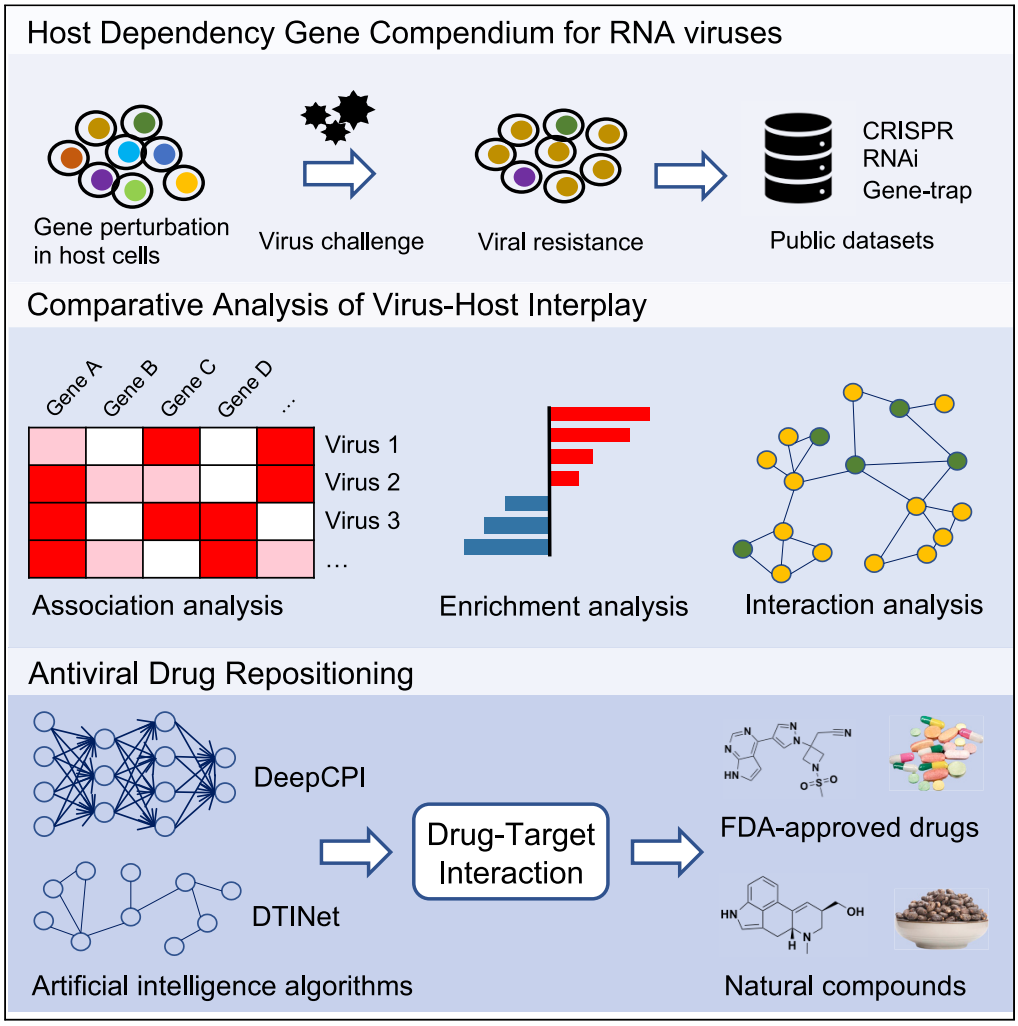
iScience

CellPress
OPEN ACCESS

Article

A computational framework of host-based drug repositioning for broad-spectrum antivirals against RNA viruses



Zexu Li, Yingjia Yao, Xiaolong Cheng, ..., Hu Zhou, Wei Li, Teng Fei

wli2@childrensnational.org (W.L.)
feiteng@mail.neu.edu.cn (T.F.)

HIGHLIGHTS

Host dependency genes for RNA viruses are systematically cataloged

Repositioned drug candidates target host dependency genes for antiviral purpose

Artificial intelligence-based algorithms help to predict drug-target interactions

Prioritized antiviral drug candidates are proposed against RNA viruses

Li et al., iScience 24, 102148
March 19, 2021
https://doi.org/10.1016/j.isci.2021.102148

iScience

CellPress
OPEN ACCESS

Article

A computational framework of host-based drug repositioning for broad-spectrum antivirals against RNA viruses

Zexu Li,[1,2,7] Yingjia Yao,[1,2,7] Xiaolong Cheng,[3,4,7] Qing Chen,[3,4] Wenchang Zhao,[1,2] Shixin Ma,[1,2] Zihan Li,[1,2] Hu Zhou,[5,6] Wei Li,[3,4,*] and Teng Fei[1,2,8,*]

SUMMARY

RNA viruses are responsible for many zoonotic diseases that post great challenges for public health. Effective therapeutics against these viral infections remain limited. Here, we deployed a computational framework for host-based drug repositioning to predict potential antiviral drugs from 2,352 approved drugs and 1,062 natural compounds embedded in herbs of traditional Chinese medicine. By systematically interrogating public genetic screening data, we comprehensively cataloged host dependency genes (HDGs) that are indispensable for successful viral infection corresponding to 10 families and 29 species of RNA viruses. We then utilized these HDGs as potential drug targets and interrogated extensive drug-target interactions through database retrieval, literature mining, and *de novo* prediction using artificial intelligence-based algorithms. Repurposed drugs or natural compounds were proposed against many viral pathogens such as coronaviruses including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), flaviviruses, and influenza viruses. This study helps to prioritize promising drug candidates for in-depth evaluation against these virus-related diseases.

INTRODUCTION

The recent outbreak and spreading of coronavirus 2019 disease (also known as COVID-19) has become a severe public health crisis that threatens not only human health but also social lifestyle and global economy (Zhu et al., 2020). RNA virus termed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the underlying pathogen for COVID-19 (Li et al., 2020). Despite some progresses in early diagnosis and clinical treatment, people still lack consistent and reliable solutions to defeat SARS-CoV-2 and halt COVID-19 pandemic globally (Altay et al., 2020). In addition to SARS-CoV-2 coronavirus, the outbreak of other pathogenic RNA viruses such as coronaviruses of various types (e.g., SARS-CoV and MERS-CoV), flaviviruses (e.g., West Nile virus, Dengue virus and Zika virus), and influenza viruses (e.g., H1N1 and H3N2 stains) also cause severe infectious diseases in human (Petersen et al., 2020; Pierson and Diamond, 2020).

Vaccination is one of the most effective approaches to prevent viral infection by conferring active immunity to the host and helping to establish herd immunity. However, it usually takes years for a successful vaccine to be developed and implemented. To achieve an immediate control of viral disease for the infected patients, therapeutic drug then serves as the primary option and is highly demanded especially for recently emerging pathogens without known therapeutic formula. Encouraging efforts have been made toward antiviral drug development or drug repositioning against the above mentioned RNA viruses and their related diseases (Dighe et al., 2019; Mottin et al., 2018; Zhang et al., 2019; Zumla et al., 2016). Most of these studies focused on various viral genes or proteins that are key mediators to complete the virus life cycle, for instance, targeting spike proteins to block cell entry or inhibiting RNA polymerase to interfere viral gene replication. Virus-centered strategy has been proved feasible in light of the successful development of antiviral drugs in recent years. This approach heavily relies on the specific knowledge about each viral pathogen and its virus-host interplay, which usually requires extensive investigation efforts and is preferable for *de novo* antiviral drug development spanning years of time (De Clercq and Li, 2016). In contrast, drug repositioning or drug repurposing that exploits existing "old" drugs for "new" purposes offers a quick

[1]College of Life and Health Sciences, Northeastern University, Shenyang 110819, People's Republic of China

[2]Key Laboratory of Data Analytics and Optimization for Smart Industry (Northeastern University), Ministry of Education, Shenyang 110819, People's Republic of China

[3]Center for Genetic Medicine Research, Children's National Hospital, 111 Michigan Avenue NW, Washington, DC 20010, USA

[4]Department of Genomics and Precision Medicine, George Washington University, 111 Michigan Avenue NW, Washington, DC 20010, USA

[5]School of Pharmaceutical Sciences, Fujian Provincial Key Laboratory of Innovative Drug Target Research, Xiamen University, Xiamen, Fujian 361102, China

[6]High Throughput Drug Screening Platform, Xiamen University, Xiamen, Fujian 361102, China

[7]These authors contributed equally

[8]Lead contact

*Correspondence: wli2@childrensnational.org (W.L.), feiteng@mail.neu.edu.cn (T.F.)

https://doi.org/10.1016/j.isci.2021.102148

solution and would be practical to respond to emerging contagious diseases, before valid vaccine and *de novo* drugs are available. For COVID-19, several known antiviral drugs or compounds previously designed for other RNA viruses have been proposed and tested in the first place, including Ebola virus-targeting drug remdesivir that demonstrated *in vitro* activity but had unsatisfactory response during following clinical trials (Wang et al., 2020a, 2020b). More rational drug repositioning strategies have been explored recently with the aim to identify potential drugs that can target important viral proteins, given the rapid progresses on SARS-CoV-2 protein structure characterization (Dai et al., 2020; Jin et al., 2020; Wu et al., 2020). However, these approaches usually neglect host effect, and the drugs proposed often exhibit significant *in vitro* activity but with less success *in vivo*.

Here, we interrogated a different drug repositioning strategy for COVID-19 and other notorious RNA virus-related diseases from the host-centered perspective. Viruses require key host genes (or factors) for infection and replication, and these host dependency genes (HDGs) serve as potential targets for drug repurposing. By a comprehensive literature collection and data mining, we cataloged HDGs and revealed their molecular features in virus-host interactions for 29 RNA virus species across 10 viral families. We then employed an integrative drug repositioning approach by combining known drug-target interactions (DTIs) from multiple databases with computational predictions for more potential DTIs. We identified candidate host-targeting drugs and natural compounds with broad-spectrum antiviral potentiality for diseases caused by pathogenic coronaviruses, flaviviruses, and influenza viruses.

RESULTS

Strategic overview of host-centered antiviral drug repositioning

Although many host genes may interplay with viral genes within the host cells, only a few of them are essential for complete infection in a virus-specific manner. Blocking these host essential or dependency genes for viral infection with targeted drugs underlies the principle of host-centered drug repositioning. In the current study, we primarily focused on RNA viruses, especially SARS-CoV-2 and other recently prevalent species (Table S1). The overall workflow of this study is illustrated in Figure 1. Firstly, we sought to systematically catalog the virus-specific HDGs by comprehensively archiving and interrogating published studies that performed functional genetic screens in human cells challenged with RNA viruses (Figure 1A). These works employed multiple genetic perturbation platforms such as gene trap, RNA interference (RNAi), or clustered regularly interspaced palindromic repeats (CRISPR) to identify HDGs whose loss of function renders host resistance to specific viral infection. When genetically perturbed cell pool is challenged by the corresponding virus, the HDG-deficient cells tend to escape from virus-inflicted cell deterioration and positively selected in the final cell pool by which the HDG could be identified. Screening data from 63 independent studies spanning 10 families and 29 species of RNA viruses were collected (Figure 1A; Table S1). With higher priority for *Coronaviridae* due to the COVID-19 pandemic, we additionally performed in-depth literature mining to include individual HDGs identified from 34 *Coronaviridae*-focused studies (Table S1). Notably, we primarily considered studies using human-derived cells or tissues as host systems to better reflect the clinically relevant host response and for appropriate drug repurposing. Next, we performed comparative analysis of the host dependency features across multiple viruses to extract consensus HDGs for the following drug repositioning (Figures 1B and 1C). To establish the targeting relationship between drugs and genes, we not only considered the known DTIs from several related databases (e.g., DGIdb3.0 and BindingDB) but also conducted *de novo* DTI prediction with independent computational methods including DeepCPI and DTINet. Top drug candidates were examined in detail, and ranked lists with two scoring systems for marketed drugs or natural compounds were recommended as potential antiviral solutions (Figure 1D).

