
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.

Computers in Biology and Medicine 138 (2021) 104889

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

# Interactome of human and SARS-CoV-2 proteins to identify human hub proteins associated with comorbidities

Nimisha Ghosh [a,b,1], Indrajit Saha [c,*,1], Nikhil Sharma [d]

[a] Department of Computer Science and Information Technology, Institute of Technical Education and Research, Siksha 'O' Anusandhan (Deemed to Be University), Bhubaneswar, Odisha, India
[b] Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Warsaw, Poland
[c] Department of Computer Science and Engineering, National Institute of Technical Teachers' Training and Research, Kolkata, West Bengal, India
[d] Department of Electronics and Communication Engineering, Jaypee Institute of Information Technology, Noida, Uttar Pradesh, India

ARTICLE INFO

*Keywords:*
COVID-19
Comorbidities
Hub proteins
Protein-protein interaction
Repurposable drugs
SARS-CoV-2

ABSTRACT

SARS-CoV-2 has a higher chance of progression in adults of any age with certain underlying health conditions or comorbidities like cancer, neurological diseases and in certain cases may even lead to death. Like other viruses, SARS-CoV-2 also interacts with host proteins to pave its entry into host cells. Therefore, to understand the behaviour of SARS-CoV-2 and design of effective antiviral drugs, host-virus protein-protein interactions (PPIs) can be very useful. In this regard, we have initially created a human-SARS-CoV-2 PPI database from existing works in the literature which has resulted in 7085 unique PPIs. Subsequently, we have identified at most 10 proteins with highest degrees viz. hub proteins from interacting human proteins for individual virus protein. The identification of these hub proteins is important as they are connected to most of the other human proteins. Consequently, when they get affected, the potential diseases are triggered in the corresponding pathways, thereby leading to comorbidities. Furthermore, the biological significance of the identified hub proteins is shown using KEGG pathway and GO enrichment analysis. KEGG pathway analysis is also essential for identifying the pathways leading to comorbidities. Among others, SARS-CoV-2 proteins viz. NSP2, NSP5, Envelope and ORF10 interacting with human hub proteins like COX4I1, COX5A, COX5B, NDUFS1, CANX, HSP90AA1 and TP53 lead to comorbidities. Such comorbidities are Alzheimer, Parkinson, Huntington, HTLV-1 infection, prostate cancer and viral carcinogenesis. Subsequently, using Enrichr tool possible repurposable drugs which target the human hub proteins are reported in this paper as well. Therefore, this work provides a consolidated study for human-SARS-CoV-2 protein interactions to understand the relationship between comorbidity and hub proteins so that it may pave the way for the development of anti-viral drugs.

## 1. Introduction

SARS-CoV-2, the virus responsible for COVID-19 has disrupted our daily lives and even after almost two years, we are still struggling in our fight against the virus. Though it originated in China, in a short time COVID-19 cases were reported from all around the globe. By September 2021, more than 229 million people have been affected by this virus with more than 4 million deaths.[2] The usual symptoms of COVID-19 range from common cough and cold, shortness of breath, fever to multiple organ failure which may eventually lead to death. Since this is a RNA virus, it shows high mutations and new strains of the virus are also in circulation right now. According to W.H.O,[3] the strains of the virus declared as variants of concern are Alpha or B.1.1.7, Beta or B.1.351, Gamma or P.1 and Delta or B.1.617.2 [1–3].

SARS-CoV-2 encompasses four structural proteins, spike glycoprotein, envelope, membrane glycoprotein and nucleocapsid, apart from non-structural proteins (NSP1-NSP16) and accessory proteins like ORF3a, ORF6, ORF7a, ORF7b, ORF8, ORF9b, ORF9c and ORF10 [4]. Viruses are incapable of living and reproducing outside a host body. Thus, they need to infiltrate a host for their survival. Protein-protein

* Corresponding author.
*E-mail address:* indrajit@nitttrkol.ac.in (I. Saha).
[1] Equally contributed.
[2] https://www.worldometers.info/coronavirus/.
[3] https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/.

https://doi.org/10.1016/j.compbiomed.2021.104889
Received 31 May 2021; Received in revised form 22 September 2021; Accepted 22 September 2021
Available online 6 October 2021

interaction (PPI) is one such way by which a virus invades a host cell [5]; SARS-CoV-2 being no exception. For SARS-CoV-2, bats are supposed to be the primary hosts and pangolins are identified to be the possible intermediate hosts from which the virus got transmitted to humans resulting in COVID-19 disease [6–8]. Furthermore, knowledge of virus invasion and pathogenesis of SARS-CoV-2 is very important to understand the comorbidities in human host. In this regard, study of PPI is crucial and helpful in drug repurposing and discovery as well. These facts have motivated us to conduct this research.

Traditionally, the collection of PPI data is mainly done through laboratory-based methods such as protein-chips [9,10], correlated mRNA expression profile [11], TAP-tagging [12,13], yeast-two hybrid [14,15] and synthetic lethal analysis [16]. However, laboratory based methods are mostly time consuming and labour-intensive. Also, due to the voluminous nature of PPI data there is a chance that PPI data generated by laboratory-based methods may not be complete [17]. Furthermore, small proteins are difficult to recognise in lab set up although they have important functional roles in many biological processes [18]. Moreover, it has been frequently observed that high false positives and false negatives occur in the prediction results of laboratory-based methods [19–21]. To mitigate these problems, a large number of computational methods have been proposed in the literature to identify protein-protein interactions. In this regard, a very popular method to predict PPI is link prediction model where it is considered that proteins interact if they are similar [22]. However, the accuracy of such models are heavily dependent on the reliability of PPI networks which may be affected due to a huge number of false-negative and false-positive PPIs. Also, in scale-free property of PPI networks [23,24], some PPI are dense while others are mostly sparse (average degree of 7 or less [25]) and link predictive models are not very efficient for sparse networks. Thus, high throughput technologies which consider biological information of proteins can be used to predict PPIs [26]. In Ref. [27], the authors have used bioinformatics and machine learning approaches to identify potential drug targets and pathways in COVID-19. In this regard, they have identified 1520 and 1733 differentially expressed genes (DEGs) from GSE152418 and CRA002390 PBMC datasets and have considered hub gene signature based on module membership (MMhub) statistics and PPI networks. Furthermore, they have demonstrated the classification performance of hub genes with more than 90% accuracy, thereby suggesting the potential of the hub genes to be biomarkers. Gupta et al. [28] have also used machine learning for prediction of new small molecule modulators of PPI. In their work, they have concluded that Random Forest predicts general PPI Modulators independent of PPI family with an AUC-ROC value $> 0.9$. They have also identified novel chemical scaffolds as inhibitors for RBD_hACE PPI which are involved in host cell entry of SARS-CoV-2.

Several public databases have been created for the experimentally determined human-virus PPI data and mostly consists of two categories [29]. The first one consists of PPI for species-specific databases encompassing only one specific viral species. It includes NCBI HIV-1 Human Interaction Database [30], HCVpro [31], DenHunt [32], DenvInt [33] and ZikaBase [34]. The second category on the other hand comprises of a wider range of virus species databases such as Viruses. STRING [35], VirusMentha [36], PHISTO [37], VirHostNet [38] and HPIDB [39]. Mostly, these public databases are created by integrating other PPI databases using automatic integration tools like PSICQUIC [40] or they may be manually collected from other public databases as well.

In order to contribute to the ongoing research pertaining to SARS-CoV-2 and PPI, in this work we have initially created a human-SARS-CoV-2 PPI database from existing works in the literature. In this regard, we have identified 7085 unique PPIs between the human and SARS-CoV-2 proteins. This consolidated database is a novel contribution of our work. Furthermore, for each virus protein we have identified at most 10 human hub proteins which have the highest degrees. These hub proteins are connected to most of the other human proteins. Consequently, if they are affected, the potential diseases in the pathways of most of the human proteins will get triggered as well, thereby leading to comorbidities. Also, the biological significance of the identified human hub proteins is reported by using KEGG which is essential for identifying the corresponding pathways related to diseases or comorbidities. Also, GO enrichment analysis is performed as well. As a consequence, it is identified that SARS-CoV-2 proteins viz. NSP2, NSP5, Envelope and ORF10 interacting with human hub proteins like COX4I1, COX5A, COX5B, NDUFS1, CANX, HSP90AA1 and TP53 can lead to comorbidities. Such comorbidities comprise of Alzheimer, Parkinson, Huntington, HTLV-1 infection, prostate cancer and viral carcinogenesis. Moreover, drug repurposing which is an effective drug discovery strategy from existing drugs is a very practical alternative to de novo drug discovery and random clinical trials. Considering this, we have also reported possible repurposable drugs like Disodium Selenite, Desipramine, Clindamycin and Vorinostat targeting the human hub proteins. To summarise, we have prepared human-SARS-CoV-2 PPI database by curating such PPIs from different existing works in the literature resulting in 7085 unique PPIs, identified human hub proteins using such PPI networks and finally identified the list of repurposable drugs for such human hub proteins as well as comorbidity issues related to such hub proteins. To the best of our knowledge, these consolidated ideas have not been addressed previously in any article. Therefore, this study mitigates the gaps in the literature through the above mentioned contributions. It is to be noted that other works like [41,42] have analysed drug repurposibility and comorbidities by considering expression data as opposed to our work which directly considers PPI data for the above analysis.

## 2. Materials and methods

In this section, the data preparation is elaborated at first which is then followed by the discussion on the pipeline of the proposed work.

### 2.1. Data preparation

For our work, initially we have prepared a consolidated human-SARS-CoV-2 PPI database taking into consideration the PPIs from Refs. [4,5,43]. There are 332 PPIs in Ref. [4] whereas [5] has reported 6489 PPIs and Li et al. [43] have reported 295 PPIs. Considering all the PPIs between human and SARS-CoV-2, 7085 unique PPIs are identified among 2204 unique human proteins and 4 structural and 25 non-structural virus proteins which include NSP1-16, Spike glycoprotein, ORF3a, ORF3b, Envelope protein, Membrane glycoprotein, ORF6, ORF7a, ORF7b, ORF8, ORF9b, ORF9c, Nucleocapsid and ORF10.

### 2.2. Pipeline of the work

The pipeline of the work is shown in Fig. 1(a). Initially, to create a consolidated human-virus PPI database, 7116 interactions are collected from the existing works in the literature which have thereafter resulted in 7085 unique PPIs. The distribution of the PPIs in the literature is shown in Fig. 1(b). Thereafter, all the human proteins for a particular virus protein are given as an input to the STRING database.[4] STRING database returns all the human-human protein interactions for those inputs and may include additional human proteins apart from the ones that are provided as inputs. It may also exclude some human proteins in the process as well. Next, for each SARS-CoV-2 protein, at most 10 human proteins viz. hub proteins are identified which have the highest degrees. It is important to note that based on their association with an individual SARS-CoV-2 protein, there are two levels of human proteins, Level 1 and Level 2 as shown in Fig. 1(c). Level 1 human proteins are those which are in the immediate vicinity or directly connected to the

[4] https://string-db.org/.

