Computational and Structural Biotechnology Journal 20 (2022) 1244–1253

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Comprehensive characterization of human–virus protein-protein interactions reveals disease comorbidities and potential antiviral drugs

Si Li [a,1], Weiwei Zhou [b,1], Donghao Li [b,1], Tao Pan [a], Jing Guo [a], Haozhe Zou [a,b], Zhanyu Tian [a], Kongning Li [a], Juan Xu [b,*], Xia Li [a,b,*], Yongsheng Li [a,*]

[a] Key Laboratory of Tropical Translational Medicine of Ministry of Education, College of Biomedical Information and Engineering, Hainan Women and Children's Medical Center, Hainan Medical University, Haikou 571199, China
[b] College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China

ARTICLE INFO

Article history:
Received 7 December 2021
Received in revised form 4 March 2022
Accepted 4 March 2022
Available online 7 March 2022

Keywords:
Protein-protein interactions
Network analysis
Antiviral therapy
HVPPI
Disease comorbidities

ABSTRACT

The protein-protein interactions (PPIs) between human and viruses play important roles in viral infection and host immune responses. Rapid accumulation of experimentally validated human–virus PPIs provides an unprecedented opportunity to investigate the regulatory pattern of viral infection. However, we are still lack of knowledge about the regulatory patterns of human–virus interactions. We collected 27,293 experimentally validated human–virus PPIs, covering 8 virus families, 140 viral proteins and 6059 human proteins. Functional enrichment analysis revealed that the viral interacting proteins were likely to be enriched in cell cycle and immune-related pathways. Moreover, we analysed the topological features of the viral interacting proteins and found that they were likely to locate in central regions of human PPI network. Based on network proximity analyses of diseases genes and human–virus interactions in the human interactome, we revealed the associations between complex diseases and viral infections. Network analysis also implicated potential antiviral drugs that were further validated by text mining. Finally, we presented the Human–Virus Protein-Protein Interaction database (HVPPI, http://bio-bigdata.hrbmu.edu.cn/HVPPI), that provides experimentally validated human–virus PPIs as well as seamlessly integrates online functional analysis tools. In summary, comprehensive understanding the regulatory pattern of human–virus interactome will provide novel insights into fundamental infectious mechanism discovery and new antiviral therapy development.

1. Introduction

Viral infectious diseases have become a major public health concern, as can be seen from the current status of the coronavirus disease 2019 global pandemic. Viruses can only reproduce themselves by attaching to and entering host cells, such as human cells, causing a number of diseases. To carry out these infective processes, viruses often interact with host proteins to hijack the host [1,2]. In this regard, knowledge of host–virus protein–protein interactions (PPIs) is critical for understanding the mechanisms of viral infections and host immune response.

Rapid identification of human–virus PPIs can shed light on potential treatment options for human infectious diseases [3,4]. Considerable efforts have been made to investigate the human–virus PPIs by multiple experimental methods, such as yeast two-hybrid assays (Y2H) and affinity purification [5]. In particular, Gordon et al. cloned, tagged and expressed 26 SARS-CoV-2 proteins individually and used mass spectrometry to measure human–virus PPIs [6]. In total, they identified 332 interactions between viral and host proteins. Further studies of these human–virus PPIs have identified druggable human proteins. Han et al. identified candidate drugs for SARS-CoV-2 by construction of a SARS-CoV-2–induced protein (SIP) network [7]. In addition, human–virus PPIs based on Y2H or affinity purification mass spectrometry (AP-MS) have been examined on several types of viruses, such as Epstein–Barr virus (EBV) [8], dengue virus (DENV) [9] and Zika virus (ZIKV) [10]. Although these experimentally validated human–virus PPIs have yielded critical insights into viral infection, limited scalability and time consumption have hampered this field. For this reason, a

* Corresponding authors at: Key Laboratory of Tropical Translational Medicine of Ministry of Education, College of Biomedical Information and Engineering, Hainan Women and Children's Medical Center, Hainan Medical University, Haikou 571199, China.
E-mail addresses: xujuanbiocc@ems.hrbmu.edu.cn (J. Xu), lixia@hrbmu.edu.cn (X. Li), liyongsheng@hainmc.edu.cn (Y. Li).
[1] Si Li, Weiwei Zhou and Donghao Li contributed equally to this work.

https://doi.org/10.1016/j.csbj.2022.03.002

number of computational methods have been proposed to identify human–viral PPIs. For example, P-HIPSTer is a useful method to predict human–viral PPIs based on structural information [11]. LSTM-PHV is another method, which combines the long short-term memory (LSTM) model with the word2vec embedding method [12]. However, these experimentally or computationally predicted PPIs are scattered in the literature, which limits the use of the valuable resource.

To solve this problem, several databases have been established to manage human–virus PPIs. These databases cover PPIs from single viral species to pan-viral species. For example, HCVpro is an HCV-specific database for human–HCV PPIs [13]. DenHunt and DenvInt are databases for PPIs between human and dengue virus [14,15]. ZikaBase is an integrated database for the ZIKV–human interactome map [16]. VirHostNet [17], VirusMentha [18] and VirusMINT [19] are useful platforms for host–virus PPIs of multiple types of viruses. Recently, HVIDB has been proposed to manage human–virus PPIs by collecting experimental and predicted PPIs [20]. This is the most comprehensive database currently. However, the current human–virus databases still need to be improved. First, the majority of these databases have been designed for specific species; otherwise they do not provide useful tools to analyze the PPI data. Second, these databases do not provide any drug-related information to facilitate further development of antiviral treatment. Finally, although these human–virus PPIs data were collected, they have not been analyzed comprehensively. We still lack knowledge about the regulatory patterns of human–virus interactions.

Here, we reported a systematic interrogation of human–virus PPIs collected from the literature and constructed the comprehensive database called HVPPI (https://bio-bigdata.hrbmu.edu.cn/HVPPI/). The current version of HVPPI includes 27,293 human–virus PPIs with diverse experimental supporting evidence, covering 140 viral proteins involving eight virus families and 6059 human proteins. Furthermore, we integrated a number of computational tools to visualize the PPIs or predict the functions of viral proteins. Moreover, we revealed diverse regulatory patterns for virus targeting and the disease-disease associations based on the data in HVPPI. Finally, candidate drugs were identified for antiviral therapy based on the network analysis.

2. Materials and methods

2.1. Collection of human–virus protein–protein interactions

We first collected the human–virus PPIs from the published literature (ending to June 2021). Several review or database papers were first manually curated by three independent researchers. Moreover, we searched the PubMed and added other literature related to human–virus PPIs. The full-text versions were read and the detailed information about the PPIs were collected, including the virus names, virus proteins, human proteins, database resource, experimental method for detecting the PPIs, and PMID. Because numbers of protein–protein interactions collected for InfluV were not provided the exact strains in literature, thus we used the InfluV (including IAV, IBV and ICV) in analysis. In addition, we found that numerous PPIs were exactly identified in H1N1; we thus included these PPIs in H1N1 category. But for the DENV, the majority of the literature did not provide the exact strains; we thus used the DENV in analysis. In total, more than 200 articles were read. All this information was integrated into HVPPI resource.

2.2. Function analysis of viral proteins

To identify the functions enriched by viral-interacting proteins, we used clusterProfiler to perform function enrichment analysis [21]. Gene ontology (GO) biological processes were considered in our analysis. We only considered the GO terms with genes ranging from 15 to 500. The biological processes with a p value <0.01 and p-adjusted value <0.05 were considered significant. Next, all the GO terms were clustered based on simplifyEnrichment R package [22]. The similarities among GO terms were calculated by the 'GO_similarity' function and the cluster results were visualized by 'simplifyGO' function.

Moreover, we obtained the immune-related pathways from one of our recent studies [23]. The hypergeometric test was used to evaluate whether the interacting proteins of a virus protein were enriched in immune-related pathways. Pathways with a p-adjusted value <0.05 were considered significant. The enrichment results were visualized by Circos plot [24].

