

H2V, a Database for Human Proteins and Genes in Response to SARS-CoV-2, SARS-CoV, and MERS-CoV Infection

Nan Zhou

Brain Hospital of Guangzhou Medical University

Jinku Bao

Sichuan University

Yuping Ning (✉ ningjeny@126.com)

Brain Hospital of Guangzhou Medical University

Database

Keywords: coronavirus, protein, knowledgebase, gene

Posted Date: August 19th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-61338/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on January 7th, 2021. See the published version at <https://doi.org/10.1186/s12859-020-03935-2>.

Abstract

The ongoing COVID-19 pandemic in the world is caused by SARS-CoV-2, a new coronavirus firstly discovered in the end of 2019. It has led to more than 10 million confirmed cases and more than 500,000 confirmed deaths across 216 countries by 1 July 2020, according to WHO statistics. SARS-CoV-2, SARS-CoV, and MERS-CoV are alike, killing people, impairing economy, and inflicting long-term impacts on the society. However, no specific drug or vaccine has been approved as a cure for these viruses. The efforts to develop antiviral measures are hampered by insufficient understanding of molecular responses of human to viral infections. In this study, we collected experimentally validated human proteins that interact with SARS-CoV-2 proteins, human proteins whose expression, translation and phosphorylation levels experience significantly changes after SARS-CoV-2 or SARS-CoV infection, human proteins that correlate with COVID-19 severity, and human genes whose expression levels significantly changed upon SARS-CoV-2 or MERS-CoV infection. A database, H2V, was then developed for easy access to these data. Currently H2V includes: 332 human-SARS-CoV-2 protein-protein interactions; 65 differentially expressed proteins, 232 differentially translated proteins, 1298 differentially phosphorylated proteins, 204 severity associated proteins, and 4012 differentially expressed genes responding to SARS-CoV-2 infection; 66 differentially expressed proteins responding to SARS-CoV infection; and 6981 differentially expressed genes responding to MERS-CoV infection. H2V can help to understand the cellular responses associated with SARS-CoV-2, SARS-CoV and MERS-CoV infection. It is expected to speed up the development of antiviral agents and shed light on the preparation for potential coronavirus emergency in the future.

Database url: <http://www.zhounan.org/h2v>

Background

Coronaviruses are single-stranded RNA viruses, and some can cross the species barrier to cause deadly and infectious respiratory disease in humans [1]. A novel coronavirus that causes viral pneumonia was reported in December 2019 [2]. The virus, now known as SARS-CoV-2, spread rapidly in the world and caused the ongoing COVID-19 pandemic. The last two coronavirus emergencies were severe acute respiratory syndrome (SARS) in 2002-2003 and Middle East respiratory syndrome (MERS) from 2012 [3]. With a case fatality of ~10%, the SARS-related coronavirus (SARS-CoV) infected 8098 people and caused 774 deaths; the MERS-related coronavirus (MERS-CoV) has a higher mortality rate of ~34%, with ~2500 confirmed infections and ~900 deaths so far [4]. The COVID-19 is much severe than the previous two outbreaks. As of 1 July 2020, > 10 million confirmed cases and > 500,000 deaths have been reported to WHO ([https:// www.who.int](https://www.who.int)) worldwide. It is urgent for the world as a whole to find effective ways to bring the COVID-19 crisis to an end.

Only supportive therapies are currently available for the treatment of diseases caused by SARS-CoV-2, SARS-CoV and MERS-CoV. All of the three viruses belong to the beta-coronavirus group. Two other beta-coronaviruses, HCoV-OC43 and HKU1, also enable to infect human beings but only cause self-limiting flu-like illness [5]. SARS-CoV-2, SARS-CoV and MERS-CoV, on the other hand, are highly pathogenic and life-

threatening. Even though the world has been repeatedly suffered from coronavirus outbreaks, there is no clinically effective prophylactics or therapeutics available.

