




# A systems-based method to repurpose marketed therapeutics for antiviral use: a SARS-CoV-2 case study

Mengran Wang, Johanna B Withers, Piero Ricchiuto, Ivan Voitalov, Michael McAnally, Helia N Sanchez, Alif Saleh, Viatcheslav R Akmaev, Susan Dina Ghiassian 

**This study describes two complementary methods that use network-based and sequence similarity tools to identify drug repurposing opportunities predicted to modulate viral proteins. This approach could be rapidly adapted to new and emerging viruses. The first method built and studied a virus–host–physical interaction network; a three-layer multimodal network of drug target proteins, human protein–protein interactions, and viral–host protein–protein interactions. The second method evaluated sequence similarity between viral proteins and other proteins, visualized by constructing a virus–host–similarity interaction network. Methods were validated on the human immunodeficiency virus, hepatitis B, hepatitis C, and human papillomavirus, then deployed on SARS-CoV-2. Comparison of virus–host–physical interaction predictions to known antiviral drugs had AUCs of 0.69, 0.59, 0.78, and 0.67, respectively, reflecting that the scores are predictive of effective drugs. For SARS-CoV-2, 569 candidate drugs were predicted, of which 37 had been included in clinical trials for SARS-CoV-2 (AUC = 0.75,  $P$ -value  $3.21 \times 10^{-3}$ ). As further validation, top-ranked candidate antiviral drugs were analyzed for binding to protein targets in silico; binding scores generated by BindScope indicated a 70% success rate.**

DOI [10.26508/lsa.202000904](https://doi.org/10.26508/lsa.202000904) | Received 8 September 2020 | Revised 4 February 2021 | Accepted 5 February 2021 | Published online 16 February 2021

## Introduction

Viruses are obligate intracellular parasites that can only replicate by entering into a host cell and hijacking host cell machinery to produce and assemble new progeny virions. The arsenal to treat viral infections focuses on viral and cellular components that are essential for the viral life cycle. There are two main categories of antiviral therapies: (1) host-targeted antivirals that modulate host proteins that interact with or are influenced by the virus and (2) virus-targeted antivirals that directly bind to and modulate the activity of viral proteins (1). Drugs that target viral proteins may directly limit viral replication and propagation (1).

The medical and scientific communities do not have the luxury of time to develop new compounds that target highly contagious

life-threatening viruses with the rapidity required to combat an ongoing pandemic (2). To facilitate the advancement of drugs into clinical trials, identification of drugs that were previously approved for other indications and exhibit a reasonable safety profile represent a resource for potential antiviral therapeutics (3, 4, 5 Preprint).

The coronavirus disease 2019 (COVID-19) pandemic, caused by the spread of SARS-CoV-2, has highlighted the need for tools to rapidly identify effective therapies against emerging pathogens. Numerous drugs have been identified that are predicted to alleviate COVID-19 symptoms and are in clinical trials to assess their safety and efficacy (6, 7). Identification of multiple treatment options is important to control the spread of the disease: patients respond differently to the same treatments because of their genetics (8), sequence variations in the viral genome may influence drug efficacy (9), and bottlenecks in drug availability may occur once an effective treatment is identified (10).

One consideration when evaluating drug repurposing opportunities is the ability of a drug to bind to multiple protein targets, called drug promiscuity. It has been previously shown that a drug's promiscuity is correlated with structural similarity and binding site similarity between the intended and unintended protein targets (11). In this study, network-based tools and bioinformatic approaches identified drug repurposing opportunities that were predicted to directly target viral proteins. The two antiviral drug discovery methods developed in this study were designed to first identify existing protein targets that are structurally similar to viral proteins then, subsequently, predict and rank the interactions likely to occur between existing compounds and viral proteins. These methods were validated using the human immunodeficiency virus 1 (HIV), hepatitis B virus (HBV), hepatitis C virus (HCV), and human papillomavirus (HPV) then used to predict candidate antiviral drug repurposing opportunities for SARS-CoV-2.

## Results

In this study, two network- and sequence-based methods were developed to identify candidate drug repurposing opportunities that directly target viral proteins. An underlying assumption behind

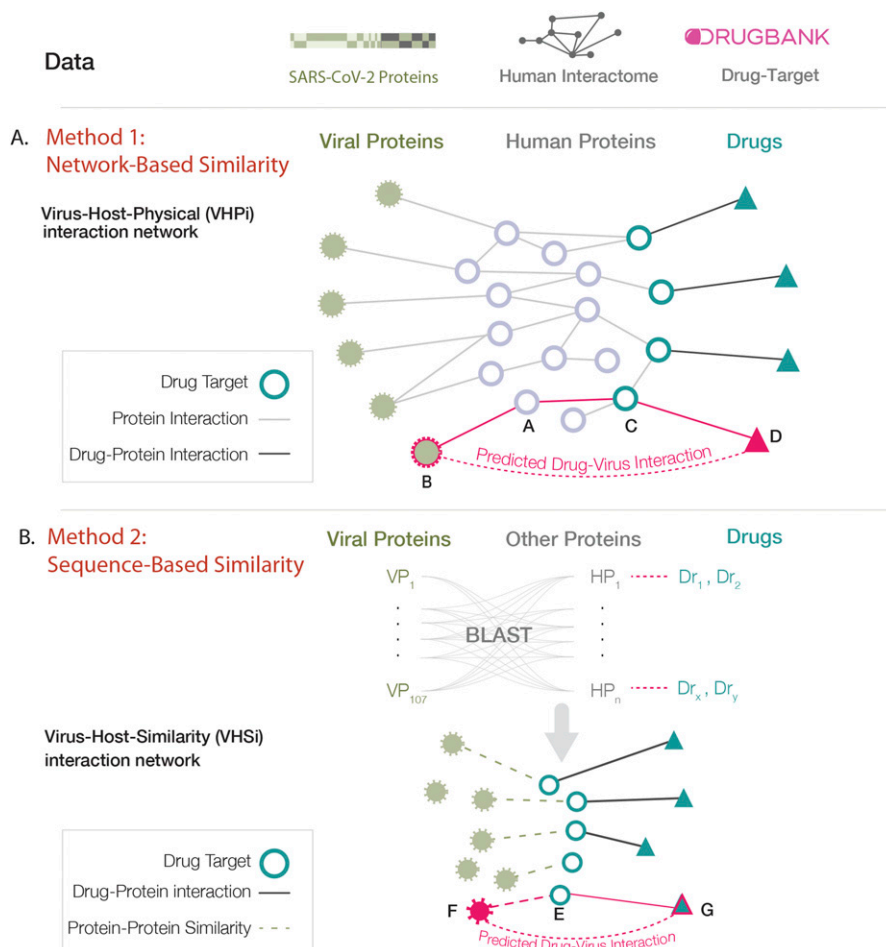
many drug–target interaction predictions is that structurally similar proteins are more likely to be targeted by similar drugs (11, 12, 13). Using complementary approaches, the methods described herein identify drugs by virtue of homology between viral proteins and proteins that are the known target of therapeutics developed for other indications. Both methods were designed to find proteins that are structurally similar to viral proteins; method 1 inferred structural similarity through common protein–protein interaction (PPI) patterns and method 2 derived structural similarity through sequence homology (Fig 1). When used in combination, these complementary methods derive high-confidence predictions of drug repurposing candidates.

### Systems-based antiviral drug ranking and identification

In the network-based method, the similarity between viral proteins and human proteins was indirectly inferred through their common interaction patterns. We constructed and studied a virus–host–physical interaction (VHPi) network; a three-layer multimodal network of (1) drug and target protein interactions from DrugBank (14,15), (2) consolidated human protein–protein pairwise interactions (Human Interactome, see the Materials and Methods section), and (3) viral–host PPIs (16). In this method, a network-based

similarity approach (17) was implemented to find human proteins that are similar to viral proteins. These network-identified proteins were predicted to have similar binding interfaces and thus are likely to interact with the same compounds.

