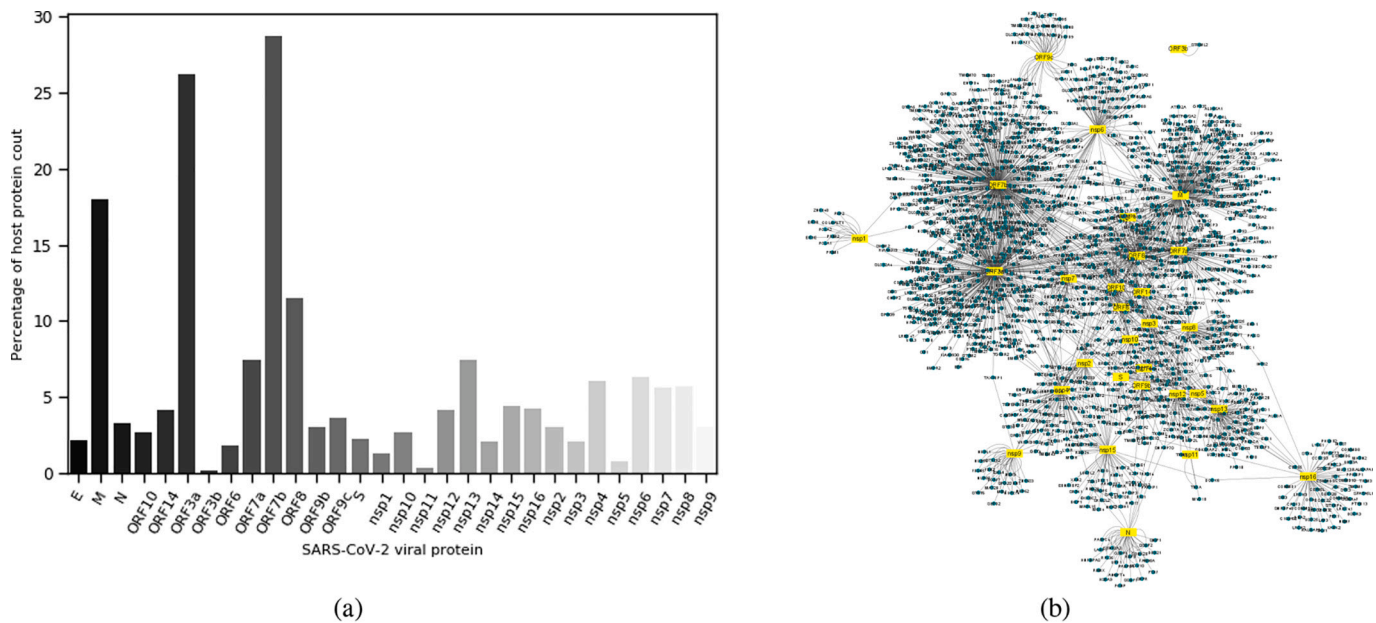


Fig. 2. The quantitative information of host-viral interactions. (a) The abundance (percentage) of collected interacting human host protein for different SARS-CoV-2 viral proteins; (b) A host-viral interaction network pattern, where few host proteins interactions can be seen associated with multiple viral proteins although most of interactions are virus protein specific unique (more details can be seen from *Supplementary-A*).

Protein-Protein Interactions (PPI) are usually modeled and analyzed with graph theory (Guzzi and Roy, 2020; Roy et al., 2019). In this formalism, the interactions are modeled as a graph whose nodes are proteins (or genes), and the edges are the interaction among them. Several studies have found that specific candidate proteins might play a crucial role (Li et al., 2013; Ferrari et al., 2018; Galicia et al., 2020; Lim et al., 2011). Protein-protein interaction networks are an essential ingredient for any systems-level understanding of cellular processes and modeling, and even drug discovery (Tucker et al., 2001; Thakur et al., 2015; Athanasios et al., 2017; Chautard et al., 2009; Nietzsche et al., 2016). The key genes/proteins involved in the different biological pathways can give valuable insight for in-depth characterization of disease progression (Lan et al., 2015; Safari-Alighiarloo et al., 2014; Wang et al., 2018; Jha et al., 2020). It is well accepted that all the viruses

have evolved to target proteins that are central and have strong control over the human interactome (Jeong et al., 2001; Bösl et al., 2019; Albert et al., 2000; Navratil et al., 2011; Halehalli and Nagarajaram, 2015). Exploring the predicted interaction networks can suggest new directions for future experimental research and provide cross-species predictions for efficient interaction mapping (Xu and Li, 2006; Safari-Alighiarloo et al., 2014).

This study aimed to identify essential human host proteins based on topology analysis of the PPI of the host proteins targeted by SARS-CoV-2 proteins. We integrated available experimentally validated host-viral PPIs and obtained a set of host proteins that were participated in the integrated network. Network centrality analysis of those candidate host proteins were performed based on their interactions within host-host PPI. We performed enrichment analysis of the central host proteins to

shed light on their significance in cellular, signaling, and disease pathways. The enriched host proteins, derived from above step were further analyzed for their possible roles in any disease pathogenesis. Finally, a set of host proteins that having role in COVID-19 related diseases were reported as vulnerable proteins. The complete workflow of the current study is illustrated in Fig. 1.

2. Materials and method

2.1. SARS-CoV-2 interacting (human) host proteins

We used recently reported host proteins that were physically verified using Affinity purification mass spectrometry (AP-MS) for their interactions with SARS-CoV-2 proteins (Gordon et al., 2020; Li et al., 2021; Stukalov et al., 2020; Cannataro and Harrison, 2021). The used host-viral protein interactions are also available in BioGRID¹ (Stark et al., 2006). A total of 2489 host-viral interactions (consisting of 1432 unique host proteins interacting with 30 SARS-CoV-2 viral proteins) were obtained. In Fig. 2(a), we provided the number of interacting host proteins for each viral protein. It is to be noted that the majority of the host proteins are targeted by any specific viral protein (Fig. 2(b)).

2.2. Reconstruction of host PPI network

Starting with the host proteins that are interacting with the virus proteins, we reconstructed the host PPI sub-network by querying the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, Version 10.0²) (Szklarczyk et al., 2010).

The host PPI sub-network is then visualized using Cytoscape,³ a general platform for complex network analysis and visualization (Shannon et al., 2003). We also used Cytoscape to obtain giant components of the host PPI network and calculated various network centrality scores.

2.3. Centrality analysis of host PPI sub-network

In network analysis, indicators of centrality identify the most critical nodes in the network (Bonacich, 1987). The centrality measure are used to characterize each node and edge in the PPI network. The degree measure is the most intuitive for topology analysis of the PPI network. Several other crucial factors that can influence network links are betweenness centrality, closeness centrality, clustering coefficient, topological coefficient, and neighborhood connectivity.

- (i) *Degree centrality*(D_c): The degree centrality (simply degree) of a node n in a network is the number of direct neighbours of n . The densely connected nodes in a PPI network are considered as hub nodes (Han et al., 2004).
- (ii) *Betweenness centrality*(B_c): Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes (Yoon et al., 2006). The betweenness centrality of a node n is defined as:

$$B_c(n) = \sum_{s \neq n \neq t} (\sigma_{st(n)} / \sigma_{st}) \quad (1)$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st(n)}$ is the number of those paths that pass through n .

- (iii) *Closeness centrality*(C_c): Closeness centrality is a way of detecting nodes that are able to spread information very efficiently through the network (Newman, 2005). It can be calculated as:

$$C_c(n) = 1 / \text{avg}(L(m, n)) \quad (2)$$

where $L(m, n)$ is the length of shortest path between node n and m , and m denotes any other nodes that are reachable to node n .