Cataloging virus-specific host dependency genes

To generate a comprehensive compendium of HDGs for RNA viruses in an efficient manner, we primarily utilized the published studies to date performing functional genetic screens. In addition, to meet the urgent need for fighting SARS-CoV-2 and COVID-19, we also included individual HDGs identified from 34 focused studies for *Coronaviridae*. We established a human-specific HDG compendium for 29 RNA virus species across 10 families (Table S1). To make the compendium as inclusive as possible, we took a union of HDGs for a given virus species across different studies and screening platforms. Phylogenetic analysis based on the sequence evolution of viral RNA-dependent RNA polymerase (RdRp) gene among these species showed that RNA viruses in the same taxonomic families tend to cluster together (Figures 2A and S1A; Table S2), indicating a potentially coherent mechanism by which different but evolutionarily close viruses employ to live. RNAi represented the mostly adopted genetic perturbation technique, accounting for 54%

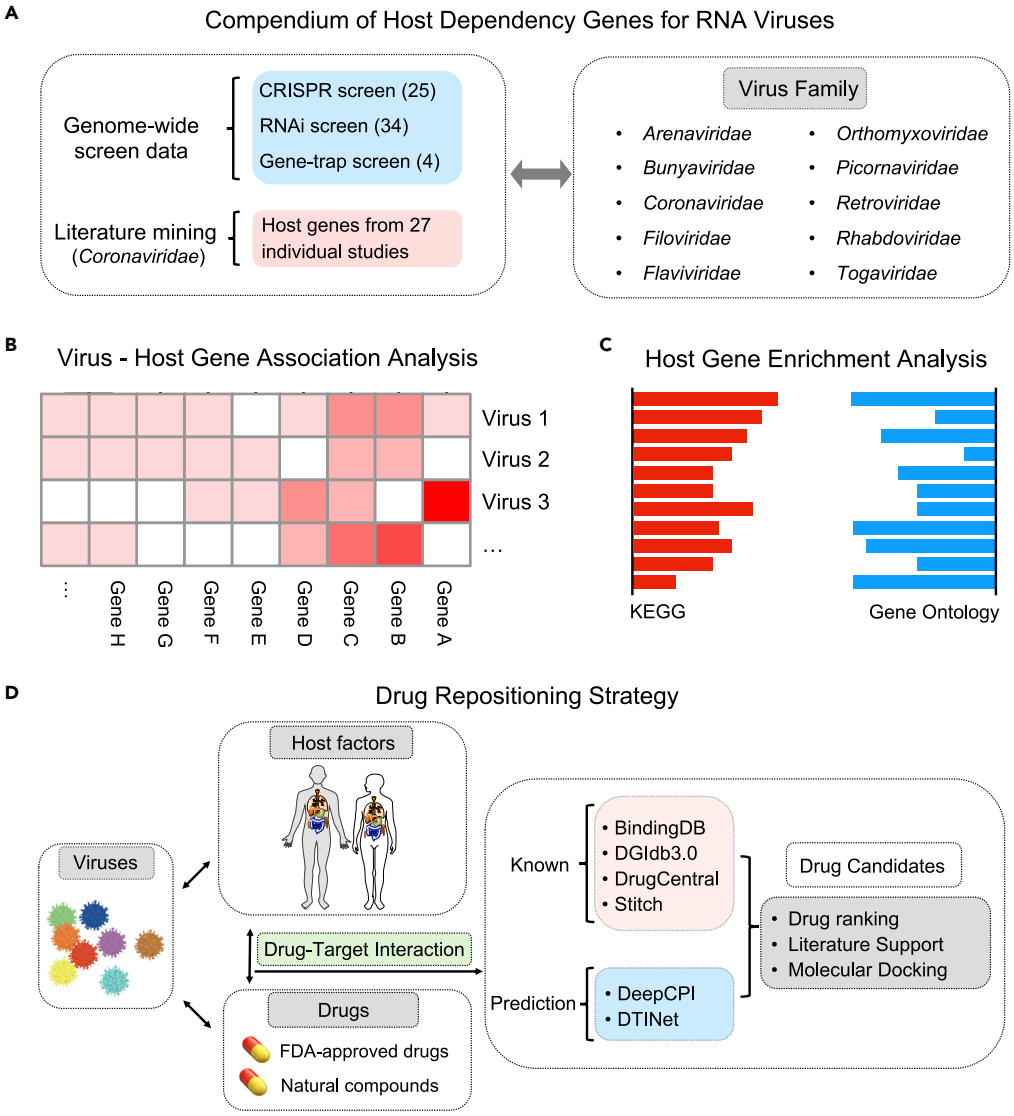


Figure 1. Strategic workflow of this study

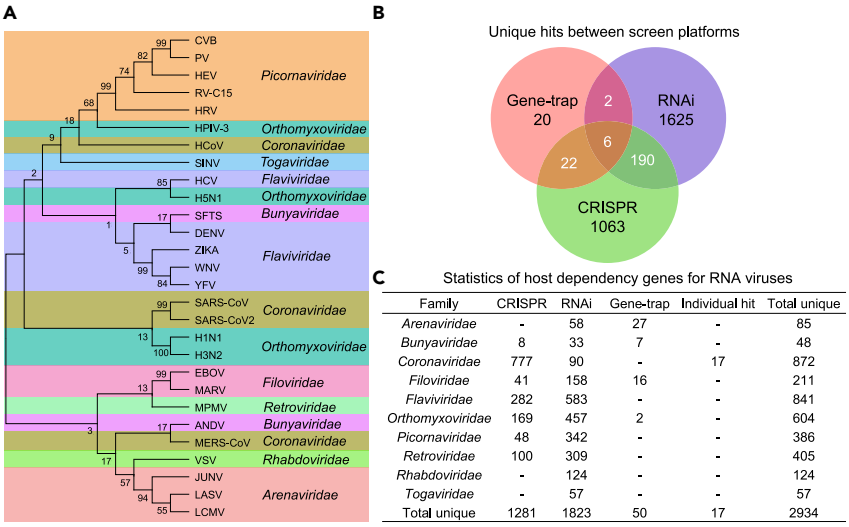
(A) Compiling of HDGs for ten families of RNA viruses. Human-specific HDGs were collected from related high-throughput genetic screening studies predominantly using CRISPR, RNAi, and haploid gene-trap techniques. For HDGs in *Coronaviridae* viruses, literatures specifically working on individual HDGs were also considered.

(B) Comparative analysis of HDGs across different RNA virus families.

(C) Functional enrichment analysis revealed molecular features of HDGs for corresponding virus families.

(D) Drug repositioning strategy in this study. We used high confident HDGs as host factors to be drugged. Two thousand three hundred fifty two FDA-approved drugs and 1,062 nature compounds selected from TCM herbs were interrogated. Potential DTIs were established by both known database information and *de novo* DTI prediction with AI-based computational methods. The top repurposed drug candidates were discussed in detail.

(34 out of 63) of all these screening studies. Most of the rest studies mainly employed the recently emerging revolutionized genome editing tool CRISPR-Cas for gene loss of function, whereas only 4 studies utilized the traditional gene-trap screening strategy in haploid cells (Table S1). Accordingly, RNAi screens identified the most HDGs, and only a fraction of them were recapitulated in CRISPR screens and gene-trap screens (Figures 2B and 2C and S1B). The low-level concordance across the three types of screens may be partially explained by (1) the unbalanced number of studies using different platforms, (2) intrinsically technical biases between screening platforms or libraries, and (3) batch effect across independent studies.



C Statistics of host dependency genes for RNA viruses

Family	CRISPR	RNAi	Gene-trap	Individual hit	Total unique
Arenaviridae	-	58	27	-	85
Bunyaviridae	8	33	7	-	48
Coronaviridae	777	90	-	17	872
Filoviridae	41	158	16	-	211
Flaviviridae	282	583	-	-	841
Orthomyxoviridae	169	457	2	-	604
Picornaviridae	48	342	-	-	386
Retroviridae	100	309	-	-	405
Rhabdoviridae	-	124	-	-	124
Togaviridae	-	57	-	-	57
Total unique	1281	1823	50	17	2934

Figure 2. Systematically cataloging HDGs for different RNA viruses

(A) The phylogenetic tree for interrogated RNA viruses was constructed with nucleic acid sequence of viral RNA polymerase RdRp gene using neighbor-joining (NJ) method.

(B) The Venn diagram of HDGs retrieved from different screening platforms.

(C) Summary and statistics of HDG compendium for all the viruses under investigation.

To further examine the variations of HDG calling across different studies, we re-analyzed a part of these CRISPR screening data where raw sequencing or count data are available using the MAGeCK-VISPR pipeline we previously developed (Li et al., 2015). Each gene is assigned a "β score" by the pipeline to indicate the function of the gene in screens. The higher the "β score", the more positive selection for the corresponding gene and the more likely for the gene to be an HDG hit in viral resistance screens. Re-analyzing of CRISPR screen data with a uniform β score criteria does not significantly affect HDG calling compared to the original analysis in corresponding studies, suggesting that computational algorithm bias here is minimal for such positive selection at least for CRISPR screens (Table S3). Different viruses across different studies exhibit variations on HDG profiles based on these re-analyzed CRISPR screen data (Figure S2A). The composite pool of HDGs identified by the re-analyzed CRISPR screens showed extensive protein-protein interactions and are enriched for infection-related pathways (Figures S2B and S2C; Table S3).

Crucial virus-host interplay revealed by functional host factors

We next sought to look into the biological features of these virus-host interactions. Comparative analysis of HDGs showed that different families of RNA virus exhibit differential profiles of HDGs and some families have fewer HDGs identified because of either fewer data sources or biological difference *per se* (Figure S3A). To minimize the analytic bias due to data insufficiency and fluctuation, we primarily focused on *Coronaviridae* (e.g., severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and SARS-CoV-2 viruses), *Flaviviridae* (e.g., dengue virus, Zika virus, and West Nile virus), and *Orthomyxoviridae* families (e.g., influenza A viruses H1N1, H3N2, and H5N1 subtypes) that have the most HDGs collected (Figure 2C; Table S1). We filtered all the HDGs by only keeping the one occurring more than once within respective families as high confidence HDGs to minimize the noises. We then depended on this refined list of HDGs for the following analysis as well as for drug repurposing. Comparative analysis indicated some common HDGs within these three viral families, posing a possibility to develop broad-spectrum antivirals when targeting these mutual targets (Figure 3A).

Pathway and functional gene category enrichment analysis with gene ontology and Kyoto Encyclopedia of Genes and Genomes tools showed that autophagy and infection-related processes are significantly enriched among HDGs for *Coronaviridae* viruses (Figures 3B and 3C and S3B; Table S4). On the other hand, *Flaviviridae* and *Orthomyxoviridae* viruses share several significantly enriched terms related to intracellular membrane system and its implicated functions (Figures 3B and S3B; Table S4). Network analysis demonstrated extensive protein-protein interactions between these HDGs and associated protein complexes with targetable HDGs highlighted (Figures 3C and S3C). These results indicated that a significant

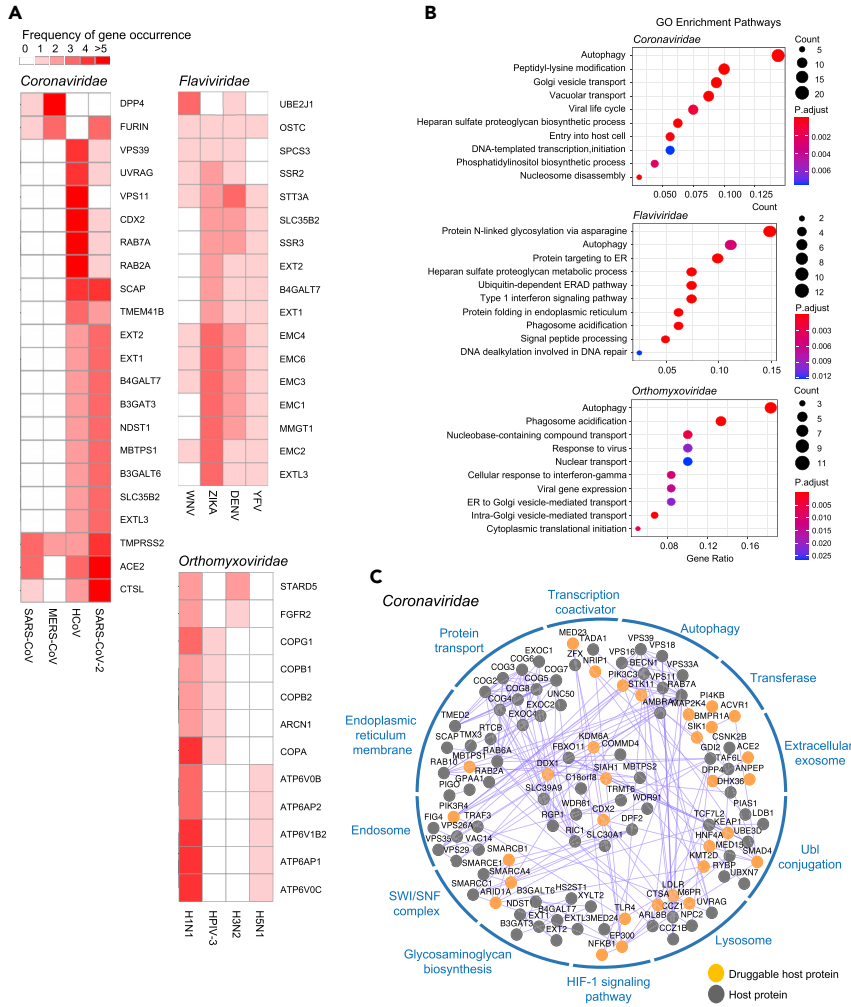


Figure 3. Characterization of HDGs for corresponding RNA virus families

(A) Subsets of shared HDGs across different RNA viruses within corresponding virus families. HDGs present in at least two different viruses within a given virus family were shown. The frequency of HDG occurrence across studies was denoted in different colors.