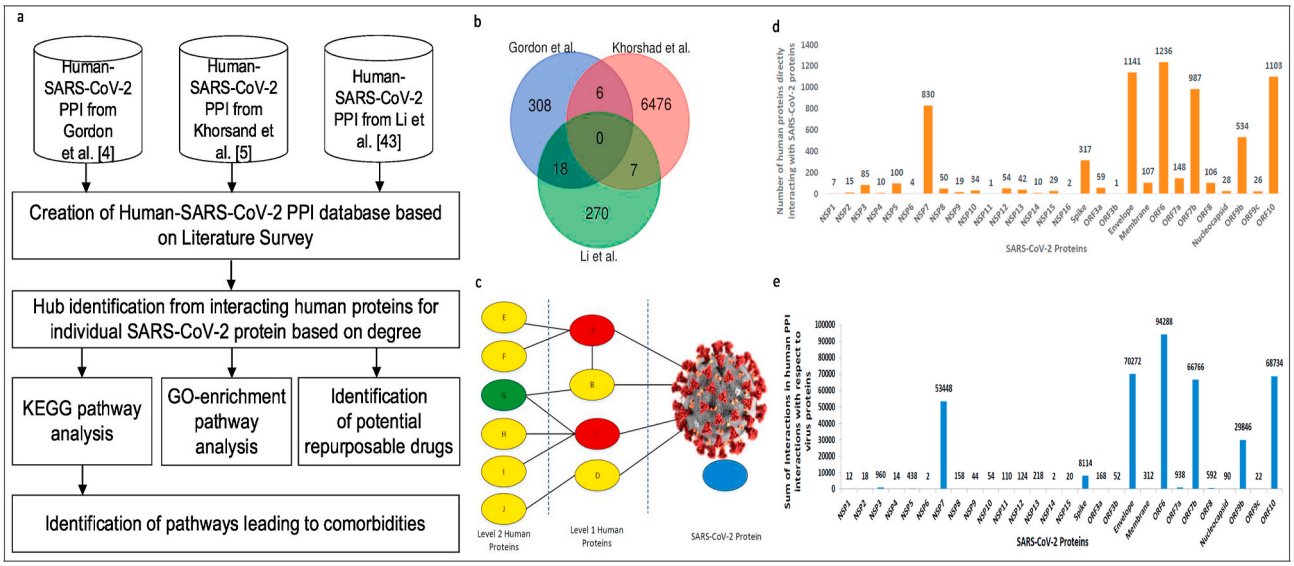


**Fig. 1.** (a) Pipeline of the work (b) Distribution of PPIs in literature (c) Representation of proteins in human-virus PPI network (d) Number of human proteins directly interacting with SARS-CoV-2 proteins and (e) Sum of interactions in human PPI interactome with respect to SARS-CoV-2 proteins.

SARS-CoV-2 protein while Level 2 are such human proteins which are indirectly connected to the virus protein through the Level 1 proteins. Among the 10 proteins as shown in the figure, A, C and G are considered to be the hub proteins as they have the highest degree among all the human proteins. Thus, a hub protein can either be a level 1 or a level 2 human protein. It is worth mentioning over here that a level 2 hub protein can be connected to the virus protein either through a hub protein or any directly connected human protein which may not be a hub protein. In this paper, the direct or level 1 hub proteins are marked in red while the indirect or level 2 hub proteins are marked in green and the rest of the human proteins are marked in yellow. SARS-CoV-2 proteins on the other hand are marked in blue throughout the paper. Once the hub proteins are identified, to understand the effects of these hub proteins on comorbidities, their pathways are explored and the

**Table 1**
Statistics of Human Proteins for each SARS-CoV-2 Protein.

| Virus | Number of Unique Human Proteins directly | Number of Unique Human Proteins | Number of Unique Human Hub Proteins (out of top 10) | Number of Unique Human Hub Proteins (out of top 10) | Number of Unique Human Proteins (other than hub proteins) |
|---|---|---|---|---|---|
| Protein | interacting with SARS-CoV-2 proteins | present in Human PPI network | directly interacting with SARS-CoV-2 proteins | indirectly interacting with SARS-CoV-2 proteins | directly connected to Hub Proteins |
| NSP1 | 7 | 4 | 4 | 0 | 0 |
| NSP2 | 15 | 9 | 8 | 1 | 0 |
| NSP3 | 85 | 72 | 10 | 0 | 32 |
| NSP4 | 10 | 6 | 5 | 1 | 0 |
| NSP5 | 100 | 87 | 10 | 0 | 35 |
| NSP6 | 4 | 2 | 1 | 1 | 0 |
| NSP7 | 830 | 788 | 9 | 1 | 469 |
| NSP8 | 50 | 39 | 10 | 0 | 16 |
| NSP9 | 19 | 13 | 9 | 1 | 1 |
| NSP10 | 34 | 27 | 10 | 0 | 10 |
| NSP11 | 1 | 11 | 1 | 9 | 1 |
| NSP12 | 54 | 32 | 9 | 1 | 15 |
| NSP13 | 42 | 29 | 10 | 0 | 11 |
| NSP14 | 10 | 2 | 2 | 0 | 0 |
| NSP15 | 29 | 11 | 8 | 2 | 1 |
| NSP16 | 2 | NA | NA | NA | NA |
| Spike glycoprotein | 317 | 302 | 10 | 0 | 158 |
| ORF3a | 59 | 44 | 9 | 1 | 16 |
| ORF3b | 1 | 11 | 1 | 9 | 1 |
| Envelope protein | 1141 | 1086 | 10 | 0 | 673 |
| Membrane glycoprotein | 107 | 81 | 9 | 1 | 36 |
| ORF6 | 1236 | 1194 | 9 | 1 | 677 |
| ORF7a | 148 | 133 | 9 | 1 | 55 |
| ORF7b | 987 | 951 | 9 | 1 | 611 |
| ORF8 | 106 | 82 | 10 | 0 | 42 |
| Nucleocapsid | 28 | 23 | 10 | 0 | 7 |
| ORF9b | 534 | 513 | 9 | 1 | 331 |
| ORF9c | 26 | 10 | 9 | 1 | 0 |
| ORF10 | 1103 | 1057 | 9 | 1 | 635 |

biological significance are demonstrated using KEGG pathway and GO enrichment analysis. KEGG pathway analysis is also important for identifying the pathways leading to comorbidities. Finally, identification of potential repurposable drugs targeting the human hub proteins to curb the effects of COVID-19 are carried out using Enrichr[5] [44,45] tool.

## 3. Results

This work is executed according to the pipeline as shown in Fig. 1(a). In this work, the primary motivations are to create a human-virus PPI interacting database and identifying the human hub proteins to understand their effects in comorbidities. In this regard, we have collected 7085 unique PPIs from the existing works in the literature, the details of which are provided in the Supplementary. Subsequently, with all the human-human interaction networks collected for each virus protein, the degree of each human protein with respect to a SARS-CoV-2 protein in the PPI network is identified. The degrees of the human proteins are provided in the Supplementary. Once the degree of each human protein for the corresponding SARS-CoV-2 protein is computed, at most top 10 human proteins are selected with the highest degrees which are then considered to be the hub proteins for each virus protein. The statistics of human proteins for each virus protein are reported in Table 1. This table shows the number of unique human proteins directly interacting with SARS-CoV-2 proteins, number of unique human proteins present in human PPI network considering proteins directly interacting with SARS-CoV-2 proteins, number of unique human hub proteins (out of top 10) directly interacting with SARS-CoV-2 proteins, number of unique human hub proteins (out of top 10) indirectly interacting with SARS-CoV-2 proteins and number of unique human proteins apart from the hub proteins directly connected to the hub proteins. As has been mentioned earlier, not all human proteins directly interacting with the SARS-CoV-2 proteins may be a part of the PPI network. This can be inferred from Table 1 as well. For example, for NSP1, 4 human proteins are present in the PPI network while 7 human proteins are directly interacting with SARS-CoV-2 proteins. The corresponding graph for the number of human proteins directly interacting with the SARS-CoV-2 proteins is shown in Fig. 1(d). The sum of interactions or the total degree of the human proteins in human PPI interactome with respect to the virus protein is shown in Fig. 1(e). For example, NSP7 has a total of 53448 human PPI interactions. It can be seen from the figure that out of the 29 virus proteins, 28 has corresponding human-human interaction networks while NSP16 does not have any associated human-human protein interactions.

All the identified human hub proteins may not be directly interacting with the SARS-CoV-2 proteins, rather they may be connected indirectly. For example, for NSP7, out of the 10 hub proteins, 9 such proteins are directly interacting with the SARS-CoV-2 protein while 1 human hub protein is indirectly interacting with the virus protein through some other human proteins. It is to be noted that for SARS-CoV-2 proteins like NSP1, NSP2, NSP4, NSP6 and NSP14 which have corresponding interacting human proteins equal to 7, 15, 10, 4 and 10 respectively have number of hub proteins equal to 4, 9, 6, 2 and 2, all less than 10. The details of the human hub proteins for each protein of SARS-CoV-2 are reported in Table 2. The table provides a list of the directly and indirectly connected hub proteins along with their respective degrees. For example, the directly connected hub proteins of NSP2 are NDUFS1, COX4I1, COX5A, COX5B, EIF4E2, FKBP15, GIGYF2 and MTCH2 with their respective degrees being 4, 3, 3, 3, 1, 1, 1 and 1 while the indirectly connected hub protein is KIAA1033 which has a degree of 1. The human-SARS-CoV-2 PPI network with only the directly and indirectly connected human hub proteins are visualised in Fig. 2 while Fig. 3 shows the individual PPI networks for all the SARS-CoV-2 proteins. The networks are created using Cytoscape [46] which is an open-source platform. As there may be a lot of human proteins directly connected to the hub proteins (for example, Envelope protein has 673 human proteins directly connected to hub proteins), for visualization purposes, for each SARS-CoV-2 protein, apart from all the hub proteins, only a handful of the human proteins are chosen from both level 1 and level 2 and shown in Fig. 3. The criteria for choosing such human proteins (excluding the hub proteins) are as follows:

- if the number of such human proteins are less than 20, then consider all such proteins,
- if the number of such human proteins are greater than 20, then consider 20 such proteins having the highest degrees.

Thus, no more than 30 human proteins ($\leq$10 hub proteins and $\leq$20 other proteins) are considered for visualization purpose in Fig. 3. The details of all the human-SARS-CoV-2 PPI corresponding to only the hub proteins for each virus protein along with the details of all such interactions for each virus protein irrespective of the hub proteins are provided in the Supplementary.