In the VHPi network, identifying a drug that is predicted to bind to a viral protein is a link prediction problem (18) (Fig 1A). In most self-organizing networks, such as social networks, two nodes (people) are more likely to interact with (know) each other if they have a greater number of common neighbors, a phenomenon known as triadic closure principle (19). Hence, most link prediction methods that have been implemented on biological networks use metrics that incorporate common neighbors (20, 21). However, these metrics do not successfully predict interactions in biological networks where interactions occur primarily through physical contacts. For example, protein A interacts with both proteins B and C, independently (Fig 1A). In this case, protein B and C have a shared neighbor–protein A–and are therefore at a network distance of two. The interaction between proteins A and B indicates that these proteins have complementary binding sites, and this assumption holds true for proteins A and C as well. Proteins with a high number of shared neighbors are more likely to have similar binding interfaces and thus tend to interact with the same proteins (nodes). Unlike the case of social networks, proteins B and C are not



**Figure 1. Complementary methods to identify drug repurposing candidates that directly target viral proteins.**

(A) A network-based approach using link prediction to identify drug repurposing opportunities that are at a path length of three from viral proteins on the virus–host–physical interaction network. (B) A sequence similarity approach that identifies drug target proteins with protein sequence homology to viral proteins.

expected to directly interact with each other. Rather proteins B and C are expected to be structurally similar. Therefore, protein B is predicted to interact with other interacting partners of protein C, in this case protein D. Note that the network distance between protein B and the predicted interacting partner, protein D, is three. Therefore, link prediction applied to biological networks involves finding path lengths of three (L3) (17). Drugs with a path of length three to the viral proteins will have a non-zero rank. In this scoring, drugs with multiple paths of length three to the viral proteins were ranked higher and promiscuous drugs (drugs with a high number of target proteins) tended to have higher scores.

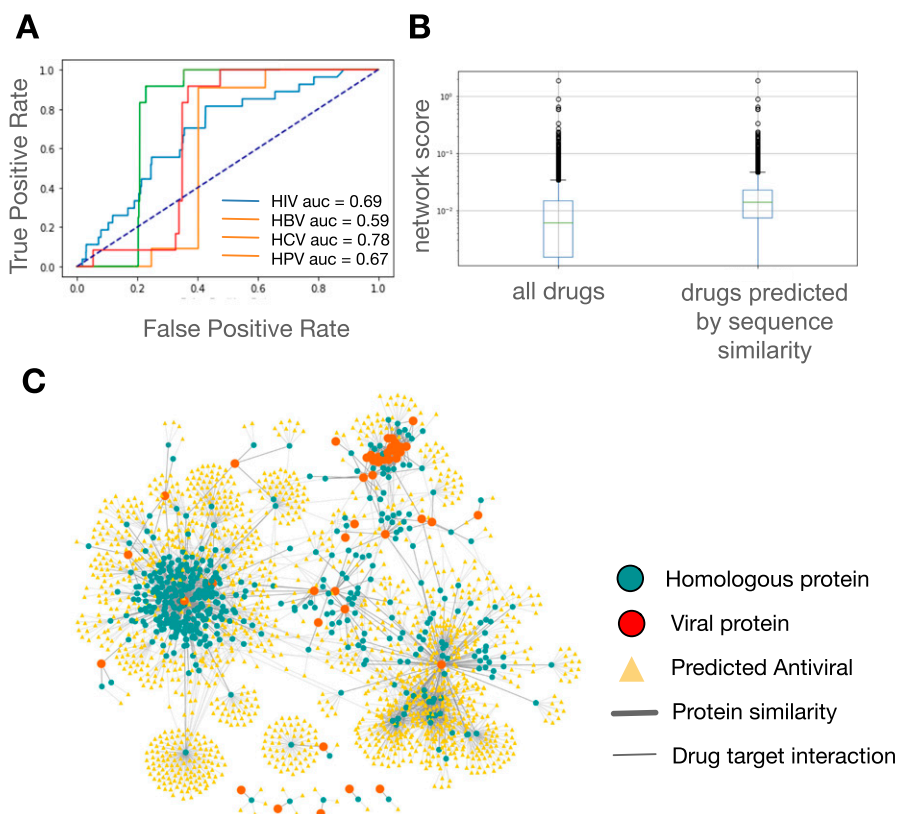
For physical PPI networks, the L3 methodology outperforms other link prediction methods (17). In principle, this can be extended to any network where the interactions are based on physical binding and has the advantage that it should not be limited to one organism or specific node type. In method 1, link prediction can be applied such that drugs known to target human proteins would be expected to interact with viral proteins if the drug is three steps away from a viral protein in the VHPi network (Fig 1A). To generate a ranking score, each drug score was penalized by the degree of the mediator nodes, that is, a protein in the path between a viral protein and predicted drug (see the Materials and Methods section, Equation (1)).

In parallel, a complementary approach was implemented in which the similarity between viral proteins and other proteins was directly derived by calculating their global sequence similarity. In this method, proteins with high sequence similarity were predicted to bind the same small molecules. BLAST homology analysis assessed the amino acid sequence similarity between viral proteins

and other protein sequences (see the Materials and Methods section). The results can be visualized by constructing a virus–host–similarity interaction (VHSi) network (Fig 1B).

### Method validation in HIV, HCV, HBV, and HPV

Four viruses were chosen to validate the methods: HIV, HBV, HCV, and HPV. The VHPi networks were consolidated from (1) drug and target protein interactions from DrugBank (14, 15), (2) consolidated human protein–protein pairwise interactions (Human Interactome, see the Materials and Methods section), and (3) virus–host protein interactions from the National Center for Biotechnology Information (NCBI) HIV-1 Interactions Database (22, 23) and published literature (24, 25, 26) (see the Materials and Methods section; Table S1). Drugs predicted to bind to viral proteins were ranked by calculating the L3 measure of the drug to a viral protein on the VHPi network. From DrugBank, 7,859 drugs (containing approved small molecule drugs, approved biologics, nutraceuticals, and experimental drugs) were considered for which 5,576 (HIV), 4,532 (HBV), 5,313 (HCV), and 5,834 (HPV) drugs were given a non-zero rank based on the L3 measure. To evaluate the performance of ranked drugs, antiviral drugs with at least one human primary target protein were curated from the DrugBank database (See the Materials and Methods section; Table S1). To assess the performance of network, mean scores in predicting antiviral drugs, an ROC curve was generated, which resulted in AUC values of 0.69, 0.59, 0.78, and 0.67 for HIV, HBV, HCV, and HPV, respectively (Fig 2A). This showed that network mean scores were



**Figure 2. Method validation in HIV, hepatitis B virus, hepatitis C virus, and HPV.**

(A) Receiver operating characteristic curve evaluating predictive power of network-based approach in ranking known drugs in HIV, hepatitis B virus, hepatitis C virus, and HPV. (B) Box and whisker plot of network mean scores generated for all drugs with a non-zero value and for drugs identified by the sequence similarity approach for HIV. The sequence similarity approach predicts drugs with high network mean scores. (C) Virus–host–similarity interaction network representation of drugs predicted by both methods for HIV.

predictive of effective drugs and that other highly ranked drugs represent candidates for further study (Table S2).

To reduce the list of predicted drugs, we applied the sequence similarity-based approach described above to the four viruses. BLASTP searches were performed on unique viral protein sequences to find homologous proteins (see the Materials and Methods section). Human proteins homologous to the viral protein sequences were identified for HIV, but not the other viruses. Drugs known to target the homologous proteins were predicted to also bind the HIV viral proteins. Unlike the network-based method where the predictions provide a continuous score for each drug, the predictions made by this complementary method are binary, that is, yes or no. Interestingly, we observed that drugs predicted through sequence-based similarity tended to have high network mean scores as predicted by the network-based method ( $P$ -value =  $3.33 \times 10^{-22}$ ) (Fig 2B). This observation confirmed the complementary nature of the two approaches. The VHSi network interactions of predicted drugs supported by both methodologies for HIV viral proteins are depicted in Fig 2C. No drug was found through BLAST search for HBV, HCV, and HPV.