- (iv) *Average shortest-path length*(S_p): The shortest-path length between two nodes (say n and m) is defined as the number of minimum steps required reach node n from m (Mao and Zhang, 2013). The average shortest path length of node n is the average value of all pair of nodes shortest path from the node n .
- (v) *Clustering coefficient*(C_{coef}): Clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together (Barabasi and Oltvai, 2004). In undirected networks, the clustering coefficient of a node n is defined as:

$$C_{coef} = 2e_n / (k_n(k_n - 1)) \quad (3)$$

where k_n is the number of neighbors of n and e_n is the number of connected pairs between all neighbors of n .

- (vi) *Topological coefficient*(T_c): Topological coefficient is a relative measure for the extent to which a node shares neighbors with other nodes (Goldberg and Roth, 2003). The topological coefficient T_c of a node n with k_n neighbors is computed as follows:

$$T_c = \text{avg}(J(nm)) / k_n \quad (4)$$

where $J(n, m)$ is defined for all nodes m that share at least one neighbor with n , and the value $J(n, m)$ is the number of neighbors shared between the nodes n and m , plus one if there is a direct link between n and m .

- (vii) *Neighborhood connectivity*(N_c): Neighborhood connectivity of a node n is defined as the average connectivity of all neighbors of n (Maslov and Sneppen, 2002). The neighborhood connectivity distribution gives the average of the neighborhood connectivities of all nodes n with k neighbors for $k = 0, 1, \dots, n-1$.

We used *NetworkAnalyzer* (Shannon et al., 2003) to calculate the above centrality scores. *NetworkAnalyzer* calculates C_c (Closeness centrality) as the reciprocal of the average shortest path length. So, high C_c means highly central, and thus low S_p .

2.4. Gene ontology and pathway enrichment analysis

We performed enrichment analysis to determine the set of significant genes/proteins in different functional and biological pathways. We used KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa and Goto, 2000) for elucidating pathway enrichment of a host protein and Gene Ontology (GO) for the assessment of protein functions (Ashburner et al., 2000). In this study, we performed GO and KEGG analysis using *Enrichr* software,⁴ a web-based suite of gene list enrichment analysis tools (Kuleshov et al., 2016).

2.5. Gene-disease association network

Complex diseases are caused by a group of genes known as disease genes. More often, a gene can participate in various disease conditions (Goh et al., 2007; W. T. C. C. Consortium, 2007). It helps unravel the disease pathogenesis, which helps in disease diagnosis, treatment, and disease prevention. We obtained gene-disease association network from

¹ <https://thebiogrid.org/>

² <http://string-db.org/>

³ <http://apps.cytoscape.org>

⁴ <https://maayanlab.cloud/Enrichr/>

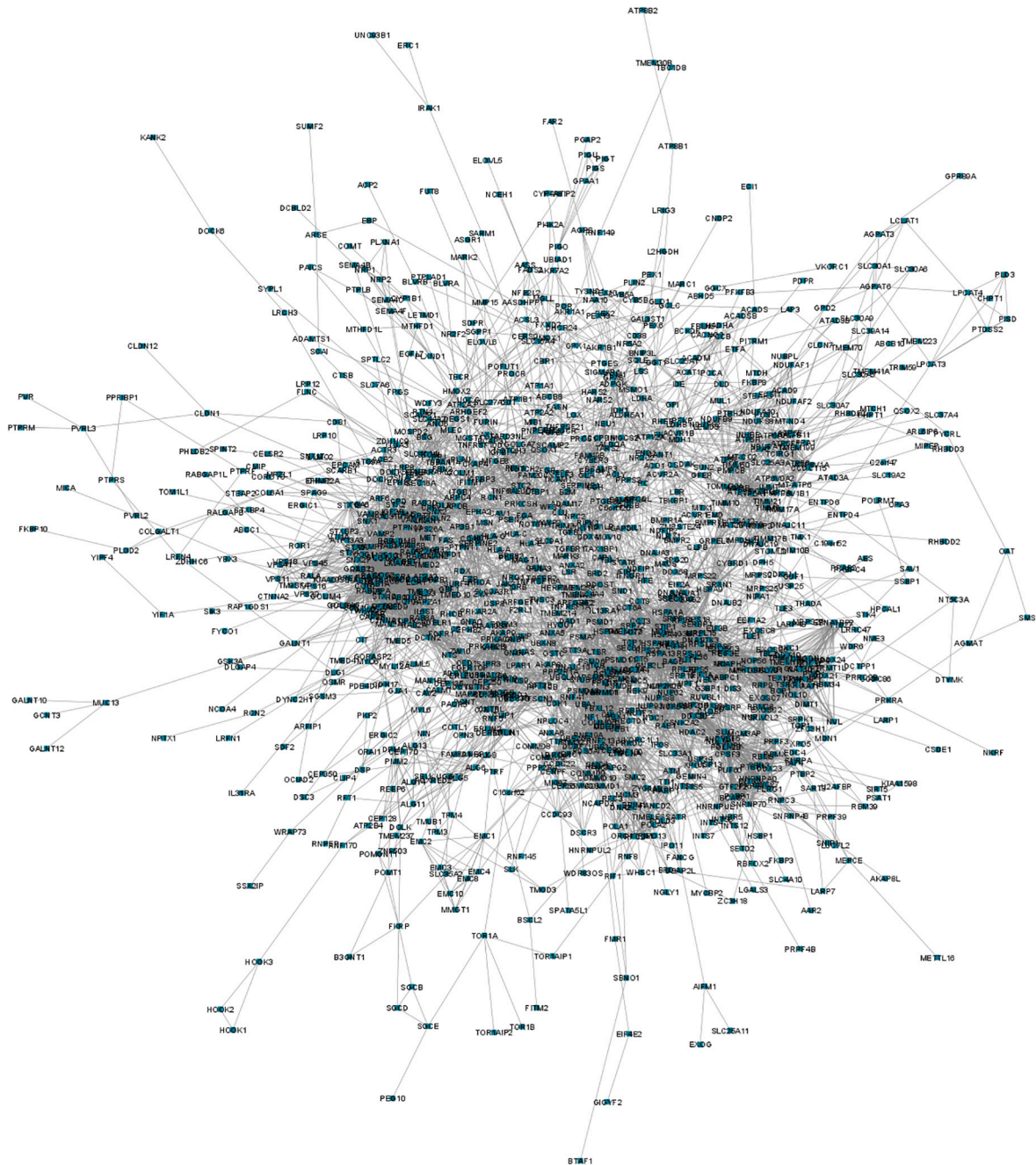


Fig. 3. The gain (connected) component of PPI network obtained from whole PPI network. The network is consisting of 1111 unique nodes (proteins) and 7043 interactions edges (more details can be seen from *Supplementary-A*).

DisGeNET (v7.0) database,⁵ which contains 1,134,942 gene-disease associations (GDAs), between 21,671 genes and 30,170 disease (Piñero et al., 2020). From this database, we considered curated gene-disease associations only.

3. Results and discussion

Here, we report the outcomes of intermediate steps to reach our objective of isolating key host proteins followed by their significance analysis.

3.1. Deriving PPI network for candidate host proteins

Our list of host proteins consists of 1432 distinct proteins that are targeted by SARS-CoV-2 during COVID-19. We rebuilt the PPI network centered around our candidate proteins using STRING DB. There are 7076 edges in the derived PPI network. We derived PPI by keeping only the interactions whose confidence scores were at least 0.7 (high confidence). The derived PPI network was then analyzed using Cytoscape. We identified the big connected component (also called gain/main component) of the PPI network. After discarding all disconnected components in the PPI network, we considered the giant component of the PPI network with 1111 nodes (approximately 78% of total candidate proteins) and 7043 edges (Fig. 3).