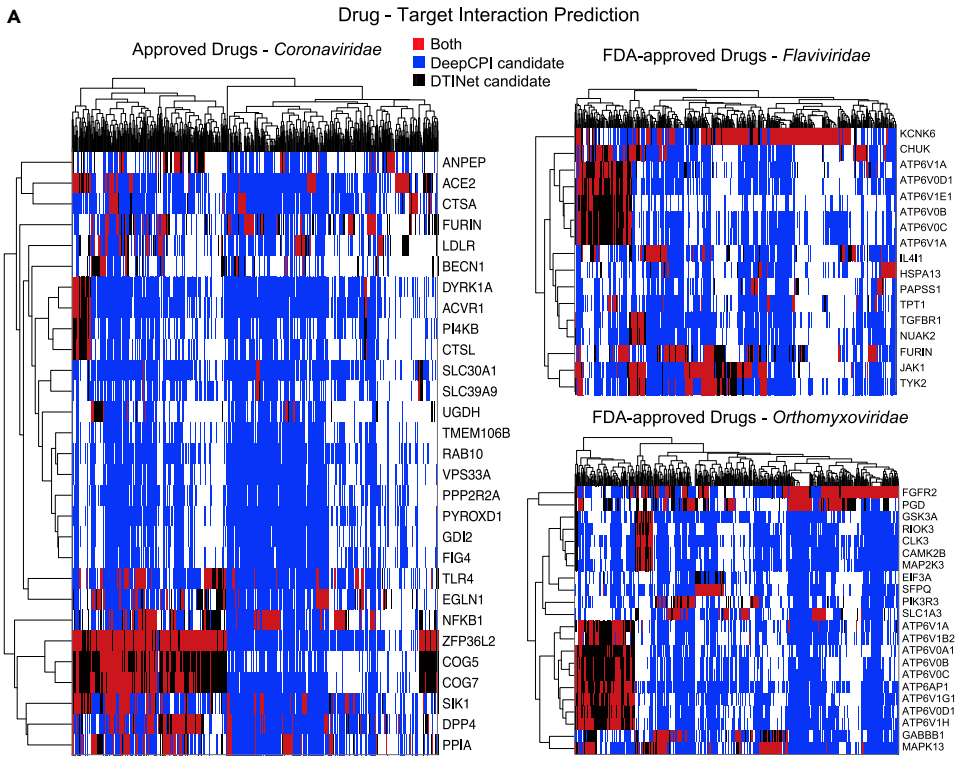
(B) Gene ontology enrichment analysis of respective HDGs for corresponding virus families. The size of the dot indicates the number of HDGs in the corresponding pathway. The color of the dot represents the value of Benjamini and Hochberg FDR-adjusted p value.

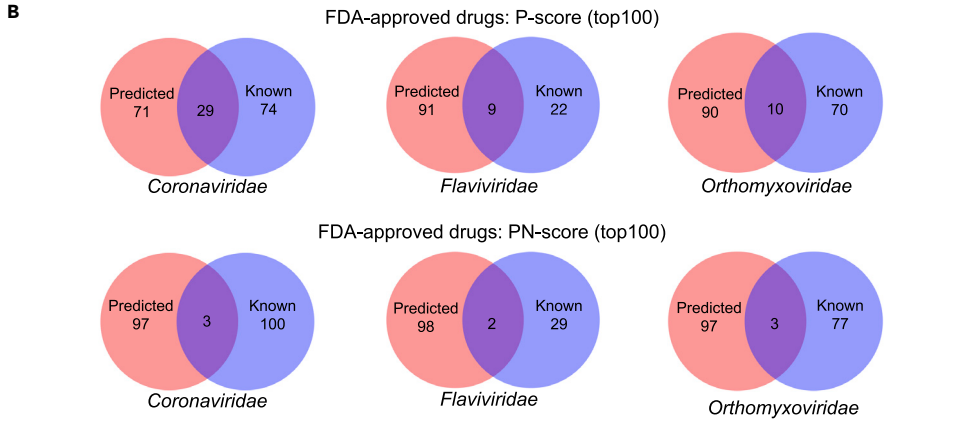
(C) Protein-protein interaction network of HDGs in *Coronaviridae* family. Each HDG is presented as a node. The edge between two nodes indicates a protein-protein interaction. The druggable HGDs with targeted drug candidates predicted in this study were highlighted.

amount of host proteins encoded by HDGs may be physically associated to collectively function in certain complexes, organelles, signaling pathways, or cellular processes that are essential for viral responses.

Mining known drug-target interactions across multiple databases

To identify potential drugs for viral HDGs, we firstly collected known DTIs from multiple public databases such as DGIdb3.0 (covering data from DrugBank, ChEMBL, therapeutic target database TTD, PharmGKB and ClinicalTrials.gov, etc.), BindingDB, DrugCentral, and Stitch (Cotto et al., 2018; Gilson et al., 2016; Kuhn et al., 2010; Ursu et al., 2019). DTIs extracted from databases depend on multiple lines of evidence ranging from approved drug description, *in vitro* binding assay, text mining and manual inspection, etc. We primarily focused on 2,352 drugs approved by the Food and Drug Administration (FDA) since their safety is validated and could be readily tested and applicable. In addition, we also included a selected list of 1,062 natural compounds that are active ingredients of traditional Chinese medicine (TCM) herbs


A
Drug - Target Interaction Prediction
Approved Drugs - Coronaviridae
Both
DeepCPI candidate
DTINet candidate
ANPEP
ACE2
CTSA
FURIN
LDLR
BECN1
DYRK1A
ACVR1
PI4KB
CTSL
SLC30A1
SLC39A9
UGDH
TMEM106B
RAB10
VPS33A
PPP2R2A
PYROXD1
GDI2
FIG4
TLR4
EGLN1
NFKB1
ZFP36L2
COG5
COG7
SIK1
DPP4
PPIA
FDA-approved Drugs - Flaviviridae
KCNK6
CHUK
ATP6V1A
ATP6V0D1
ATP6V1E1
ATP6V0B
ATP6V0C
ATP6V1A
IL4I1
HSPA13
PAPSS1
TPT1
TGFBR1
NUAK2
FURIN
JAK1
TYK2
FDA-approved Drugs - Orthomyxoviridae
FGFR2
PGD
GSK3A
RIOK3
CLK3
CAMK2B
MAP2K3
EIF3A
SFPQ
PIK3R3
SLC1A3
ATP6V1A
ATP6V1B2
ATP6V0A1
ATP6V0B
ATP6V0C
ATP6AP1
ATP6V1G1
ATP6V0D1
ATP6V1H
GABBR1
MAPK13


B
FDA-approved drugs: P-score (top100)
Predicted 71
29
Known 74
Coronaviridae
Predicted 91
9
Known 22
Flaviviridae
Predicted 90
10
Known 70
Orthomyxoviridae
FDA-approved drugs: PN-score (top100)
Predicted 97
3
Known 100
Coronaviridae
Predicted 98
2
Known 29
Flaviviridae
Predicted 97
3
Known 77
Orthomyxoviridae

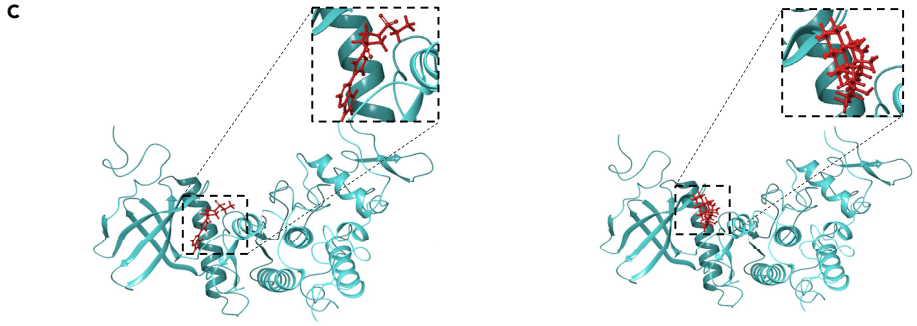

C
Baricitinib --- DYRK1A-A chain 6eis
Docking score: -9.431
Glide energy: -48.038 kcal/mol
Solanocapsine --- DYRK1A-A chain 6eis
Docking score: -4.128
Glide energy: -34.175 kcal/mol

Figure 4. Drug repositioning using multiple prediction models

(A) The heatmap showing DTI prediction by DeepCPI or DTINet methods. Each row represents a targetable HDG, and each column represents an FDA-approved drug. The top predicted DTI is color coded according to the color legend.

(B) The Venn diagram of FDA-approved drugs repurposed from known and *de novo* prediction sources of the top 100 hits by two ranking methods (P-score; PN-score) for the three indicated virus families.

(C) Molecular docking analysis showing the potential binding pocket of the repurposed drug baricitinib and natural compound solanocapsine with targeted host factors DYRK1A.

and pass special criteria for favorable druggability (Methods). With this database retrieval approach and further manual inspection, we investigated known drug-gene pairs and introduced 31~103 FDA-approved drugs targeting HDGs for respective *Coronaviridae*, *Flaviviridae*, and *Orthomyxoviridae* viral families (Tables S5 and S6).

Predicting drug-target interactions

We employed two machine learning methods (DeepCPI and DTINet) to predict more potential drug-gene interactions *in silico*. DeepCPI, a high-throughput computational framework combining feature embedding and deep learning technique to predict compound-protein interactions (Wan et al., 2019), was adopted to extensively exploit potential DTIs between HDGs and FDA-approved drugs or natural compounds. Another independent method termed DTINet, a network-based machine learning pipeline for DTI prediction on a large scale (Luo et al., 2017), was also utilized. DeepCPI can be applied for the DTI prediction without much prior knowledge of drugs and targets, and it is superior in terms of computational speed and easy installment. In contrast, DTINet requires a heterogeneous network that is constructed using the known information from four domains such as drugs, proteins, diseases, and side effects. Although using different principles and strategies, both methods have been shown to perform well during cross-validation with large-scale DTI data (Luo et al., 2017; Wan et al., 2019) and may complement with each other for better DTI prediction. Either method depends on a calculated score (DeepCPI score or DTINet score) to quantify the confidence of predicted interaction for a given drug-target pair. We took the intersection of the prediction results from both methods with normalized *Z* score cutoffs for FDA-approved drug repurposing (Figure 4A). On the other hand, we primarily relied on DeepCPI results for natural compound analysis since DTINet does not perform well due to insufficient modeling data for natural compounds. Compared to known DTIs retrieved from databases, more DTIs are predicted for the three viral families with some consensus (Figure 4B; Table S5).

Prioritizing candidate drugs and natural compounds

In addition to known DTI annotation, we prioritized potential drug candidates primarily by the predicted DTIs. Since one drug may target multiple HDG targets that may produce enhanced antiviral response, we first ranked these repurposed drugs mainly according to their targeting number and potency of HDG targets reflected by DTI prediction scores. For FDA-approved drugs, the two DTI prediction scores from both DeepCPI and DTINet algorithms were considered and transformed into a joint P-score to rank the predicted drugs (Methods; Table S5). This ranking system puts emphasis on the HDG targeting effect but overlooks the negative impact from promiscuous non-HDG targets and cytotoxic essential gene targets. Thus, we also provided a second ranking system by incorporating the promiscuousness and cytotoxicity effect with a joint PN-score (Methods; Table S5). As shown in Tables 1, 2, S7, S8, S11, and S12, the two ranking systems shared a great portion of common hits from the top candidate drugs for antiviral purpose against *Coronaviridae*, *Flaviviridae*, and *Orthomyxoviridae* viruses, in agreement with the fact that the promiscuousness and side effect of approved drug are already tested clinically and usually controllable.