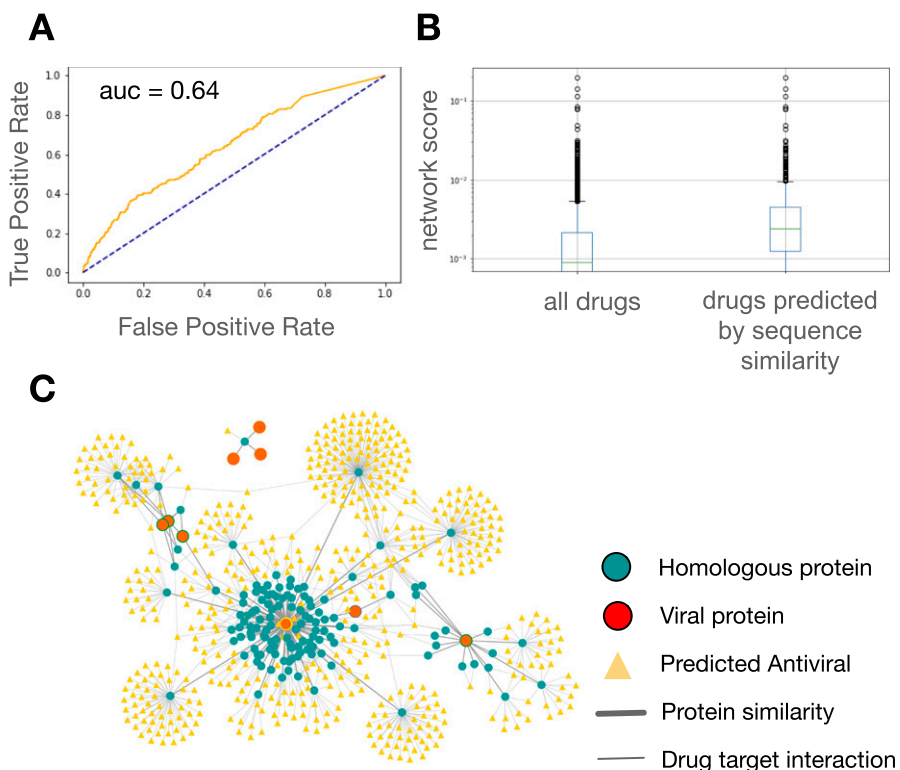
### COVID-19: prediction and in silico validation of candidate antiviral drugs

The SARS-CoV-2 VHPi network was consolidated to identify drugs that are predicted to bind viral proteins based on link prediction. From DrugBank, 7,859 drugs were considered and 4,486 were given a ranking based on the L3 measure of the drug to a viral protein on the VHPi network. The highest-ranked drug for each viral protein

type (structural, non-structural, and accessory) is shown in Fig S1. The antimalarial drug arteminol was predicted to target the nucleocapsid structural protein, the tyrosine kinase inhibitor fostatinib used to treat chronic immune thrombocytopenia was predicted to target the non-structural protein Nsp13, and the coenzyme NADH was predicted to bind to the viral protein ORF9C.

As there are currently no approved drugs for COVID-19, we used a list of drugs currently in clinical trials for preliminary validation (Table S3). A list of 322 drugs in clinical trials for COVID-19 (as of 1 January 2021) was curated from ClinicalTrials.gov as a surrogate for effective drugs (see the Materials and Methods section). The drug rankings provided by the network-based approach were able to predict the drugs in clinical trials with an AUC of 0.64 (Fig 3A).

To refine the drug predictions, sequence-based similarity analyses of SARS-CoV-2 viral proteins were performed. To facilitate further in silico methodology validation, the similarity analyses were restricted to those proteins for which a structure was available in the Protein Data Bank (PDB). Similar to HIV, the drugs predicted by sequence similarity tended to have high scores as predicted by the network-based method ( $P$ -value =  $2.57 \times 10^{-38}$ ) (Fig 3B). The interaction patterns of the human proteins with sequence homology to SARS-CoV-2 viral proteins were visualized on the VHSi network (Fig 3C). The network visualization represents 569 SARS-CoV-2 candidate antiviral drugs predicted by the sequence similarity method that had a non-zero network mean score according to the network-based method (Table S4). Among these 569 predicted drugs, 37 of them have been included in clinical trials for SARS-CoV-2 (AUC = 0.75,  $P$ -value  $3.21 \times 10^{-3}$ ). Gene ontology analysis of human proteins that are homologous to SARS-CoV-2 proteins indicated an



**Figure 3. Method implementation on SARS-CoV-2.** (A) Receiver operating characteristic curve evaluating predictive power of network-based approach in ranking drugs currently under investigation for COVID-19 in clinical trials. (B) Box and whisker plot of network mean scores generated for all drugs with a non-zero value and for drugs identified by the sequence similarity approach. (C) Visualization of predicted antiviral drugs and SARS-CoV-2 Protein Data Bank structures on the virus-host-similarity interaction network.

enrichment for proteins involved in neurotransmission and mitochondrial function (Table S5).

Computational prediction of drug–target interactions has become an essential step in the drug discovery process. The BindScope web application for active-inactive classification of compounds based on deep convolutional neural networks (27), was used to evaluate the binding affinity of the candidate drug–protein pairs predicted in this study. The interactions are scored based on probability where values close to one imply strong binding affinity and those close to zero imply low binding affinity. Binding scores were determined for the drug–targets pairs predicted by both the network-based and sequence similarity methods introduced in this study (Table 1). The results showed that 70% of the predicted drug–virus pairs have a binding score over 0.6.

## Discussion

Using network topological and biological properties of viral proteins, two complementary methods were developed for identification of candidate antiviral therapies. Host proteins that interact with viral proteins were studied by mapping the proteins onto a consolidated Human Interactome map of PPIs, observing their underlying interaction patterns, and studying their biological properties. Four well-studied viruses were used to validate the drug prediction methods. Candidate antivirals ranked by the network-based method predicted known HIV, HBC, HCV, and HPV drugs with AUC values of 0.69, 0.59, 0.78, and 0.67, respectively, reflecting that the network mean scores were predictive of effective drugs. Finally, the novel SARS-CoV-2 drug–virus protein interactions that were predicted by both methods were validated *in silico* using BindScope and resulted in a 70% success rate of identifying candidate antiviral therapies that bind directly to viral proteins with high affinity. This suggests that the methods described herein have the potential to identify candidate drug repurposing opportunities that directly bind to viral proteins.

Numerous antiviral drug repurposing candidates identified by this study were predicted to bind to viral proteins using BindScope. The application of deep neural networks through tools such as BindScope has opened a new path to perform molecular docking using methods that have been previously trained and validated to predict drug–target interactions. Many of the candidate antiviral drugs were predicted to target SARS-CoV-2, which reflects the ability of the methods to identify promiscuous drugs. However, inclusion of the sequence similarity-based method eliminated many drugs that bind indiscriminately and lack specificity for the particular virus of interest.

Gene ontology analysis of the proteins homologous to SARS-CoV-2 proteins identified enrichment of terms associated with mitochondria and synaptic transmission. Previous studies suggested a link between SARS-CoV-2 pathogenesis and manipulation of mitochondrial function (16, 28, 29). Homology to human mitochondrial proteins may enable viral proteins to bind to and manipulate mitochondrial protein function to promote viral entry and modulate host responses to viral infection (29). The enrichment of gene ontology terms associated with neurological processes is consistent with the neuro-invasive potential of SARS-CoV-2 (30, 31, 32). Patients with COVID-19 show signs of neurologic involvement including headache, loss of taste and smell, or, less frequently, encephalopathy and acute cerebrovascular disease (33, 34). The homology between human proteins and those encoded by SARS-CoV-2, suggests that, as is common among viruses, mimicry of host proteins can promote infection and pathogenesis. Although the similarity between viral and host proteins can make it difficult for the immune system to recognize and clear invading pathogens, this same characteristic could be exploited to identify drug repurposing opportunities.