⁵ <https://www.disgenet.org/>

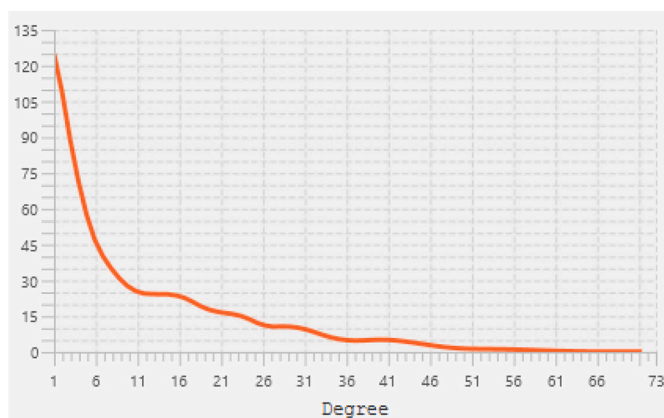


Fig. 4. The degree distribution of all 1111 nodes (proteins) in the gain component of PPI network. The X-axis indicates degree distribution, whereas Y-axis shows relative frequency distributions.

Table 1

Correlation analysis among all centrality parameters computed for 1111 proteins. The correlation score (>0.5) among the three centrality measures (D_c, B_c, C_c) are similar (highlighted in bold).

	D_c	B_c	C_{coef}	T_c	N_c
B_c	0.603	1			
C_{coef}	0.209	-0.168	1		
T_c	-0.32	-0.283	0.451	1	
N_c	0.557	0.1101	0.346	-0.139	1
C_c	0.759	0.5146	0.213	-0.329	0.684

3.2. Network topology analysis of gain component

We performed the topological analysis of the gain component using *NetworkAnalyzer* (Shannon et al., 2003). The degree distribution of all the candidate proteins in the gain component showed that the majority of the proteins in the gain component exhibit a higher degree of connectivity (Fig. 4). Few proteins with degree (shown within parentheses) more than 50 are *CDK1*(73), *PPP2R1A*(65), *NOP56*(60), *POLR2B*(60), *RAB1A*(59), *RBX1*(58), *SKIV2L2*(57), *NAPA*(57), *RPS14*(56), *STX5*(54), *TGOLN2*(54), *TCEB1*(53), *DCTN2*(53), *TCEB2*(52), *HSPA9*(51), and *GNB2L1*(50).

The histogram analysis of all the centrality measures (discussed in Section 2.3) showing non-random distribution (Supplementary-B). Out of seven centrality measures we considered only those measures which were aligning in their decisions. To select the candidate centrality measures, we computed Pearson’s correlation among all centrality scores (Table 1). The correlation between degree centrality (D_c) and closeness centrality (C_c) observed to be the highest ($r = 0.759$) in comparison to other measures. Although, we observed correlation between D_c and neighbourhood centrality (N_c) is the third-highest ($r = 0.557$), but N_c and B_c showed less correlative ($r = 0.1101$). Overall, we observed that the correlation scores among the three centrality measures (D_c, B_c, C_c) are quite similar. Therefore, we selected them in subsequent analysis. We identified 373 proteins in these criteria, which are considered highly central proteins (above the median score for all three selected parameters). When we considered all measures, we found only six common proteins (*GEMIN4, DDX20, GOLGA3, FKBP15, PMPCA, and AK4*) above the median score in each category of centrality measurement, which was the reason for selecting three centrality measures for our downstream analysis.

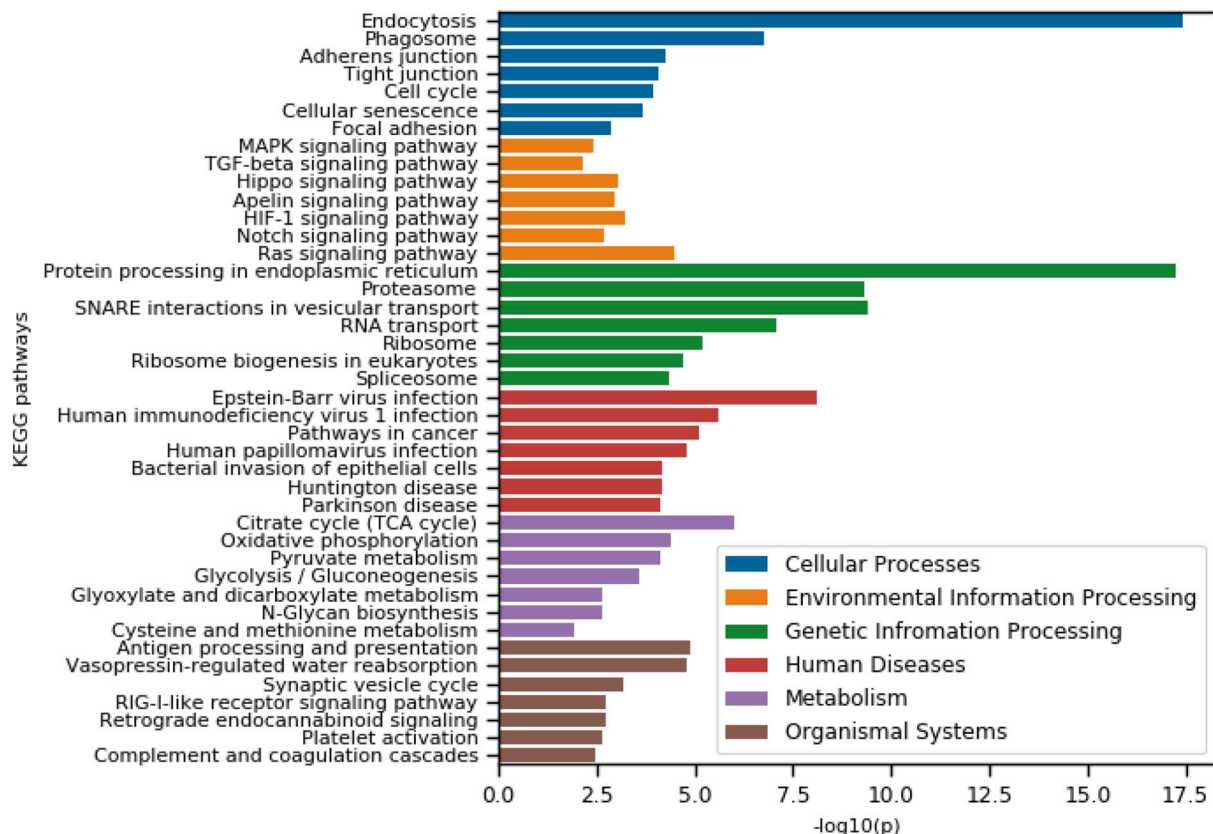


Fig. 5. The top 7 enriched pathways from each category of KEGG pathways. In each category, pathways are shown ordered by $-\log_{10}(p)$ value.