Of note, among these top drug candidates for *Coronaviridae* viruses (Tables 1 and 2), baricitinib, a Janus kinase (JAK) inhibitor approved for rheumatoid arthritis treatment, has been shown to lower the cytokine effect and reduce the viral load in patients with COVID-19 by targeting Janus kinase/signal transducers and activators of transcription (JAK/STAT) signaling and numb-associated kinases, respectively (Stebbing et al., 2020). Several clinical trials have been launched globally to evaluate the therapeutic effect of baricitinib (ClinicalTrials.gov Identifier: NCT04358614, NCT04320277, and NCT04321993). Molecular docking analysis of baricitinib and its predicted targets showed high binding affinity between them, further supporting their potential interactions (Figures 4C and S4; Table S15). Another JAK inhibitor tofacitinib among the top ten repurposed *Coronaviridae*-targeting drugs is also being evaluated for COVID-19 treatment in an active clinical trial (NCT04415151). Moreover, tofacitinib was previously shown to be a potent inhibitor for immunodeficiency virus type 1 (HIV-1)

Table 1. Joint P-score ranking: the top ten repurposed FDA-approved drugs against *Coronaviridae* viruses

Drug candidate	Approved indication	PubChem CID	Top 10 predicted host targets	Known interaction	Joint P-score
Fostamatinib	Chronic immune thrombocytopenia	11671467	PI4KB DYRK1A ACVR1 CTSL SIK1 COG5 COG7 ZFP36L2 ACE2	SYK	1.078
Baricitinib	Rheumatoid arthritis	44205240	CTSL PI4KB DYRK1A ACVR1 ACE2 COG5 SIK1 COG7 ZFP36L2	JAK1 JAK2 JAK3 TYK2	0.407
Simvastatin	Hypercholesterolemia	54454	ANPEP COG7 DPP4 COG5 ZFP36L2 CTSL PI4KB DYRK1A ACVR1 ACE2	HMGCR	0.363
Tofacitinib	Rheumatoid arthritis	9926791	DYRK1A ACVR1 CTSL ACE2 SIK1 COG7 COG5 ZFP36L2 DPP4	JAK3 JAK2 JAK1	0.362
Etoricoxib	Rheumatoid arthritis	123619	DYRK1A ACVR1 CTSL PI4KB ACE2 COG5 SIK1 COG7 ZFP36L2	COX2	0.340
Bivalirudin	Angina	16129704	CTSL PI4KB ACVR1 DYRK1A SIK1 LDLR COG7 COG5 ZFP36L2	F2	0.334
Flurbiprofen	Arthritis	3394	CTSL PI4KB ACVR1 DYRK1A ACE2 SIK1 ZFP36L2	COX1 COX2	0.311
Lusutrombopag	Thrombocytopenia	49843517	PI4KB CTSL DYRK1A ACVR1 SIK1 COG7 COG5 ZFP36L2	MPL	0.309
Bosutinib	Chronic Myelogenous Leukemia	5328940	BECN1 COG5 COG7 LDLR FURIN SIK1 PPIA DPP4 ZFP36L2 ACE2	SRC ABL1	0.253
Hydroxychloroquine	Rheumatoid arthritis	3652	ACE2 DYRK1A ACVR1 SIK1 ZFP36L2	N/A	0.252

N/A: not applicable.

replication *in vitro*, further supporting its antiviral activity (Gavegnano et al., 2014). Interestingly, hydroxychloroquine also stands out among the top ten candidates, consistent with its *in vitro* antiviral activity for SARS-CoV-2 albeit less effective *in vivo* (Maisonnasse et al., 2020; Wang et al., 2020a). Taken together, our analysis provides encouraging repurposing candidates for antiviral application.

Candidate natural compounds were also ranked with two systems by either DeepCPI P-score or DeepCPI PN-score (Methods; Tables 3 and 4; Tables S5, S9, S10, S13, and S14). We also summarized the herbs that include the corresponding compound as part of their active ingredients. Among the top ten predicted natural compounds against *Coronaviridae* viruses (Tables 3 and 4), some of them, such as lysergol, solanocapsine, picrasidine D, and effusol, have emerged under both ranking methods. Interestingly, sophocarpine has been reported to exhibit antiviral activity against enterovirus 71 (Jin et al., 2017). Moreover, Asari Radix et Rhizoma (Xi Xin) and Codonopsis Radix (Dang Shen), the TCM herbs that contain selected compounds picrasidine D and sophocarpine, respectively, are included in the current TCM formula to treat COVID-19 in China according to Chinese National Health Commission Guidelines for COVID-19 Treatment, eighth edition (http://www.nhc.gov.cn/yzygj/). Representative molecular docking analysis was also performed for compound solanocapsine and its predicted targets, and again, high-affinity interaction modules can be generated between the compound and predicted targets (Figures 4C and S4; Table S15). These results further supported the validity of our repurposing strategy, and it is worthy to evaluate these drug candidates for corresponding antiviral purposes in depth. In addition to HDGs, we also applied our drug prediction pipeline onto 38 SARS-CoV-2 viral proteins with DeepCPI algorithm. A tentative list of repurposed drugs with direct antiviral functions was provided and ready for further experimental validation (Table S5).

DISCUSSION

Given the limited number of *de novo* antiviral drugs approved during recent years, drug repositioning or repurposing has become a pivotal approach to combat pathogenic viruses and related diseases. In particular, when confronted with an emergent pandemic such as current COVID-19 caused by coronavirus SARS-CoV-2, people highly demand quick and effective solutions for disease control and therapeutic treatment. By systematically compiling the HDGs for RNA viruses and thoroughly digging tentative DTIs, we took host-centered angle to prioritize the potential FDA-approved drugs and natural products as repurposed antiviral candidates against a plethora of RNA viruses, including recently prevailing coronaviruses, Zika virus,

Table 2. Joint PN-score ranking: the top ten repurposed FDA-approved drugs against *Coronaviridae* viruses

Drug candidate	Approved indication	PubChem CID	Top 10 predicted host targets	Known interaction	Joint PN-score
Baricitinib	Rheumatoid arthritis	44205240	CTSL PI4KB DYRK1A ACVR1 ACE2 COG5 SIK1 COG7 ZFP36L2	JAK1 JAK2 JAK3 TYK2	0.155
Lusutrombopag	Thrombocytopenia	49843517	PI4KB CTSL DYRK1A ACVR1 SIK1 COG7 COG5 ZFP36L2	MPL	0.127
Bivalirudin	Thrombocytopenia	16129704	CTSL PI4KB ACVR1 DYRK1A SIK1 LDLR COG7 COG5 ZFP36L2	F2	0.126
Etoricoxib	Rheumatoid arthritis	123619	DYRK1A ACVR1 CTSL PI4KB ACE2 COG5 SIK1 COG7 ZFP36L2	PTGS2	0.123
Semaglutide	Type 2 diabetes	56843331	DPP4 COG5 COG7 ZFP36L2 ANPEP	GLP1R	0.108
Fostamatinib	Chronic immune thrombocytopenia	11671467	PI4KB DYRK1A ACVR1 CTSL SIK1 COG5 COG7 ZFP36L2 ACE2	SYK	0.105
Histrelin	Prostate cancer	25077993	BECN1 EGLN1 COG5 COG7 ZFP36L2	GNRH1	0.087
Lopromide	X-ray contrast agent	3736	GDI2 RAB10 PYROXD1 FIG4 PPP2R2A VPS33A TMEM106B	PGP	0.081
Hydroxychloroquine	Rheumatoid arthritis	3652	ACE2 DYRK1A ACVR1 SIK1 ZFP36L2	N/A	0.079
Vildagliptin	Type 2 diabetes	6918537	DPP4 COG7 ZFP36L2	DPP4	0.073

N/A: not applicable.

dengue viruses, influenza viruses, etc. These recommended drugs or natural compounds are readily tested in the laboratory and clinical settings for their antiviral uses.

Compared to virus-centered antiviral strategy that targets viral genes to directly interfere with virus reproduction and infection, a host-centered antiviral approach has several advantages such as (1) functional host genes are more conserved and evolutionally stable than viral genes, which makes host-targeting drugs more tolerant to frequent viral mutations than those virus-targeting counterparts; 2) different viruses may share a similar set of host genes during certain stages of viral life cycle, which underlines the basis of developing broad-spectrum antivirals so that one host-targeting drug may treat multiple virus infection; and (3) there are significantly more targeted drugs approved for host genes than those for viral genes, thus likely increasing the success rate of drug repurposing by adopting host-centered strategy. Previous studies have extensively tried targeting host genes for developing antiviral solutions (Ackerman et al., 2018; Bosl et al., 2019; Li et al., 2019; Loganathan et al., 2020; Luo et al., 2017; Saiz et al., 2018; Zhou et al., 2020). Host receptors mediating viral entrance into the cells represent the most popular host targets for drugs to block viral infection. A wider range of host genes identified through protein-protein interaction with viral genes serves as the predominant source of host factors to be targeted. In addition, targeting the host transcriptome change resulted from viral infection can be viewed as another host-based drug repositioning strategy. Recent studies also identified SARS-CoV-2-associated human proteins, changed transcriptome, and proteome of human cells in response to SARS-CoV-2 infection to facilitate drug repurposing (Bojkova et al., 2020; Gordon et al., 2020). However, most of these host targets are not essentially required or functionally redundant for complete viral reproduction and infection, even though they are closely associated to the viral components or processes. In principal, effective host drugs should target those functional host genes or related processes on which the virus depends to hinder viral functions within a cell. Therefore, our work particularly focused on those HDGs identified primarily by recent genome-wide screening studies for multiple RNA viruses, which may greatly improve the success rate of drug repositioning compared to previous host-based approaches.

Table 3. DeepCPI P-score ranking: the top ten repurposed natural compounds against *Coronaviridae* viruses

Drug candidate	TCMSP[a] MOL ID	PubChem CID	Herb	Top 10 predicted host targets	DeepCPI P-score
Lysergol	MOL005261	14,987	Pharbitidis Semen	ANGPT2 BECN1 COG2 COG6 PCBD1 RTCB TMED2 UGDH VPS29	4.059
Atropine	MOL002219	174174	Lycii Cortex, Hyoscyami Semen	COG2 HNF4A KDM6A KEAP1 PGGT1B RABL3 RAD54L2 SIRT6 SMARCB1 UGDH	3.842
Solanocapsine	MOL007356	73,419	Solanum Nigrum	ANGPT2 BECN1 COG2 COG7 CTSL EXOC1 KDM6A PCBD1 PGGT1B	3.587
Costaclavine	MOL008145	160462	Ricini Semen	ANGPT2 BECN1 COG6 DOHH PCBD1 RTCB SAR1A TMED2 VPS26A VPS29	3.473
Chanoclavine	MOL005260	5281381	Semen Pharbitidis	ANGPT2 ANPEP COG2 PCBD1 RTCB SAR1A TMED2 UGDH VPS26A VPS29	3.299
Triptofordin B1	MOL003232	122391803	Tripterygii Radix	BECN1 EP300 HIRA KEAP1 PCBD1 PGGT1B TADA1 TOM1 VPS11 VPS29	3.175
Picrasidine D	MOL012140	5316876	Asari Radix et Rhizoma	HIRA EP300 HNF4A RAD54L2 SCAP SIRT6 SMARCB1 TADA1 TOM1 WDR91	3.141
9alpha-hydroxysophoramine	MOL006570	50695119	Sophorae Flavescentis Radix	CTSL DPF2 KEAP1 PIK3C3 RAD54L2 SCAP SIRT6 SMARCA4 UGDH WDR91	2.976
Effusol	MOL007910	100801	Junci Medulla	AKAP6 ANGPT2 COG6 HIRA KDM6A PIAS1 RLF RTCB SMARCB1	2.886
Sophocarpine	MOL003627	115269	Codonopsis Radix	DDX1 DPF2 GDI2 PIK3C3 SCAP SMARCA4 SMARCC1 TMEM106B TMPRSS2 UGDH	2.833

[a]TCMSP database: Traditional Chinese Medicine Systems Pharmacology Database.

Given a set of host genes, how to evaluate the potential drug effect on specific genes becomes the major challenge for successful drug repurposing. Experimental evaluation of physical interaction strength and kinetics between a drug and a target is an ideal way to establish a definite drug-target relationship. Nevertheless, it tends to be exhausting and impractical when dealing with multiple drugs versus multiple targets. Although drug-related databases have annotated some DTIs from multiple lines of evidence including experimental data, marketed drug description, and literature mining, more systematic and logic approaches to define DTIs especially in a high-throughput manner are still highly demanded. Artificial intelligence such as machine learning and deep learning has been implemented in several computational tools to predict the potential DTIs at a large scale (D'Souza et al., 2020; Rifaioglu et al., 2019; Zhou et al., 2019). In addition to database-retrieved information, here we applied two independent computational pipelines to predict *de novo* DTIs with quantitative measures. We expect to improve DTI identification with these combinatorial approaches by prioritizing the consensus results. Furthermore, quantitative evaluation of DTI with interaction scores enables a likelihood ranking of potential drug candidates, which may provide better guidance for the following in-depth evaluation.

The repurposed drug candidates recommended by this study not only cover FDA-approved drugs but also include natural compounds especially present in TCM herbs. The active ingredients from the TCM herbs provide a wealth of resource by which new drugs for specific diseases can be discovered, including for antiviral purposes. As our approach is primarily based on targeting HDGs, the viral families that share common druggable host targets may occasionally result in similar repurposed drug or compounds (Tables 1, 2, 3, and 4, S7–S14). Although vaccination is a major strategy to build immune barrier among the population against viral spread, effective drugs are still quite crucial for those individuals already infected by the virus, especially for those detrimental ones investigated here. The fundamental difference of this study with previous drug repositioning work largely lies in target selection, DTI determination, and final repurposed drug candidates.