## 4. Discussion

### 4.1. KEGG pathway analysis

KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis reveals the potential diseases that can develop in humans due to SARS-CoV-2. Hub proteins are the ones which are connected to most of the other human proteins in the PPI network. Thus, instead of considering all the human proteins that have been returned by the STRING database, for the KEGG pathway analysis only the hub proteins and those human proteins which are directly connected to the hub proteins are considered. Table 3 reports such maximum 5 significant KEGG pathways (if there are any) with the corresponding human hub proteins related to them and their FDR corrected p-values. The detailed pathways are provided in the Supplementary. These results are collected from STRING database. Fig. 4 shows the KEGG pathways for NSP2, NSP5, Envelope protein and ORF10. For better visualization, maximum top 30 pathways are shown in the figures. The size of the bubbles in the figures are based on the corresponding number of human hub proteins associated with each pathway; lesser the number of hub proteins, smaller are the size of the bubbles while their colours are based on the FDR-corrected p-values. It can be seen from Fig. 4 that the most significant pathways corresponding to hub proteins for a SARS-CoV-2 protein are involved in various diseases. For example, the human hub proteins targeted by NSP2 are enriched in pathways relating to *hsa05010: Alzheimer's disease*, *hsa05012: Parkinson's disease* and *hsa05016: Huntington's disease* with the respective FDR corrected p-value being 4.51E-06 for all the three pathways while the corresponding hub proteins targeted by NSP2 are COX4I1, COX5A, COX5B and NDUFS1. SARS-CoV-2 can aggravate cancer pathways as well. For example, human hub protein CANX targeted by NSP5 is enriched in pathway for *hsa05166: HTLV-I infection* (FDR-corrected p-value 3.20E-03) which is associated with aggressive adult T-cell lymphoma, GAPDH targeted by Envelope protein is enriched in pathway for *hsa04066: HIF-1 signaling pathway* (FDR-corrected p-value 2.50E-04) while HSP90AA1 is enriched in pathways for *hsa04151: PI3K-Akt signaling pathway* (FDR-corrected p-value 1.30E-03), *hsa05215: Prostate cancer* (FDR-corrected p-value 1.24E-02), *hsa05200: Pathways in cancer* (FDR-corrected p-value 1.26E-02) and EEF2 is responsible for *hsa04010: AMPK signaling pathway* (FDR-corrected p-value 4.80E-04). Furthermore, RPN1, SEC61A1, CANX and HSP90B1 all targeted by NSP5 are enriched in the pathway for *hsa04141: Protein processing in endoplasmic reticulum* (FDR-corrected p-value 2.37E-07) and there are studies [47,48] which show that prolonged endoplasmic reticulum stress is responsible for the development and progression of many diseases like atherosclerosis, neurodegeneration, liver disease, type 2 diabetes and cancer. Moreover, TP53 targeted by ORF10 is enriched in

[5] https://maayanlab.cloud/Enrichr/.

**Table 2**
Details of Human Hub Proteins for each SARS-CoV-2 Protein.

| Virus | Human hub proteins (out of top 10) | Degree of Human hub proteins | Human hub proteins (out of top 10) indirectly | Degree of Human hub proteins indirectly |
|---|---|---|---|---|
| Protein | directly interacting with SARS-CoV-2 proteins | directly interacting with SARS-CoV-2 proteins | interacting with SARS-CoV-2 proteins | interacting with SARS-CoV-2 proteins |
| NSP1 | POLA1, POLA2, PRIM1, PRIM2 | 3, 3, 3, 3 | NA | NA |
| NSP2 | NDUFS1, COX4I1, COX5A, COX5B, EIF4E2, FKBP15, GIGYF2, MTCH2 | 4, 3, 3, 3, 1, 1, 1, 1 | KIAA1033 | 1 |
| NSP3 | RPL8, RPSA, RPL12, EEF1A1, RPL6, RPL15, RPS11, RPS16, RPL11, RPS15A | 31, 30, 29, 28, 28, 27, 27, 27, 26, 26 | NA | NA |
| NSP4 | TIMM10, TIMM10B, TIMM9, RAD23A, XPC | 3, 3, 3, 1, 1 | C19orf52 | 3 |
| NSP5 | CCT5, RPN1, CCT3, SEC61A1, CANX, CCT7, EEF1G, HSP90B1, PSMC6, PSMD14 | 16, 16, 15, 15, 14, 13, 13, 12, 12, 12 | NA | NA |
| NSP6 | ATP6AP1 | 1 | ATP5L | 1 |
| NSP7 | EEF2, HNRNPA1, EIF4A3, HSPA8, RPL4, RPS20, RPSA, RPS3, RPL8 | 259, 251, 243, 236, 232, 231, 222, 221, 219 | NHP2L1 | 238 |
| NSP8 | NOP58, MPHOSPH10, EXOSC3, DDX10, NGDN, XPO1, EXOSC2, EXOSC5, KPNA2, SRP54 | 13, 10, 9, 8, 8, 8, 7, 7, 7, 7 | NA | NA |
| NSP9 | NUP214, NUP54, NUP62, NUP88, HSPA1A, NEK9, FBN1, FBLN5, FBN2 | 6, 6, 6, 6, 5, 4, 2, 1, 1 | NUPL1 | 4 |
| NSP10 | ALDH18A1, NMD3, XPOT, XPO5, AASDHPPT, ALDH7A1, AP2A2, DIS3, GALK1, GSPT2 | 8, 4, 4, 3, 2, 2, 2, 2, 2, 2 | NA | NA |
| NSP11 | TBCA | 10 | TBCD, TBCE, TUBA1A, TUBA4A, TUBB1, TUBB2A, TUBB2B, TUBB4A, TUBB4B | 10, 10, 10, 10, 10, 10, 10, 10, 10 |
| NSP12 | PABPC1, HSPA8, NCL, PCBP1, MATR3, RBMX, STAU1, DDX1, RPS19, | 11, 10, 10, 8, 7, 7, 7, 5, 5 | C14orf166 | 4 |
| NSP13 | AKAP9, PCNT, CDK5RAP2, CEP135, PRKAR2B, CEP250, PRKACA, CNTRL, FGFR1OP, NIN | 17, 17, 14, 12, 12, 11, 11, 10, 10, 10 | NA | NA |
| NSP14 | PRDX4, PRDX5 | 1, 1 | NA | NA |
| NSP15 | POLR1B, URB1, TRMT2A, CDK12, CTCF, NUTF2, NXF1, REXO4 | 3, 3, 2, 1, 1, 1, 1, 1 | KIAA0020, TCEB3 | 5,1 |
| NSP16 | NA | NA | NA | NA |
| Spike glycoprotein | HSPA8, CCT4, RPL8, RPS3, RPSA, CCT5, CCT8, EEF1A1, CCT7, RPLP0 | 101, 89, 89, 88, 86, 85, 84, 84, 83, 83 | NA | NA |
| ORF3a | HYOU1, P4HB, PDIA6, PPIB, FKBP10, PDIA4, SERPINH1, EDEM3, TXNDC5 | 14, 13, 10, 10, 8, 8, 8, 7, 7 | ERO1L | 7 |
| ORF3b | STOML2 | 10 | PHB2, YME1L1, PARL, PHB, SMDT1, ATP5A1, MRPL40, HSPA1A, HSPA1L | 6, 6, 5, 5, 5, 4, 4, 3, 2 |
| Envelope protein | HSPA8, GAPDH, EEF2, RPS27A, HNRNPA1, HSP90AA1, EIF4A3, RPL4, CCT2, RPS3 | 301, 286, 279, 268, 260, 256, 254, 253, 251, 248 | NA | NA |
| Membrane glycoprotein | KPNA2, POLR2A, IPO5, NOP58, SMC4, NUP133, TOP2A, POLR2B, SNRNP70 | 12, 12, 10, 10, 10, 9, 9, 8, 8 | ATP5O | 7 |
| ORF6 | HSPA8, EEF2, EIF4A3, RPL4, HNRNPA1, RPS3, RPSA, RPS20, RPS27A | 320, 312, 307, 306, 305, 302, 302, 299, 299 | NHP2L1 | 309 |
| ORF7a | POLR2A, DHX9, SNRPD2, CPSF1, SRSF3, CPSF4, FIP1L1, NOP58, PRPF8 | 32, 29, 28, 22, 22, 21, 21, 21, 20 | SKIV2L2 | 22 |
| ORF7b | HSPA8, EEF2,HNRNPA1, EIF4A3, GAPDH, RPL4, RPS20, RPSA, RPS27A | 284, 273, 273, 264, 258, 248, 248, 245, 244 | NHP2L1 | 262 |
| ORF8 | CANX, HSP90B1, CALR, PDIA6, CCT5, CCT7, HYOU1, CCT3, SEC61A1, RPN1 | 24, 23, 21, 20, 19, 19, 19, 18, 18, 17 | NA | NA |
| Nucleocapsid | PABPC1, ELAVL1, G3BP1, UPF1, HNRNPDL, HSPA4, MOV10, PABPC4, G3BP2, LARP1 | 10, 9, 8, 8, 5, 5, 5, 5, 4, 4 | NA | NA |
| ORF9b | HSPA8, RPS20, RPSA, RPL4, RPS3, CCT2, EEF1A1, HNRNPA1, RPS27A | 185, 177, 177, 176, 175, 173, 173, 172, 172 | GNB2L1 | 171 |
| ORF9c | ECSIT, ACAD9, NDUFAF1, NDUFB9, GPAA1, PIGO, PIGS, BCS1L, DPY19L1 | 4, 3, 3, 3, 2, 2, 2, 1, 1 | FAM134C | 1 |
| ORF10 | HSPA8, RPS27A, EEF2, EIF4A3, HNRNPA1, RPL4, RPS20, RPS3, TP53 | 282, 272, 269, 264, 260, 256, 252, 251, 251 | NHP2L1 | 265 |

pathway relating to *hsa05203: Viral carcinogenesis* (FDR-corrected p-value 1.70E-04). Other significant pathways found for the human proteins with FDR corrected p-values within 5% statistical significance are *Influenza A*, *Measles*, *Epstein-Barr Virus infection* and *Vibrio cholerae infection*.