The life cycle of SARS-CoV-2, SARS-CoV, and MERS-CoV includes several key steps: viral entry, genomic RNA replication, mRNA translation, protein processing, and virion assembly and release [6]. The interplay between human and the viruses at the viral entry stage has been well documented. To enter the human cell, both SARS-CoV-2 and SARS-CoV bind their S proteins to the cell surface receptor ACE2, angiotensin-converting enzyme 2 [7]. MERS-CoV enters the human cell via binding another receptor, dipeptidyl peptidase 4 (DPP4) [3]. Hoffmann and colleagues have also proved that the binding of SARS-CoV-2 S protein to human ACE2 additionally depends on TMPRSS2 and showed that the cell entry of SARS-CoV-2 can be blocked by serine protease inhibitor camostat mesylate [8]. More details about the interplay between human and other activities regarding the virus life cycle remain to be revealed. There is no doubt that the human body must respond to virus infection and the response can be detected on the molecular level by genome- and proteome-wide measurements.

SARS-CoV-2 keeps spreading rapidly in some parts of the world. To pull life back to normal track, specific drugs with clinical efficacy that target SARS-CoV-2 are urgently required but unavailable yet. There is also no cure for SARS and MERS, indicating our understanding of these dangerous coronaviruses is very limited. As long as knowledge of cellular responses to viral infections would help to combat the COVID-19 pandemic, we developed H2V, a database for human proteins and genes that respond to SARS-CoV-2, SARS-CoV and MERS-CoV infection in the present study.

Construction And Content

Data collection

In the study, human proteins/genes that respond to viral infections are defined as interactors of human-virus protein-protein interactions (PPIs), differentially expressed genes (DEGs), differentially expressed proteins (DEPs), differentially translated proteins (DTPs), differentially phosphorylated proteins (DPPs), and severity associated proteins (SAPs). The source data for the discovery of response proteins and genes were collected from published studies.

High-confidence human-SARS-CoV-2 protein-protein interactions (PPIs) were collected from Gordon and colleagues' research [9]. In their study, physical PPIs were measured by affinity-purification mass spectrometry. The collected data was used in this study without any post-processing.

DEPs are proteins whose abundance undergo significantly change after virus infection. DTPs are proteins who experience significant translation changes upon virus infection. Translatome and proteome proteomics data in human cells at 2, 6, 10, and 24 hours post SARS-CoV-2 infection were collected from Bojkova and colleagues' study [10]. Proteins were selected as significantly changed if they have a fold change (infection vs control) of > 2 and a p value of < 0.05 at any of the time points.

DEGs are genes whose expression levels change significantly after virus infection. DEGs in response to SARS-CoV-2 infection were collected from Blanco-Melo and colleagues' study [11]. Within the downloaded data, missing records were removed and genes with a fold change (infection vs control) of > 2 and a p value of < 0.05 were chosen as DEGs.

SAPs are proteins that can differentiate critical COVID-19 cases from moderate cases or healthy cohorts. Original data was from Shen and colleagues' work [12]. From the collected data, SAPs were chosen as proteins whose expression levels changed > 2 folds (infection vs control) with a p value of < 0.05 .

Phosphorylation changes of kinases reflect the signaling pathways that virus relies on to survive in the human body. Phosphorylation dynamics data in Vero E6 cells at 0, 2, 4, 8, 12, and 24 hours after SARS-CoV-2 infection was collected from Bouhaddou and colleagues' study [13]. Since Vero E6 cells are from monkey, the original authors had mapped phosphorylation sites and protein identifiers to their respective human protein orthologs. From the downloaded data, DPPs were selected as proteins with a fold change (infection vs control) of > 2 and a p value of < 0.05 at any time.

DEPs in response to SRAR-CoV infection were collected from Jiang and colleagues' article [14]. The data was used in our study without any post processing.

RNA-seq data at 6 and 24 hours after MERS-CoV infection in Calu-3 cells was collected from Zhang and colleagues' research [15]. Gene expression was measured by Salmon in RaNA-seq and DEGs were detected by DESeq2 in RaNA-seq [16–18]. DEGs were selected with a fold change (infection vs control) of > 2 and a p value of < 0.05 at any of the two sampling times.

Genome assembly MN985325.1 from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) was used to annotate SARS-CoV-2 genes. Drugs that target H2V proteins were collected from the DrugBank database [19]. Post processing of data was performed by R (<https://www.r-project.org/>).