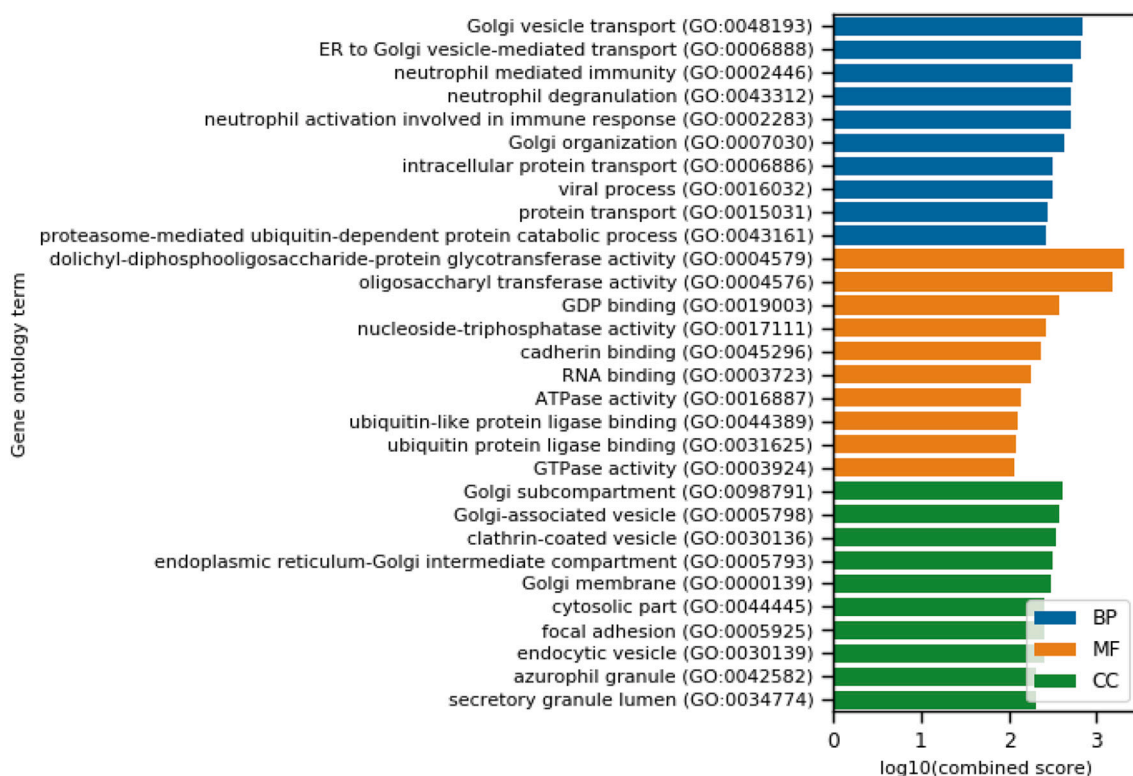


Fig. 6. The top 10 enriched terms in each category of gene ontology (BP-Biological process, MF-Molecular function, CC-Cellular component). In each category, terms are shown ordered by log10(combined score) value.

3.3. Pathway and gene ontology enrichment analysis of highly central proteins

We performed KEGG pathway analysis of selected 373 highly central proteins. We obtained a total of 84 enriched KEGG pathways within the significant level ($adj - p < .05$). The enriched pathways are involved in Cellular Processes (9), Environmental Information Processing (9), Genetic Information Processing (13), Human Disease (31), Organismal Systems (15), and Metabolism (7). The top seven pathways in each category were shown in Fig. 5.

Our current study mainly focused on the proteins that are involved in cellular process, signaling transduction, and human disease (viral and

neurodegenerative) pathways. The most affected thirty-one (31) pathways were considered in the context of COVID-19 disease (Seif et al., 2020; Ganesan et al., 2019; Luo et al., 2020; Grimes and Grimes, 2020). There are nine enriched pathways in cellular process (Endocytosis, Phagosome, Adherens junction, Tight junction, Cell cycle, Cellular senescence, Focal adhesion, Regulation of actin cytoskeleton, Lysosome), nine enriched pathways in Environmental Information Processing-signaling transduction (Ras signaling pathway, HIF-1 signaling pathway, Hippo signaling pathway, Apelin signaling pathway, MAPK signaling pathway, TGF-beta signaling pathway, AMPK signaling pathway, NF-kappa B signaling pathway), nine enriched pathways from human viral disease sub-category (Human immunodeficiency virus-1

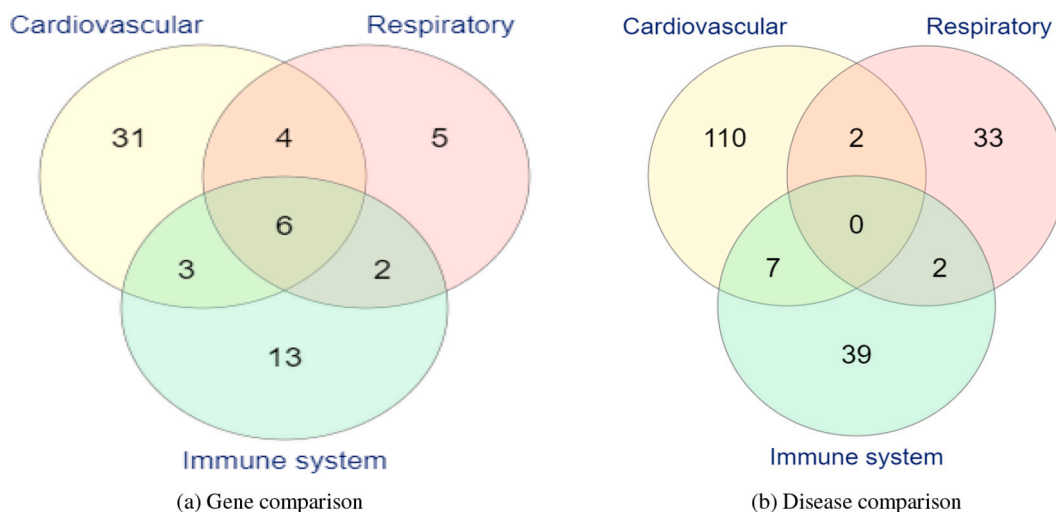


Fig. 7. Comparison of three groups of disease categories (Cardiovascular, Respiratory, Immune system) using venn-diagram. (a) based on number of proteins count in each category; (b) based on number of disease associated proteins (curated from database) among the observed proteins in each category.

Table 2

The shortlisted 64 host proteins with their degree in host-host PPI, disease count (out of 204), disease type count (out of 3), and pathway count (out of 31). Some of the proteins that are known to be targeted by other viruses were also highlighted.