Among the top predicted HIV-1 drug/target combinations are drugs approved by the FDA for treatment of HIV (etravirine, zidovudine, and lamivudine) and have completed Phase III clinical trials (bictegravir) (35). A fifth drug, capravirine, was predicted in this study to target HIV-1 Gag protein; however, its development ceased because of safety concerns and because it did not provide sufficiently substantial improvement over already available treatment options. Among the top predicted SARS-CoV-2-specific drug repurposing candidates were fostamatinib, glutathione,

**Table 1. Binding scores for drug–targets pairs predicted by network-based and sequence similarity methods.**

Drug name	Drugbank ID	Viral protein name	Viral protein target PDB	Network mean score	BindScope score
Arteminol	<a href="#">DB11638</a>	NSP8	7BTF_chainB	0.51001897	0.951099694
NADH	<a href="#">DB00157</a>	NSP13	7BTF_chainC	0.35027915	0.908634782
Fostamatinib	<a href="#">DB12010</a>	NSP13	7BTF_chainC	1.23132454	0.878183186
Arteminol	<a href="#">DB11638</a>	N	<a href="#">6YI3</a>	0.38998956	0.793080389
Fostamatinib	<a href="#">DB12010</a>	ORF9b	<a href="#">6Z4U</a>	0.57350959	0.535310388
NADH	<a href="#">DB00157</a>	ORF8	QHD43422	0.43256958	0.407533526
Fostamatinib	<a href="#">DB12010</a>	NSP7	<a href="#">6WCF</a>	0.31270554	0.27964434
Fostamatinib	<a href="#">DB12010</a>	NSP1	QHD43415_1	0.42845161	0.1326
NADH	<a href="#">DB00157</a>	NSP7	<a href="#">6WCF</a>	0.46385305	0.034212183
Fostamatinib	<a href="#">DB12010</a>	NSP2	QHD43415_2	0.39158755	0.0123
Fostamatinib	<a href="#">DB12010</a>	ORF8	QHD43422	0.4730801	0.01216421
Fostamatinib	<a href="#">DB120109</a>	N	<a href="#">6YI3</a>	0.30990618	0.003463759

Colors indicate strength of binding predictions from BindScope from poor (red) to good (green).



and valproic acid. Fostamatinib, arteminol, and bosutinib were identified in other bioinformatic studies as a potential treatment option to target human proteins associated with COVID-19 (36, 37). Furthermore, fostamatinib is used to treat chronic immune thrombocytopenia, a condition associated with severe COVID-19 (38). Glutathione is an antioxidant that helps defend against oxidative damage of cells from reactive oxygen species and regulates metabolic pathways vital to whole-body homeostasis (39, 40). Glutathione deficiency in patients has been linked to severe manifestations of COVID-19, which has been attributed to indirect effects of glutathione deficiency on enhancing SARS-CoV-2-induced lung damage (41). The results of this study suggest that glutathione may also interact with viral proteins. Valproic acid is a short chain fatty acid with antiviral activity against several viruses (42, 43) and might also be effective against SARS-CoV-2 (44, 45, 46). Therefore, the network- and sequence similarity-based methods identified compounds that may limit viral activities.

Viral proteins often share common tertiary folds and short segments of protein sequence homology with proteins encoded by the host that they infect. These similarities can be exploited to identify drugs that may limit viral replication. However, subtle differences in amino acid sequences relative to the human protein or mutation of viral genomes can limit the efficacy and affinity of a drug for a viral protein by altering binding affinity or binding site availability. World-wide mutations of SARS-CoV-2 have been studied extensively in other publications, and the mutation rate of the virus is considered to be low (47, 48, 49, 50). Mutations that result in amino acid changes are the most common mutation type and every viral protein has at least one frequently recurring mutation (48). Preclinical testing of any candidate viral protein drug target should include assessments of these common protein sequence variations. The SARS-CoV-2 protein with the top-ranked drug repurposing opportunities was Nsp15. Nsp15 is of particular interest as a drug target because it is an endoribonuclease common to all coronaviruses, is necessary for SARS-CoV-2 replication and interferes with host immune responses (51, 52, 53, 54, 55, 56, 57).

Many of the predicted drug repurposing candidates identified by these methods were molecules that promiscuously bind to many proteins and are required for a sizable fraction of all human enzymatic activities (e.g., NADH, zinc, pyridoxal phosphate, and serine) (58, 59, 60, 61). The ability to identify promiscuous drugs is both the greatest strength and greatest weakness of the methods described herein. Therefore, experimental evidence is required to assess whether any of the candidate antiviral drugs reported in this study possess the ability to affect the SARS-CoV-2 replication cycle. This work does not directly test the efficacy of these drugs in experimental models or in clinical trials. It does, however, identify drugs that represent promising repurposing candidates for further study.

## Materials and Methods

### Human interactome

The Human Interactome was consolidated as previously described (62, 63) from 21 public databases containing different types of experimentally derived PPIs data:

1. Binary PPIs, derived from high-throughput yeast-two hybrid (Y2H) experiments (HI-Union [2016] (64)), 3D protein structures

(Interactome3D (65), Instruct (66), Insider (67)) or literature curation (PINA [2014] (68), MINT [2019] (69), LitBM17 [2013] (64), Interactome3D, Instruct, Insider, BioGrid [2019] (70), HINT [2019] (71), HIPPIE (72), APID (73), InWeb (74), IntAct (75))

2. PPIs identified by affinity purification followed by mass spectrometry present in BioPlex2 (2017) (76), QUBIC (77), CoFrac (78), HINT, HIPPIE, APID, LitBM17, and InWeb
3. Kinase-substrate interactions from KinomeNetworkX (79) and PhosphoSitePlus (80)
4. Signaling interactions from SignalLink (2019) (81) and InnateDB (2019) (82); and
5. Regulatory interactions derived by the ENCODE consortium (2012).

The curated list of molecular interactions provided by Alonso-López et al (73) was used for differentiating binary interactions among the experimental methods present in the literature curation databases. For InWeb, interactions with a curation score <0.175 (75<sup>th</sup> percentile) were not considered. All proteins were mapped to their corresponding Entrez ID (NCBI), with unmapped proteins removed. The resulting interactome includes 18,505 proteins and 327,924 interactions.

The network analyses were limited to the largest connected component, containing 18,446 proteins and 322,159 interactions.

### Viral-host interactions curation and analysis

The map of virus-host interactions for SARS-CoV-2 was generated using an affinity purification mass spectrometry experiment (16). Virus-host interactions with high-confidence scores were selected, resulting in 332 interactions among 26 viral proteins and 332 unique human proteins (Table S3).

The entire list of submitted HIV-1 virus-host interactions was downloaded from NCBI website (Table S1).

As the network-based link prediction method was based on physical interactions of the nodes (proteins), interactions were limited to physical interaction annotations only. The interaction types considered as physical included:

1. Physical binding interactions: interacts with, associates with, complexes with, fractionates with, interacts, binds, and
2. Substrate-enzyme interactions: deglycosylated, glycosylated by, modified by, processed by, sulfated by, methylated by, palmitoylated by, is polyubiquitinated by, phosphorylated by, recruited by, ubiquitinated by, isomerized by, myristoylated by, acetylated by, sumoylated by, phosphorylates, acetylates, dephosphorylated by, and ubiquitinates.