Implementation

H2V was developed with mainstream web developing techniques. The user interface was developed with HTML5, CSS3, and JavaScript. Bootstrap v4 (<https://getbootstrap.com/>) was used for layout design. DataTables (<https://datatables.net/>) was used to organize data in table on the web page. Cytoscape.js was used for network visualization of PPIs [20]. Plotly (<https://plotly.com/>) was used to create interactive plots. PHP (<https://www.php.net/>), python (<https://www.python.org/>), and bash scripts were used for server-side development. Data is managed by SQLite (<https://www.sqlite.org/>). NCBI's sequence viewer (<https://www.ncbi.nlm.nih.gov/projects/sviewer/>) was embedded in H2V, for easily browsing SARS-CoV-2 gene information within our database. Drug information is not stored in H2V, instead it is automatically retrieved on request from the DrugBank database via UniProt's REST API [21]. H2V is deployed in an Amazon AWS host running Ubuntu 16.04.

Utility And Discussion

Statistics of H2V data

Overall, H2V contains 332 PPIs, 131 DEPs, 232 DTPs, 1298 DPPs, 204 SAPs, and 10993 DEGs entries (Table 1). Specifically: there are 332 PPIs, 65 DEPs, 232 DTPs, 1298 DPPs, 204 SAPs, and 4012 DEGs in response to SARS-CoV-2 infection; there are 66 DEPs in response to SARS-CoV infection; and there are 6981 DEGs in response to MERS-CoV infection. The human-SARS-CoV-2 PPIs also include 27 SARS-CoV-2 proteins, with nsp5 mutant C145A considered as a unique protein.

Among the human proteins that respond to SARS-CoV-2 infection (Figure 1a): 10 proteins intersect between DEPs and DTPs; 3 proteins intersect between DEPs and SAPs; 6 proteins intersect between DEPs and DPPs; 5 proteins intersect between DTPs and SAPs; 26 proteins intersect between DTPs and DPPs; 5 proteins intersect between SAPs and DPPs; DEPs, DTPs, and SAPs share 1 protein; DEPs, SAPs, and DPPs share 1 protein as well. There is only 1 common protein between DEPs in response to SARS-CoV-2 and SARS-CoV infection (Figure 1b). The set of intersecting genes is larger than that of intersecting proteins. As shown in Figure 1c, there are 1497 intersecting genes between DEGs in response to SARS-CoV-2 and MERS-CoV infection.

Proteins/genes responding to SARS-CoV-2

The “SARS2” drop-down menu in the navigation bar of the website provides links to human proteins and genes that respond to SARS-CoV-2 infection. PPIs are shown in a table or as a network (Figure 2a-b). In the table of PPIs (Figure 2a), human-SARS-CoV-2 PPIs are put in rows, with other protein information shown in columns. For SARS-CoV-2 proteins, the corresponding genes are put in the “SARS2 gene” column. By clicking on a SARS-CoV-2 gene, its annotation will be shown in the embedded NCBI sequence viewer on the same web page (Figure 2c). UniProt identifiers, protein names, gene names, and HGNC identifiers of human proteins are also displayed in the PPIs table. The external links to the UniProt database and HGNC database can help users to study the proteins of their interests. The “Drug” column in the PPIs table can be used to retrieve drugs of a protein target from the DrugBank database. This feature would facilitate rapid discovery of candidate antiviral drugs. DEPs, DTPs, and DPPs are listed in a table in a similar format, so DPPs are taken as an example and demonstrated in Figure 2d. Compared to the PPIs table, SARS-CoV-2 columns are removed but a “Temporal profile” column is added. Clicking on the “Show” button of a protein in this column will display a plot showing how the protein phosphorylation changes over time and when the change reaches significance (Figure 2e). DEPs and DTPs also have a similar profile to show how the expression or translation of a protein dynamically changes over time, but only the DPPs profile is shown as an example. Because time series data is not available for SAPs, the log₂ fold changes and p values are put in columns “log₂FC” and “P-value” (Figure 2f). DEGs are also shown in a table, with links to the HGNC database, log₂ fold changes, and p values put in columns (Figure 2g).

Proteins responding to SARS-CoV

All of the collected human proteins that respond to SARS-CoV infection are DEPs. The DEPs table can be accessed via the “DEPs” link on the “SARS1” drop-down menu in the navigation bar of the website. As can be seen from Figure 3, DEPs are listed in table rows, and the correlated information is listed in table columns.