Gene/ Protein	Degree in PPI network	Pathway count	Disease category	Disease count	Known target virus
<i>ADAM17</i>	10	2	Immune	1	
<i>ALDOA</i>	26	1	Cardiovascular	1	
<i>AP3B1</i>	17	1	Respiratory	2	
<i>AREG</i>	20	2	Respiratory, Cardiovascular, Immune	10	
<i>ATM</i>	29	6	Cardiovascular, Immune	8	
<i>ATP2A2</i>	10	1	Cardiovascular	11	
<i>ATP6</i>	19	5	Cardiovascular	1	Human SARS coronavirus, Bovine papillomavirus type 1, Human papillomavirus type 16
<i>ATR</i>	17	5	Respiratory	2	Human adenovirus 5
<i>B2M</i>	25	4	Cardiovascular, Immune	4	Hepatitis C virus genotype 1b (isolate Con1)
<i>CANX</i>	31	2	Cardiovascular	1	
<i>CAPZB</i>	34	1	Cardiovascular	1	
<i>CAV1</i>	22	2	Respiratory, Cardiovascular, Immune	17	Poliovirus type 1 (strain Sabin)
<i>CD44</i>	19	1	Immune	1	
<i>COX2</i>	9	3	Cardiovascular	1	
<i>CRKL</i>	11	5	Cardiovascular	4	
<i>DDX58</i>	10	6	Respiratory, Cardiovascular	2	
<i>ENO1</i>	13	1	Cardiovascular, Immune	2	
<i>EPHA2</i>	9	2	Cardiovascular	5	
<i>FASN</i>	9	1	Cardiovascular	10	
<i>GAPDH</i>	23	2	Cardiovascular	1	Hepatitis C virus genotype 1b (isolate Con1), Epstein-Barr virus (strain GD1)
<i>GLA</i>	16	1	Cardiovascular	5	
<i>GNAQ</i>	14	5	Cardiovascular	1	
<i>GUSB</i>	16	1	Immune	1	
<i>HDAC2</i>	12	5	Respiratory	2	Human herpesvirus 1 (strain 17), Human papillomavirus type 16, Human papillomavirus type 31
<i>HLA-A</i>	15	8	Immune	5	Epstein-Barr virus (strain GD1), Human papillomavirus type 16
<i>HLA-C</i>	14	8	Immune	7	
<i>HMGCR</i>	11	1	Immune	4	
<i>HSPA1A</i>	30	5	Cardiovascular	1	Epstein-Barr virus (strain GD1)
<i>HSPA5</i>	46	1	Respiratory, Cardiovascular	3	Epstein-Barr virus (strain GD1)
<i>IFIH1</i>	10	3	Respiratory, Cardiovascular, Immune	10	Sendai virus (strain Fushimi)
<i>IGF2R</i>	31	2	Respiratory	1	
<i>ITCH</i>	46	1	Immune	1	Epstein-Barr virus (strain B95-8)
<i>ITGA6</i>	10	3	Immune	1	
<i>ITGB1</i>	29	5	Cardiovascular	10	Hepatitis C virus genotype 1b (isolate Con1)
<i>JAK2</i>	22	2	Cardiovascular	6	
<i>LDHA</i>	14	1	Cardiovascular	4	
<i>LDLR</i>	20	2	Cardiovascular	4	
<i>MET</i>	18	4	Respiratory	1	
<i>NDUFS2</i>	20	3	Cardiovascular	3	
<i>NF1</i>	16	2	Cardiovascular	3	
<i>NOTCH1</i>	23	3	Cardiovascular	4	Hepatitis C virus genotype 1b (isolate Con1)
<i>NOTCH2</i>	11	2	Cardiovascular	1	
<i>NOTCH3</i>	15	3	Cardiovascular	2	
<i>PARP1</i>	14	1	Respiratory, Cardiovascular, Immune	17	Human herpesvirus 1 (strain 17)
<i>PCNA</i>	25	3	Immune	1	Human herpesvirus 1 (strain 17)
<i>PDIA3</i>	14	3	Cardiovascular	1	
<i>PLAU</i>	24	1	Respiratory, Cardiovascular, Immune	59	
<i>PPP1CB</i>	14	4	Cardiovascular	2	
<i>PRKDC</i>	15	1	Respiratory, Immune	3	Human herpesvirus 1 (strain 17)
<i>PSMC5</i>	34	1	Immune	7	Human adenovirus 5, Human adenovirus 12, Simian virus 40
<i>PSMD6</i>	24	1	Immune	2	
<i>PTPN11</i>	22	1	Cardiovascular	6	
<i>RAB5A</i>	40	3	Cardiovascular	1	
<i>RAB7A</i>	41	2	Cardiovascular	1	
<i>RAB8A</i>	40	3	Immune	1	
<i>RBX1</i>	58	4	Immune	5	
<i>SERPINE1</i>	16	4	Cardiovascular	8	
<i>SLC9A1</i>	12	2	Cardiovascular	7	
<i>SORT1</i>	14	1	Cardiovascular	3	
<i>STAT6</i>	11	1	Respiratory, Immune	4	
<i>TGFB1</i>	29	7	Respiratory, Cardiovascular, Immune	29	Hepatitis C virus genotype 1b (isolate Con1)
<i>TGFBR1</i>	17	9	Respiratory, Cardiovascular	6	
<i>TGFBR2</i>	16	8	Respiratory, Cardiovascular	13	
<i>XPO1</i>	25	2	Cardiovascular	2	

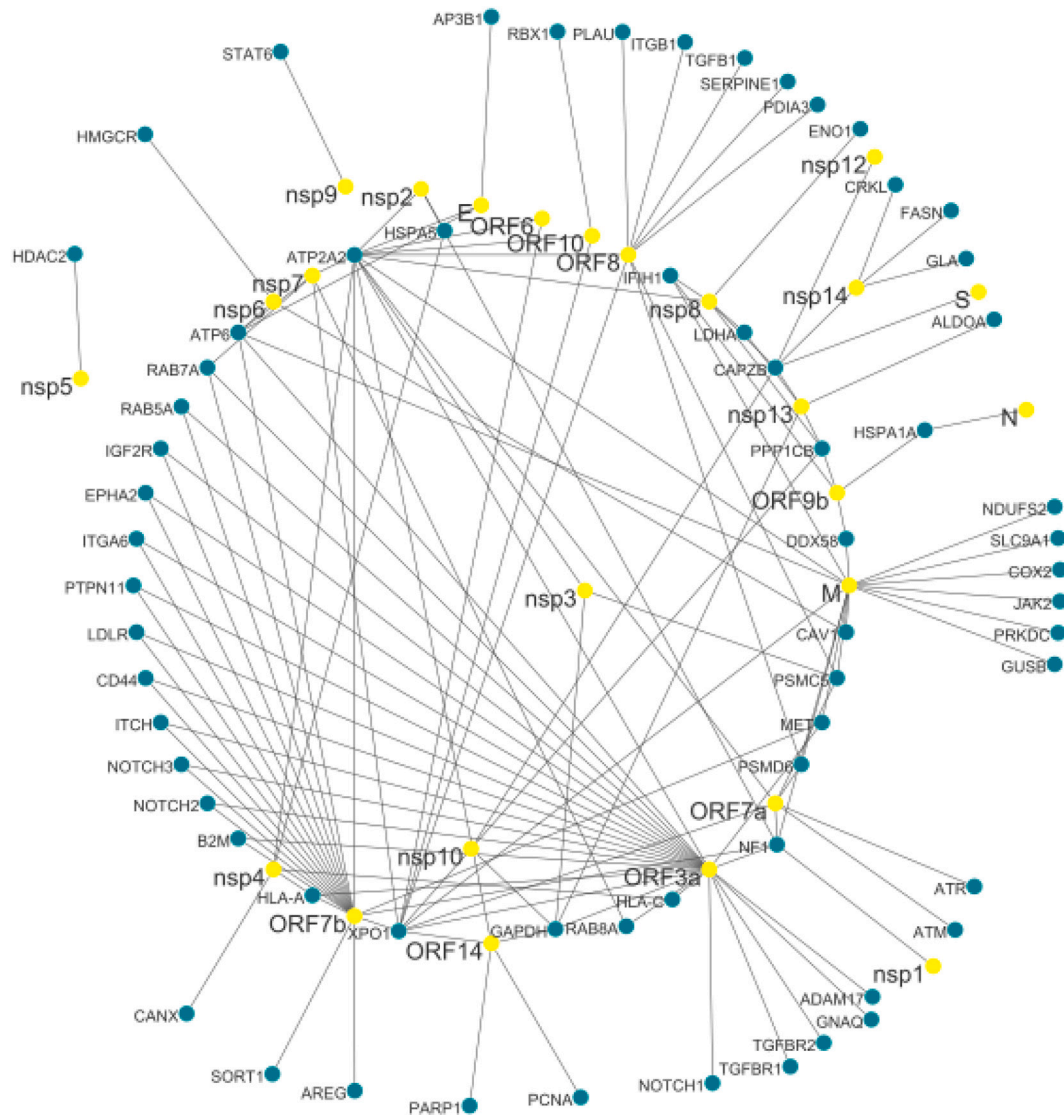


Fig. 8. The interaction network represents the most influential host protein and viral protein. The network is consisting of sixty-four (64) host proteins interacting with twenty-five (25) SARS-CoV-2 viral proteins. The yellow colour node represents the viral proteins in the network, whereas the green one represents the host proteins.

infection, Human papillomavirus infection, Human cytomegalovirus infection, Hepatitis B, Human T-cell leukemia virus-1 infection, Influenza A, Hepatitis C, Measles) and four enriched pathways from *neurodegenerative disease* with sub-category (Huntington disease, Parkinson’s disease, Alzheimer’s disease, Prion diseases). A total of 141 distinct proteins (out of 373) were obtained from these pathways, which were then ranked based on their presence in selected enriched pathways. We found that 79 proteins were associated with our candidate pathways. All these proteins were then further studied for disease-gene association in the next.