In summary, our study presents a host-based strategy by focusing on HDGs for a series of RNA viruses to identify potential candidate drugs or natural compounds against related viral diseases, with special emphasis on drug

Table 4. DeepCPI PN-score ranking: the top ten repurposed natural compounds against *Coronaviridae* viruses

Drug candidate	TCMSP[a] MOL ID	PubChem CID	Herb	Top 10 predicted host targets	DeepCPI PN-score
Solanocapsine	MOL007356	73,419	Solanum Nigrum	ANGPT2 BECN1 COG2 COG7 CTSL EXOC1 KDM6A PCBD1 PGGT1B	0.378
Vitexifolin C	MOL011912	11033408	Viticis Fructus	EIF4G2 HIRA KDM6A	0.374
Dehydroeffusal	MOL007904	101191858	Junci Medulla	DPF2 KDM6A PIAS1 RLF	0.341
Lysergol	MOL005261	14,987	Pharbitidis Semen	ANGPT2 BECN1 COG2 COG6 PCBD1 RTCB TMED2 UGDH VPS29	0.338
Picrasidine D	MOL012140	5316876	Asari Radix et Rhizoma	HIRA HNF4A RAD54L2 SCAP SIRT6 SMARCB1 TADA1 TOM1 WDR91	0.334
Isolimonic acid	MOL013443	131752314	Aurantii Fructus Immaturus	ANGPT2 ANPEP BECN1 COG2 COG7 EXOC1 RAB6A SMARCB1 UGDH	0.329
Methyl 15-hydroxydehydroabietate	MOL012165	11573479	Solidaginis Herba	AKAP6 B4GALT7	0.305
Neotigogenin	MOL008519	12304433	Trigonellae Semen	AKAP6 ANGPT2 BECN1 COG2 COG4 EXOC1 PIAS1 VPS11 VPS29	0.301
Cyclopamine	MOL009027	442972	Fritiliariae Irrhosae Bulbus	ANGPT2 BECN1 DPF2 EP300 KDM6A LDLR PIAS1 RLF VPS11	0.298
Effusol	MOL007910	100801	Junci Medulla	AKAP6 ANGPT2 COG6 HIRA KDM6A PIAS1 RLF RTCB SMARCB1	0.291

[a]TCMSP database: Traditional Chinese Medicine Systems Pharmacology Database.

repositioning scheme toward SARS-CoV-2 and COVID-19. This work not only reveals key essential features of viral infection from the host perspective but also provides reasonable and promising antiviral drug candidates for further evaluations in hope of finally controlling these detrimental viral diseases.

Limitations of the study

There are several limitations in the current study. Firstly, we were unable to perform experimental evaluations of these proposed drugs for their antiviral effect at current stage, due to the restricted access to those highly pathogenic viruses. Secondly, the compiling of HDGs may not be complete enough for some viruses to infer the whole host dependency basis and perform appropriate drug repurposing since the currently available data for HDGs are still limited despite the studies collected in this work. Thirdly, we mainly relied on DeepCPI, DTINet, and database-retrieved information followed by manual inspection to assign drug-gene pairing relationship. Further application of more other computational DTI prediction tools may compensate or improve the outcomes of drug selection.

Resource availability

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Teng Fei (feiteng@mail.neu.edu.cn).

Material availability

The study did not generate any unique reagents.

Data and code availability

This published article includes all data sets generated or analyzed during this study.

METHODS

All methods can be found in the accompanying Transparent methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102148.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (31871344, 32071441), the Fundamental Research Funds for the Central Universities (N2020001, N182005005), the 111 Project (B16009), XingLiao Talents Program (XLYC1807212) (to T.F.), the startup fund of the Center for Genetic Medicine Research (to W.L.), and the China Postdoctoral Science Foundation (2020T130012ZX) (to Y.Y.). The computational processing of CRISPR screens is partly supported by COVID-19 HPC Consortium (to W.L.).

AUTHOR CONTRIBUTIONS

Z.L., Y.Y., W.L., and T.F. conceived and designed the research. Z.L., Y.Y., X.C., and Q.C. performed the research. All the authors analyzed the data. Z.L., Y.Y., X.C., Q.C., W.L., and T.F. wrote the manuscript with the input of all the other authors. T.F. and W.L. supervised the study.

DECLARATION OF INTERESTS

W.L. reports serving as a consultant of Tavros Therapeutics. Other authors declare no conflict of interest.

Received: September 15, 2020
Revised: January 11, 2021
Accepted: February 1, 2021
Published: March 19, 2021

REFERENCES

Ackerman, E.E., Kawakami, E., Katoh, M., Watanabe, T., Watanabe, S., Tomita, Y., Lopes, T.J., Matsuoka, Y., Kitano, H., Shoemaker, J.E., et al. (2018). Network-Guided discovery of influenza virus replication host factors. mBio *9*, e02002.

Altay, O., Mohammadi, E., Lam, S., Turkez, H., Boren, J., Nielsen, J., Uhlen, M., and Mardinoglu, A. (2020). Current status of COVID-19 therapies and drug repositioning applications. iScience *23*, 101303.

Bojkova, D., Klann, K., Koch, B., Widera, M., Krause, D., Ciesek, S., Cinatl, J., and Münch, C. (2020). Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. Nature *583*, 469–472.

Bosl, K., Ianevski, A., Than, T.T., Andersen, P.I., Kuivanen, S., Teppor, M., Zusinaite, E., Dumpis, U., Vitkauskiene, A., Cox, R.J., et al. (2019). Common nodes of virus-host interaction revealed through an integrated network analysis. Front. Immunol. *10*, 2186.

Cotto, K.C., Wagner, A.H., Feng, Y.Y., Kiwala, S., Coffman, A.C., Spies, G., Wollam, A., Spies, N.C., Griffith, O.L., and Griffith, M. (2018). DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. Nucleic Acids Res. *46*, D1068–D1073.

D'Souza, S., Prema, K.V., and Balaji, S. (2020). Machine learning models for drug-target interactions: current knowledge and future directions. Drug Discov. Today *25*, 748–756.

Dai, W., Zhang, B., Jiang, X.M., Su, H., Li, J., Zhao, Y., Xie, X., Jin, Z., Peng, J., Liu, F., et al. (2020). Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. Science *368*, 1331–1335.

De Clercq, E., and Li, G. (2016). Approved antiviral drugs over the past 50 years. Clin. Microbiol. Rev. *29*, 695–747.

Dighe, S.N., Ekwudu, O., Dua, K., Chellappan, D.K., Katavic, P.L., and Collet, T.A. (2019). Recent update on anti-dengue drug discovery. Eur. J. Med. Chem. *176*, 431–455.

Gavegnano, C., Detorio, M., Montero, C., Bosque, A., Planelles, V., and Schinazi, R.F. (2014). Ruxolitinib and tofacitinib are potent and selective inhibitors of HIV-1 replication and virus reactivation in vitro. Antimicrob. Agents Chemother. *58*, 1977–1986.

Gilson, M.K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L., and Chong, J. (2016). BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Res. *44*, D1045–D1053.

Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature *583*, 459–468.

Jin, Z., Yang, L., Ding, G., Yang, G., Han, Y., Zhang, X., and Li, W. (2017). Sophocarpine against enterovirus 71 in vitro. Exp. Ther. Med. *14*, 3792–3797.

Jin, Z., Zhao, Y., Sun, Y., Zhang, B., Wang, H., Wu, Y., Zhu, Y., Zhu, C., Hu, T., Du, X., et al. (2020). Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur. Nat. Struct. Mol. Biol. *27*, 529–532.

Kuhn, M., Szklarczyk, D., Franceschini, A., Campillos, M., von Mering, C., Jensen, L.J., Beyer, A., and Bork, P. (2010). Stitch 2: an interaction network database for small molecules and proteins. Nucleic Acids Res. *38*, D552–D556.

Li, C.C., Wang, X.J., and Wang, H.R. (2019). Repurposing host-based therapeutics to control coronavirus and influenza virus. Drug Discov. Today *24*, 726–736.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. New Engl. J. Med. *382*, 1199–1207.

Li, W., Koster, J., Xu, H., Chen, C.H., Xiao, T., Liu, J.S., Brown, M., and Liu, X.S. (2015). Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. Genome Biol. *16*, 281.

Loganathan, T., Ramachandran, S., Shankaran, P., Nagarajan, D., and Mohan, S.S. (2020). Host transcriptome-guided drug repurposing for COVID-19 treatment: a meta-analysis based approach. PeerJ *8*, e9357.

Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., Peng, J., Chen, L., and Zeng, J. (2017). A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. Nat. Commun. *8*, 573.

Maisonnasse, P., Guedj, J., Contreras, V., Behillil, S., Solas, C., Marlin, R., Naninck, T., Pizzorno, A., Lemaitre, J., Gonçalves, A., et al. (2020). Hydroxychloroquine use against SARS-CoV-2 infection in non-human primates. Nature *585*, 584–587.

Mottin, M., Borba, J., Braga, R.C., Torres, P.H.M., Martini, M.C., Proenca-Modena, J.L., Judice, C.C., Costa, F.T.M., Ekins, S., Perryman, A.L., et al. (2018). The A-Z of Zika drug discovery. Drug Discov. Today *23*, 1833–1847.

Petersen, E., Koopmans, M., Go, U., Hamer, D.H., Petrosillo, N., Castelli, F., Storgaard, M., Al Khalili, S., and Simonsen, L. (2020). Comparing SARS-

CoV-2 with SARS-CoV and influenza pandemics. Lancet Infect. Dis. *20*, e238–e244.

Pierson, T.C., and Diamond, M.S. (2020). The continued threat of emerging flaviviruses. Nat. Microbiol. *5*, 796–812.

Rifaioglu, A.S., Atas, H., Martin, M.J., Cetin-Atalay, R., Atalay, V., and Dogan, T. (2019). Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. Brief Bioinform *20*, 1878–1912.

Saiz, J.C., Oya, N.J., Blazquez, A.B., Escribano-Romero, E., and Martin-Acebes, M.A. (2018). Host-directed antivirals: a realistic alternative to fight Zika virus. Viruses *10*, 453.

Stebbing, J., Krishnan, V., de Bono, S., Ottaviani, S., Casalini, G., Richardson, P.J., Monteil, V., Lauschke, V.M., Mirazimi, A., Youhanna, S., et al. (2020). Mechanism of baricitinib supports artificial intelligence-predicted testing in COVID-19 patients. EMBO Mol. Med. e12697.

Ursu, O., Holmes, J., Bologa, C.G., Yang, J.J., Mathias, S.L., Stathias, V., Nguyen, D.T., Schurer, S., and Oprea, T. (2019). DrugCentral 2018: an update. Nucleic Acids Res. *47*, D963–D970.

Wan, F., Zhu, Y., Hu, H., Dai, A., Cai, X., Chen, L., Gong, H., Xia, T., Yang, D., Wang, M.W., et al. (2019). DeepCPI: a deep learning-based framework for large-scale in silico drug screening. Genomics Proteomics Bioinformatics *17*, 478–495.

Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., Shi, Z., Hu, Z., Zhong, W., and Xiao, G. (2020a). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res. *30*, 269–271.

Wang, Y., Zhang, D., Du, G., Du, R., Zhao, J., Jin, Y., Fu, S., Gao, L., Cheng, Z., Lu, Q., et al. (2020b). Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. Lancet *395*, 1569–1578.

Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., Wang, Q., Xu, Y., Li, M., Li, X., et al. (2020). Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. Acta Pharm. Sin. B *10*, 766–788.

Zhang, J., Hu, Y., Musharrafieh, R., Yin, H., and Wang, J. (2019). Focusing on the influenza virus polymerase complex: recent progress in drug discovery and assay development. Curr. Med. Chem. *26*, 2243–2263.

Zhou, L., Li, Z., Yang, J., Tian, G., Liu, F., Wen, H., Peng, L., Chen, M., Xiang, J., and Peng, L. (2019). Revealing drug-target interactions with computational models and algorithms. Molecules *24*, 1714.

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., and Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov. *6*, 14.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. *382*, 727–733.