2.3. Topological features of proteins

The topological features of proteins were calculated based on igraph package (https://igraph.org/). Here, human PPIs were obtained from PCNet [25]. The human PPIs included 2,724,723 interactions among 19,779 proteins. Degree, betweenness, and closeness of each protein were calculated. We next compared the topological features between viral-interacting proteins and other proteins by Wilcoxon's rank-sum test.

2.4. Disease associations based on PPIs

To evaluate the associations between viral infection and other human complex diseases, we first collected the disease-related genes for 299 diseases from a recent article [26]. There were 3173 genes for 299 human complex diseases and these diseases were clustered into 10 clusters [27]. We next evaluated the network-based overlap between the disease proteins and host protein targets of each virus [28]. We first calculated the S_{vb} metric, where $S_{vb} < 0$ suggests a network-based overlap between the viral targets v and genes associated with disease b. S_{vb} was calculated as follows:.

$$S_{vb} = d_{vb} - \frac{d_{vv} + d_{bb}}{2}$$

where S_{vb} compares the shortest distances between proteins within viral interacting proteins d_{vv} or diseases proteins d_{bb}, to the shortest distances d_{vb} between viral interacting proteins and disease proteins. Proteins associated to both conditions have $d_{vb} = 0$.

2.5. Prioritization of potential antivirus drugs based on PPIs

Human PPIs have provided a rational method for predicting the potential drugs or small molecules for treatment of viral infections. We prioritized the candidate drugs based on the network-based method [3]. Given V, the host gene encoded proteins that interacted with a specific virus, and T, the targets of a drug of interest, we calculated the network proximity measure of V with the target proteins T of each candidate drug as follows:.

$$d_{VT} = \frac{1}{||V|| + ||T||}\left(\sum_{i \in V} min_{j \in T} d(i,j) + \sum_{j \in T} min_{i \in V} d(i,j)\right)$$

where $d(i,j)$ is the shortest distance between protein i and j in the human PPI network. Next, the network proximity measure was further transformed into Z-score based on the permutation tests:.

$$Z_{d_{VT}} = \frac{d_{VT} - \bar{d}_r}{\sigma_r}$$

where $\bar{d}_r$ and σ_r are the mean network proximity measures and standard deviation of 1000 times permutation, respectively. We

randomly selected gene-encoded proteins from the whole human proteome, which were with similar degree distributions to V and T for calculating the d_{VT}. The p value was calculated as the proportion of random conditions that were with lower d_{VT} than observed. Drugs with Z-score < −1.5 and p value < 0.001 were identified as potential antivirus drugs.

2.6. Construction of HVPPI resource

All the PPI data in HVPPI were stored and managed using the MySQL (version 5.5.21). The HVPPI web interface was built in JavaServer Pages (JSP). The data processing programs in HVPPI were written in Java (version 1.7.0_80), and the web services were built based on Apache Tomcat. The web interface was developed using HTML5 with JavaScript. The HVPPI database is freely accessible at https://bio-bigdata.hrbmu.edu.cn/HVPPI/.

3. Results

3.1. Comprehensively curated human–virus protein–protein interactions resource

To collect the experimentally validated human–virus PPIs, we manually read more than 200 published articles. The details of the interactions, including viral protein, human protein, experimental methods, and PubMed ID were all collected (Fig. 1A). In total, we obtained 27,293 human–virus PPIs with diverse experimental supporting evidence, involving eight virus families (including 13 viruses), 140 viral proteins, and 6059 human proteins. All

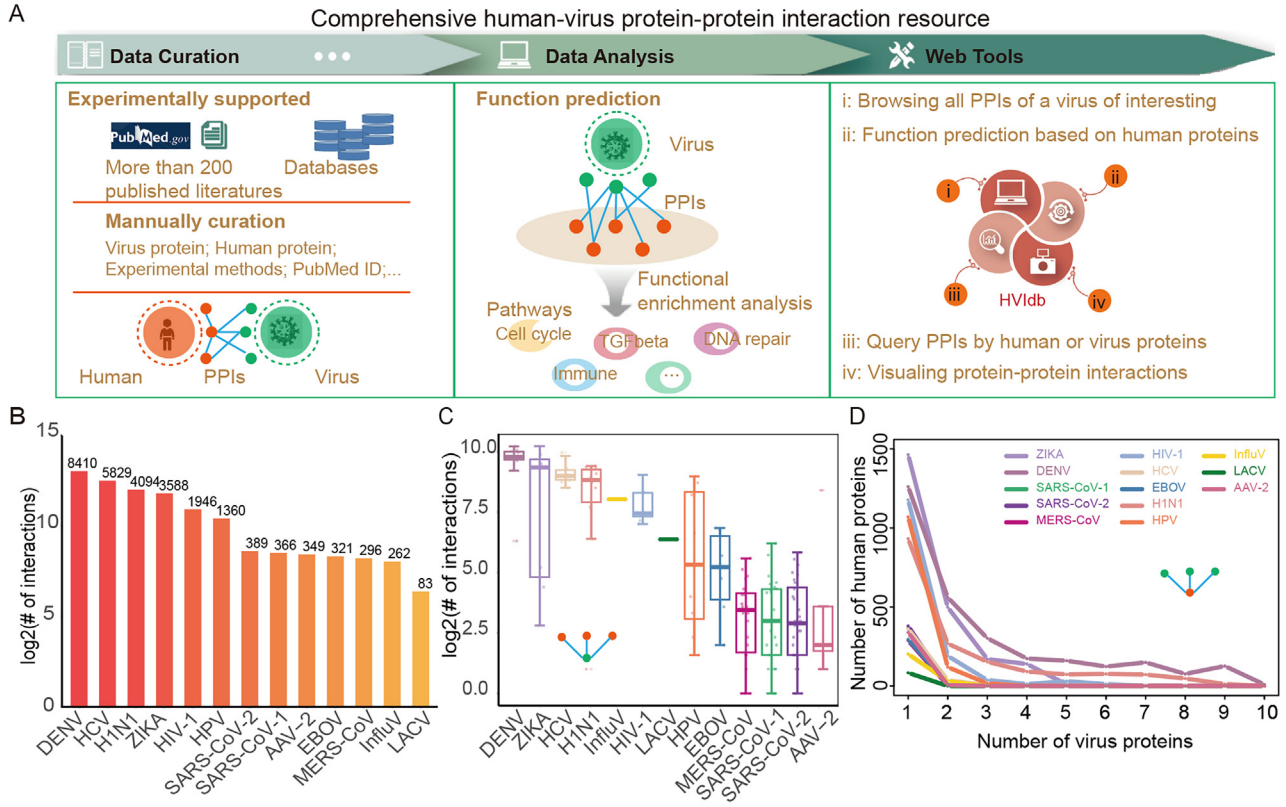


Fig. 1. Overview of human–virus protein–protein interaction. A. Construction of the HVPPI resource for human–virus protein–protein interactions. B. Number of PPIs across 13 viruses. C. Box plots showing the number of interacting human proteins for each viral protein. D. Line charts showing the number of human proteins that interact with a different number of viral proteins across viruses.