We also performed gene set enrichment analysis (Gene ontology). It is observed that the selected genes are primarily involved in Biological processes (*Supplementary-C*). The top ten terms in each category of gene ontology (Biological processes (BF), Molecular Function (MF), and Cellular Component (CC)) were shown in *Fig. 6* that includes neutrophil-mediated immunity (GO:0002446), neutrophil activation involved in immune response (GO:0002283) and viral process (GO:0016032) from BP category; dolichyl-diphosphooligosaccharide-protein glycotransferase activity (GO:0004579), GDP binding (GO:0019003), cadherin binding (GO:0045296) and ATPase activity (GO:0016887) from MF category; and focal adhesion (GO:0005925) from CC category.

3.4. Analysis of disease-gene associations

The shortlisted 141 proteins involved in four significant pathways (cellular process, signaling transduction, viral and neurodegenerative) were further screened by looking into their association with COVID-19 related diseases. We particularly focused on three highly influential diseases during COVID-19, namely cardiovascular, respiratory tract (Wu et al., 2020; Clerkin et al., 2020; Konturek et al., 2020; Cannataro et al., 2005) and immune system disease (Melenotte et al., 2020; Chowdhury et al., 2020). To obtain disease-gene association, we used DisGeNET database (Piñero et al., 2020) and selected *CURETED* source only. We found a total of 64 proteins (out of 141) playing roles in various COVID-19 related diseases such as *Asthma, Pneumonitis, Pneumonia, Influenza, Lung diseases, Cardiomyopathies, Coronary, Arteriosclerosis, Coronary Artery Disease, Heart failure, HIV Infections*, and others. (*Supplementary-D*). We compared proteins involved in all three disease categories and individual diseases in each category (*Fig. 7*). A total of 119, 37, and 48 unique diseases (total count-204) and 44, 17, and 24 distinct proteins are associated with the Cardiovascular, Respiratory, and Immune system disease category. Interestingly, we found few proteins that are associated with all three disease categories (*AREG, CAV1, IFIH1, PARP1,*

PLAU, TGFBI, ATM, B2M, DDX58, ENO1, HSPA5, PRKDC, STAT6, TGFBR1, and TGFBR2). The top few proteins which are associated with ten or more diseases are PLAU(59), TGFBI(29), CAV1(17), PARP1(17), TGFBR2(13), ATP2A2(11), AREG(10), FASN(10), IFIH1(10), and ITGB1(10). Finally, we listed a total of 64 unique genes/proteins and their various quantitative scores (degree, disease count (out of 204), disease type count (out of 3), and pathway count (out of 31)) in Table 2.

3.5. Viral proteins targeting key host proteins

We next looked into the source viral-host PPI network (refer Fig. 2) to identify the viral proteins that were targeting shortlisted 64 disease-associated proteins. We found 25 SARS-CoV-2 proteins were interacting 64 key host proteins (Fig. 8). Among 25 SARS-CoV-2 viral proteins, eight are accessory proteins (*Orf3a*, *Orf7b*, *Orf6*, *Orf7a*, *Orf7b*, *Orf8*, *Orf9b*, *Orf10*), four structural proteins (*E*, *M*, *N*, *S*) and thirteen non-structural poly-proteins (*nsp1*, *nsp10*, *nsp12*, *nsp13*, *nsp14*, *nsp2*, *nsp3*, *nsp4*, *nsp5*, *nsp6*, *nsp7*, *nsp8*, *nsp9*). It was observed that several host proteins were interacting with a single viral protein. Very few host proteins were interacting with more than one viral protein. The viral protein, *Orf7b* interacts with a maximum number of target host proteins, followed by *Orf3a* and *M* protein. Further, five host proteins were found to interact with both *Orf3a* and *Orf7b*.

We looked further for any other viruses that are targeting our 64 host proteins. We mined VirusMINT (Chatr-Aryamontri et al., 2009), a virus-host association database, to find the other related viral diseases. We found that the majority of the highlighted host proteins were also targeted by Hepatitis C virus genotype 1b, Poliovirus Type 1, Human herpesvirus 1, Human papillomavirus type 16 & 31, Simian virus 40, Sendai virus, Human adenovirus 5 & 12, Epstein-Barr virus, Human SARS coronavirus. Bovine papillomavirus type 1, and Epstein-Barr virus (Table 2). Further, a few genes (AREG, PLAU, MET, NF1, JAK2, SERPINE1) were observed to be important drug targets and associated with various signaling pathways as reported in recent research (Das et al., 2021). These proteins might be highly essential and need to put utmost importance on developing host-directed antiviral therapies for COVID-19.

4. Conclusion

In this study, we analyzed the human (host) protein-protein interaction network targeted by SARS-CoV-2. Using topology and pathway enrichment analysis of important host PPI network, we reported 64 potential key SARS-CoV-2 interacting host proteins associated with disease pathways, like respiratory, cardiovascular, and immune system diseases. Among them, we highlighted a set of highly central host proteins, such as *RBX1*, *HSPA5*, *ITCH*, *RAB7A*, *RAB5A*, *RAB8A*, *PSMCS5*, *CAPZB*, *CANX*, *IGF2R*, and *HSPA1A*, which might influence the whole PPI network. Some of the key host proteins are known to target different viruses and may be highly important for therapeutic solution for controlling COVID-19 severeness. We strongly believe that the highlighted key proteins are promising drug targets, which might play a crucial role in inhibiting COVID-19 progression.

Author contributions

Jayanta Kumar Das and Swarup Roy conceived and designed the study. Swarup Roy and Pietro H. Guzzi supervised the study. Jayanta Kumar Das performed the computational analysis and wrote the original draft, which is finally edited by Swarup Roy and Pietro H. Guzzi. All authors approved the final version of the draft.

Declaration of Competing Interest

We would like to declare that we do not have any conflicts of interest in connection with the submitted work.

Acknowledgements

This work has been partially supported by DST-ICPS Data Science program [DST/ICPS/Cluster/Data Science/General] from Govt. of India and VQA PON from Italian Ministry of Economic Development (MISE). The authors thank to Dr. Lawrence M. Noguee and Dr. Mohan Kumar Krishnan for their constructive suggestions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.meegid.2021.104921>.