### 4.2. Gene ontology (GO) enrichment analysis

GO enrichment analysis is performed to understand the significance of the roles that the different interacting human proteins play in biological activities. Similar to KEGG pathways, the GO enrichment results are collected from STRING database as well and considered only for the hub proteins and their direct connections. The result of the analysis for

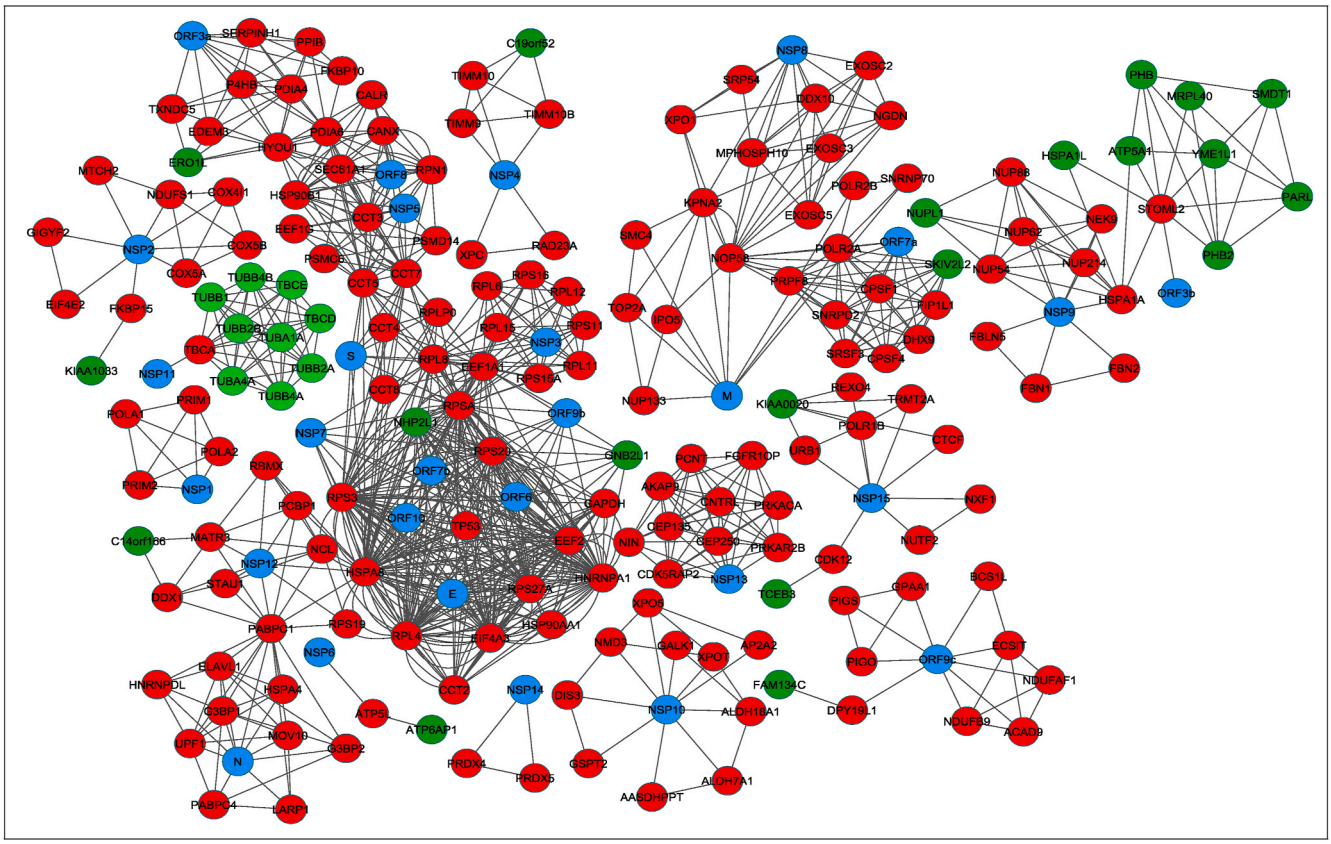


**Fig. 2.** Human-SARS-CoV-2 PPI network with only directly and indirectly connected human hub proteins. Nodes marked in blue represent the SARS-CoV-2 proteins, nodes marked in red represent the human hub proteins directly connected to SARS-CoV-2 proteins and green represents the human hub proteins indirectly connected to SARS-CoV-2 proteins.

biological processes for NSP2, NSP5, Envelope protein and ORF10 considering at least one hub protein are reported in Fig. 5. For better readability, only the top 30 pathways are shown in the figures. The detailed analysis for all the GO pathways (biological, molecular and cellular) are provided in the Supplementary. Some significant biological pathways for human hub proteins COX5B, COX5A and COX4I1 targeted by NSP2 are: *GO:0006123: mitochondrial electron transport, cytochrome c to oxygen* (FDR-corrected p-value 1.11E-05), hub proteins HSP90B1, PSMC6 and PSMD14 targeted by NSP5: *GO:0030163: protein catabolic process* (FDR-corrected p-value 1.34E-06), HSPA8, RPS27A, HNRNPA1, EIF4A3, RPL4 and RPS3 targeted by Envelope protein: *GO:0016071: mRNA metabolic process* (FDR-corrected p-value 4.13E-123) and HSPA8, RPS27A, NHP2L1, EIF4A3, HNRNPA1, RPL4, RPS20, RPS3 and TP53 targeted by ORF10: *GO:0016071: mRNA metabolic process* (FDR-corrected p-value 6.46E-132).

### 4.3. Repurposable drugs

Till now, no efficacious drug has been discovered to combat SARS-CoV-2. The traditional mechanism of drug development is expensive and time-consuming, thereby making drug repurposing a viable option for effective drug identification for COVID-19. In this regard, human hub proteins corresponding to each SARS-CoV-2 protein can be considered to be good candidates as targets for drug repurposing. Such drugs that interact with the hub proteins are identified using DSigDB in Enrichr tool. For each virus protein, the results for at most top 5 drugs (if any) along with their Drug Bank ID as collected from Drug Bank,[6] their FDR corrected p-values and the possible treatments are reported in Table 4.

As can be seen from Table 4, several drugs are identified which can be used for treating cancer. For example, Tanespimycin (FDR corrected p-value 4.44E-03 and Drug Bank ID DB05134) which targets human hub protein like HSP90AA1 corresponding to Envelope protein is used for treating several types of cancer, solid tumors or chronic myelogenous leukemia. As previously discussed, HSP90AA1 which is targeted by SARS-CoV-2 Envelope protein triggers PI3K-Akt signaling pathway whose aberrant activation promotes the survival and growth of tumor cells in many human cancers. Other drugs like Phenethyl isothiocyanate, 4-Hydroxytamoxifen, Daunorubicin, Camptothecin, Vorinostat, Diindolylmethane etc. are also used for the treatment of various types of cancer. It is worth noting that identified drugs like Resveratrol known for the treatment of high cholesterol, cancer and heart disease and Niclosamide used for treating tapeworm infection are under trials for the treatment of COVID-19 [49,50]. Please note that all the hub proteins involved for KEGG pathway analysis may not have corresponding drugs with FDR corrected p-value less than 5%. Thus, only those hub proteins are reported in Table 4 for which there are corresponding relevant drugs. For example, for NSP2, the hub proteins with corresponding KEGG pathways having FDR corrected p-values less than 5% are NDUFS1, COX4I1, COX5A and COX5B while the hub proteins with relevant drugs having FDR corrected p-values less than 5% are NDUFS1, COX5A and COX5B. Fig. 6 provides a glimpse of the common hub proteins and drugs among multiple SARS-CoV-2 proteins. For example, RPSA is a hub protein common to NSP3, NSP7 and Spike glycoprotein and the corresponding drug that targets RPSA is Disodium Selenite. Please note that though RPSA is also targeted by ORF9b as shown in Table 4, it is not shown in the figure as Disodium Selenite is not a relevant drug for RPSA in ORF9b as the corresponding FDR corrected p-value of Disodium Selenite is not less than 5% in this case. Other drugs like Desipramine, Clindamycin and Vorinostat used as antidepressants,

[6] https://go.drugbank.com/drugs.

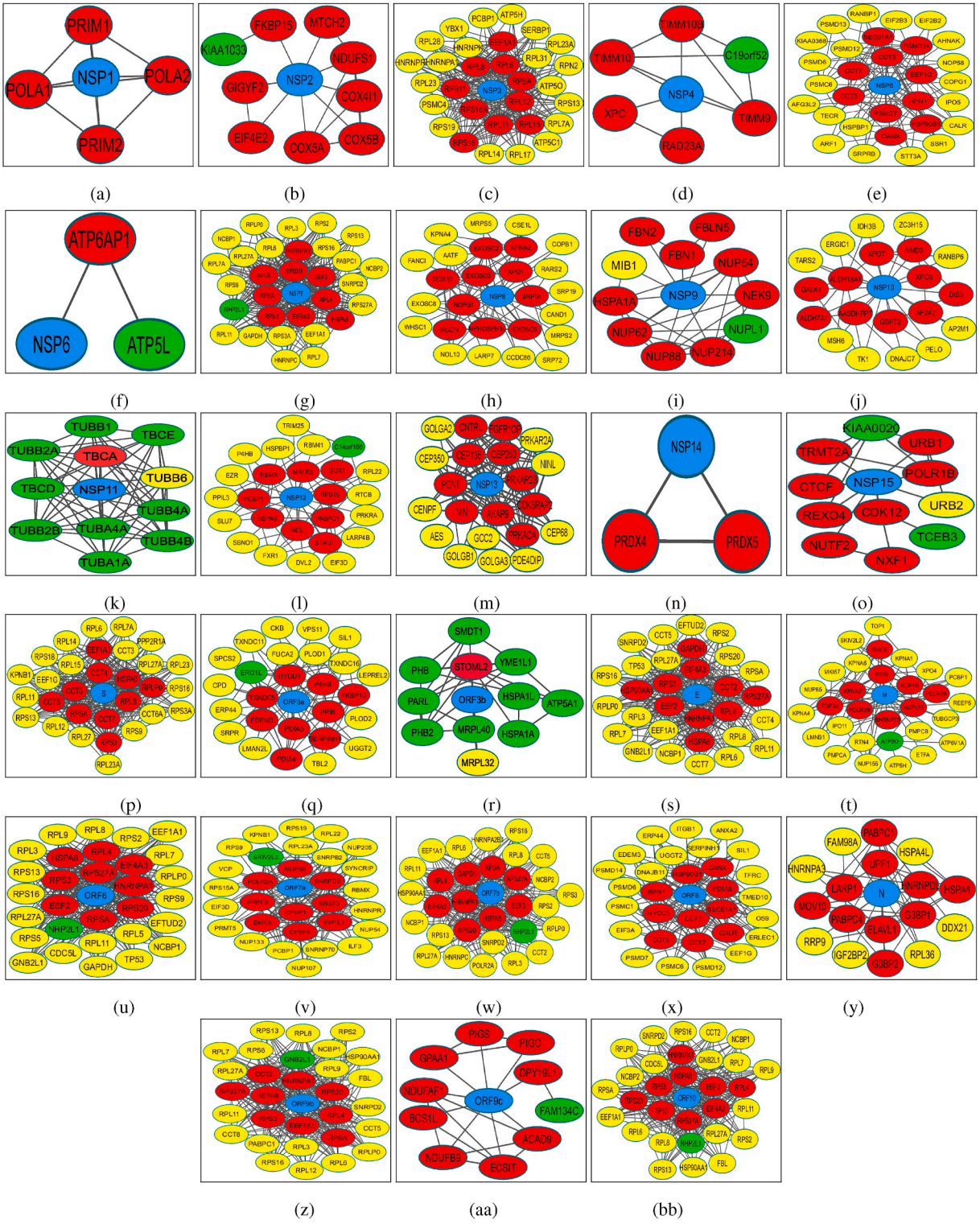


**Fig. 3.** A glance into human-SARS-CoV-2 PPI network for (a) NSP1 (b) NSP2 (c) NSP3 (d) NSP4 (e) NSP5 (f) NSP6 (g) NSP7 (h) NSP8 (i) NSP9 (j) NSP10 (k) NSP11 (l) NSP12 (m) NSP13 (n) NSP14 (o) NSP15 (p) Spike glycoprotein (q) ORF3a (r) ORF3a (s) Envelope protein (t) Membrane glycoprotein (u) ORF6 (v) ORF7a (w) ORF7b (x) ORF8 (y) Nucleocapsid (z) ORF9b (aa) ORF9c and (bb) ORF10. In these figures, nodes marked in blue represent the SARS-CoV-2 proteins, nodes marked in red represent the human hub proteins directly connected to SARS-CoV-2 proteins, green represents the human hub proteins indirectly connected to SARS-CoV-2 proteins and yellow represents other human proteins directly connected to hub proteins.