Virus-host interactions included in analyses had at least two sources of evidence (<https://www.ncbi.nlm.nih.gov/genome/viruses/retroviruses/hiv-1/interactions/>). Virus-host interactions for HBV, HCV, and HPV were as reported in the literature (24, 25, 26) (Table S1).

### Method 1: link prediction method (L3)

The link prediction problem between two nodes describes the existence of network paths of length three (L3) between the two nodes. To account for biases resulting from high-degree nodes (i.e., those with a large number of connections on

the Human Interactome), a degree-normalized version of L3 was used where the score of each drug–target pair was penalized by the degree of the nodes in the path of length three between them:

$$P_{d,t} = \sum_{ij} \frac{a_{d,i} a_{i,j} a_{j,t}}{\sqrt{k_i k_j}} \quad (1)$$

Where  $k_i$  represents the degree of node  $i$  and  $a_{i,j} = 1$  if nodes  $i$  and  $j$  interact, and zero otherwise.

### Drug protein interaction (DPI) data sources

The protein targets for the drugs evaluated in this study were obtained from DrugBank Release Version 5.1.6 (<https://github.com/emreg00/drugbox>) (14, 15). Only drugs with at least one reported target were included in this analysis. For network-based rankings, drugs were required to have at least one protein target in the human interactome (6,128 drugs in total).

### Curation of COVID-19 drugs in clinical trial and known HIV drugs

Using the [ClinicalTrials.gov](https://www.clinicaltrials.gov) Advanced Search tool, all studies with the condition or disease corresponding to either COVID-19, SARS-CoV-2, or 2019-nCoV were retrieved. Condition terms were those designated by [ClinicalTrials.gov](https://www.clinicaltrials.gov) to include all COVID-19 studies. Similar to the clinical trials curation procedure from a previous study (5 Preprint), we selected clinical studies of interventional type, and whose first submission year was not earlier than 2020. Studies that are either terminated, suspended, no longer available or withdrawn were excluded from the analysis. For each remaining study, drug names were retrieved by parsing the “intervention\_name” and “mesh\_term” fields corresponding to interventions used in a study. To search these fields, we manually constructed a list of regular expressions containing synonyms, possible misspellings or non-English names of drugs. Drugs presented in the “intervention\_name” and “mesh\_term” fields were then checked against the compiled list of regular expressions, and each study was mapped to a list of drugs used in it. Finally, we compiled a list of potential COVID-19 drugs that were used in clinical trials at least once for all studies submitted as of 1 January 2021.

The list of known drugs for HIV were selected manually from the literature and the corresponding targets were curated from DrugBank. A list of 54 “known HIV drugs” was generated from approved HIV drugs and drugs currently in late stage clinical trials (Phase II and Phase III) curated from the DrugBank database on 22 May 2020. A distinction was made between drugs with primary targets (polypeptides), which are thought to modulate the therapeutically beneficial effects of each drug, and drugs with secondary targets (enzymes, transporters, and carriers), which mediate pharmacologically relevant effects (83). Of the 54 curated known HIV drugs, 27 had at least one human primary target. Drugs targeting HBV, HCV, and HPV were obtained from DrugBank.

### Method 2: curation of protein FASTA sequences from PDB and BLAST sequence similarity analysis

160 SARS-CoV-2 protein structures and corresponding amino acid sequences were obtained from the PDB. Redundant amino acid sequences were removed, resulting in 149 unique sequences.

HIV-1, HBV, HCV, and HPV protein structures and corresponding amino acid sequences were obtained from the PDB. After filtering the “Entity Macromolecule Type” to “Polypeptide only” to remove nucleic acid sequences, a list of 3,421 amino acid sequences was generated for HIV-1. A list of 30, 238, 47 amino acid sequences was generated for HBV, HCV, and HPV, respectively. BLAST search found only a few homologous proteins and none of them are known drug targets.

Because of the large number of existing HIV amino acid sequences, redundant sequences, as measured by pairwise sequence similarities, were removed. The 3,421 sequences were split into four batches, each with a similar number of sequences, to generate a pairwise distance matrix for each batch using the MSA package in R (84). An empirical cutoff of 0.8 was selected from the distribution plot and used to remove one of the two paired sequences with distance below 0.8 (high similarity). After processing all four batches, the total number of amino acid sequences was reduced to 363. MSA was run again on the reduced list of 363 sequences from all four batches. Using the same distance cutoff of 0.8, the final number of unique sequences was reduced to 199.

The protein FASTA sequences were used as input for the similarity searching program BLASTP from NCBI with standard general parameters and “max target sequences” set to 20,000 non-redundant (nr) protein sequences. Because finding distantly related protein sequences is more challenging than finding closely related sequences, the BLOSUM62 matrix (85) was used. NCBI provides a unique accession ID for each homolog protein found in the BLAST search results. NCBI accession IDs were converted to a UniProt ID using the ID Mapping database ([https://ftp.expasy.org/databases/uniprot/current\\_release/knowledgebase/idmapping/](https://ftp.expasy.org/databases/uniprot/current_release/knowledgebase/idmapping/)).

BLASTP was performed with the final list of unique viral protein sequences querying human (specifying homo sapiens in “Organism”) and non-human (excluding homo sapiens in “Organism”) sequences to find proteins with similarity to the viral proteins (see Supplemental Data 1).

### Evaluating performance of drug predictions

Each drug was assigned by method 1 (network-based similarity method, Fig 1) a score specific to each viral protein. The drug’s mean score across all viral proteins was used as the final score and method validation. Performance of the drug predictions was evaluated by area under the receiver operating characteristic curve (AUC), which was calculated by comparing network mean scores provided by network-based method 1 and the label indicating whether the predicted drug was used in a clinical trial for COVID-19 or a known antiviral drug with at least one human primary target (see Supplemental Data 1).

### DPI scores

BindScope predicts DPIs for a given protein PDB structure and small molecule (e.g., drug) (27). The binding pocket and ligand pose are featurized by voxelizing according to different pharmacophoric-like properties, then trained on a three-dimensional convolutional neural network to predict the binding likelihood. BindScope was

trained on ligand conformations from the latest iteration of the Directory of Useful Decoys, Enhanced (DUD-E) (86), and docked using Smina (87). In their article, the authors evaluated BindScope performance using AUC and per receptor fivefold cross-validation. The yielded AUC values ranged from 0.496 to 0.997 with 0.885 as the average.

To use BindScope, PDB structures were downloaded from the PDB database, water molecules were removed and one polypeptide chain was extracted for the analysis. The protein hydrogens were added to the PDB structure in BindScope. Ligand structures were downloaded from DrugBank in SDF format.

## Data Availability

<https://github.com/ScipherMedicine/NetworkMedicine>.

## Supplementary Information

Supplementary Information is available at <https://doi.org/10.26508/lsa.202000904>.

## Acknowledgements

We thank Mohamed Alshalalfa, Istvan A Kovacs, Razzi Movassaghi, and Onur Varol for their insightful consultations, Enrico Maiorino, and Asher Ameli for converting the network-based script from C to python and Alice Grishchenko for helping generate the flowchart visualization.

## Author Contributions

M Wang: data curation, formal analysis, validation, visualization, writing—original draft, review, and editing.

JB Withers: data curation, validation, investigation, and writing—original draft, review, and editing.

P Ricchiuto: data curation, formal analysis, validation, and writing—original draft, review, and editing.

I Voitalov: data curation, investigation, methodology, and writing—review and editing.

M McAnally: data curation.

HN Sanchez: data curation.

A Saleh: resources and writing—original draft.

VR Akmaev: resources and writing—original draft.

SD Ghiassian: conceptualization, data curation, formal analysis, supervision, investigation, methodology, and writing—original draft, review, and editing.

## Conflict of Interest Statement

M Wang, JB Withers, I Voitalov, M McAnally, HN Sanchez, A Saleh, VR Akmaev, and SD Ghiassian are full-time employees and shareholders of Scipher Medicine Corporation. P Ricchiuto is a consultant of Scipher Medicine Corporation.