References

- Albert, R., Jeong, H., Barabási, A.-L., 2000. Error and attack tolerance of complex networks. *Nature* 406, 378–382.
- Antonelli, L., Guarracino, M.R., Maddalena, L., Sangiovanni, M., 2019. Integrating imaging and omics data: a review. *Biomed. Signal Process. Control* 52, 264–280.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al., 2000. Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.
- Athanasios, A., Charalampous, V., Vasileios, T., et al., 2017. Protein-protein interaction (ppi) network: recent advances in drug discovery. *Curr. Drug Metab.* 18, 5–10.
- Barabási, A.-L., Oltvai, Z.N., 2004. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113.
- Beck, B.R., Shin, B., Choi, Y., Park, S., Kang, K., 2020. Predicting commercially available antiviral drugs that may act on the novel coronavirus (sars-cov-2) through a drug-target interaction deep learning model. *Comput. Struct. Biotechnol. J.* 18, 784–790.
- Bonacich, P., 1987. Power and centrality: a family of measures. *Am. J. Sociol.* 92, 1170–1182.
- Bösl, K., Ianevski, A., Than, T.T., Andersen, P.I., Kuivanen, S., Teppor, M., Zusinaite, E., Dumpis, U., Vitkauskiene, A., Cox, R.J., et al., 2019. Common nodes of virus–host interaction revealed through an integrated network analysis. *Front. Immunol.* 10, 2186.
- Cannataro, M., Guzzi, P.H., Mazza, T., Tradigo, G., Veltri, P., 2005. Preprocessing of mass spectrometry proteomics data on the grid. 18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05) 554–559. <https://doi.org/10.1109/CBMS.2005.87>.
- Cannataro, M., Harrison, A., 2021. Bioinformatics Helping to Mitigate the Impact of Covid-19—Editorial.
- Chatr-Aryamontri, A., Ceol, A., Peluso, D., Nardozza, A., Panni, S., Sacco, F., Tinti, M., Smolyar, A., Castagnoli, L., Vidal, M., et al., 2009. Virusmint: a viral protein interaction database. *Nucleic Acids Res.* 37, D669–D673.
- Chautard, E., Thierry-Mieg, N., Ricard-Blum, S., 2009. Interaction networks: from protein functions to drug discovery. a review. *Pathol. Biol.* 57, 324–333.
- Cho, Y.-R., Mina, M., Lu, Y., Kwon, N., Guzzi, P.H., 2013. M-finder: uncovering functionally associated proteins from interactome data integrated with go annotations. *Proteome Sci.* 11, 1–12.
- Chowdhury, M.A., Hossain, N., Kashem, M.A., Shahid, M.A., Alam, A., 2020. Immune response in covid-19: a review. *J. Infect. Public Health* 13 (11), 1619–1629.
- Clerkin, K.J., Fried, J.A., Raikhelkar, J., Sayer, G., Griffin, J.M., Masoumi, A., Jain, S.S., Burkhoff, D., Kumaraiah, D., Rabbani, L., et al., 2020. Covid-19 and cardiovascular disease. *Circulation* 141, 1648–1655.
- Das, Jayanta Kumar, Roy, Swarup, Chakraborty, Subhadip, et al., 2021. A Scheme for Inferring Viral-Host Associations based on Codon Usage Patterns Identifies the Most Affected Signaling Pathways during COVID-19. *Journal of Biomedical Informatics* 118, 103801. <https://doi.org/10.1016/j.jbi.2021.103801>.
- De Groot, R.J., Baker, S.C., Baric, R.S., Brown, C.S., Drosten, C., Enjuanes, L., Fouchier, R. A., Galiano, M., Gorbalenya, A.E., Memish, Z.A., et al., 2013. Commentary: Middle east respiratory syndrome coronavirus (mers-cov): announcement of the coronavirus study group. *J. Virol.* 87, 7790–7792.
- Ferrari, R., Kia, D.A., Tomkins, J.E., Hardy, J., Wood, N.W., Lovering, R.C., Lewis, P.A., Manzoni, C., 2018. Stratification of candidate genes for parkinson's disease using weighted protein-protein interaction network analysis. *BMC Genomics* 19, 1–8.
- Galicía, J.C., Guzzi, P.H., Giorgi, F.M., Khan, A.A., 2020. Predicting the response of the dental pulp to sars-cov2 infection: a transcriptome-wide effect cross-analysis. *Genes Immun.* 21, 360–363.
- Ganesan, H., Balasubramanian, V., Iyer, M., Venugopal, A., Subramaniam, M.D., Cho, S.-G., Vellingiri, B., 2019. mtor signalling pathway—a root cause for idiopathic autism? *BMB Rep.* 52, 424.
- Goh, K.-I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., Barabási, A.-L., 2007. The human disease network. *Proc. Natl. Acad. Sci.* 104, 8685–8690.
- Goldberg, D.S., Roth, F.P., 2003. Assessing experimentally derived interactions in a small world. *Proc. Natl. Acad. Sci.* 100, 4372–4376.
- 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* 1–13.
- Grimes, J.M., Grimes, K.V., 2020. p38 mapk inhibition: a promising therapeutic approach for covid-19. *J. Mol. Cell. Cardiol.* 144, 63–65.
- Guzzi, P.H., Roy, S., 2020. Biological Network Analysis: Trends, Approaches, Graph Theory, and Algorithms. Academic Press.