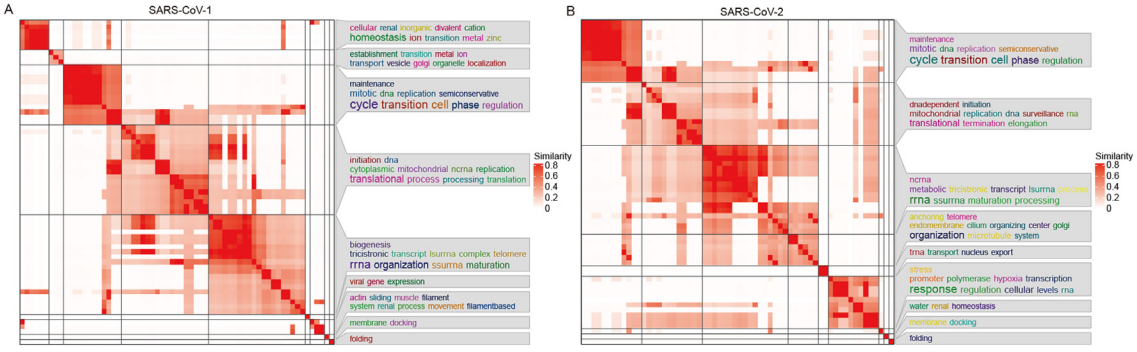


Fig. 2. Function enrichment of viral-interacting human proteins. (A) for SARS-CoV-1 and (B) for SARS-CoV-2. GO terms were clustered into groups based on gene similarities.

these interactions were integrated into a web-based resource named HVPPI. Users can browse and query of viral proteins or human proteins to get the detailed interactions in this platform (Fig. 1A). We also provided a number of useful tools for predicting the function of viral proteins based on the interacting human proteins.

Next, we counted the number of interactions in each virus. A higher number of interactions were identified in DENV, HCV, H1N1, and ZIKA, and the number of interactions was the highest in DENV (Fig. 1B). In particular, 389 interactions were validated in SARS-CoV-2. Given that there are different numbers of proteins in viruses, we also calculated the number of interacting human proteins for each viral protein. We found that DENV, ZIKA, HCV, and H1N1 proteins were likely to interact with a higher number of human proteins (Fig. 1C). Moreover, we found that proteins in several viruses, such as ZIKA, HPV and EBOV, interacted with different numbers of human proteins, while proteins in DENV, HCV, and HIV interacted with a similar number of human proteins (Fig. 1C). We calculated the number of viral proteins interacting with each human protein. The majority of human proteins only interacted with one viral protein across viruses (Fig. 1D); however, we also identified that some human proteins can interact with more than seven DENV proteins. Function analysis revealed that these genes were significantly enriched in 'viral gene expression', 'translational initiation' and 'RNA catabolic process' (Figure S1). These results suggest that multiple viral proteins might co-regulate the RNA synthesis-related pathways to facilitate the virus replication in host.

3.2. Viral proteins widely interact with cell cycle and immune-related genes

To identify the functions of these viral proteins, we next characterized the function of human proteins interacting with viral proteins. We first performed functional enrichment analysis for each virus and found that these virus-interacting proteins (VIPs) were significantly enriched in cell cycle and immune response-related functions. For example, the majority of SARS-CoV-1-interacting human proteins were significantly enriched in cell cycle-related functions (Fig. 2A). Particularly, four interacting proteins of nsp1 in SARS-CoV-1 were all annotated in 'DNA replication', 'G1/S transition of mitotic cell cycle' and 'telomere maintenance' (Table S1). Moreover, SARS-CoV-2-interacting proteins were also significantly enriched in cell cycle and DNA replication-related functions (Fig. 2B). For example, four of six nsp1-interacting genes were significantly enriched in 'DNA replication initiation', 'cell cycle G1/S phase transition' and 'telomere maintenance' (Table S1).

Moreover, we also found that viral-interacting proteins were significantly enriched in a number of immune-related functions, such as 'regulation of type I interferon production' in DENV and ZIKA, and 'regulation of innate immune response' in HCV and HPV (Figure S2). Thus, we next focused on 17 immune-related pathways collected in the literature and performed functional enrichment analysis for viral-interacting proteins. We found that six types of virus-interacting proteins were enriched in at least one immune-related pathway (Fig. 3A). Particularly, the majority of proteins were significantly enriched in 'antigen processing and presentation', 'antimicrobials' and 'TCR (T cell receptor) signaling pathway'. For example, human proteins that interact with eight DNEV proteins were significantly enriched in antigen processing and presentation pathway (Fig. 3B). Antigen presentation is closely associated with vital immune process, which is essential for triggering the T cell immune response [29]. Several critical genes, such as CD8A (CD8a Molecule), IFNG (Interferon Gamma), CD4 (CD4 Molecule), and CD74 (CD74 Molecule), were found to interact with viral proteins and play important roles in immune regulation (Fig. 3B). In addition, antibiotics have commonly been used to treat

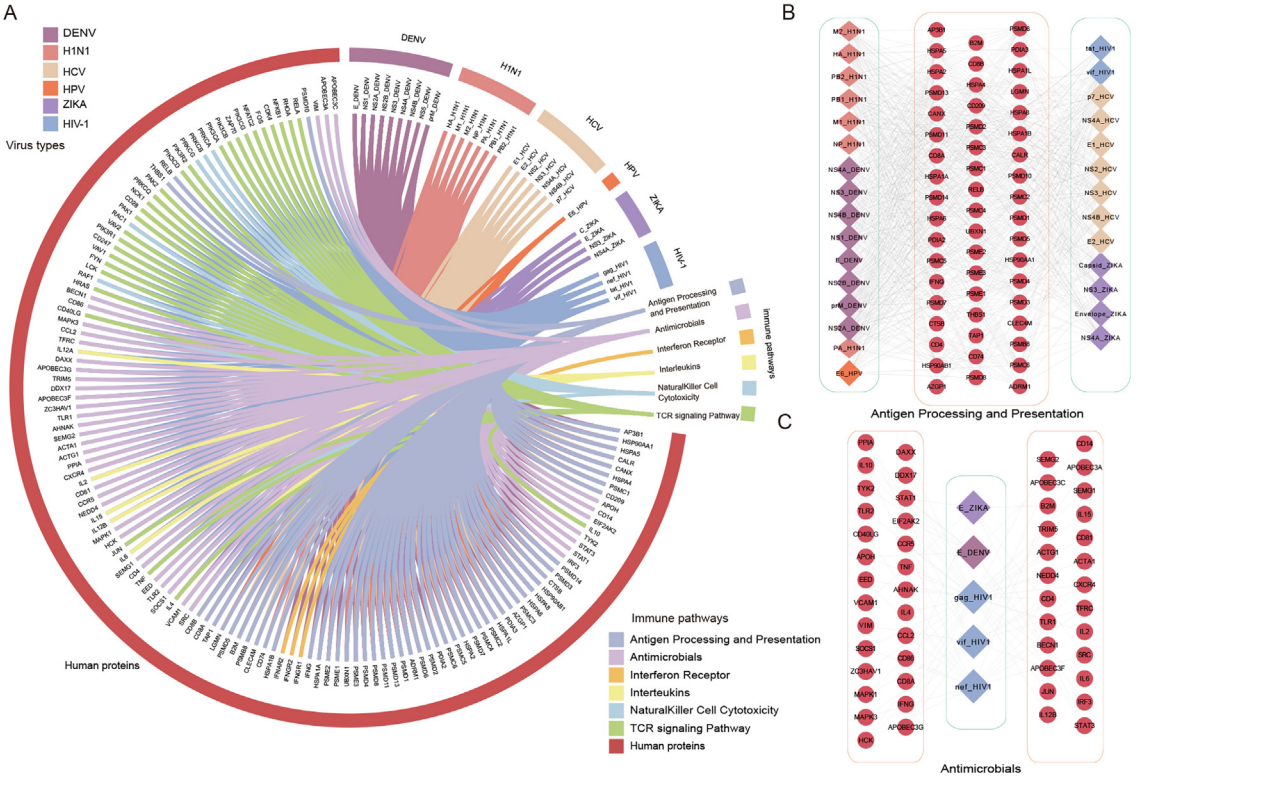


Fig. 3. Virus-interacting proteins enriched in immune-related pathways. A. Circos plot showing the viral-interacting proteins annotated in different immune-related pathways. B. Network visualization of viral–human protein–protein interactions in antigen processing and presentation pathway. C. Network visualization of viral–human protein–protein interactions in antimicrobials pathway.