Genes responding to MERS-CoV

All of the collected human genes that respond to MERS-CoV infection are DEGs. The DEGs table can be accessed by clicking on the “DEGs” link on the “MERS-CoV” drop-down menu in the navigation bar of the website. In the table, gene names, HGNC identifiers, and buttons to temporal profiles are given (Figure 4a). Users can click on the “Show” button of a DEG to see how its expression changes with time after MERS-CoV infection (Figure 4b).

Application case

H2V provides two easy-to-use utilities, one is “Drug finder” and the other is “Data animation”. The former one facilitates to find drugs that can target a given protein based on its UniProt accession number. If drug exists, drug name and DrugBank identifier will be displayed below the utility on the same page. For example, searching 9BYF1 will find drugs moexipril and SPP1148 (Figure 5a).

The data animation utility provides a concrete perception of the proteomic dynamics over a period (Figure 5b-5g). Compared to 2 h after SARS-CoV-2 infection, the number of DEPs and DTPs increased at 24 h after virus infection. The number of DPPs at 24 h is also larger than that at 0 h after SARS-CoV-2 infection. These trends give us a clear picture that the human body responds to SARS-CoV-2 infection by continuously activating protein biogenesis.

Conclusions

We have developed the first database for human proteins and genes that respond to SARS-CoV-2, SARS-CoV, and MERS-CoV infection. The database will help to understand the cellular details of how the human body responds to the coronavirus infection. We have to acknowledge that the present release of our database may omit some data which should be included, we will keep updating the database and offer missing data in future releases. Apart from providing knowledge about human response to viral infections, another key feature of the database is that drugs of any protein of interest can be found with ease. This provides valuable hints on drug development. In summary, the database will help to design effective and specific drugs and preventive vaccines targeting SARS-CoV-2, SARS-CoV, and MERS-CoV.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

N.Z. and J.B. collected and analyzed data. N.Z. developed the website and wrote manuscript draft. J.B. and Y.N. conceived and supervised the study. All author took roles in reviewing and revising the manuscript.

Acknowledgements

We thank the authors who generated the raw data that have been used in this study.

References

1. Weiss SR, Navas-Martin S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol Mol Biol Rev.* 2005;69:635–64. doi:10.1128/MMBR.69.4.635-664.2005.
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med.* 2020;382:727–33. doi:10.1056/NEJMoa2001017.
3. Zhou N, Zhang Y, Zhang J-C, Feng L, Bao J-K. The receptor binding domain of MERS-CoV: The dawn of vaccine and treatment development. *J Formos Med Assoc.* 2014;113:143–7. doi:https://doi.org/10.1016/j.jfma.2013.11.006.
4. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Velesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell.* 2020;181:281–92.e6. doi:https://doi.org/10.1016/j.cell.2020.02.058.
5. Liu DX, Liang JQ, Fung TS. Human Coronavirus-229E, -OC43, -NL63, and -HKU1. *Ref Modul Life Sci.* 2020;:B978-0-12-809633-8\$421501-X. doi: 10.1016/B978-0-12-809633-8.21501-X .
6. Dai L, Zheng T, Xu K, Han Y, Xu L, Huang E, et al. A Universal Design of Betacoronavirus Vaccines against COVID-19, MERS, and SARS. *Cell.* 2020. doi:https://doi.org/10.1016/j.cell.2020.06.035.

7. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579:270–3. doi:10.1038/s41586-020-2012-7.
8. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*. 2020;181:271–80.e8. doi:https://doi.org/10.1016/j.cell.2020.02.052.
9. 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. doi:10.1038/s41586-020-2286-9.
10. Bojkova D, Klann K, Koch B, Widera M, Krause D, Ciesek S, et al. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature*. 2020. doi:10.1038/s41586-020-2332-7.
11. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, Uhl S, Hoagland D, Møller R, et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*. 2020;181:1036–45.e9. doi:https://doi.org/10.1016/j.cell.2020.04.026.
12. Shen B, Yi X, Sun Y, Bi X, Du J, Zhang C, et al. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell*. 2020. doi:https://doi.org/10.1016/j.cell.2020.05.032.
13. Bouhaddou M, Memon D, Meyer B, White KM, Rezelj VV, Marrero MC, et al. The Global Phosphorylation Landscape of SARS-CoV-2 Infection. *Cell*. 2020. doi:https://doi.org/10.1016/j.cell.2020.06.034.
14. Jiang X-S, Tang L-Y, Dai J, Zhou H, Li S-J, Xia Q-C, et al. Quantitative Analysis of Severe Acute Respiratory Syndrome (SARS)-associated Coronavirus-infected Cells Using Proteomic Approaches. *Mol & Cell Proteomics*. 2005;4:902 LP – 913. doi:10.1074/mcp.M400112-MCP200.
15. Zhang X, Chu H, Wen L, Shuai H, Yang D, Wang Y, et al. Competing endogenous RNA network profiling reveals novel host dependency factors required for MERS-CoV propagation. *Emerg Microbes Infect*. 2020;9:733–46. doi:10.1080/22221751.2020.1738277.
16. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14:417–9. doi:10.1038/nmeth.4197.
17. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
18. Prieto C, Barrios D. RaNA-Seq: interactive RNA-Seq analysis from FASTQ files to functional analysis. *Bioinformatics*. 2019;36:1955–6. doi:10.1093/bioinformatics/btz854.
19. 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*. 2017;46:D1074–82. doi:10.1093/nar/gkx1037.
20. Franz M, Lopes CT, Huck G, Dong Y, Sumer O, Bader GD. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*. 2015;:btv557.
21. Consortium TU. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2018;47:D506–15. doi:10.1093/nar/gky1049.

Tables

Table 1
Statistics of data in H2V.

Virus	Category	# Database entries	# Unique ^a
SARS-CoV-2	PPIs	332	332
	DEPs	65	65
	DTPs	232	232
	DPPs	1298	779
	SAPs	204	130
	DEGs	4012	4012
SARS-CoV	DEPs	66	66
MERS-CoV	DEGs	6981	6981

^a The number of unique proteins or genes. The number of DPPs entries is larger than the number of unique proteins because a protein can have multiple phosphorylation sites. The number of SAPs entries is larger than the number of unique proteins because the same protein can appear in multiple comparison groups.