Zumla, A., Chan, J.F., Azhar, E.I., Hui, D.S., and Yuen, K.Y. (2016). Coronaviruses - drug discovery and therapeutic options. Nat. Rev. Drug Discov. *15*, 327–347.

iScience, Volume *24*

Supplemental Information

A computational framework of host-based drug repositioning for broad-spectrum antivirals against RNA viruses

Zexu Li, Yingjia Yao, Xiaolong Cheng, Qing Chen, Wenchang Zhao, Shixin Ma, Zihan Li, Hu Zhou, Wei Li, and Teng Fei

Supplemental Figure 1

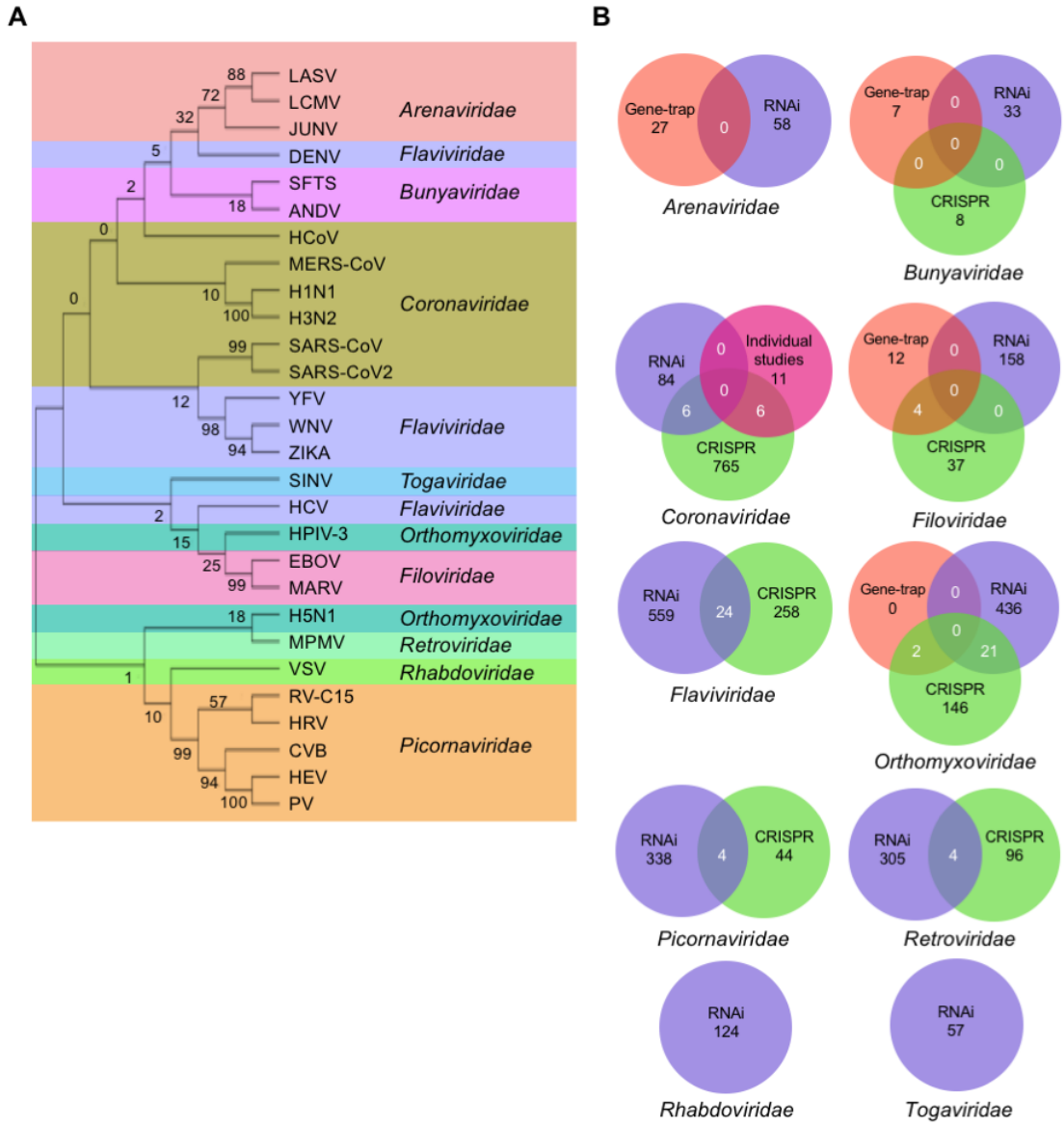


Figure S1. HDG collection for different RNA viruses, Related to Figure 2

(A) The phylogenetic tree for interrogated RNA viruses was constructed with protein sequence of viral RNA polymerase RdRp gene using maximum parsimony method.

(B) The venn diagrams of HDGs for indicated RNA virus families retrieved from different screening platforms.

Supplemental Figure 2

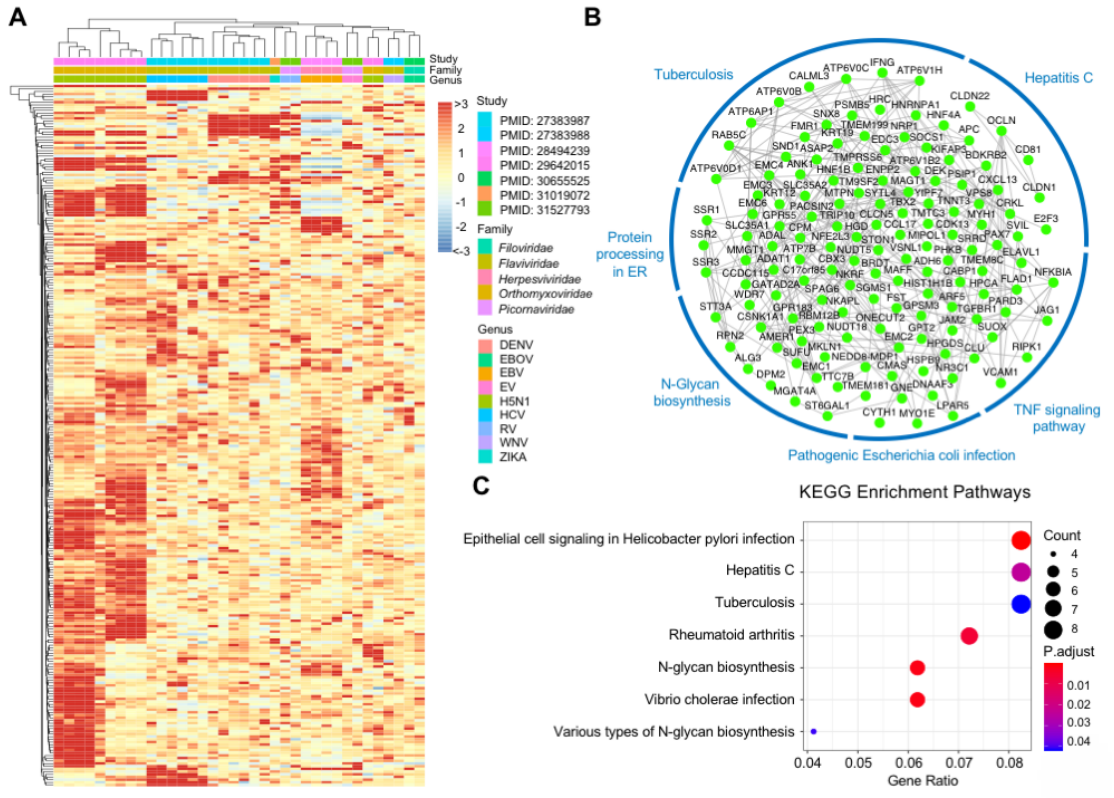


Figure S2. Re-analysis of CRISPR screening data for HDGs, Related to Figure 3

(A) The heatmap clustering of corresponding gene's β score calculated by MAGeCK-VISPR for multiple CRISPR screen studies related to HDG identification. HDG would have a high β score indicating a positive selection against corresponding virus challenge.

(B) The protein-protein interaction network for all the HDGs identified from re-analyzed CRISPR screens.

(C) Functional category enrichment analysis by KEGG for HDGs identified from re-analyzed CRISPR screens.

Supplemental Figure 3

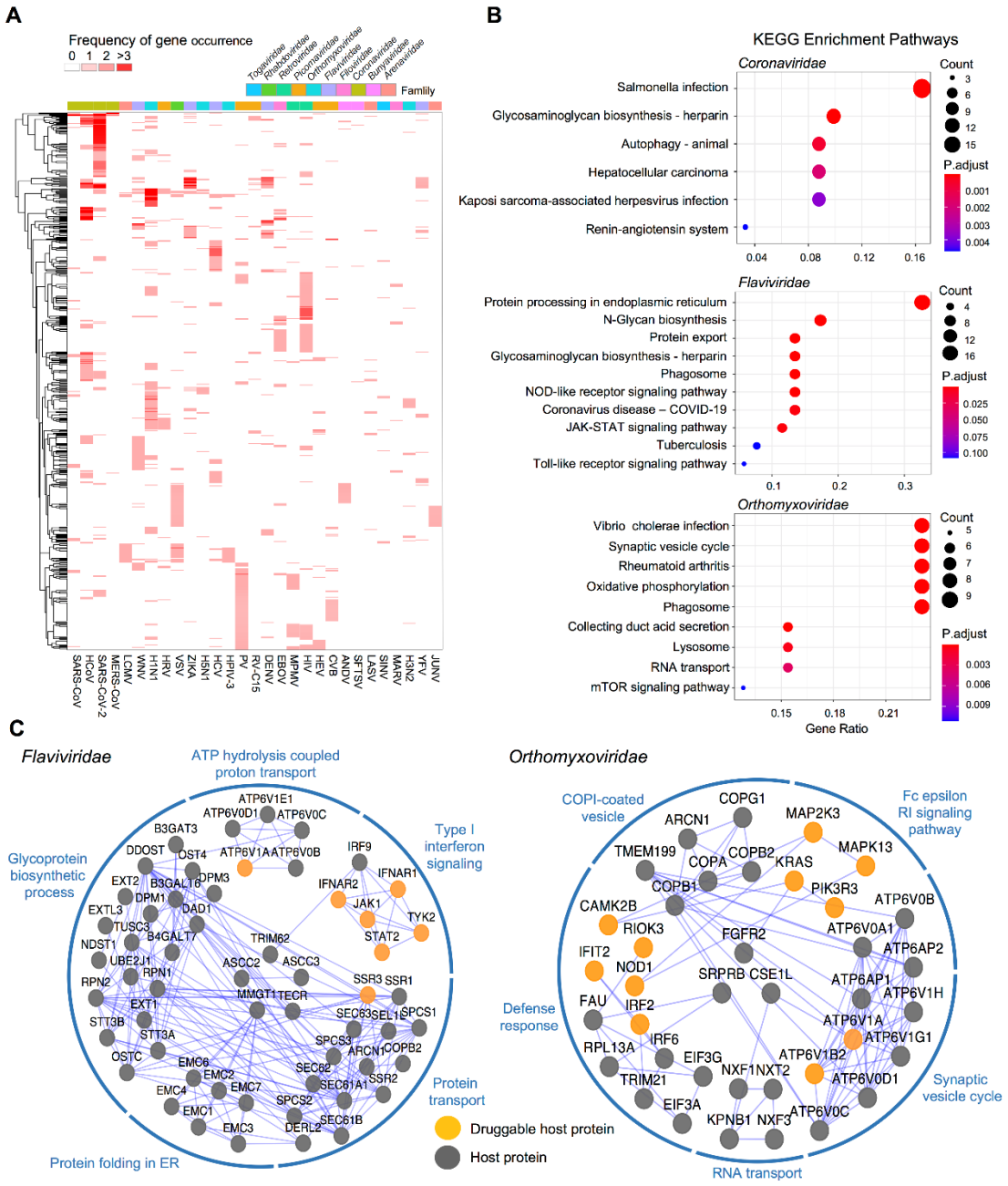


Figure S3. Comparative analysis and characterization of HDGs for indicated RNA virus families, Related to Figure 3

(A) The landscape of all the collected HDGs for indicated RNA viruses. The occurrence frequency of each HDG across studies was indicated by color legend.

(B) KEGG enrichment analysis of HDGs for the three indicated virus families. The size of the dot indicates the number of HDGs in the corresponding terms. The color of the dot represents the value of Benjamini and Hochberg FDR–adjusted p-value.

(C) The protein-protein interaction network of HDGs for *Flaviviridae* and *Orthomyxoviridae* virus families. Each HDG is presented as a node. The edge

between two nodes indicates a protein-protein interaction. The druggable HGDs with targeted drug candidates predicted in this study were highlighted.