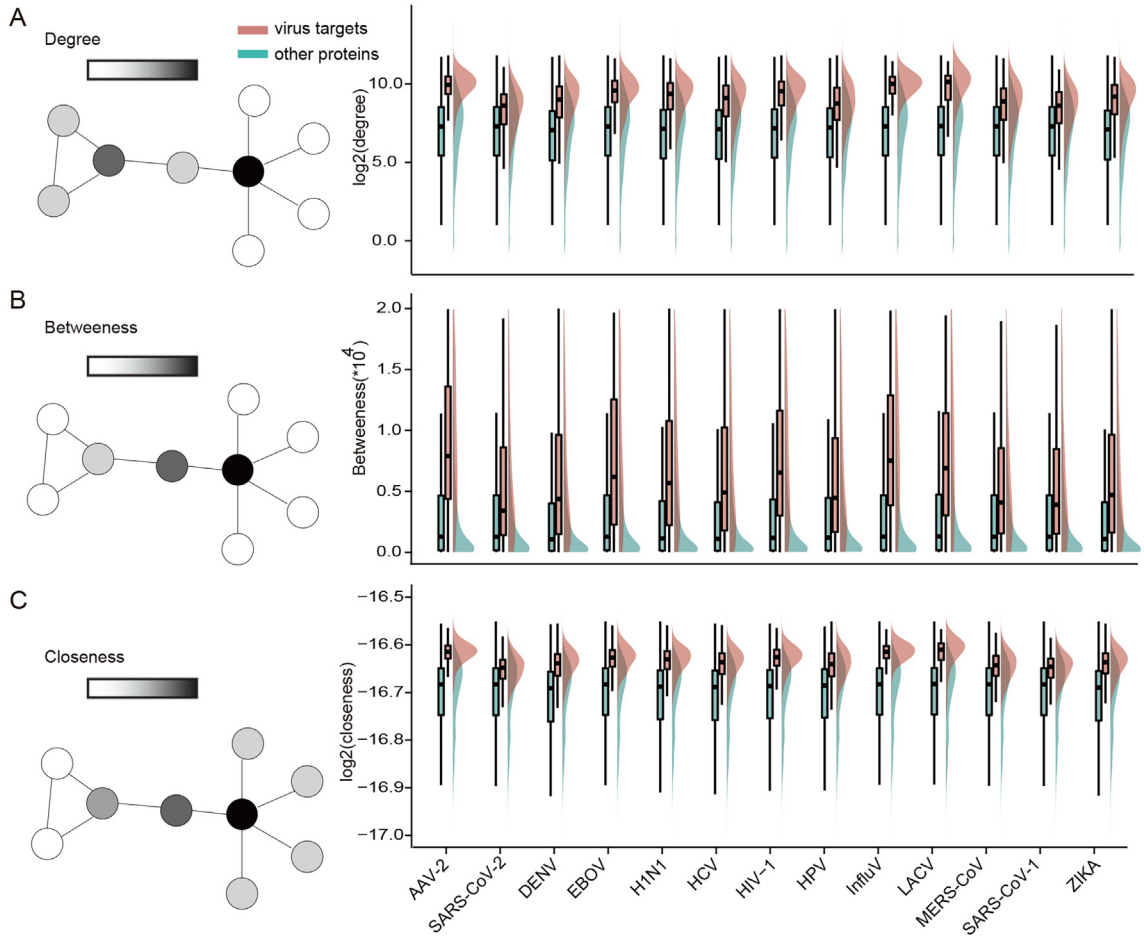


Fig. 4. Topological features of viral-interacting proteins in human PPIs. A. Degree distributions of viral interacting proteins and other proteins. B. Betweenness distributions of viral interacting proteins and other proteins. C. Closeness distributions of viral interacting proteins and other proteins.

patients with acute viral infection in hospital [30]. We found that several HIV, DENV, and ZIKA proteins interacting with a number of human proteins were enriched in antimicrobials pathway (Fig. 3C). These genes included CCR5 (C–C Motif Chemokine Receptor 5), IL10 (Interleukin 10), PPIA (Peptidylprolyl Isomerase A) and B2M (Beta-2-Microglobulin), which have been demonstrated to play important roles in immunology [31,32]. Taken together, these results suggest that the global virus–human map for various types of viruses can provide a more comprehensive view of the targeting pathways involved in viral pathogenesis.

3.3. Viral proteins are likely to interact with central proteins in PPI network

The locations of proteins in human PPI network represent their functional importance [33–35]. Thus, we next investigated the location of viral targeting proteins in the context of human PPI network. Degree is one of the most important topological features of a network and indicates local centrality of protein in the network (Fig. 4A). The degree is defined as the number of interacting proteins. We divided all the proteins in PPI network into two groups: viral-targeting proteins and other proteins. We found that viral-targeting proteins have significantly higher degrees in PPI network than other proteins (Fig. 4A, all p-values < 0.01, Wilcoxon rank-sum tests). Moreover, betweenness centrality is another widely used measure for centrality in a network based on shortest paths [36]. The betweenness centrality of each protein is defined as the number of shortest paths that pass through the protein in the network. We next compared the betweenness of viral targeting proteins with others and found that they were likely to have higher betweenness across all viruses (Fig. 4B, all p-values < 0.01, Wilcoxon rank-sum tests). Closeness is a way of detecting proteins that are able to spread biological information very efficiently through the PPIs. Protein with a high closeness score has the shortest paths to all other proteins. When we compared the closeness of viral targeting proteins with other proteins, we found that they were also likely to have higher closeness scores (Fig. 4C, all p values < 0.01, Wilcoxon rank-sum tests). Taken together, all these results suggest that viruses are likely to interact with proteins that are located in the center of the network or with those that play important roles in information spread.

3.4. Human–virus protein–protein interactions reveal disease associations

Emerging evidence has revealed complications and comorbidities of viral infections [28,37]. We thus systematically evaluated the disease associations with viral infection based on the state-of-art network proximity measure. We found that central nervous system diseases, brain diseases, and neurodegenerative diseases frequently showed close association with SARS-CoV-1 (Fig. 5A). In addition, SARS-CoV-2 infection showed significant association with metabolic diseases, central nervous system diseases, and congenital diseases (Fig. 5B). Network visualization can further show the associations between viral infection and other diseases in detail. For instance, we found that several SARS-CoV-1-