**Table 3**
Details of KEGG Pathways corresponding to Human Hub Proteins for each SARS-CoV-2 Protein.

| Virus Protein | Human hub proteins | KEGG Pathways related to Comorbidities | FDR corrected p-value | Virus Protein | Human hub proteins | KEGG Pathways related to Comorbidities | FDR corrected p-value |
|---|---|---|---|---|---|---|---|
| NSP1 | POLA1, POLA2, PRIM1, PRIM2 | DNA replication | 5.98E-11 | Spike glycoprotein | HSPA8, RPL8, RPS3, RPSA, EEF1A1, RPLP0 | Ribosome | 7.36E-30 |
| | | Pyrimidine metabolism | 1.51E-09 | | | Protein processing in endoplasmic reticulum | 1.96E-17 |
| | | Purine metabolism | 8.63E-09 | | | RNA transport | 1.03E-09 |
| | | Metabolic pathways | 1.68E-05 | | | Epstein-Barr virus infection | 1.06E-09 |
| | | | | | | Legionellosis | 1.50E-04 |
| NSP2 | NDUFS1, COX4I1, COX5A, COX5B | Alzheimers disease | 4.51E-06 | ORF3a | HYOU1, P4HB, PDIA6, PDIA4, EDEM3, ERO1L, TXNDC5 | Protein processing in endoplasmic reticulum | 4.96E-12 |
| | | Huntingtons disease | 4.51E-06 | | | | |
| | | Non-alcoholic fatty liver disease (NAFLD) | 4.51E-06 | | | Vibrio cholerae infection | 5.30E-03 |
| | | Oxidative phosphorylation | 4.51E-06 | | | | |
| | | Parkinsons disease | 4.51E-06 | | | | |
| NSP3 | RPL8, RPSA, RPL12, EEF1A1, RPL6, RPL15, RPS11, RPS16, RPL11, RPS15A | Ribosome | 1.72E-26 | ORF3b | HSPA1A, HSPA1L | Estrogen signaling pathway | 9.10E-03 |
| | | | | | | Measles | 9.10E-03 |
| | | RNA transport | 5.20E-03 | | | Influenza A | 9.10E-03 |
| | | | | | | Epstein-Barr virus infection | 9.80E-03 |
| | | | | | | MAPK signaling pathway | 1.80E-02 |
| NSP4 | XPC, RAD23A | Nucleotide excision repair | 1.80E-04 | Envelope protein | GAPDH, EEF2, HSP90AA1 | HIF-1 signaling pathway | 2.50E-04 |
| | | | | | | AMPK signaling pathway | 4.80E-04 |
| | | | | | | PI3K-Akt signaling pathway | 1.30E-03 |
| | | | | | | Prostate cancer | 1.24E-02 |
| | | | | | | Pathways in cancer | 1.26E-02 |
| NSP5 | RPN1, SEC61A1, CANX, HSP90B1, PSMC6, PSMD14 | Protein processing in endoplasmic reticulum | 2.37E-07 | Membrane glycoprotein | POLR2A, NUP133, POLR2B, ATP5O | RNA transport | 1.10E-02 |
| | | Proteasome | 1.10E-04 | | | Huntington's disease | 1.56E-02 |
| | | Epstein-Barr virus infection | 1.10E-04 | | | RNA polymerase | 2.44E-02 |
| | | HTLV-I infection | 3.20E-03 | | | Oxidative phosphorylation | 2.58E-02 |
| | | Vibrio cholerae infection | 3.39E-02 | | | Alzheimer's disease | 4.06E-02 |
| NSP6 | ATP5L, ATP6AP1 | Oxidative phosphorylation | 4.60E-04 | ORF6 | HSPA8, NHP2L1, EIF4A3 | RNA transport | 1.79E-39 |
| | | | | | | Epstein-Barr virus infection | 8.05E-18 |
| | | Metabolic pathways | 2.05E-02 | | | mRNA surveillance pathway | 1.51E-16 |
| | | | | | | Influenza A | 3.30E-04 |
| | | | | | | Legionellosis | 1.09E-02 |
| NSP7 | HNRNPA1, EIF4A3, NHP2L1, HSPA8, RPL4, RPS20, RPSA, RPS3, RPL8 | Ribosome | 1.08E-104 | ORF7a | CPSF1, SRSF3, CPSF4, FIP1L1, PRPF8 | Spliceosome | 1.48E-05 |
| | | Spliceosome | 1.00E-42 | | | | |
| | | Epstein-Barr virus infection | 1.18E-07 | | | mRNA surveillance pathway | 3.03E-02 |
| | | Influenza A | 1.39E-05 | | | | |
| | | Legionellosis | 3.60E-04 | | | | |
| NSP8 | NOP58, MPHOSPH10, EXOSC3, XPO1, EXOSC2, EXOSC5, SRP54 | RNA degradation | 4.54E-05 | ORF7b | HSPA8, GAPDH | Epstein-Barr virus infection | 6.52E-21 |
| | | Protein export | 4.54E-05 | | | Influenza A | 3.35E-05 |
| | | Ribosome biogenesis in eukaryotes | 6.10E-04 | | | Legionellosis | 1.70E-03 |
| | | | | | | Longevity regulating pathway - multiple species | 1.16E-02 |
| | | | | | | HIF-1 signaling pathway | 3.92E-02 |
| NSP9 | NUP214, NUP54, NUP62, NUP88, HSPA1A, NUPL1 | RNA transport | 2.24E-07 | ORF8 | CANX, HSP90B1, CALR, PDIA6, HYOU1, SEC61A1, RPN1 | Protein processing in endoplasmic reticulum | 3.18E-15 |
| | | | | | | Phagosome | 3.03E-05 |

(*continued on next page*)

**Table 3** (*continued*)

| Virus Protein | Human hub proteins | KEGG Pathways related to Comorbidities | FDR corrected p-value | Virus Protein | Human hub proteins | KEGG Pathways related to Comorbidities | FDR corrected p-value |
|---|---|---|---|---|---|---|---|
| | | Epstein-Barr virus infection | 0.0336 | | | Antigen processing and presentation | 5.80E-03 |
| | | | | | | N-Glycan biosynthesis | 4.66E-02 |
| | | | | | | Vibrio cholerae infection | 4.66E-02 |
| NSP10 | ALDH18A1, ALDH7A1, AP2A2, GALK1 | Arginine and proline metabolism | 2.00E-02 | Nucleocapsid | PABPC1, UPF1, PABPC4 | mRNA surveillance pathway | 6.50E-04 |
| | | Biosynthesis of amino acids | 2.00E-02 | | | RNA transport | 1.70E-03 |
| | | Endocrine and other factor-regulated calcium reabsorption | 2.00E-02 | | | RNA degradation | 7.00E-03 |
| | | Synaptic vesicle cycle | 2.00E-02 | | | | |
| | | Metabolic pathways | 3.70E-02 | | | | |
| NSP11 | TUBA1A, TUBA4A, TUBB1, TUBB2A, TUBB2B, TUBB4A, TUBB4B | Pathogenic Escherichia coli infection | 4.51E-18 | ORF9b | HSPA8, RPS20, RPSA, RPL4, RPS3, EEF1A1, HNRNPA1, RPS27A | Ribosome | 3.33E-81 |
| | | Gap junction | 9.27E-17 | | | Spliceosome | 1.40E-26 |
| | | Phagosome | 3.12E-15 | | | Epstein-Barr virus infection | 1.86E-11 |
| | | Apoptosis | 3.20E-03 | | | Legionellosis | 8.90E-03 |
| | | Tight junction | 3.90E-03 | | | Antigen processing and presentation | 1.75E-02 |
| NSP12 | PABPC1, HSPA8, NCL, PCBP1, RBMX | Spliceosome | 9.50E-04 | ORF9c | NDUFAF1, NDUFB9, GPAA1, PIGO, PIGS | Glycosylphosphatidylinositol (GPI)-anchor biosynthesis | 3.12E-06 |
| | | RNA transport | 2.23E-02 | | | Metabolic pathways | 1.29E-02 |
| | | Protein processing in endoplasmic reticulum | 2.23E-02 | | | Thermogenesis | 1.94E-02 |
| | | Pathogenic Escherichia coli infection | 2.23E-02 | | | | |
| NSP13 | PRKAR2B, PRKACA | Insulin signaling pathway | 2.72E-02 | ORF10 | HSPA8, TP53 | Epstein-Barr virus infection | 2.07E-20 |
| | | | | | | Herpes simplex infection | 4.58E-05 |
| | | | | | | Viral carcinogenesis | 1.70E-03 |
| | | | | | | Huntingtons disease | 8.80E-03 |
| | | | | | | Influenza A | 2.21E-02 |

antibiotic and for treating Cutaneous T-cell lymphoma (CTCL) respectively are also relevant drugs for the human hub proteins targeted by multiple SARS-CoV-2 proteins. Apart from the discussed hub proteins, it is to be noted that as per https://cancer.sanger.ac.uk/cosmic/, other identified hub proteins like XPC in NSP4, RPN1 in NSP5, XPO1 in NSP8, NUP214 in NSP9, PABPC1 and PCBP1 in NSP12, PRKACA in NSP13, SRSF3 and FIP1L1 in ORF7a and CALR in ORF8 are also cancer related human proteins.

## 5. Conclusion

Comorbidity in COVID-19 patients is one of the primary reasons which have led to so many deaths around the globe. SARS-CoV-2, the virus causing COVID-19, sneaks its way into human body by interacting with the human proteins. In this work, we have identified human and SARS-CoV-2 protein-protein interactions to identify human hub proteins associated with comorbidities. In this regard, we have initially collected 7116 human-SARS-CoV-2 PPI from different works in the literature resulting in identifying 7085 unique PPIs. This can be considered to be a novel and significant contribution of our work. Thereafter, we have considered at most top 10 human hub proteins based on their degrees. Moreover, biological significance of the identified human proteins is demonstrated using KEGG which is essential for identifying the pathways related to diseases or comorbidities. Also, GO Enrichment analysis is performed as well. SARS-CoV-2 proteins like NSP2, NSP5, Envelope and ORF10 interacting with human hub proteins COX4I1, COX5A, COX5B, NDUFS1, CANX, HSP90AA1 and TP53 can lead to comorbidities like Alzheimer, Parkinson, Huntington's, HTLV-1 infection, prostate cancer and viral carcinogenesis. Furthermore, possible repurposable drugs like Disodium Selenite, Desipramine, Clindamycin and Vorinostat targeting the human hub proteins are reported in this paper for future reference for researchers. Also, reported drugs like Resveratrol and Niclosamide are under trials for the treatment of COVID-19. This work provides a consolidated study for human-SARS-CoV-2 protein interactions to understand the association between comorbidity and human hub proteins and we hope it will also be helpful in drug repurposing and discovery as well. To summarise, we have prepared human-SARS-CoV-2 PPI database by curating such PPIs from different works in the literature resulting in 7085 unique PPIs, identified human hub proteins using such PPI networks and identified a list of repurposable drugs for such human hub proteins as well as comorbidity issues related to such hub proteins.