## References

1. Lou Z, Sun Y, Rao Z (2014) Current progress in antiviral strategies. *Trends Pharmacol Sci* 35: 86–102. doi:10.1016/j.tips.2013.11.006
2. Everts M, Cihlar T, Bostwick JR, Whitley RJ (2017) Accelerating drug development: Antiviral therapies for emerging viruses as a model. *Annu Rev Pharmacol Toxicol* 57: 155–169. doi:10.1146/annurev-pharmtox-010716-104533
3. Lima WG, Brito JCM, Overhage J, Nizer W (2020) The potential of drug repositioning as a short-term strategy for the control and treatment of COVID-19 (SARS-CoV-2): A systematic review. *Arch Virol* 165: 1729–1737. doi:10.1007/s00705-020-04693-5
4. Rameshrad M, Ghafoori M, Mohammadpour AH, Nayeri MJD, Hosseinzadeh H (2020) A comprehensive review on drug repositioning against coronavirus disease 2019 (COVID19). *Naunyn Schmiedebergs Arch Pharmacol* 393: 1137–1152. doi:10.1007/s00210-020-01901-6
5. Gysi DM, Do Valle I, Zitnik M, Ameli A, Gan X, Varol O, Sanchez H, Baron RM, Ghiassian D, Loscalzo J, et al (2020) Network medicine framework for identifying drug repurposing opportunities for COVID-19. ArXiv Preprint posted April 15, 2020.
6. Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB (2020) Pharmacologic treatments for coronavirus disease 2019 (COVID-19): A review. *JAMA* 323: 1824–1836. doi:10.1001/jama.2020.6019
7. Fragkou PC, Belhadi D, Peiffer-Smadja N, Moschopoulos CD, Lescure FX, Janocha H, Karofylakis E, Yazdanpanah Y, Mentre F, Skevaki C, et al (2020) Review of trials currently testing treatment and prevention of COVID-19. *Clin Microbiol Infect* 26: 988–998. doi:10.1016/j.cmi.2020.05.019
8. Kenney AD, Dowdle JA, Bozzacco L, McMichael TM, St Gelais C, Panfil AR, Sun Y, Schlesinger LS, Anderson MZ, Green PL, et al (2017) Human genetic determinants of viral diseases. *Annu Rev Genet* 51: 241–263. doi:10.1146/annurev-genet-120116-023425
9. Houldcroft CJ, Beale MA, Breuer J (2017) Clinical and biological insights from viral genome sequencing. *Nat Rev Microbiol* 15: 183–192. doi:10.1038/nrmicro.2016.182
10. Ledford H (2020) Dozens of coronavirus drugs are in development—What happens next? *Nature* 581: 247–248. doi:10.1038/d41586-020-01367-9
11. Haupt VJ, Daminelli S, Schroeder M (2013) Drug promiscuity in PDB: Protein binding site similarity is key. *PLoS One* 8: e65894. doi:10.1371/journal.pone.0065894
12. Liu H, Sun J, Guan J, Zheng J, Zhou S (2015) Improving compound-protein interaction prediction by building up highly credible negative samples. *Bioinformatics* 31: i221–i229. doi:10.1093/bioinformatics/btv256
13. Alaimo S, Pulvirenti A, Giugno R, Ferro A (2013) Drug-target interaction prediction through domain-tuned network-based inference. *Bioinformatics* 29: 2004–2008. doi:10.1093/bioinformatics/bt307
14. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, et al (2018) DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res* 46: D1074–D1082. doi:10.1093/nar/gkx1037
15. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, et al (2014) DrugBank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Res* 42: D1091–D1097. doi:10.1093/nar/gkt1068
16. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, O’Meara MJ, Rezelj VV, Guo JZ, Swaney DL, et al (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583: 459–468. doi:10.1038/s41586-020-2286-9
17. Kovacs IA, Luck K, Spirohn K, Wang Y, Pollis C, Schlabach S, Bian W, Kim DK, Kishore N, Hao T, et al (2019) Network-based prediction of protein interactions. *Nat Commun* 10: 1240. doi:10.1038/s41467-019-09177-y
18. Namata G, Getoor L (2017) Link prediction *Encyclopedia of Machine Learning and Data Mining*. Boston, MA: Springer: 753–758. doi:10.1007/978-1-4899-7687-1\_486
19. Keskin O, Tuncbag N, Gursoy A (2016) Predicting protein-protein interactions from the molecular to the proteome level. *Chem Rev* 116: 4884–4909. doi:10.1021/acs.chemrev.5b00683