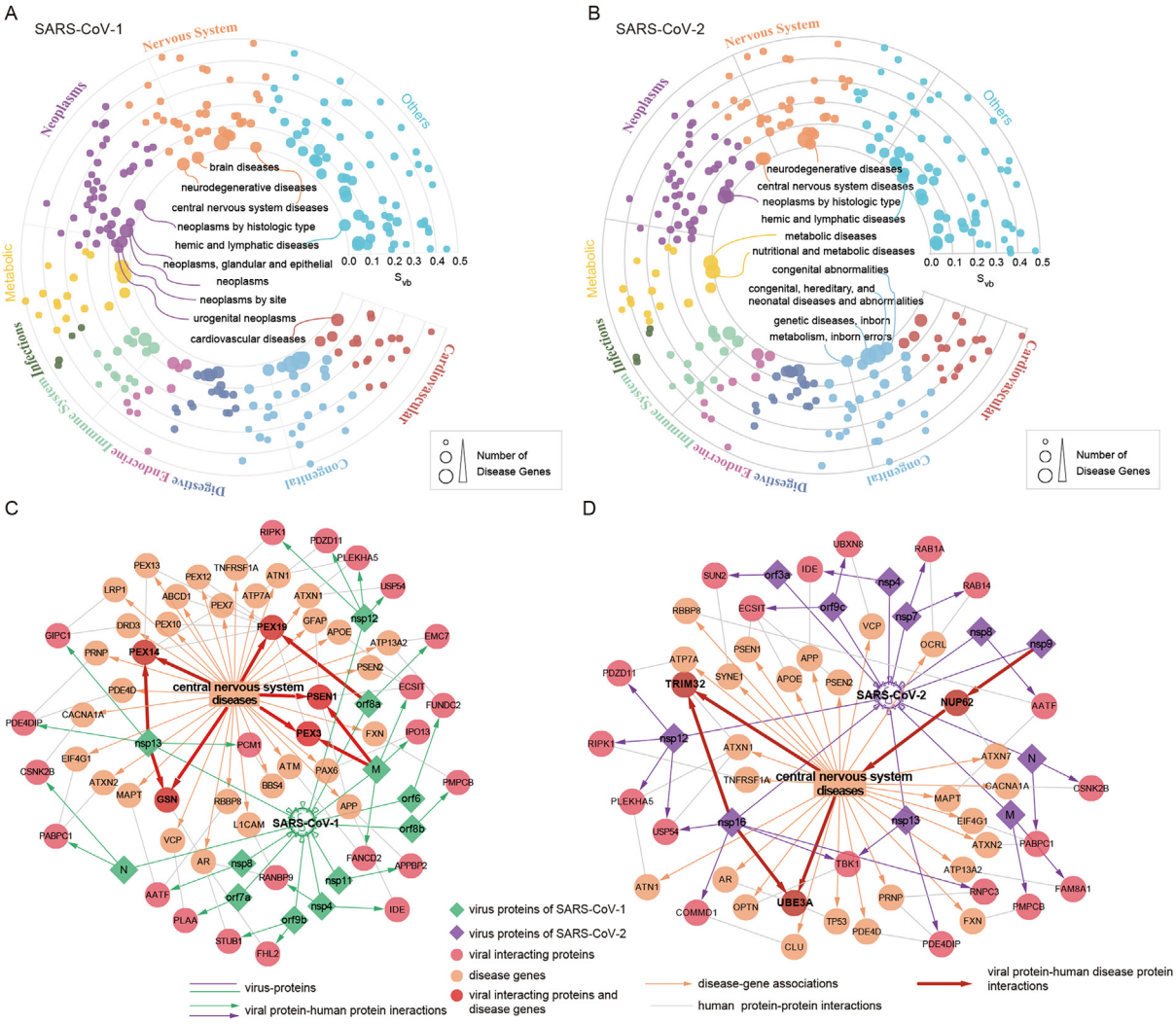


Fig. 5. Association of viral infections and human complex diseases. A. Disease comorbidity measured by the network overlap between SARS-CoV-1 targets and 299 diseases. The dots represent diseases whose radius reflects the number of associated diseases genes. The diseases closest to the center, whose names are marked, are expected to have higher comorbidity with viral infection. (B) for SARS-CoV-2. C. Network visualization showing the protein–protein interactions among viral-interacting proteins and diseases-associated proteins. (C) for SARS-CoV-1 and central nervous systems disease and (D) for SARS-CoV-2 and central nervous systems disease.

interacting proteins, such as PEX14 (Peroxisomal Biogenesis Factor 14), PEX3 (Peroxisomal Biogenesis Factor 3), PSEN1 (Presenilin 1) and GSN (Gelsolin) directly interact with the central nervous system disease-associated proteins (Fig. 5C). PSEN1 is a protein-coding gene that is associated with Alzheimer disease [38]. M protein of SARS-CoV-1 can directly interact with PSEN1, suggesting that PSEN1 plays important roles in mediating viral infection-related central nervous system diseases.

Moreover, there is emerging evidence that SARS-CoV-2 can cause neurological complications [39,40]. A recent study has found that approximately 36% of SARS-CoV-2 infection patients from China had neurological manifestations [41]. Network visualization showed that the SARS-CoV-2-interacting proteins can significantly interact with central nervous system diseases-associated proteins (Fig. 5D). In particular, nsp16 protein directly interacts with disease proteins, tripartite motif-containing 32 (TRIM32) and ubiquitin protein ligase E3A (UBE3A). Previous studies have demonstrated that TRIM32 plays important roles in cancer and antiviral immunity processes [42,43]. UBE3A is a gene responsible for the pathogenesis of intellectual disability, delayed development and severe speech impairment [44]. In addition, UBE3A is also known as an important regulator of the immune system in the brain tissue. These results indicate that UBE3A plays a critical role in SARS-CoV-2 associated central nervous system diseases. Moreover, we also revealed the close association between SARS-CoV-1- or SARS-CoV-2-associated brain diseases and neurodegenerative diseases (Figure S3). All these results suggest that the comprehensive interactome map helps to understand the association between viral infections and other human complex diseases.

3.5. Network analyses identify potential drugs for antiviral therapy

Knowledge of the complex interactions between viral-interacting proteins and human diseases-associated proteins implies possibilities of predicting novel drugs. For example, the drugs or small molecules that target other human diseases could potentially target viral infection via the shared human PPI networks. Thus, we developed a network-based model for predicting the potential antiviral drugs based on the known drug–target network and the manually curated human–virus interactome collected in this study (Fig. 6A). First, Food and Drug Administration (FDA)-approved drug–target interactions were downloaded from DrugBank [45], including 2170 drugs and 2694 targets. Based on network analysis, we prioritized candidate drugs in each virus (Fig. 6B). In particular, we identified 46 candidate drugs for treatment of SARS-CoV-2, including 45 small molecule drugs and one

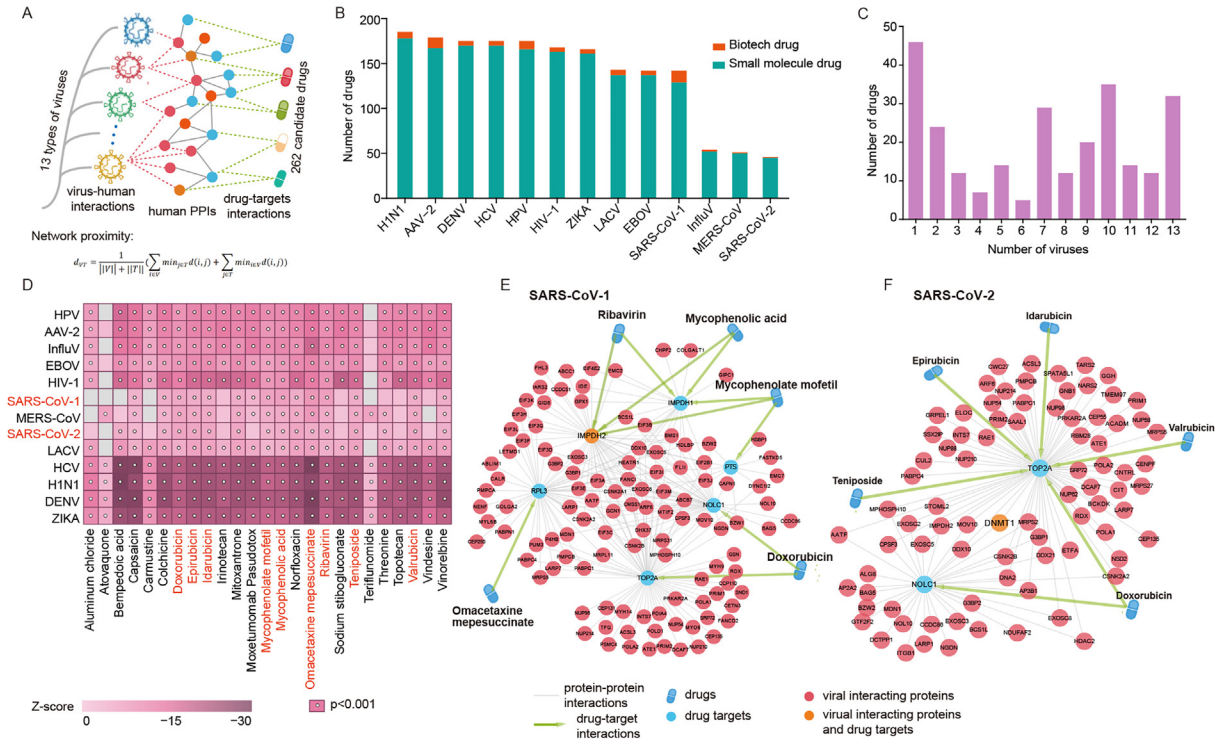


Fig. 6. Prioritization of potential antiviral drugs. A. Network-based method for prioritization of potential drugs. B. Number of potential drugs prioritized across viruses. C. Number of viruses that can potentially be targeted by drugs. D. Heat map showing the Z-scores of different drugs across viruses. (E) and (F). Network visualization of protein–protein interactions among viral-interacting proteins and drug targets. (E) for SARS-CoV-1 and (F) for SARS-CoV-2.

biotech drug (Table S2). We found that 28/46 (60.87%) drugs supported by the literature.