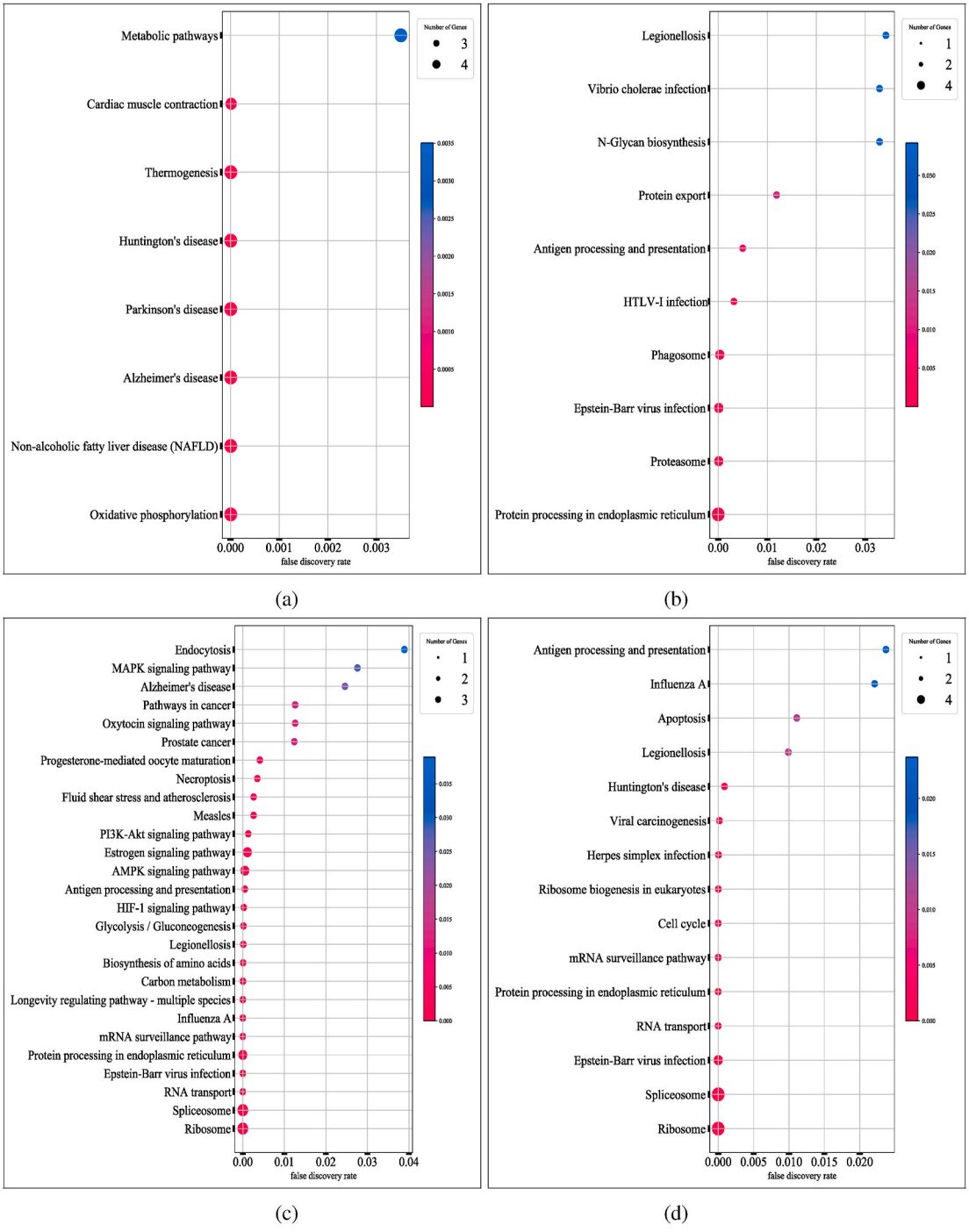


**Fig. 4.** Significant KEGG pathways corresponding to Hub Proteins for (a) NSP2 (b) NSP5 (c) Envelope protein and (d) ORF10.

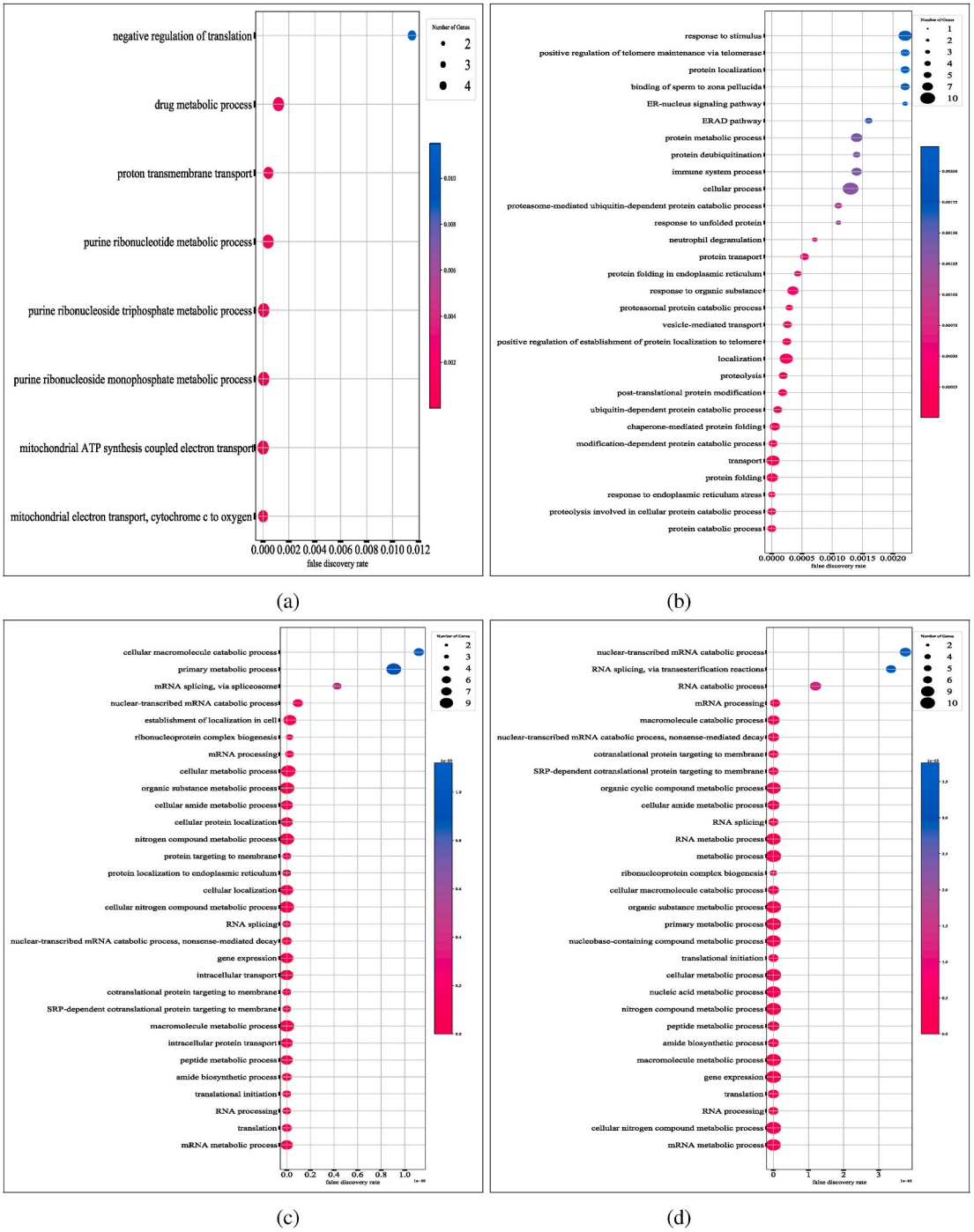


**Fig. 5.** Significant GO Biological Processes corresponding to Hub Proteins for (a) NSP2 (b) NSP5 (c) Envelope protein and (d) ORF10.

**Table 4**
Details of Drugs corresponding to Human Hub Proteins for each SARS-CoV-2 Protein.

| Virus Protein | Human hub proteins | Drugs | FDR corrected p-value | Drug Bank ID | Treatment | Virus Protein | Human hub proteins | Drugs | FDR corrected p-value | Drug Bank ID | Treatment |
|---|---|---|---|---|---|---|---|---|---|---|---|
| NSP1 | POLA1, POLA2, PRIM1, PRIM2 | Dasatinib | 3.96E-05 | DB01254 | Lymphoblastic or chronic myeloid leukemia | ORF3b | HSPA1A, HSPA1L | D-Penicillamine | 1.89E-02 | DB00859 | Wilson's disease |
| | | Resveratrol | 1.38E-03 | DB02709 | High cholesterol, cancer, heart disease | | | Dronabinol | 2.62E-02 | DB00470 | Treat nausea and vomiting caused by chemotherapy |
| | | Demecolcine | 4.67E-03 | DB13318 | Chemotherapy | | | Chlortetracycline | 3.42E-02 | DB09093 | Antibiotic |
| | | Fluorouracil | 8.35E-03 | DB00544 | Cancer | | | Aspirin | 3.91E-02 | DB00945 | Fever and pain |
| | | Troglitazone | 3.41E-02 | DB00197 | Type 2 Diabetes | | | Lomustine | 3.92E-02 | DB01206 | Brain Tumour, Hodgkin's lymphoma |
| NSP2 | NDUFS1, COX5A, COX5B | Vitinoin | 1.31E-02 | DB00755 | Eczema and certain types of promyelocytic leukemia | Envelope Protein | GAPDH, EEF2, HSP90AA1 | Idebenone | 3.38E-03 | DB09081 | Alzheimer's disease and Leber's disease |
| | | | | | | | | Tanespimycin | 4.44E-03 | DB05134 | Several types of cancer, solid tumors or chronic myelogenous leukemia |
| | | 3′-Azido-3′-deoxythymidine | 2.00E-02 | DB00495 | HIV | | | Ivermectin | 5.67E-03 | DB00602 | Anti-parasite |
| | | | | | | | | Alvespimycin | 5.67E-03 | DB12442 | Antitumour in cancer |
| | | | | | | | | Disodium selenite | 0.035247665 | DB11127 | Prevents Cancer |
| NSP3 | RPL8, RPSA, RPL12, EEF1A1, RPL6, RPL15, RPS16, RPS15A | Disodium selenite | 2.00E-03 | DB11127 | Prevents Cancer | Membrane glycoprotein | NUP133, POLR2B | Calcitrol | 1.64E-02 | DB00136 | Treat hyperparathyroidism |
| | | Artesunate | 8.35E-03 | DB09274 | Malaria | | | Fulvestrant | 3.45E-02 | DB00947 | Breast Cancer |
| | | Vorinostat | 2.43E-02 | DB02546 | Cutaneous T-cell lymphoma (CTCL) | | | | | | |
| NSP5 | SEC61A1, CANX, HSP90B1, PSMD14 | Clindamycin | 9.20E-03 | DB01190 | Antibiotic | ORF6 | HSPA8, NHP2L1 | Disodium selenite | 2.00E-03 | DB11127 | Prevents Cancer |
| NSP6 | ATP5L, ATP6AP1 | Niclosamide | 5.24E-04 | DB06803 | Tapeworm infection | ORF7a | CPSF1, SRSF3, PRPF8 | Clindamycin | 8.97E-03 | DB01190 | Antibiotic |
| | | Vanadium | 5.24E-04 | DB13971 | Diabetes, low blood sugar, high cholesterol, heart disease, tuberculosis, syphilis, preventing cancer | | | | | | |
| | | Phenethyl isothiocyanate | 3.37E-02 | DB12695 | Leukemia, Lung Cancer | | | | | | |
| NSP7 | NHP2L1, HSPA8, RPL4, RPSA, RPS3, RPL8 | Disodium selenite | 8.28E-05 | DB11127 | Prevents Cancer | ORF7b | HSPA8, GAPDH | Disodium selenite | 2.14E-02 | DB11127 | Prevents Cancer |
| | | | | | | | | Ellagic acid | 2.76E-02 | DB08846 | Follicular Lymphoma |
| | | | | | | | | Diltiazem | 2.76E-02 | DB00343 | Treat high blood pressure and to control angina (chest pain) |
| NSP8 | MPHOSPH10, XPO1, EXOSC2, EXOSC5 | 4-Hydroxytamoxifen | 8.13E-03 | DB04468 | Breast cancer | ORF8 | CANX, HSP90B1, CALR, PDIA6, | Loxapine | 1.74E-04 | DB00408 | Schizophrenia |