Supplemental Figure 4

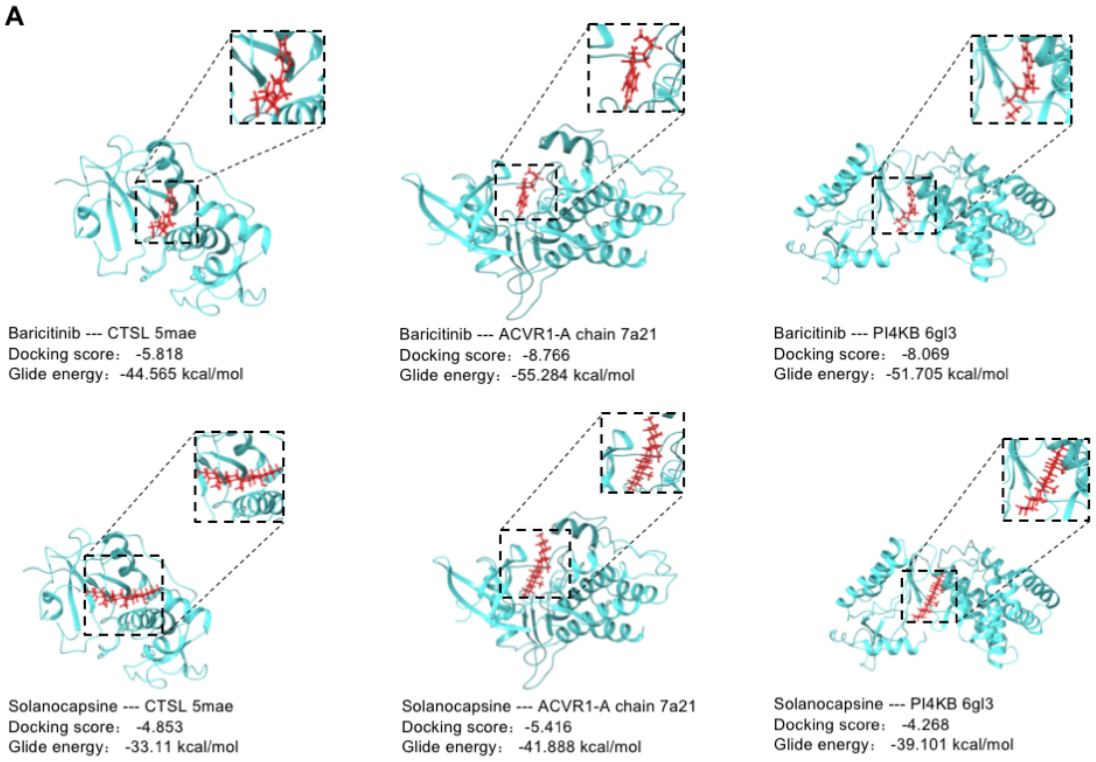


Figure S4. Molecular docking analysis for indicated drug-target pairs, Related to Figure 4

Molecular docking analysis showing the potential binding pockets between the repurposed drug Baricitinib and natural compound Solanocapsine with their corresponding targeted host factors.

Supplemental Tables

Table S1. Compendium of host dependency genes for multiple RNA viruses, Related to Figure 1 (attached dataset)

Table S2. Sequence sources for phylogenetic analysis, Related to Figure 2 (attached dataset)

Table S3. Re-analysis of CRISPR screening data, Related to Figure 3 (attached dataset)

Table S4. Functional gene enrichment analysis of host dependency genes, Related to Figure 3 (attached dataset)

Table S5. List of drug-target interactions and repurposed drug candidates, Related to Figure 4 (attached dataset)

Table S6. Summary of host dependency genes with repurposed drugs or natural compounds, Related to Figure 4

Virus Family	Druggable host dependence gene		High confidence host dependence gene
	Predicted	Known*	
Coronaviridae	29	32	166
Flaviviridae	79	8	81
Orthomyxoviridae	22	13	63

* Interaction Database: BindingDB, DGIdb3.0, DrugCentral, Stitch, ChEMBL

Table S7. Joint P-score ranking: the top ten repurposed FDA-approved drugs against *Flaviviridae* viruses, Related to Table 1

Drug candidate	Approved Indication	PubChem CID	Top10 Predicted Host Targets	Known interaction	Joint P-score
Alatrofloxacin	Bacterial infection	3086677	EMC2 RABGEF1 ARCN1 KRT31 COPB2 TRIM62 UGP2 SPCS3 DPM1 TEAD3	N/A	1.578
Fostamatinib	Chronic immune thrombocytopenia	11671467	NUAK2 TGFBR1 TYK2 JAK1 CHUK	SYK	1.240
Ozenoxacin	Impetigo	9863827	ARCN1 EMC2 RABGEF1 UGP2 COPB2 RPS6KL1 TRIM62 DPM1 KRT31 CKAP5	GyrA	1.191
Grepafloxacin	Bacterial infection	72474	UGP2 CKAP5 IFNAR1 SPCS3 SEC62 NOP56 UBE2J1 RPS6KL1 SSR2 OST4	N/A	1.091
Roxithromycin	Bacterial infection	6915744	SSR1 SEC62 SEC63 SSR2 CD81 IFNAR1 EMC3 DPM3 STT3A B3GALT6	N/A	0.810
Gramicidin D	Bacterial infection	45267103	DERL2 B4GALT7 EMC4 MMGT1 SEC61B OST4 DDOST DAD1 TUSC3 VMP1	N/A	0.732
Triclosan	Bacterial infection	5564	SEC61A1 VMP1 OST4 CD81 TUSC3 DAD1 DERL2 TMEM41B MMGT1 B3GALT6	N/A	0.700
Tofacitinib	Rheumatoid arthritis	9926791	TYK2 JAK1 NUAK2 TGFBR1 CHUK	JAK3 JAK2 JAK1	0.697
Omadacycline	Bacterial infection	54697325	UGP2 NOP56 CKAP5 ARCN1 COPB2 RPN1 CD81 SEC62 SPCS3 RPN2	N/A	0.600
Lasofoxifene	Osteoporosis	216416	ATP6V1E1 ATP6V0C ATP6V0D1 ATP6V0B ATP6V1A CHUK	ESR1、ESR2	0.597

N/A: Not applicable

Table S8. Joint PN-score ranking: the top ten repurposed FDA-approved drugs against *Flaviviridae* viruses, Related to Table 2

Drug candidate	Approved Indication	PubChem CID	Top10 Predicted Host Targets	Known interaction	Joint PN-score
Alatrofloxacin	Bacterial infection	3086677	EMC2 RABGEF1 ARCN1 KRT31 COPB2 TRIM62 UGP2 SPCS3 DPM1 TEAD3	N/A	1.566
Ozenoxacin	Impetigo	9863827	ARCN1 EMC2 RABGEF1 UGP2 COPB2 RPS6KL1 TRIM62 DPM1 KRT31 CKAP5	N/A	1.181
Grepafloxacin	Bacterial infection	72474	UGP2 CKAP5 IFNAR1 SPCS3 SEC62 NOP56 UBE2J1 RPS6KL1 SSR2 OST4	N/A	1.083
Roxithromycin	Bacterial infection	6915744	SSR1 SEC62 SEC63 SSR2 CD81 IFNAR1 EMC3 DPM3 STT3A B3GALT6	N/A	0.804
Gramicidin D	Bacterial infection	45267103	DERL2 B4GALT7 EMC4 MMGT1 SEC61B OST4 DDOST DAD1 TUSC3 VMP1	N/A	0.727
Triclosan	Bacterial infection	5564	SEC61A1 VMP1 OST4 CD81 TUSC3 DAD1 DERL2 TMEM41B MMGT1 B3GALT6a	N/A	0.695
Omadacycline	Bacterial infection	54697325	UGP2 NOP56 CKAP5 ARCN1 COPB2 RPN1 CD81 SEC62 SPCS3 RPN2	N/A	0.596
Lasofoxifene	Osteoporosis	216416	ATP6V1E1 ATP6V0C ATP6V0D1 ATP6V0B ATP6V1A CHUK	ESR1、ESR2	0.550
Candicidin	Bacterial infection	10079874	KAT8 WDR7 STAT2 UBE2J1 USP11 ASCC2 SEC63 RABGEF1 SEL1L TRIM62	N/A	0.536
Troleandomycin	Bacterial infection	202225	CKAP5 USP11 SEC62 SEC63 EMC7 B3GALT6 SSR1 SSR2 OSTC SSR1	N/A	0.392

N/A: Not applicable

Table S9. DeepCPI P-score ranking: the top ten repurposed natural compounds against *Flaviviridae* viruses, Related to Table 3

Drug candidate	TCMSP* MOL ID	PubChem CID	Herb	Top10 Predicted Host Targets	DeepCPI P-score
Lysergol	MOL005261	14987	Semen Pharbitidis	EMC2 ARCN1 RPN1 ATP6V1E1 TEAD3 COPB2 AAR2 RABGEF1 ATP6V0D1 IRF9	3.918
Atropine	MOL002219	174174	Lycii Cortex, Hyoscyami Semen	TEAD3 AAR2 PAPSS1 ARCN1 RPN1 COPB2 ATP6V0D1 IRF9 ASCC2 KAT8	3.591
Costaclavine	MOL008145	160462	Ricini Semen	ATP6V1E1 EMC2 ATP6V0D1 SEC61A1 RPN1 ARCN1 RPN2 HSPA13 DPM3 AAR2	3.545
Solanocapsine	MOL007356	73419	Solanum Nigrum	EMC2 RABGEF1 ATP6V1E1 TEAD3 KRT31 STAT2 ATP6V0D1 TRIM62 DPM1 KAT8	3.389
Chanoclavine	MOL005260	5281381	Semen Pharbitidis	RPN2 ATP6V1E1 ARCN1 RPN1 AAR2 EMC2 TEAD3 ATP6V0D1 KAT8 HSPA13	3.321
Triptofordin B1	MOL003232	122391803	Tripterygii Radix	IRF9 SEC61A1 RPN2 TEAD3 ASCC2 TMEM41B B4GALT7 IL4I1 STAT2 NDST1	2.931
Penniclavin	MOL005257	115247	Semen Pharbitidis	EMC2 ATP6V0D1 ARCN1 HSPA13 ATP6V0C SPCS3 TPT1 TMEM41B RPN2 UGP2	2.620
Picrasidine D	MOL012140	5316876	Asari Radix et Rhizoma	KAT8 RABGEF1 AAR2 TEAD3 STAT2 ARCN1 ASCC2 IRF9 TRIM62 RPS6KL1	2.806
elymoclavine	MOL005267	16758153	Semen Pharbitidis	EMC2 ATP6V0D1 ARCN1 SEC61A1 HSPA13 RPN1 DPM3 SPCS3 TPT1 TMEM41B	2.519
13,14-dehydrosophoridine	MOL006573		Sophorae Flavescentis Radix	IRF9 ARCN1 AAR2RABGEF1 STAT2 COPB2 TRIM62 PAPSS1 KAT8 RABEPK	2.614

*TCMSP database: Traditional Chinese Medicine Systems Pharmacology Database

Table S10. DeepCPI PN-score ranking: the top ten repurposed natural compounds against *Flaviviridae* viruses, Related to Table 4

Drug candidate	TCMSP* MOL ID	PubChem CID	Herb	Top10 Predicted Host Targets	DeepCPI PN-score
N-[6-(9-acridinylamino)hexyl]benzamide	MOL005935	146515	Sophora Japonica L.	NDST1 TMEM41B DERL2 VMP1 MMGT1 EMC3 ATP6V0D1 RPN2 TECR EMC1	0.535
14b-pregnane	MOL009604	176992	Lycii Fructus	ATP6V0D1 SPCS2 SPCS3 UGP2 ASCC3 MMGT1 CHUK VMP1 EMC2 DPM1	0.465
(S)-Canadine	MOL001455	21171	Chelidonii Herba	SEC61A1 ATP6V0C RPN2 TMEM41B ATP6V0B DERL2 HSPA13 VMP1 MMGT1 SPCS3	0.460
(R)-Canadine	MOL002903	443422	Coptidis Rhizoma	SEC61A1 ATP6V0C RPN2 TMEM41B ATP6V0B DERL2 HSPA13 VMP1 MMGT1 SPCS3	0.460
Androstane	MOL003790	6857536	Styrax	MMGT1 SPCS3 SPCS2 VMP1 CHUK ASCC3 HSPA13 UGP2 SEC61A1 TMEM41B	0.407
Narcein	MOL009329	8564	Papaveris Pericarpium	DERL2 OST4 NDST1 EXT2 TECR MMGT1 ATP6V0B SSR2 CD81 STT3A	0.404
malkangunin	MOL005360	90473155	Stemonae Radix	EMC1 PAPSS1 COPB2 EXTL3 ATP6V1A AAR2 ASCC3 SEL1L NDST1 USP11	0.390
alpha-berbine	MOL008182	164543	Dichroae Radix	SEC61A1 TMEM41B ATP6V0C ATP6V0B DERL2 SEC61B OST4 MMGT1 CD81 SPCS3	0.363
yohimbine	MOL008488	8969	Uncariae Ramulus Cumuncis	SEC61A1 DERL2 TMEM41B ATP6V0C VMP1 ATP6V0B TECR MMGT1 SPCS3 SEC61B	0.359
coryincine	MOL008635	3058605	Uncariae Ramulus Cumuncis	SEC61A1 DERL2 TMEM41B ATP6V0C VMP1 ATP6V0B TECR MMGT1 SPCS3 SEC61B	0.359

*TCMSP database: Traditional Chinese Medicine Systems Pharmacology Database

Table S11. Joint P-score ranking: the top ten repurposed FDA-approved drugs against *Orthomyxoviridae* viruses, Related to Table 1