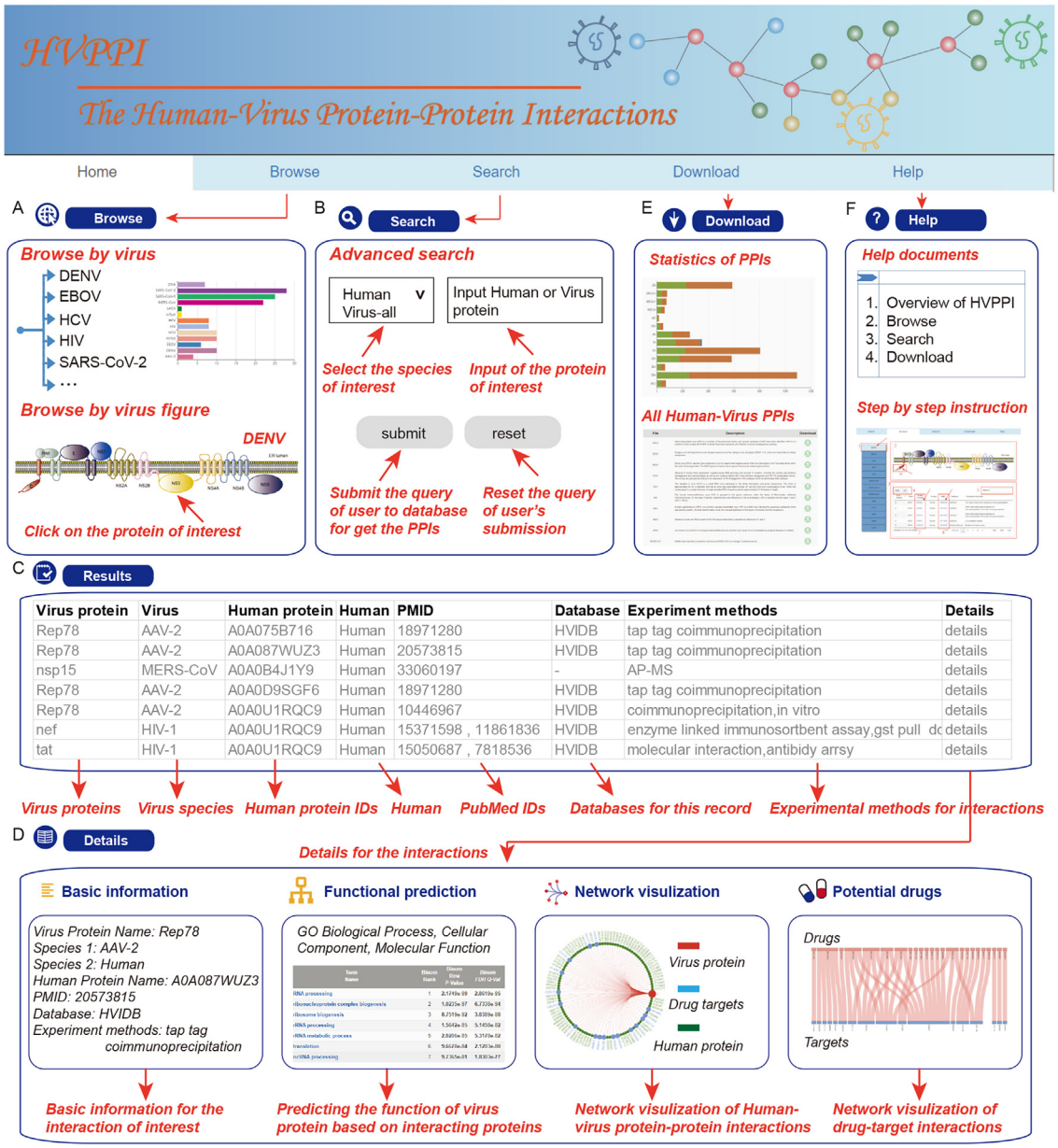
Next, we calculated the number of viruses for each drug that can be used for treatment. We found that most of the prioritized antiviral drugs were effective against a small number of viruses (Fig. 6C), which is consistent with previous observation [46]. In contrast, we identified 32 candidate drugs in 13 viruses that could be used to develop broad-spectrum antiviral agents. In addition, we found that the top five candidate drugs in each virus showed consistent z-scores across viruses. For example, epirubicin, norfloxacin and ribavirin showed significantly larger z-scores across all viruses (Fig. 6D), and there is evidences for their broad-spectrum antiviral effects [47,48]. Next, we analyzed the network of two coronaviruses in detail. For SARS-CoV-1, the top five identified drugs were ribavirin, mycophenolic acid, mycophenolate mofetil, omacetaxine mepesuccinate and doxorubicin (Fig. 6E). In particular, Inosine Monophosphate Dehydrogenase 2 (IMPDH2) interacts with viral proteins and can be targeted by candidate drugs. It has been demonstrated that IMPDH2 knockdown or chemical inhibition using ribavirin and mycophenolic acid abolishes NF-κB activation and cytokine induction [49]. IMDPH2 inhibitors efficiently block coronaviruses infection [49]. Moreover, valrubicin, idarubicin, epirubicin, teniposide, and doxorubicin were identified as top five drugs for SARS-CoV-2, which were all targeting DNA Topoisomerase II Alpha (TOP2A) (Fig. 6F). TOP2A has been found to play important roles in proliferation, clone formation and invasion of virus [50]. Thus, TOP2A could be used for the development of therapeutic intervention. Moreover, DNA Methyltransferase 1 (DNMT1) was identified to interact with viral proteins and can be targeted by drugs. It has been demonstrated that SARS-CoV-2 infection significantly reduced the levels of DNA methyltransferases [51]. Taken together, these results suggest that network-based analysis offers a powerful method for rapid identification of candidate antiviral drugs.

3.6. HVPPI: a web-based resource for Human–Virus Protein-Protein interactions

In order to facilitate the analysis of human–virus PPIs, we constructed the web-based resource, HVPPI (https://bio-bigdata.hrbmu.edu.cn/HVPPI). Currently, HVPPI provides 27,293 human–virus PPIs with diverse experimental supporting evidence, involving 140 viral proteins from eight virus families and 6059 human proteins. HVPPI provides multiple browsing/querying modules, allowing users to easily access the interactome of proteins of interest (Fig. 7). Users can browse through virus names or they can click the proteins in virus genome structure figure (Fig. 7A). In addition, entries can be searched by selecting the species, inputting the names or IDs of proteins, and clicking the 'submit' button for querying the PPIs (Fig. 7B). HVPPI provides the PPIs about the queried proteins, including viral proteins, virus names, human proteins, PMID, database resources, experimental methods and details (Fig. 7C). In the 'Detail' pages, the basic information for the PPIs and several functional modules are provided. The users can predict the function of viral proteins by their interactions with human proteins. Protein–protein interactions and drug–target interactions are visualized by network (Fig. 7D). Users can download all PPIs from the Download page for local use and the Help page provides a step-by-step instruction on the usage of HVPPI (Fig. 7E-F).