(*continued on next page*)

**Table 4** (*continued*)

| Virus Protein | Human hub proteins | Drugs | FDR corrected p-value | Drug Bank ID | Treatment | Virus Protein | Human hub proteins | Drugs | FDR corrected p-value | Drug Bank ID | Treatment |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | HYOU1, SEC61A1 | | | | |
| | | Pergolide | 1.53E-02 | DB01186 | Parkinson's Disease | | | Desipramine | 2.54E-04 | DB01151 | Antidepressant |
| | | Daunorubicin | 3.99E-02 | DB00694 | Acute myeloid leukemia(AML) | | | Chlorprothixene | 2.54E-04 | DB01239 | Schizophrenia |
| | | Camptothecin | 4.70E-02 | DB04690 | Leukemia | | | Clindamycin | 1.55E-03 | DB01190 | Antibiotic |
| | | | | | | | | Nilutamide | 3.19E-03 | DB00665 | Prostate cancer |
| NSP9 | NUP62, HSPA1A | Gefitinib | 4.25E-02 | DB00317 | Lung cancer. | Nucleocapsid | UPF1 | Latamoxef | 6.89E-04 | DB04570 | Antibiotic |
| | | Tanespimycin | 4.25E-02 | DB05134 | Cancer | | | | | | |
| | | Pyrvinium | 4.65E-02 | DB06816 | Tapeworm infection | | | | | | |
| NSP11 | TUBA1A, TUBA4A, TUBB1, TUBB2A, TUBB2B, TUBB4A, TUBB4B | Vinblastine | 6.18E-18 | DB00570 | Cancer | ORf9b | HSPA8, RPSA, EEF1A1, HNRNPA1 | Emetine | 1.24E-02 | DB13393 | Anticoronaviral agent |
| | | Docetaxel | 6.80E-18 | DB01248 | Cancer | | | Vorinostat | 2.80E-02 | DB02546 | Cutaneous T-cell lymphoma (CTCL) |
| | | Paclitaxel | 1.34E-16 | DB01229 | Cancer | | | Sulforaphane | 3.10E-02 | DB12422 | Autism Spectrum Disorder |
| | | Vincristine sulfate | 1.34E-16 | DB00541 | Cancer | | | Mitoxantrone | 3.42E-02 | DB01204 | Chemotherapy |
| | | Sulforaphane | 1.84E-07 | DB12422 | Prevention of prostate cancer and other types of cancer | | | | | | |
| NSP12 | PABPC1, HSPA8, NCL, PCBP1 | Disodium selenite | 3.42E-02 | DB11127 | Prevents Cancer | ORF9c | NDUFAF1, NDUFB9 | Metformin | 2.31E-02 | DB00331 | Type 2 Diabetes |
| | | Fulvestrant | 3.42E-02 | DB00947 | Breast Cancer | | | | | | |
| | | Temozolomide | 3.42E-02 | DB00853 | Brain Tumour | | | | | | |
| | | Hexachlorophene | 4.80E-02 | DB00756 | Antibacterial | | | | | | |
| Spike glycoprotein | HSPA8, RPL8, RPS3, RPSA, EEF1A1, RPLP0 | Disodium selenite | 2.45E-02 | DB11127 | Prevents Cancer | ORF10 | HSPA8, NHP2L1, HNRNPA1, RPL4, RPS3, TP53 | Disodium selenite | 9.07E-03 | DB11127 | Prevents Cancer |
| | | Vorinostat | 3.89E-02 | DB02546 | Cutaneous T-cell lymphoma (CTCL) | | | | | | |
| | | Diindolylmethane | 3.89E-02 | DB11875 | Breast, uterine, and colorectal cance | | | | | | |
| ORF3a | HYOU1, PDIA6, PDIA4, ERO1L, TXNDC5 | Geldanamycin | 2.37E-04 | DB02424 | Antiviral | | | | | | |
| | | Tanespimycin | 4.19E-04 | DB05134 | Several types of cancer, solid tumors or chronic myelogenous leukemia. | | | | | | |
| | | Emetine | 1.37E-03 | DB13393 | Anticoronaviral agent | | | | | | |
| | | Deptropine | 7.67E-03 | DB13466 | Chronic bronchitis, bronchiectasis, bronchial asthma and vasomotor rhinitis. | | | | | | |
| | | Desipramine | 7.67E-03 | DB01151 | Antidepressant | | | | | | |

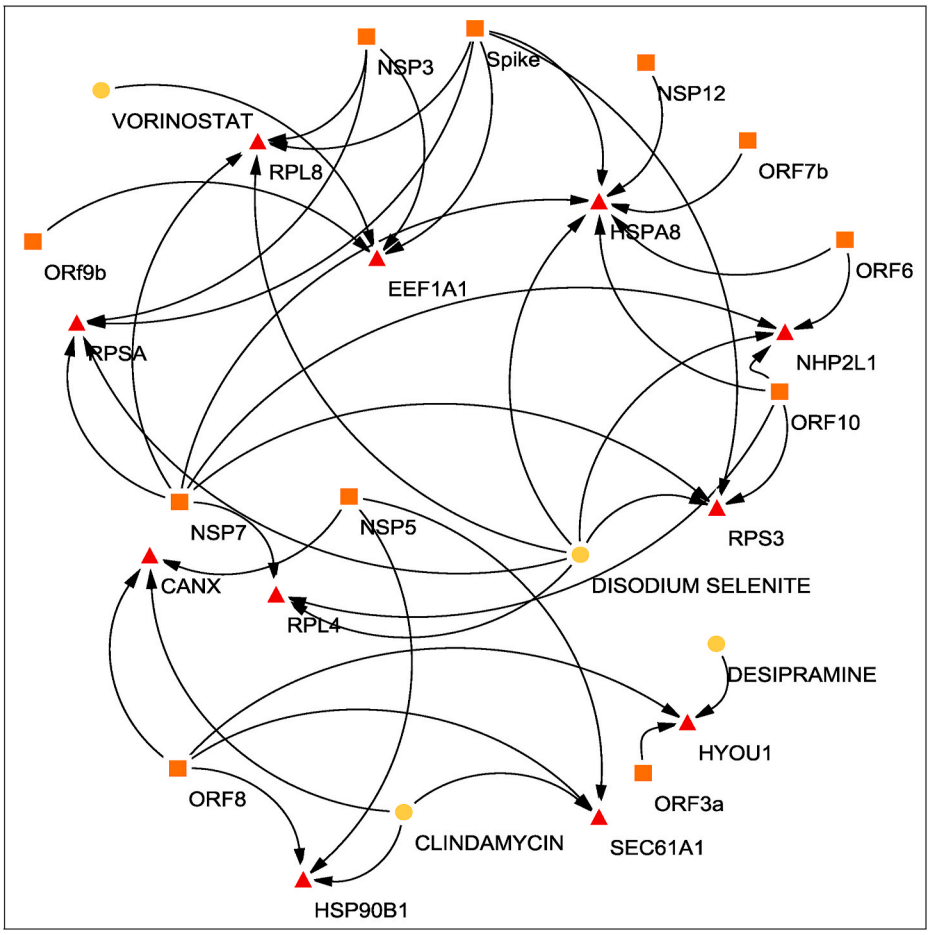


**Fig. 6.** Drug Protein Interaction Network. In this figure, rectangle represents the SARS-CoV-2 proteins, triangle represents the human hub proteins and the circle denotes the various drugs that interact with the hub proteins.

## Ethics approval and consent to participate

The ethical approval or individual consent was not applicable.

## Availability of data and materials

The supplementary of this work is available at "http://www.nittrkol.ac.in/indrajit/projects/COVID-PPI/".

## Consent for publication

Not applicable.

## Funding

This work has been partially supported by CRG short term research grant on COVID-19 (CVD/2020/000991) from Science and Engineering Research Board (SERB), Department of Science and Technology, Govt. of India.

## Author contributions

**Nimisha Ghosh**: Formal analysis; Methodology, Coding; Visualization; Writing - original draft & editing, **Indrajit Saha**: Conceptualization; Data curation; Supervision; Funding acquisition; Formal analysis; Investigation; Methodology; Project administration; Resources; Validation; Visualization; Writing - review & editing, **Nikhil Sharma**: Methodology; Visualization; Writing - review & editing.

## Declaration of competing interest

The authors declare that they have no conflict of interest.

## Acknowledgment

We are thankful to the reviewers for improving the quality of the paper by providing constructive comments.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2021.104889.

## References


[1] E. Boehm, I. Kronig, R.A. Neher, et al., Novel SARS-CoV-2 variants: the pandemics within the pandemic, Clin. Microbiol. Infect. 27 (8) (2021) 1109–1117, https://doi.org/10.1016/j.cmi.2021.05.022.

[2] F. González-Candelas, M.-A. Shaw, T. Phan, et al., One year into the pandemic: short-term evolution of SARS-CoV-2 and emergence of new lineages, Infect. Genet. Evol. 92 (2021), 104869, https://doi.org/10.1016/j.meegid.2021.104869.

[3] D. Planas, D. Veyer, A. Baidaliuk, et al., Reduced sensitivity of SARS-CoV-2 variant delta to antibody neutralization, Infection, Genetics and Evolution 596 (2021) 276–280, https://doi.org/10.1038/s41586-021-03777-9.