20. Granovetter M (1983) The strength of weak ties: A network theory revisited. *Sociol Theory* 1: 201–233. doi:10.2307/202051
21. Fuxman Bass JI, Diallo A, Nelson J, Soto JM, Myers CL, Walhout AJ (2013) Using networks to measure similarity between genes: Association index selection. *Nat Methods* 10: 1169–1176. doi:10.1038/nmeth.2728
22. Fu W, Sanders-Beer BE, Katz KS, Maglott DR, Pruitt KD, Ptak RG (2009) Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Res* 37: D417–D422. doi:10.1093/nar/gkn708
23. Ptak RG, Fu W, Sanders-Beer BE, Dickerson JE, Pinney JW, Robertson DL, Rozanov MN, Katz KS, Maglott DR, Pruitt KD, et al (2008) Cataloguing the HIV type 1 human protein interaction network. *AIDS Res Hum Retroviruses* 24: 1497–1502. doi:10.1089/aid.2008.0113
24. Wu ZJ, Zhu Y, Huang DR, Wang ZQ (2010) Constructing the HBV-human protein interaction network to understand the relationship between HBV and hepatocellular carcinoma. *J Exp Clin Cancer Res* 29: 146. doi:10.1186/1756-9966-29-146
25. Kwofle SK, Schaefer U, Sundararajan VS, Bajic VB, Christoffels A (2011) HCVpro: Hepatitis C virus protein interaction database. *Infect Genet Evol* 11: 1971–1977. doi:10.1016/j.meegid.2011.09.001
26. Farooq QUA, Shaikat Z, Zhou T, Aiman S, Gong W, Li C (2020) Inferring Virus-Host relationship between HPV and its host Homo sapiens using protein interaction network. *Sci Rep* 10: 8719. doi:10.1038/s41598-020-65837-w
27. Skalic M, Martinez-Rosell G, Jimenez J, De Fabritiis G (2019) PlayMolecule BindScope: Large scale CNN-based virtual screening on the web. *Bioinformatics* 35: 1237–1238. doi:10.1093/bioinformatics/bty758
28. Gatti P, Ilamathi HS, Todkar K, Germain M (2020) Mitochondria targeted viral replication and survival strategies-prospective on SARS-CoV-2. *Front Pharmacol* 11: 578599. doi:10.3389/fphar.2020.578599
29. Singh KK, Chaubey G, Chen JY, Suravajhala P (2020) Decoding SARS-CoV-2 hijacking of host mitochondria in COVID-19 pathogenesis. *Am J Physiol Cell Physiol* 319: C258–C267. doi:10.1152/ajpcell.00224.2020
30. Zhang BZ, Chu H, Han S, Shuai H, Deng J, Hu YF, Gong HR, Lee AC, Zou Z, Yau T, et al (2020) SARS-CoV-2 infects human neural progenitor cells and brain organoids. *Cell Res* 30: 928–931. doi:10.1038/s41422-020-0390-x
31. Li YC, Bai WZ, Hashikawa T (2020) The neuroinvasive potential of SARS-CoV2 may play a role in the respiratory failure of COVID-19 patients. *J Med Virol* 92: 552–555. doi:10.1002/jmv.25728
32. Hu J, Jolkkonen J, Zhao C (2020) Neurotropism of SARS-CoV-2 and its neuropathological alterations: Similarities with other coronaviruses. *Neurosci Biobehav Rev* 119: 184–193. doi:10.1016/j.neubiorev.2020.10.012
33. Ellul MA, Benjamin L, Singh B, Lant S, Michael BD, Easton A, Kneen R, Defres S, Sejvar J, Solomon T (2020) Neurological associations of COVID-19. *Lancet Neurol* 19: 767–783. doi:10.1016/s1474-4422(20)30221-0
34. Ahmed MU, Hanif M, Ali MJ, Haider MA, Kherani D, Memon GM, Karim AH, Sattar A (2020) Neurological manifestations of COVID-19 (SARS-CoV-2): A review. *Front Neurol* 11: 518. doi:10.3389/fneur.2020.00518
35. Wohl DA, Yazdanpanah Y, Baumgarten A, Clarke A, Thompson MA, Brinson C, Hagins D, Ramgopal MN, Antinori A, Wei X, et al (2019) Bictegravir combined with emtricitabine and tenofovir alafenamide versus dolutegravir, abacavir, and lamivudine for initial treatment of HIV-1 infection: Week 96 results from a randomised, double-blind, multicentre, phase 3, non-inferiority trial. *Lancet HIV* 6: e355–e363. doi:10.1016/s2352-3018(19)30077-3
36. Saha S, Halder AK, Bandyopadhyay SS, Chatterjee P, Nasipuri M, Basu S (2020) Is Fostamatinib a possible drug for COVID-19: A computational study. *OSF Preprints Web* doi:10.31219/osf.io/7hgpj (Preprint posted May 11, 2020)
37. Stolfi P, Manni L, Soligo M, Vergni D, Tieri P (2020) Designing a network proximity-based drug repurposing strategy for COVID-19. *Front Cell Dev Biol* 8: 545089. doi:10.3389/fcell.2020.545089
38. Lippi G, Plebani M, Henry BM (2020) Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A meta-analysis. *Clinica Chim Acta* 506: 145–148. doi:10.1016/j.cca.2020.03.022
39. Wu G, Fang YZ, Yang S, Lupton JR, Turner ND (2004) Glutathione metabolism and its implications for health. *J Nutr* 134: 489–492. doi:10.1093/jn/134.3.489
40. Forman HJ, Zhang H, Rinna A (2009) Glutathione: Overview of its protective roles, measurement, and biosynthesis. *Mol Aspects Med* 30: 1–12. doi:10.1016/j.mam.2008.08.006
41. Polonikov A (2020) Endogenous deficiency of glutathione as the most likely cause of serious manifestations and death in COVID-19 patients. *ACS Infect Dis* 6: 1558–1562. doi:10.1021/acsinfectdis.0c00288
42. Crespillo AJ, Praena B, Bello-Morales R, Lerma L, Vazquez-Calvo A, Martin-Acebes MA, Tabares E, Sobrino F, Lopez-Guerrero JA (2016) Inhibition of herpes virus infection in oligodendrocyte cultured cells by valproic acid. *Virus Res* 214: 71–79. doi:10.1016/j.virusres.2016.01.009
43. Vazquez-Calvo A, Saiz JC, Sobrino F, Martin-Acebes MA (2011) Inhibition of enveloped virus infection of cultured cells by valproic acid. *J Virol* 85: 1267–1274. doi:10.1128/jvi.01717-10
44. Unal G, Turan B, Balcioglu YH (2020) Immunopharmacological management of COVID-19: Potential therapeutic role of valproic acid. *Med Hypotheses* 143: 109891. doi:10.1016/j.mehy.2020.109891
45. Chiquete E, Toapanta-Yanchapaxi L, Cantu-Brito C (2020) Methods of an open-label proof-of-concept trial of intravenous valproic acid for severe COVID-19. medRxiv 2020.04.26.20079988. doi:10.101/2020/04/26/2-79988 (Preprint posted May 1, 2020).
46. Bhavesh NS, Patra A (2020) Virtual screening and molecular dynamics simulation suggest valproic acid Co-A could bind to SARS-CoV2 RNA depended RNA polymerase. *Preprints* 2020030393. 10.20944/preprints202003.0393.v1 (Preprint posted March 26, 2020).
47. van Dorp L, Richard D, Tan CCS, Shaw LP, Acman M, Balloux F (2020) No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nat Commun* 11: 5986. doi:10.1038/s41467-020-19818-2
48. Mercatelli D, Giorgi FM (2020) Geographic and genomic distribution of SARS-CoV-2 mutations. *Front Microbiol* 11: 1800. doi:10.3389/fmicb.2020.01800
49. Kaushal N, Gupta Y, Goyal M, Khaiboullina SF, Baranwal M, Verma SC (2020) Mutational frequencies of SARS-CoV-2 genome during the beginning months of the outbreak in USA. *Pathogens* 9: 565. doi:10.3390/pathogens9070565
50. Turakhia Y, De Maio N, Thornlow B, Gozashli L, Lanfear R, Walker CR, Hinrichs AS, Fernandes JD, Borges R, Slodkovic G, et al (2020) Stability of SARS-CoV-2 phylogenies. *PLoS Genet* 16: e1009175. doi:10.1371/journal.pgen.1009175
51. Yuen CK, Lam JY, Wong WM, Mak LF, Wang X, Chu H, Cai JP, Jin DY, To KK, Chan JF, et al (2020) SARS-CoV-2 nsp13, nsp14, nsp15 and orf6 function as potent interferon antagonists. *Emerg Microbes Infect* 9: 1418–1428. doi:10.1080/22221751.2020.1780953
52. Krishnan DA, Sangeetha G, Vajravijayan S, Nandhagopal N, Gunasekaran K (2020) Structure-based drug designing towards the identification of potential anti-viral for COVID-19 by targeting endoribonuclease NSP15. *Inform Med Unlocked* 20: 100392. doi:10.1016/j.imu.2020.100392
53. Khan RJ, Jha RK, Singh E, Jain M, Amara GM, Singh RP, Muthukumar J, Singh AK (2020) Identification of promising antiviral drug candidates against non-structural protein 15 (NSP15) from SARS-CoV-2: An in silico assisted drug-repurposing study. *J Biomol Struct Dyn* 1–11. doi:10.1080/07391102.2020.1814870
54. Sharma A, Goyal S, Yadav AK, Kumar P, Gupta L (2020) In-silico screening of plant-derived antivirals against main protease, 3CL(pro) and endoribonuclease, NSP15 proteins of SARS-CoV-2. *J Biomol Struct Dyn* 1–15. doi:10.1080/07391102.2020.1808077
55. Kumar S, Kashyap P, Chowdhury S, Kumar S, Panwar A, Kumar A (2020) Identification of phytochemicals as potential therapeutic agents that binds to Nsp15 protein target of coronavirus (SARS-CoV-2) that are capable of inhibiting virus replication. *Phytomedicine* 153317. doi:10.1016/j.phymed.2020.153317
56. Saeed M, Saeed A, Alam MJ, Alreshidi M (2020) Identification of persuasive antiviral natural compounds for COVID-19 by targeting