4. Discussion

In this study, we comprehensively collected the human–virus PPIs and analyzed the regulatory patterns of interacting with viruses. We found that viral interacting proteins were significantly enriched in cell cycle and immune-related functions. Viral proteins were likely to interact with human proteins located in the central regions of the PPI network. Based on network analysis, we also



Virus protein	Virus	Human protein	Human	PMID	Database	Experiment methods	Details
Rep78	AAV-2	A0A075B716	Human	18971280	HVIDB	tap tag coimmunoprecipitation	details
Rep78	AAV-2	A0A087WUZ3	Human	20573815	HVIDB	tap tag coimmunoprecipitation	details
nsp15	MERS-CoV	A0A0B4J1Y9	Human	33060197	-	AP-MS	details
Rep78	AAV-2	A0A0D9SGF6	Human	18971280	HVIDB	tap tag coimmunoprecipitation	details
Rep78	AAV-2	A0A0U1RQC9	Human	10446967	HVIDB	coimmunoprecipitation,in vitro	details
nef	HIV-1	A0A0U1RQC9	Human	15371598 , 11861836	HVIDB	enzyme linked immunosortbent assay,gst pull dc	details
tat	HIV-1	A0A0U1RQC9	Human	15050687 , 7818536	HVIDB	molecular interaction,antibidy arrsy	details

Fig. 7. Usage of HVPPI resource for human–virus PPIs. A. Users can browse the PPIs in different viruses by virus names or by clicking on the virus figure. B. Users can query of PPIs by species of interest, and protein of interest. C. The results page for human–virus PPIs. D. Detail pages for human–virus PPIs, including basic information, function prediction, network visualization, and potential drugs. E. Download pages of HVPPI. F. Help pages of HVPPI.

revealed the associations between viral infection and human complex diseases, and we prioritized several candidate drugs.

Recent studies have implied that SARS-CoV-2 infection can affect multiple organs in addition to lungs [52]. Emerging studies have also revealed that the disease-related genes are likely to show tissue-specific expression patterns [23,53]. We next investigated the association between viral-interacting proteins and tissue-specific expressed proteins. We observed that SARS-CoV-1-interacting proteins were likely to interact with genes specifically expressed in adrenal gland, heart, liver, muscle and brain (Figure S4). In addition, SARS-CoV-2-interacting proteins were likely to interact with genes specifically expressed in liver, brain and heart (Figure S4). These results are consistent with previous observations [54].

There were several potential limitations of this study. First, although we collected PPI data from databases and the literature to construct the human–virus interactome, they are still incomplete. Moreover, for specific virus families, limited papers have been taken into account. Considering the increasing interest of the research community on virus infection, we further efforts will be devoted to identification of HVPPIs in the future. We will continually update the web resource for collecting the comprehensive HVPPIs. Second, different datasets from multiple sources may be different in quality and bias, which may have influenced the revealed regulatory patterns in our analysis. Moreover, our network analysis can only be applied to diseases with known associated genes and not to diseases lacking disease-related genes, such as rare diseases. We also identified several candidate antiviral drugs; these drug candidates need to be validated using experimental methods and clinical trials with larger number of patients before recommendation for the use in patients with viral infection.

In summary, our study provides a comprehensive, integrative network pipeline for understanding viral infection and reveals the regulatory patterns of viruses. The comprehensive human–

virus interactome helps with the prediction of drug candidates for antiviral therapy. The network-based medicine strategy applied in this study could provide novel insights for developing effective treatment strategies for emerging infectious diseases.

Funding

This work was supported by Hainan Province Science and Technology Special Fund [ZDYF2021SHFZ051], Hainan Provincial Natural Science Foundation of China [820MS053], the Major Science and Technology Program of Hainan Province [ZDKJ202003], Marshal Initiative Funding of Hainan Medical University [JBGS202103], National Natural Science Foundation of China [31871338, 31970646, 61873075, 32060152, 32070673, 32170676], HMU Marshal Initiative Funding [HMUMIF-21024], project supported by Hainan Province Clinical Medical Center [QWYH202175], the National Key R&D Program of China [2018YFC2000100], Natural Science Foundation for Distinguished Young Scholars of Heilongjiang Province [JQ2019C004], Heilongjiang Touyan Innovation Team Program, project from Hainan Provincial Key Laboratory of Carcinogenesis and Intervention [JCKF2021003] and Innovation Research Fund for Graduate Students [Qhys2021-377, Qhys2021-348, Qhys2021-350, Qhys2021-351, HYYB2021A01, HYYS2021A31, HYYS2020-28, Hys2020-369].

Author contributions

YL, XL and JX came up with the design and conception. Material preparation, data collection and analysis were performed by SL, WZ, DL, TP, JG, HZ, ZT and KL. WZ and DL constructed the website of HVIdb. The first draft of the manuscript was written by YL, JX and SL. SL, WZ and DL contributed equally to this study and shared co-first authors. All authors read and approved the final manuscript.

Declaration of Competing Interest

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

Acknowledgement

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.03.002.