Figures

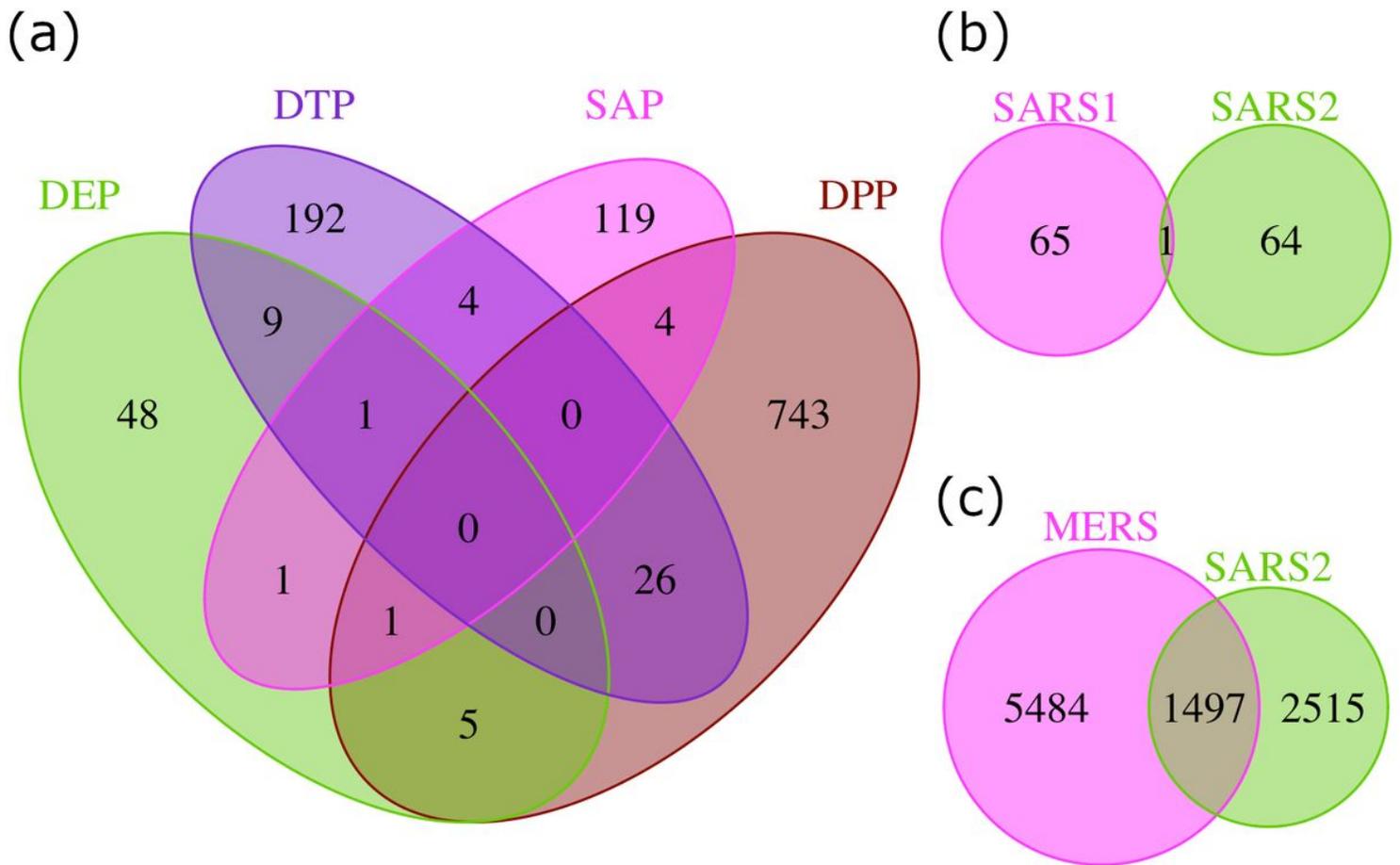


Figure 1

Venn diagrams showing intersections of human proteins or genes among different categories. (a) The intersection of DEPs, DPPs, DTPs, and SAPs in response to SARS-CoV-2 infection. (b) The intersection of DEPs in response to SARS-CoV-2 and SARS-CoV infection. (c) The intersection of DEGs in response to SARS-CoV-2 and MERS-CoV infection. SARS2, SARS-CoV-2; SARS1, SARS-CoV; MERS, MERS-CoV.