- endoribonuclease NSP15: A structural-bioinformatics approach. *Molecules* 25: 5657. doi:[10.3390/molecules25235657](https://doi.org/10.3390/molecules25235657)
57. Batool A, Bibi N, Amin F, Kamal MA (2020) Drug designing against NSP15 of SARS-COV2 via high throughput computational screening and structural dynamics approach. *Eur J Pharmacol* 892: 173779. doi:[10.1016/j.ejphar.2020.173779](https://doi.org/10.1016/j.ejphar.2020.173779)
  58. McCall KA, Huang C, Fierke CA (2000) Function and mechanism of zinc metalloenzymes. *J Nutr* 130: 1437S–1446S. doi:[10.1093/jn/130.5.1437s](https://doi.org/10.1093/jn/130.5.1437s)
  59. Percudani R, Peracchi A (2003) A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Rep* 4: 850–854. doi:[10.1038/sj.embor.embor914](https://doi.org/10.1038/sj.embor.embor914)
  60. Kalhan SC, Hanson RW (2012) Resurgence of serine: An often neglected but indispensable amino acid. *J Biol Chem* 287: 19786–19791. doi:[10.1074/jbc.r112.357194](https://doi.org/10.1074/jbc.r112.357194)
  61. Canto C, Menzies KJ, Auwerx J (2015) NAD(+) metabolism and the control of energy homeostasis: A balancing act between mitochondria and the nucleus. *Cell Metab* 22: 31–53. doi:[10.1016/j.cmet.2015.05.023](https://doi.org/10.1016/j.cmet.2015.05.023)
  62. Mellors T, Withers JB, Ameli A, Jones A, Wang M, Zhang L, Sanchez HN, Santolini M, Do Valle I, Sebek M, et al (2020) Clinical validation of a blood-based predictive test for stratification of response to tumor necrosis factor inhibitor therapies in rheumatoid arthritis patients. *Netw Syst Med* 3: 91–104. doi:[10.1089/nsm.2020.0007](https://doi.org/10.1089/nsm.2020.0007)
  63. Rolland T, Tasan M, Charlotiaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, et al (2014) A proteome-scale map of the human interactome network. *Cell* 159: 1212–1226. doi:[10.1016/j.cell.2014.10.050](https://doi.org/10.1016/j.cell.2014.10.050)
  64. Luck K, Kim D-K, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-Laborie FJ, Charlotiaux B, et al (2020) A reference map of the human binary protein interactome. *Nature* 580: 402–408. doi:[10.1038/s41586-020-2188-x](https://doi.org/10.1038/s41586-020-2188-x)
  65. Mosca R, Ceol A, Aloy P (2013) Interactome3D: Adding structural details to protein networks. *Nat Methods* 10: 47–53. doi:[10.1038/nmeth.2289](https://doi.org/10.1038/nmeth.2289)
  66. Meyer MJ, Das J, Wang X, Yu H (2013) INstruct: A database of high-quality 3D structurally resolved protein interactome networks. *Bioinformatics* 29: 1577–1579. doi:[10.1093/bioinformatics/btt181](https://doi.org/10.1093/bioinformatics/btt181)
  67. Meyer MJ, Beltran JF, Liang S, Fragoza R, Rumack A, Liang J, Wei X, Yu H (2018) Interactome INSIDER: A structural interactome browser for genomic studies. *Nat Methods* 15: 107–114. doi:[10.1038/nmeth.4540](https://doi.org/10.1038/nmeth.4540)
  68. Cowley MJ, Pinese M, Kassahn KS, Waddell N, Pearson JV, Grimmond SM, Biankin AV, Hautaniemi S, Wu J (2012) PINA v2.0: Mining interactome modules. *Nucleic Acids Res* 40: D862–D865. doi:[10.1093/nar/gkr967](https://doi.org/10.1093/nar/gkr967)
  69. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, Sacco F, Palma A, Nardoza AP, Santonico E, et al (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* 40: D857–D861. doi:[10.1093/nar/gkr930](https://doi.org/10.1093/nar/gkr930)
  70. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L, et al (2015) The BioGRID interaction database: 2015 update. *Nucleic Acids Res* 43: D470–D478. doi:[10.1093/nar/gku1204](https://doi.org/10.1093/nar/gku1204)
  71. Das J, Yu H (2012) HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC Syst Biol* 6: 92. doi:[10.1186/1752-0509-6-92](https://doi.org/10.1186/1752-0509-6-92)
  72. Alanis-Lobato G, Andrade-Navarro MA, Schaefer MH (2017) HIPPIE v2.0: Enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Res* 45: D408–D414. doi:[10.1093/nar/gkw985](https://doi.org/10.1093/nar/gkw985)
  73. Alonso-López D, Campos-Laborie FJ, Gutierrez MA, Lambourne L, Calderwood MA, Vidal M, De Las Rivas J (2019) APID database: Redefining protein-protein interaction experimental evidences and binary interactomes. *Database (Oxford)* 2019: baz005. doi:[10.1093/database/baz005](https://doi.org/10.1093/database/baz005)
  74. Li T, Wernersson R, Hansen RB, Horn H, Mercer J, Slodkovic G, Workman CT, Rigina O, Rapacki K, Staerfeldt HH, et al (2017) A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat Methods* 14: 61–64. doi:[10.1038/nmeth.4083](https://doi.org/10.1038/nmeth.4083)
  75. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N, et al (2014) The MintAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* 42: D358–D363. doi:[10.1093/nar/gkt1115](https://doi.org/10.1093/nar/gkt1115)
  76. Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, Colby G, Gebreab F, Gygi MP, Parzen H, et al (2017) Architecture of the human interactome defines protein communities and disease networks. *Nature* 545: 505–509. doi:[10.1038/nature22366](https://doi.org/10.1038/nature22366)
  77. Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak IA, Weisswange I, Mansfeld J, Buchholz F, et al (2015) A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 163: 712–723. doi:[10.1016/j.cell.2015.09.053](https://doi.org/10.1016/j.cell.2015.09.053)
  78. Wan C, Borgeson B, Phanse S, Tu F, Drew K, Clark G, Xiong X, Kagan O, Kwan J, Bezginov A, et al (2015) Panorama of ancient metazoan macromolecular complexes. *Nature* 525: 339–344. doi:[10.1038/nature14877](https://doi.org/10.1038/nature14877)
  79. Cheng F, Jia P, Wang Q, Zhao Z (2014) Quantitative network mapping of the human kinome interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer therapy. *Oncotarget* 5: 3697–3710. doi:[10.18632/oncotarget.1984](https://doi.org/10.18632/oncotarget.1984)
  80. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E (2015) PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Res* 43: D512–D520. doi:[10.1093/nar/gku1267](https://doi.org/10.1093/nar/gku1267)
  81. Fazekas D, Koltai M, Turei D, Modos D, Palfy M, Dul Z, Zsakai L, Szalay-Beko M, Lenti K, Farkas IJ, et al (2013) Signalink 2: A signaling pathway resource with multi-layered regulatory networks. *BMC Syst Biol* 7: 7. doi:[10.1186/1752-0509-7-7](https://doi.org/10.1186/1752-0509-7-7)
  82. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock RE, Brinkman FS, Lynn DJ (2013) InnateDB: Systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res* 41: D1228–D1233. doi:[10.1093/nar/gks1147](https://doi.org/10.1093/nar/gks1147)
  83. Yera ER, Cleves AE, Jain AN (2011) Chemical structural novelty: On-targets and off-targets. *J Med Chem* 54: 6771–6785. doi:[10.1021/jm200666a](https://doi.org/10.1021/jm200666a)
  84. Bodenhofer U, Bonatesta E, Horejs-Kainrath C, Hochreiter S (2015) msa: An R package for multiple sequence alignment. *Bioinformatics* 31: 3997–3999. doi:[10.1093/bioinformatics/btv494](https://doi.org/10.1093/bioinformatics/btv494)
  85. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* 89: 10915–10919. doi:[10.1073/pnas.89.22.10915](https://doi.org/10.1073/pnas.89.22.10915)
  86. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK (2012) Directory of Useful Decoys, Enhanced (DUD-E): Better ligands and Decoys for better benchmarking. *J Med Chem* 55: 6582–6594. doi:[10.1021/jm300687e](https://doi.org/10.1021/jm300687e)
  87. Koes DR, Baumgartner MP, Camacho CJ (2013) Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J Chem Inf Model* 53: 1893–1904. doi:[10.1021/ci300604z](https://doi.org/10.1021/ci300604z)



**License:** This article is available under a Creative Commons License (Attribution 4.0 International, as described at <https://creativecommons.org/licenses/by/4.0/>).