References

[1] Dyer MD, Murali TM, Sobral BW. The landscape of human proteins interacting with viruses and other pathogens. PLoS Pathog 2008;4:e32.
[2] Lian X, Yang X, Yang S, Zhang Z. Current status and future perspectives of computational studies on human-virus protein-protein interactions. Brief Bioinform 2021.
[3] Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov 2020;6:14.
[4] Fiscon G, Conte F, Farina L, Paci P. SAveRUNNER: a network-based algorithm for drug repurposing and its application to COVID-19. PLoS Comput Biol 2021;17:e1008686.
[5] Nicod C, Banaei-Esfahani A, Collins BC. Elucidation of host-pathogen protein-protein interactions to uncover mechanisms of host cell rewiring. Curr Opin Microbiol 2017;39:7–15.
[6] Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 2020;583:459–68.
[7] Han N, Hwang W, Tzelepis K, Schmerer P, Yankova E, MacMahon M, et al. Identification of SARS-CoV-2-induced pathways reveals drug repurposing strategies. Sci Adv 2021;7.
[8] Calderwood MA, Venkatesan K, Xing L, Chase MR, Vazquez A, Holthaus AM, et al. Epstein-Barr virus and virus human protein interaction maps. Proc Natl Acad Sci U S A 2007;104:7606–11.
[9] Khadka S, Vangeloff AD, Zhang C, Siddavatam P, Heaton NS, Wang L, et al. A physical interaction network of dengue virus and human proteins. Mol Cell Proteomics 2011;10(M111):012187.
[10] Shah PS, Link N, Jang GM, Sharp PP, Zhu T, Swaney DL, et al. Comparative flavivirus-host protein interaction mapping reveals mechanisms of dengue and Zika virus pathogenesis. Cell 2018;175(1931–1945):e1918.
[11] Lasso G, Mayer SV, Winkelmann ER, Chu T, Elliot O, Patino-Galindo JA, et al. A structure-informed atlas of human-virus interactions. Cell 2019;178(1526–1541):e1516.
[12] Tsukiyama S, Hasan MM, Fujii S, Kurata H. LSTM-PHV: prediction of human-virus protein-protein interactions by LSTM with word2vec. Brief Bioinform 2021.
[13] Kwofie SK, Schaefer U, Sundararajan VS, Bajic VB, Christoffels A. HCVpro: hepatitis C virus protein interaction database. Infect Genet Evol 2011;11:1971–7.
[14] Karyala P, Metri R, Bathula C, Yelamanchi SK, Sahoo L, Arjunan S, et al. DenHunt - a comprehensive database of the intricate network of dengue-human interactions. PLoS Negl Trop Dis 2016;10:e0004965.
[15] Dey L, Mukhopadhyay A. DenvInt: a database of protein-protein interactions between dengue virus and its hosts. PLoS Negl Trop Dis 2017;11:e0005879.
[16] Gurumayum S, Brahma R, Naorem LD, Muthaiyan M, Gopal J, Venkatesan A. ZikaBase: an integrated ZIKV- human interactome map database. Virology 2018;514:203–10.
[17] Guirimand T, Delmotte S, Navratil V. VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. Nucleic Acids Res 2015;43:D583–587.
[18] Calderone A, Licata L, Cesareni G. VirusMentha: a new resource for virus-host protein interactions. Nucleic Acids Res 2015;43:D588–592.
[19] Chatr-aryamontri A, Ceol A, Peluso D, Nardozza A, Panni S, Sacco F, et al. VirusMINT: a viral protein interaction database. Nucleic Acids Res 2009;37:D669–673.
[20] Yang X, Lian X, Fu C, Wuchty S, Yang S, Zhang Z. HVIDB: a comprehensive database for human-virus protein-protein interactions. Brief Bioinform 2021;22:832–44.
[21] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284–7.
[22] Zuguang Gu DH: simplifyEnrichment: an R/Bioconductor package for Clustering and Visualizing Functional Enrichment Results. *BioRxiv* 2021.
[23] Lv D, Xu K, Jin X, Li J, Shi Y, Zhang M, et al. LncSpA: LncRNA spatial atlas of expression across normal and cancer tissues. Cancer Res 2020;80:2067–71.
[24] Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res 2009;19:1639–45.
[25] Mohammadi S, Davila-Velderrain J, Kellis M. Reconstruction of cell-type-specific interactomes at single-cell resolution. Cell Syst 2019;9(559–568):e554.
[26] Morselli Gysi D, do Valle I, Zitnik M, Ameli A, Gan X, Varol O, Ghiassian SD, Patten JJ, Davey RA, Loscalzo J, Barabasi AL. Network medicine framework for identifying drug-repurposing opportunities for COVID-19. Proc Natl Acad Sci U S A 2021;118.
[27] Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, et al. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Science 2015;347(1257601).
[28] Zhou Y, Hou Y, Shen J, Mehra R, Kallianpur A, Culver DA, et al. A network medicine approach to investigation and population-based validation of disease manifestations and drug repurposing for COVID-19. PLoS Biol 2020;18:e3000970.
[29] Purcell AW, Croft NP, Tscharke DC. Immunology by numbers: quantitation of antigen presentation completes the quantitative milieu of systems immunology! Curr Opin Immunol 2016;40:88–95.
[30] Shiley KT, Lautenbach E, Lee I. The use of antimicrobial agents after diagnosis of viral respiratory tract infections in hospitalized adults: antibiotics or anxiolytics? Infect Control Hosp Epidemiol 2010;31:1177–83.
[31] Xu H, Wang B, Ono M, Kagita A, Fujii K, Sasakawa N, et al. Targeted disruption of HLA genes via CRISPR-Cas9 generates iPSCs with enhanced immune compatibility. Cell Stem Cell 2019;24(566–578):e567.
[32] Saraiva M, O'Garra A. The regulation of IL-10 production by immune cells. Nat Rev Immunol 2010;10:170–81.
[33] Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. Proc Natl Acad Sci U S A 2007;104:8685–90.
[34] Shao T, Wang G, Chen H, Xie Y, Jin X, Bai J, et al. Survey of miRNA-miRNA cooperative regulation principles across cancer types. Brief Bioinform 2019;20:1621–38.
[35] Li Y, Huo C, Pan T, Li L, Jin X, Lin X, et al. Systematic review regulatory principles of non-coding RNAs in cardiovascular diseases. Brief Bioinform 2019;20:66–76.

[36] Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. Proc Natl Acad Sci U S A 2004;101:3747–52.
[37] Puccioni-Sohler M, Rosadas C, Cabral-Castro MJ. Neurological complications in dengue infection: a review for clinical practice. Arq Neuropsiquiatr 2013;71:667–71.
[38] Oksanen M, Petersen AJ, Naumenko N, Puttonen K, Lehtonen S, Gubert Olive M, et al. PSEN1 Mutant iPSC-derived model reveals severe astrocyte pathology in Alzheimer's disease. Stem Cell Rep 2017;9:1885–97.
[39] De Felice FG, Tovar-Moll F, Moll J, Munoz DP, Ferreira ST. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the central nervous system. Trends Neurosci 2020;43:355–7.
[40] Meinhardt J, Radke J, Dittmayer C, Franz J, Thomas C, Mothes R, et al. Olfactory transmucosal SARS-CoV-2 invasion as a port of central nervous system entry in individuals with COVID-19. Nat Neurosci 2021;24:168–75.
[41] Mao L, Jin H, Wang M, Hu Y, Chen S, He Q, et al. Neurologic manifestations of hospitalized patients with coronavirus disease 2019 in Wuhan, China. JAMA Neurol 2020;77:683–90.
[42] Yu Y, Huang X, Liu J, Zhang J, Hu Y, Yang Y, et al. Fish TRIM32 functions as a critical antiviral molecule against iridovirus and nodavirus. Fish Shellfish Immunol 2017;60:33–43.
[43] Bodda C, Reinert LS, Fruhwurth S, Richardo T, Sun C, Zhang BC, et al. HSV1 VP1-2 deubiquitinates STING to block type I interferon expression and promote brain infection. J Exp Med 2020;217.
[44] Furumai R, Tamada K, Liu X, Takumi T. UBE3A regulates the transcription of IRF, an antiviral immunity. Hum Mol Genet 2019;28:1947–58.
[45] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 2018;46:D1074–82.
[46] Bekerman E, Einav S. Infectious disease. Combating emerging viral threats. Science 2015;348:282–3.
[47] Ferrazzi E, Peracchi M, Biasolo MA, Faggionato O, Stefanelli S, Palu G. Antiviral activity of gyrase inhibitors norfloxacin, coumermycin A1 and nalidixic acid. Biochem Pharmacol 1988;37:1885–6.
[48] Janowski AB, Dudley H, Wang D. Antiviral activity of ribavirin and favipiravir against human astroviruses. J Clin Virol 2020;123:104247.
[49] Li T, Kenney AD, Liu H, Fiches GN, Zhou D, Biswas A, Que J, Santoso N, Yount JS, Zhu J: SARS-CoV-2 Nsp14 activates NF-kappaB signaling and induces IL-8 upregulation. *bioRxiv* 2021.
[50] Song J, Ma Q, Hu M, Qian D, Wang B, He N. The inhibition of miR-144-3p on cell proliferation and metastasis by targeting TOP2A in HCMV-positive glioblastoma cells. Molecules 2018;23.
[51] Muhammad JS, Saheb Sharif-Askari N, Cui ZG, Hamad M, Halwani R. SARS-CoV-2 infection-induced promoter hypomethylation as an epigenetic modulator of heat shock protein A1L (HSPA1L) gene. Front Genet 2021;12:622271.
[52] Lamers MM, Beumer J, van der Vaart J, Knoops K, Puschhof J, Breugem TI, et al. SARS-CoV-2 productively infects human gut enterocytes. Science 2020;369:50–4.
[53] Xu K, Cai Y, Zhang M, Zou H, Chang Z, Li D, et al. Pan-cancer characterization of expression and clinical relevance of m(6)A-related tissue-elevated long non-coding RNAs. Mol Cancer 2021;20:31.
[54] Puelles VG, Lutgehetmann M, Lindenmeyer MT, Sperhake JP, Wong MN, Allweiss L, et al. Multiorgan and Renal Tropism of SARS-CoV-2. N Engl J Med 2020;383:590–2.