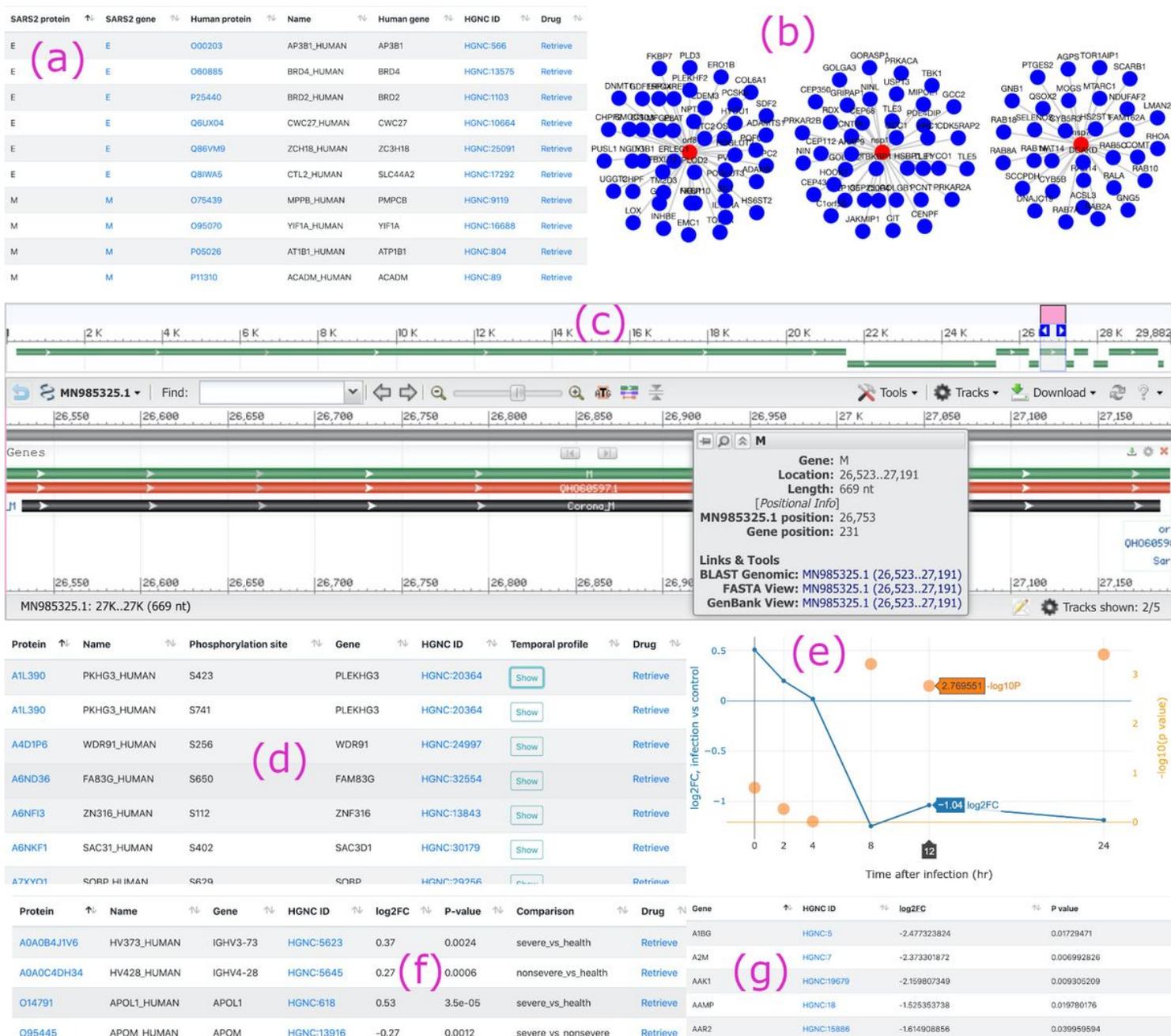


Figure 2

Overview of proteins and genes in response to SARS-CoV-2 infection. (a) A table of PPIs. (b) Network visualization of PPIs, with SARS-CoV-2 proteins shown as red nodes, human proteins shown as blue nodes, and interactions indicated by gray lines. (c) NCBI sequence viewer embedded on the PPIs page, with SARS-CoV-2 M gene annotation shown as an example. (d) A table of DPPs. (e) Dynamic phosphorylation profile over time, with fold change and p value showing on the respective left and right y axes. (f) A table of SAPs. (g) A table of DEGs.

Protein	Name	log2FC	Gene	HGNC ID	Drug
O14775	GNB5_HUMAN	1.15	GNB5	HGNC:4401	Retrieve
O15146	MUSK_HUMAN	-1.12	MUSK	HGNC:7525	Retrieve
O60701	UGDH_HUMAN	-1.83	UGDH	HGNC:12525	Retrieve
O75439	MPPB_HUMAN	1.32	PMPCB	HGNC:9119	Retrieve
O95292	VAPB_HUMAN	1.03	VAPB	HGNC:12649	Retrieve
P02795	MT2_HUMAN	-1.33	MT2A	HGNC:7406	Retrieve
P07900	HS90A_HUMAN	-1.06	HSP90AA1	HGNC:5253	Retrieve
P08559	ODPA_HUMAN	1.47	PDHA1	HGNC:8806	Retrieve
P08865	RSSA_HUMAN	-1.04	RPSA	HGNC:6502	Retrieve
P09382	LEG1_HUMAN	-1.05	LGALS1	HGNC:6561	Retrieve

Figure 3

Overview of DEPs responding to SARS-CoV infection.



Figure 4

Overview of DEGs responding to MERS-CoV infection. (a) A table of DEPs. (b) Gene expression profile over time, with fold change and p value showing on the respective left and right y axes.

Find drugs for a protein

(a)

WARNING: Do not use any drug on the page for any treatment purpose by yourself. If you are worried about your health, please seek help from the doctor.

Uniprot Accession Q9BYF1

Find

Reset

Finish loading data. Drugs will be shown as following. If you do not see anything, no drug exists.