- Guzzi, P.H., Mercatelli, D., Ceraolo, C., Giorgi, F.M., 2020. Master regulator analysis of the sars-cov-2/human interactome. *J. Clin. Med.* 9, 982.
- Halehalli, R.R., Nagarajaram, H.A., 2015. Molecular principles of human virus protein-protein interactions. *Bioinformatics* 31, 1025–1033.
- Han, J.-D.J., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P., et al., 2004. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 430, 88–93.
- Hoffmann, H.-H., Sánchez-Rivera, F.J., Schneider, W.M., Luna, J.M., Soto-Feliciano, Y. M., Ashbrook, A.W., Le Pen, J., Leal, A.A., Ricardo-Lax, I., Michailidis, E., et al., 2021. Functional interrogation of a sars-cov-2 host protein interactome identifies unique and shared coronavirus host factors. *Cell Host Microbe* 29 (2), 267–280.
- Jeong, H., Mason, S.P., Barabási, A.-L., Oltvai, Z.N., 2001. Lethality and centrality in protein networks. *Nature* 411, 41–42.
- Jha, M., Roy, S., Kalita, J.K., 2020. Prioritizing disease biomarkers using functional module based network analysis: a multilayer consensus driven scheme. *Comput. Biol. Med.* 126, 104023.
- Kanehisa, M., Goto, S., 2000. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.
- Konturek, P.C., Harsch, I., Neurath, M., Zopf, Y., 2020. Covid-19—more than respiratory disease: a gastroenterologist's perspective. *J. Physiol. Pharmacol.* 71, 765–767.
- Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–W97.
- Kumar Das, J., Tradigo, G., Veltri, P., Guzzi, P.H., Roy, S., 2021. Data science in unveiling covid-19 pathogenesis and diagnosis: evolutionary origin to drug repurposing. *Brief. Bioinform.* 22, 855–872.
- Lan, W., Wang, J., Li, M., Peng, W., Wu, F., 2015. Computational approaches for prioritizing candidate disease genes based on ppi networks. *Tsinghua Sci. Technol.* 20, 500–512.
- Li, W., Chen, L., He, W., Li, W., Qu, X., Liang, B., Gao, Q., Feng, C., Jia, X., Lv, Y., et al., 2013. Prioritizing disease candidate proteins in cardiomyopathy-specific protein-protein interaction networks based on “guilt by association” analysis. *PLoS One* 8, e71191.
- Li, J., Guo, M., Tian, X., Wang, X., Yang, X., Wu, P., Liu, C., Xiao, Z., Qu, Y., Yin, Y., et al., 2021. Virus-host interactome and proteomic survey reveal potential virulence factors influencing sars-cov-2 pathogenesis. *Medicine* 2 (1), 99–112.
- Li, J., Guo, M., Tian, X., Wang, X., Yang, X., Wu, P., Liu, C., Xiao, Z., Qu, Y., Yin, Y., et al., 2021. Virus-host interactome and proteomic survey reveal potential virulence factors influencing sars-cov-2 pathogenesis. *Medicine* 2, 99–112.
- Lim, D., Kim, N.-K., Park, H.-S., Lee, S.-H., Cho, Y.-M., Oh, S.-J., Kim, T.-H., Kim, H., 2011. Identification of candidate genes related to bovine marbling using protein-protein interaction networks. *Int. J. Biol. Sci.* 7, 992.
- Liu, Y., Gayle, A.A., Wilder-Smith, A., Rocklöv, J., 2020. The reproductive number of covid-19 is higher compared to sars coronavirus. *J. Travel Med.* 27 (2) <https://doi.org/10.1093/jtm/taaa021>.
- Luo, W., Li, Y.-X., Jiang, L.-J., Chen, Q., Wang, T., Ye, D.-W., 2020. Targeting jak-stat signaling to control cytokine release syndrome in covid-19. *Trends Pharmacol. Sci.* 41 (8), 531–543.
- Mao, G., Zhang, N., 2013. Analysis of average shortest-path length of scale-free network. *J. Appl. Math.* 2013.
- Maslov, S., Sneppen, K., 2002. Specificity and stability in topology of protein networks. *Science* 296, 910–913.
- Melenotte, C., Silvin, A., Goubet, A.-G., Lahmar, I., Dubuisson, A., Zumla, A., Raoult, D., Merad, M., Gachot, B., Hénon, C., et al., 2020. Immune responses during covid-19 infection. *Oncotarget* 9, 1807836.
- Memišević, V., Zavaljevski, N., Rajagopala, S.V., Kwon, K., Pieper, R., DeShazer, D., Reifman, J., Wallqvist, A., 2015. Mining host-pathogen protein interactions to characterize burkholderia mallei infectivity mechanisms. *PLoS Comput. Biol.* 11, e1004088.
- Messina, F., Giombini, E., Agrati, C., Vairo, F., Bartoli, T.A., Al Moghazi, S., Piacentini, M., Locatelli, F., Kobinger, G., Maeurer, M., et al., 2020. Covid-19: viral-host interactome analyzed by network based-approach model to study pathogenesis of sars-cov-2 infection. *J. Transl. Med.* 18, 1–10.
- Milano, M., Cannataro, M., 2020. Statistical and network-based analysis of italian covid-19 data: communities detection and temporal evolution. *Int. J. Environ. Res. Public Health* 17, 4182.
- Navratil, V., de Chasse, B., Combe, C.R., Lotteau, V., 2011. When the human viral infectome and disease networks collide: towards a systems biology platform for the aetiology of human diseases. *BMC Syst. Biol.* 5, 13.
- Newman, M.E., 2005. A measure of betweenness centrality based on random walks. *Soc. Networks* 27, 39–54.
- Nietzsche, M., Landgraf, R., Tohge, T., Börnke, F., 2016. A protein-protein interaction network linking the energy-sensor kinase snrk1 to multiple signaling pathways in arabidopsis thaliana. *Curr. Plant Biol.* 5, 36–44.
- Ortuso, F., Mercatelli, D., Guzzi, P.H., Giorgi, F.M., 2021. Structural genetics of circulating variants affecting the sars-cov-2 spike/human ace2 complex. *J. Biomol. Struct. Dyn.* 1–11.
- Pearlman, S., Netland, J., 2009. Coronaviruses post-sars: update on replication and pathogenesis. *Nat. Rev. Microbiol.* 7, 439–450.
- Piñero, J., Ramirez-Anguita, J.M., Sañch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., Furlong, L.L., 2020. The disgenet knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* 48, D845–D855.
- Roy, S., Manners, H.N., Elmsallati, A., Kalita, J.K., 2019. Alignment of protein-protein interaction networks. In: Ranganathan, S., Gribskov, M., Nakai, K., Schönbach, C. (Eds.), *Encyclopedia of Bioinformatics and Computational Biology*, 1. Elsevier, pp. 997–1015. <https://doi.org/10.1016/b978-0-12-809633-8.20429-9>.
- Safari-Alighiarloo, N., Taghizadeh, M., Rezaei-Tavirani, M., Goliaei, B., Peyvandi, A.A., 2014. Protein-protein Interaction Networks (ppi) and Complex Diseases, *Gastroenterology and Hepatology From Bed to Bench*, 7, p. 17.
- Seif, F., Aazami, H., Khoshmirsafa, M., Kamali, M., Mohsenzadegan, M., Pornour, M., Mansouri, D., 2020. Jak inhibition as a new treatment strategy for patients with covid-19. *Int. Arch. Allergy Immunol.* 181, 467–475.
- Servick, K., 2020. For survivors of severe covid-19, beating the virus is just the beginning. *Science* 8. <https://doi.org/10.1126/science.abc1486>.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
- Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A., Tyers, M., 2006. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34, D535–D539.
- Stukalov, A., Girault, V., Grass, V., Bergant, V., Karayel, O., Urban, C., Haas, D.A., Huang, Y., Oubraham, L., Wang, A., et al., 2020. Multi-level proteomics reveals host-perturbation strategies of sars-cov-2 and sars-cov. *Biorxiv*. <https://doi.org/10.1101/2020.06.17.156455>.
- Surveillances, V., 2020. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (covid-19)—China, 2020. *China CDC Week.* 2, 113–122.
- Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Müller, J., Bork, P., et al., 2010. The string database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 39, D561–D568.
- Tang, Y.-W., Schmitz, J.E., Persing, D.H., Stratton, C.W., 2020. Laboratory diagnosis of covid-19: current issues and challenges. *J. Clin. Microbiol.* 58.
- Thakur, S., Dhiman, M., Tell, G., Mantha, A.K., 2015. A review on protein-protein interaction network of ape1/ref-1 and its associated biological functions. *Cell Biochem. Funct.* 33, 101–112.
- Thiel, V., Ivanov, K.A., Putics, A., Hertzog, T., Schelle, B., Bayer, S., Weißbrich, B., Snijder, E.J., Rabenau, H., Doerr, H.W., et al., 2003. Mechanisms and enzymes involved in sars coronavirus genome expression. *J. Gen. Virol.* 84, 2305–2315.
- Tucker, C.L., Gera, J.F., Uetz, P., 2001. Towards an understanding of complex protein networks. *Trends Cell Biol.* 11, 102–106.
- W. T. C. Consortium, et al., 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661.
- Wang, Y., Zhang, Y., Huang, Q., Li, C., 2018. Integrated bioinformatics analysis reveals key candidate genes and pathways in breast cancer. *Mol. Med. Rep.* 17, 8091–8100.
- Weber, G.M., Hong, C., Palmer, N.P., Avillach, P., Murphy, S.N., Gutiérrez-Sacristán, A., Xia, Z., Serret-Larmande, A., Neuraz, A., Omenn, G.S., et al., 2021. International comparisons of harmonized laboratory value trajectories to predict severe covid-19: leveraging the 4ce collaborative across 342 hospitals and 6 countries: a retrospective cohort study. *medRxiv*. <https://doi.org/10.1101/2020.12.16.20247684>.
- Whetton, A.D., Preston, G.W., Abubeker, S., Geifman, N., 2020. Proteomics and informatics for understanding phases and identifying biomarkers in covid-19 disease. *J. Proteome Res.* 19, 4219–4232.
- Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.-L., Abiona, O., Graham, B. S., McLellan, J.S., 2020. Cryo-em structure of the 2019-ncov spike in the prefusion conformation. *Science* 367, 1260–1263.
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., et al., 2020. A new coronavirus associated with human respiratory disease in china. *Nature* 579, 265–269.
- Xu, J., Li, Y., 2006. Discovering disease-genes by topological features in human protein-protein interaction network. *Bioinformatics* 22, 2800–2805.
- Yoon, J., Blumer, A., Lee, K., 2006. An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. *Bioinformatics* 22, 3106–3108.
- Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., Cheng, F., 2020. Network-based drug repurposing for novel coronavirus 2019-ncov/sars-cov-2. *Cell Discov.* 6, 1–18.