[4] D.E. Gordon, G.M. Jang, M. Bouhaddou, et al., A SARS-CoV-2 protein interaction map reveals targets for drug repurposing, Nature 583 (2020) 459–468, https://doi.org/10.1038/s41586-020-2286-9.

[5] B. Khorsand, A. Savadi, M. Naghibzadeh, SARS-CoV-2-human protein-protein interaction network, Informatics in Medicine Unlocked 20 (2020) 100413, https://doi.org/10.1016/j.imu.2020.100413.

[6] J. Zhao, W. Cui, B.p. Tian, The potential intermediate hosts for SARS-CoV-2, Front. Microbiol. 11 (2020) 2400, https://doi.org/10.3389/fmicb.2020.580137.

[7] K.G. Andersen, W.I.L.A. Rambaut, et al., The proximal origin of SARS-CoV-2, Nat. Med. 26 (2020) 450–452, https://doi.org/10.1038/s41591-020-0820-9.

[8] E.A. Karlsson, V. Duong, The continuing search for the origins of SARS-CoV-2, Cell 184 (17) (2021) 4373–4374, https://doi.org/10.1016/j.cell.2021.07.035.

[9] A.H.Y. Tong, B. Drees, G. Nardelli, et al., A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules, Science 295 (5553) (2002) 321–324, https://doi.org/10.1126/science.1064987.

[10] H. Zhu, M. Bilgin, R. Bangham, et al., Global analysis of protein activities using proteome chips, Science 293 (5537) (2001) 2101–2105, https://doi.org/10.1126/science.1062191.

[11] H. Ge, Z. Liu, G. Church, et al., Correlation between transcriptome and interactome mapping data from saccharomyces cerevisiae, Nat. Genet. 29 (4) (2001) 482–486, https://doi.org/10.1038/ng776.

[12] T. Bürckstümmer, K.L. Bennett, A. Preradovic, et al., An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells, Nat. Methods 3 (12) (2006) 1013–1019, https://doi.org/10.1038/nmeth968.

[13] Y. Ho, A. Gruhler, A. Heilbut, et al., Systematic identification of protein complexes in saccharomyces cerevisiae by mass spectrometry, Nature 415 (6868) (2002) 180–183, https://doi.org/10.1038/415180a.

[14] S. Fields, R. Sternglanz, The two-hybrid system: an assay for protein-protein interactions, Trends Genet. 10 (8) (1994) 286–292, https://doi.org/10.1016/0168-9525(90)90012-u.

[15] T. Ito, T. Chiba, R. Ozawa, et al., A comprehensive two-hybrid analysis to explore the yeast protein interactome, Proc. Natl. Acad. Sci. Unit. States Am. 98 (8) (2001) 4569–4574, https://doi.org/10.1073/pnas.061034498.

[16] A.H.Y. Tong, M. Evangelista, A. Parsons, et al., Systematic genetic analysis with ordered arrays of yeast deletion mutants, Science 294 (5550) (2001) 2364–2368, https://doi.org/10.1126/science.1065810.

[17] V.S. Rao, K. Srinivas, G.N. Sujini, et al., Protein-protein interaction detection: methods and analysis, International Journal of Proteomics (2014), https://doi.org/10.1155/2014/147648.

[18] X. Yang, S. Yang, Q. Li, et al., Prediction of human-virus protein-protein interactions through a sequence embedding-based machine learning method, Comput. Struct. Biotechnol. J. 18 (2020) 153–161, https://doi.org/10.1016/j.csbj.2019.12.005.

[19] A.-C. Gingras, M. Gstaiger, B. Raught, et al., Analysis of protein complexes using mass spectrometry, Nat. Rev. Mol. Cell Biol. 8 (8) (2007) 645–654, https://doi.org/10.1038/nrm2208.

[20] H. Huang, J.S. Bader, Precision and recall estimates for two-hybrid screens, Bioinformatics 25 (3) (2009) 372–378, https://doi.org/10.1093/bioinformatics/btn640.

[21] J. Piehler, New methodologies for measuring protein interactions in vivo and in vitro, Curr. Opin. Struct. Biol. 15 (1) (2005) 4–14, https://doi.org/10.1016/j.sbi.2005.01.008.

[22] I.A. Kovács, K. Luck, K. Spirohn, et al., Network-based prediction of protein interactions, Nat. Commun. 10 (1) (2019) 1–8, https://doi.org/10.1038/s41467-019-09177-y.

[23] L. Hu, J. Zhang, X. Pan, et al., HiSCF: leveraging higher-order structures for clustering analysis in biological networks, Bioinformatics (2020), https://doi.org/10.1093/bioinformatics/btaa775.

[24] M. Li, H. Gao, J. Wang, et al., Control principles for complex biological networks, Briefings Bioinf. 20 (6) (2018) 2253–2266, https://doi.org/10.1093/bib/bby088.

[25] Y.-K. Lei, Z.-H. You, Z. Ji, et al., Assessing and predicting protein interactions by combining manifold embedding with multiple information integration, BMC Bioinf. 13 (2012) 1–18.

[26] L. Hu, X. Wang, Y.-A. Huang, et al., A survey on computational models for predicting protein–protein interactions, Briefings Bioinf. (03 2021), https://doi.org/10.1093/bib/bbab036.

[27] M.R. Auwul, M.R. Rahman, E.E. Gov, et al., Bioinformatics and machine learning approach identifies potential drug targets and pathways in COVID-19, Briefings Bioinf. (04 2021), https://doi.org/10.1093/bib/bbab120.

[28] P. Gupta, D. Mohanty, SMMPPI: a machine learning-based approach for prediction of modulators of protein–protein interactions and its application for identification of novel inhibitors for RBD:hACE2 interactions in SARS-CoV-2, Briefings Bioinf. (2021), https://doi.org/10.1093/bib/bbab111, 4.

[29] X. Lian, X. Yang, S. Yang, et al., Current status and future perspectives of computational studies on human–virus protein–protein interactions, Briefings Bioinf. (03 2021), https://doi.org/10.1093/bib/bbab029.

[30] D. Ako-Adjei, W. Fu, C. Wallin, et al., HIV-1, human interaction database: current status and new features, Nucleic Acids Res. 43 (2015) 566–570, https://doi.org/10.1093/nar/gku1126.

[31] S.K. Kwofie, U. Schaefer, V.S. Sundararajan, et al., HCVpro: hepatitis C virus protein interaction database, Infect. Genet. Evol. 11 (8) (2011) 1971–1977, https://doi.org/10.1016/j.meegid.2011.09.001.

[32] P. Karyala, R. Metri, C. Bathula, et al., Denhunt - a comprehensive database of the intricate network of dengue-human interactions, PLoS Neglected Trop. Dis. 10 (9) (2016) 1–20, https://doi.org/10.1371/journal.pntd.0004965.

[33] L. Dey, A. Mukhopadhyay, Denvint: a database of protein–protein interactions between dengue virus and its hosts, PLoS Neglected Trop. Dis. 11 (10) (2017) 1–12, https://doi.org/10.1371/journal.pntd.0005879.

[34] S. Gurumayum, R. Brahma, L.D. Naorem, et al., Zikabase: an integrated ZIKV-human interactome map database, Virology 514 (2018) 203–210, https://doi.org/10.1016/j.virol.2017.11.007.

[35] H.V. Cook, N.T. Doncheva, D. Szklarczyk, et al., Viruses.STRING: a virus-host protein-protein interaction database, Viruses 10 (2018) 1–11, https://doi.org/10.3390/v10100519.

[36] A. Calderone, L. Licata, G. Cesareni, et al., Virusmentha: a new resource for virus-host protein interactions, Nucleic Acids Res. 43 (2015) 588–592, https://doi.org/10.1093/nar/gku830.

[37] S.D. Tekir, T. Çakır, E. Ardıç, et al., PHISTO: pathogen–host interaction search tool, Bioinformatics 29 (10) (2013) 1357–1358, https://doi.org/10.1093/bioinformatics/btt137.

[38] T. Guirimand, S. Delmotte, V. Navratil, VirHostNet 2.0: surfing on the web of virus/host molecular interactions data, Nucleic Acids Res. 43 (2015) 583–587, https://doi.org/10.1093/nar/gku1121.

[39] M.G. Ammari, C.R. Gresham, F.M. McCarthy, et al., Hpidb 2.0: a curated database for host-pathogen interactions, Database (03 2016) 2016, https://doi.org/10.1093/database/baw103.

[40] B. Aranda, H. Blankenburg, S. Kerrien, et al., PSICQUIC and PSISCORE: accessing and scoring molecular interactions, Nat. Methods 8 (3) (2011) 528–529, https://doi.org/10.1038/nmeth.1637.

[41] P. Krishnamoorthy, A.S. Raj, S. Roy, et al., Comparative transcriptome analysis of sars-cov, mers-cov, and sars-cov-2 to identify potential pathways for drug repurposing, Comput. Biol. Med. 128 (2021) 104123, https://doi.org/10.1016/j.compbiomed.2020.104123.

[42] D. Das, S. Podder, Unraveling the molecular crosstalk between atherosclerosis and covid-19 comorbidity, Comput. Biol. Med. 134 (2021) 104459, https://doi.org/10.1016/j.compbiomed.2021.104459.

[43] J. Li, M. Guo, X. Tian, et al., Virus-host interactome and proteomic survey reveal potential virulence factors influencing SARS-CoV-2 pathogenesis, Med 2 (1) (2021) 99–112, https://doi.org/10.1016/j.medj.2020.07.002, e7.

[44] E. Chen, C. Tan, Y. Kou, et al., Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool, BMC Bioinf. 14 (2013) 128, https://doi.org/10.1186/1471-2105-14-128.

[45] M. Kuleshov, M. Jones, A. Rouillard, et al., Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, Nucleic Acids Res. 44 (2016) W90, https://doi.org/10.1093/nar/gkw377. –W97.

[46] M. Smoot, K. Ono, J. Ruscheinski, et al., Cytoscape 2.8: new features for data integration and network visualization, Bioinformatics 27 (2011) 431–432, https://doi.org/10.1093/bioinformatics/btq675.

[47] L.D. Eizirik, A.K. Cardozo, M. Cnop, The role for endoplasmic reticulum stress in diabetes mellitus, Endocr. Rev. 29 (1) (2008) 42–61, https://doi.org/10.1210/er.2007-0015.

[48] L. Ozcan, I. Tabas, Role of endoplasmic reticulum stress in metabolic disease and other disorders, Annu. Rev. Med. 63 (2012) 317–328, https://doi.org/10.1146/annurev-med-043010-144749.

[49] S. Pasquereau, Z. Nehme, S.H. Ahmad, et al., Resveratrol inhibits HCoV-229E and SARS-CoV-2 coronavirus replication in vitro, Viruses 13 (2021) 354, https://doi.org/10.3390/v13020354.

[50] K. Kunzelmann, Getting hands on a drug for covid-19: inhaled and intranasal Niclosamide, Lancet 4 (2021) 100094, https://doi.org/10.1016/j.lanepe.2021.100094.