DB00691: Moexipril

DB05203: SPP1148

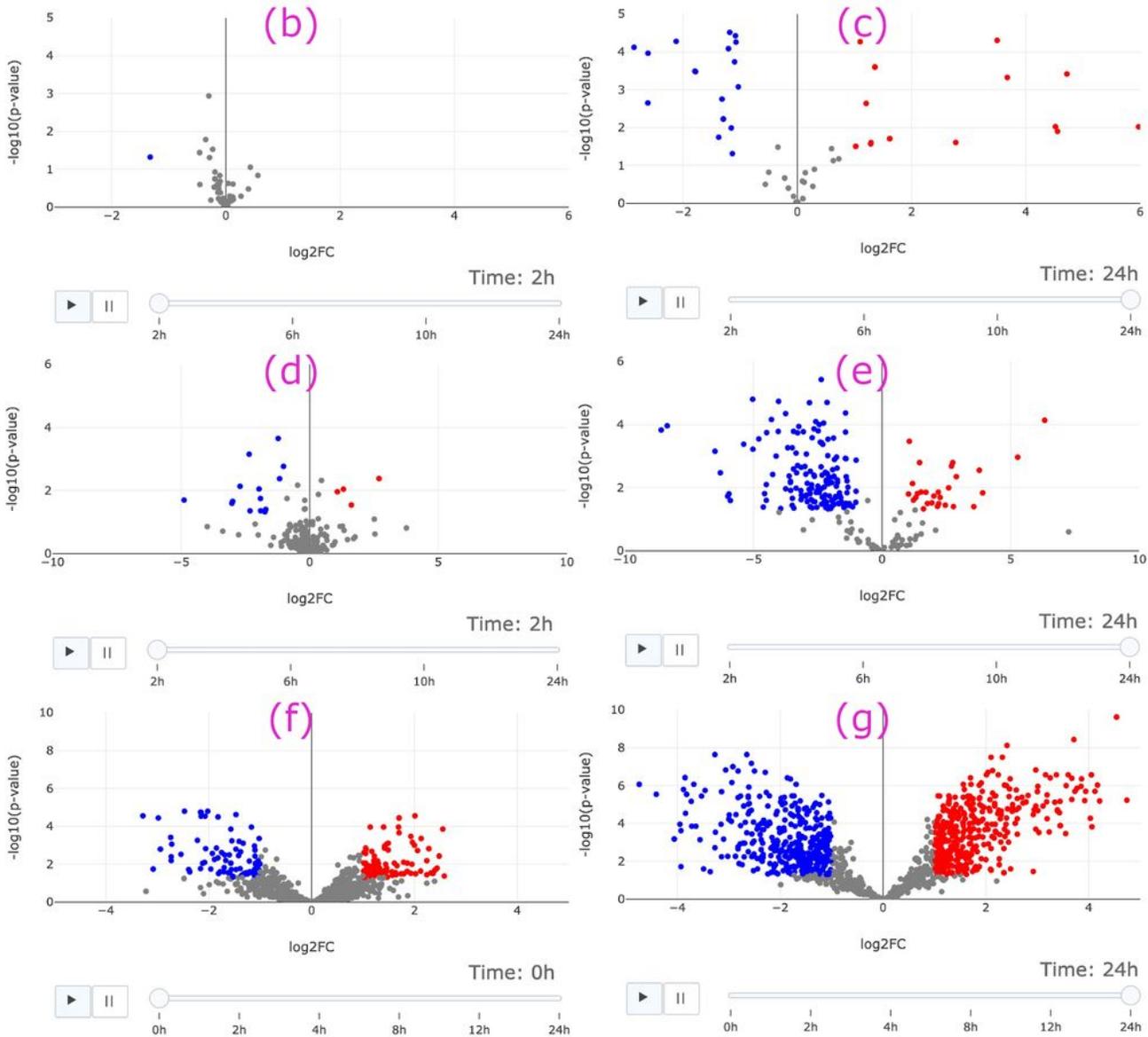


Figure 5

Overview of utilities. (a) Drug finder. (b-g) Animation of all DEPs (b and c), DTPs (d and e), and DPPs (f and g) that respond to SARS-CoV-2 infection from the first sampling time point (b, d, and f) to the last time point (c, e, and g).