Drug candidate	Approved Indication	PubChem CID	Top10 Predicted Host Targets	Known interaction	Joint P-score
Fostamatinib	Chronic immune thrombocytopenia	11671467	CAMK2B CLK3 FGFR2 GSK3A MAP2K3 MAPK13 RIOK3	SYK	2.724
Lasofoxifene	Osteoporosis	216416	ATP6AP1 ATP6V0A1 ATP6V0B ATP6V0C ATP6V0D1 ATP6V1A ATP6V1B2 ATP6V1G1 ATP6V1H	ESR1 ESR2	1.259
Baricitinib	Rheumatoid arthritis	44205240	CAMK2B CLK3 GSK3A MAP2K3 MAPK13 RIOK3	JAK1 JAK2 JAK3 TYK2	0.869
Tofacitinib	Rheumatoid arthritis	9926791	CAMK2B CLK3 GSK3A MAP2K3 RIOK3 MAPK13 FGFR2	JAK3 JAK2 JAK1	0.859
Lusutrombopag	Thrombocytopenia	49843517	CAMK2B CLK3 GSK3A MAP2K3 MAPK13 RIOK3	MPL	0.814
Bosutinib	Chronic Myelogenous Leukemia	5328940	MAPK13 ATP6V1H ATP6V1G1 ATP6AP1 ATP6V0D1 ATP6V1A	SRC ABL1	0.799
Raloxifene	Breast cancer	5035	ATP6AP1 ATP6V0B ATP6V0C ATP6V1A ATP6V1B2 ATP6V1G1 ATP6V1H	ESR1、ESR2	0.787
Etoricoxib	Rheumatoid arthritis	123619	CAMK2B CLK3 GSK3A MAP2K3 RIOK3 MAPK13	COX2	0.770
Hydroxychloroquine	Rheumatoid arthritis	3652	CAMK2B CLK3 GSK3A MAP2K3 RIOK3 MAPK13	N/A	0.667
Eltrombopag	Thrombocytopenia	135449332	CAMK2B CLK3 GSK3A MAP2K3 MAPK13 ATP6V1A FGFR2	CD110	0.637

N/A: Not applicable

Table S12. Joint PN-score ranking: the top ten repurposed FDA-approved drugs against *Orthomyxoviridae* viruses, Related to Table 2

Drug candidate	Approved Indication	PubChem CID	Top10 Predicted Host Targets	Known interaction	Joint PN-score
Fostamatinib	Chronic immune thrombocytopenia	11671467	CAMK2B CLK3 FGFR2 GSK3A MAP2K3 MAPK13 RIOK3	SYK	1.751
Lasofoxifene	Osteoporosis	216416	ATP6AP1 ATP6V0A1 ATP6V0B ATP6V0C ATP6V0D1 ATP6V1A ATP6V1B2 ATP6V1G1 ATP6V1H	ESR1、ESR2	1.212
Raloxifene	Breast cancer	5035	ATP6AP1 ATP6V0B ATP6V0C ATP6V1A ATP6V1B2 ATP6V1G1 ATP6V1H	ESR1、ESR2	0.724
Lusutrombopag	Thrombocytopenia	49843517	CAMK2B CLK3 GSK3A MAP2K3 MAPK13 RIOK3	MPL	0.632
Baricitinib	Rheumatoid arthritis	44205240	CAMK2B CLK3 GSK3A MAP2K3 MAPK13 RIOK3	JAK1 JAK2 JAK3 TYK2	0.617
Etoricoxib	Rheumatoid arthritis	123619	CAMK2B CLK3 GSK3A MAP2K3 RIOK3 MAPK13	COX2	0.553
Vancomycin	Bacterial infection	14969	ATP6V1A ATP6AP1 ATP6V0B ATP6V0C ATP6V0D1 ATP6V1G1 ATP6V1H ATP6V1B2 PGD MAPK13	N/A	0.526
Eltrombopag	Thrombocytopenia	135449332	CAMK2B CLK3 GSK3A MAP2K3 MAPK13 ATP6V1A FGFR2	CD110	0.495
Hydroxychloroquine	Rheumatoid arthritis	3652	CAMK2B CLK3 GSK3A MAP2K3 RIOK3 MAPK13	N/A	0.495
Ibandronate	Osteoporosis	60852	ATP6V0A1 ATP6V1H	FPPS	0.459

N/A: Not applicable

Table S13. DeepCPI P-score ranking: the top ten repurposed natural compounds against *Orthomyxoviridae* viruses, Related to Table 3

Drug candidate	TCMSP* MOL ID	PubChem CID	Herb	Top10 Predicted Host Targets	DeepCPI P-score
Lysergol	MOL005261	14987	Semen Pharbitidis	ARCN1 ATP6V0D1 COPB2 COPG1 FAU IFIT5 IVNS1ABP KPNB1 NXT2 PIK3R3	4.410
Atropine	MOL002219	174174	Lycii Cortex, Hyoscyami Semen	ARCN1 ATP6V0D1 COPB2 COPG1 EIF3G IFIT5 IRF6 IVNS1ABP PGD PIK3R3	3.988
Costaclavine	MOL008145	160462	Ricini Semen	ARCN1 ATP6V0D1 COPG1 EIF3A FAU IFIT5 KPNB1 NXT2 PIK3R3 SLC1A3	3.935
Chanoclavine	MOL005260	5281381	Semen Pharbitidis	ARCN1 ATP6V1B2 COPG1 EIF3A EIF3G FAU IFIT5 KPNB1 NXT2 PIK3R3	3.868
Solanocapsine	MOL007356	73419	Solanum Nigrum	ATP6V0D1 COPB2 COPG1 EIF3A IFIT5 KPNB1 NXT2 PIK3R3 RIOK3 TRIM21	3.770
Penniclavin	MOL005257	115247	Semen Pharbitidis	EIF3A KPNB1 ATP6V0D1 SLC1A3 ARCN1 IFIT5 STARD5 FAU IFIT2 ATP6V0C	3.151
Picrasidine D	MOL012140	5316876	Asari Radix et Rhizoma	IRF6 TRIM21 ARCN1 COPG1 PHF3 AKAP13 EIF3A PIK3R3 SF3A1 NXF1	3.149
elymoclavine	MOL005267	440904	Semen Pharbitidis	ARCN1 ATP6V0D1 EIF3A FAU IFIT2 IFIT5 KPNB1 PIK3R3 SLC1A3 STARD5	3.099
(-)-9alpha-hydroxysophoramine	MOL006570	50695119	Sophorae Flavescentis Radix	IRF6 IVNS1ABP EIF3G PIK3R3 COPG1 BUB3 NXF3 SF3A1 EIF3A ARCN1	3.006
Triptofordin B1	MOL003232	122391803	Tripterygii Radix	IVNS1ABP IRF6 COPG1 ATP6V0A1 TRIM21 NXT2 ATP6AP2 EIF3A NXF1 SLC1A3	2.984

*TCMSP database: Traditional Chinese Medicine Systems Pharmacology Database

Table S14. DeepCPI PN-score ranking: the top ten repurposed natural compounds against *Orthomyxoviridae* viruses, Related to Table 4

Drug candidate	TCMSP* MOL ID	PubChem CID	Herb	Top10 Predicted Host Targets	DeepCPI PN-score
Deoxycamptothecine	MOL008209	169724	Andrographis Herba	ARCN1 ATP6V0D1 ATP6V1G1 ATP6V1H EIF3A EIF3G FAU IFIT2 IFIT5 NXT2	0.908
Lysergol	MOL005261	14987	Semen Pharbitidis	ARCN1 ATP6V0D1 COPB2 COPG1 FAU IFIT5 IVNS1ABP KPNB1 NXT2 PIK3R3	0.690
Chanoclavine	MOL005260	5281381	Semen Pharbitidis	ARCN1 ATP6V1B2 COPG1 EIF3A EIF3G FAU IFIT5 KPNB1 NXT2 PIK3R3	0.685
Costaclavine	MOL008145	160462	Ricini Semen	ARCN1 ATP6V0D1 COPG1 EIF3A FAU IFIT5 KPNB1 NXT2 PIK3R3 SLC1A3	0.608
elymoclavine	MOL005267	440904	Semen Pharbitidis	ARCN1 ATP6V0D1 EIF3A FAU IFIT2 IFIT5 KPNB1 PIK3R3 SLC1A3 STARD5	0.580
Solanocapsine	MOL007356	73419	Solanum Nigrum	ATP6V0D1 COPB2 COPG1 EIF3A IFIT5 KPNB1 NXT2 PIK3R3 RIOK3 TRIM21	0.562
Penniclavin	MOL005257	115247	Semen Pharbitidis	EIF3A KPNB1 ATP6V0D1 SLC1A3 ARCN1 IFIT5 STARD5 FAU IFIT2 ATP6V0C	0.531
Vitexifolin C	MOL011912	11033408	Viticis Fructus	KPNB1 FAU ATP6V0D1 MAP2K3 FGFR2 EIF3A IFIT2 COPB2 ATP6V0B CMAS	0.428
Atropine	MOL002219	174174	Lycii Cortex, Hyoscyami Semen	ARCN1 ATP6V0D1 COPB2 COPG1 EIF3G IFIT5 IRF6 IVNS1ABP PGD PIK3R3	0.426
Isolimonic acid	MOL013443	131752314	Aurantii Fructus Immaturus	IFIT5 ARCN1 NXF1 RIOK3 KPNB1 FAU USP46 EIF3A IFIT2 ATP6V0D1	0.420

*TCMSP database: Traditional Chinese Medicine Systems Pharmacology Database

Table S15. Key parameters of molecular docking analysis, Related to Figure 4

PubChem CID	Drug or natural compound	PDB ID	Target	Docking score	Glide energy (kcal/mol)
44205240	Baricitinib	5MAE	CTSL	-5.818	-44.565
44205240	Baricitinib	6EIS	DYRK1A-A chain	-9.431	-48.038
44205240	Baricitinib	6EIS	DYRK1A-B chain	-8.288	-48.185
44205240	Baricitinib	6EIS	DYRK1A-C chain	-8.289	-46.526
44205240	Baricitinib	6EIS	DYRK1A-D chain	-6.793	-44.422
44205240	Baricitinib	6GL3	PI4KB	-8.069	-51.705
44205240	Baricitinib	7A21	ACVR1-A chain	-8.766	-55.284
44205240	Baricitinib	7A21	ACVR1-B chain	-8.413	-50.806
73419	Solanocapsine	5MAE	CTSL	-4.853	-33.11
73419	Solanocapsine	6EIS	DYRK1A-A chain	-4.128	-34.175
73419	Solanocapsine	6EIS	DYRK1A-B chain	-4.297	-37.025
73419	Solanocapsine	6EIS	DYRK1A-C chain	-4.961	-29.42
73419	Solanocapsine	6EIS	DYRK1A-D chain	-5.011	-34.605
73419	Solanocapsine	6GL3	PI4KB	-4.268	-39.101
73419	Solanocapsine	7A21	ACVR1-A chain	-5.416	-41.888
73419	Solanocapsine	7A21	ACVR1-B chain	-5.413	-42.105
14987	Lysergol	6FYV	CLK4	-7.132	-23.175
5316876	Picrasidine D	6FYV	CLK4	-6.939	-31.901
154417	Hyoscyamine	6FYV	CLK4	-6.279	-18.605
14987	Lysergol	5T4E	DPP4	-6.196	-31.612
154417	Hyoscyamine	5T4E	DPP4	-6.064	-29.032

Table S16. 2D structures of the top drug candidates, Related to Tables 1-4 (attached dataset)

Transparent Methods

Host dependency gene collection and literature mining

By systematically searching the literature to date, studies performing genetic screening for human-specific HDGs corresponding to RNA viruses were collected. Screens for DNA viruses or in non-human cells were not included with an exception for SARS-CoV-2 virus-related screens. We collected all the recently published viral resistance CRISPR screens against SARS-CoV-2 virus, with 5 studies in human cells and 1 study in Vero-E6 cells (Table S1). Under this criteria, data from 63 studies with different genetic perturbation techniques (CRISPR knockout, RNAi and haploid gene-trap mutagenesis) were collected. These studies identified virus-specific HDGs for 29 RNA viruses spanning 10 RNA virus families. Due to the high interest for *Coronaviridae* virus family, we collected additional 34 individual gene-focused non-screening studies to include as many *Coronaviridae* HDGs as possible. A gene is defined as a HDG when it meets any of the following criteria: 1) Its loss-of-function impedes or reduces viral infection or activity by experimental evidence in non-screen studies; 2) It has been clearly classified into HDG group in screen studies; 3) When HDG group is not specified in screen studies, we took the top ~5% of all the interrogated genes in the positive selection list as HDGs with a custom log fold change cutoff in CRISPR knockout or RNAi screens challenged by the corresponding virus. The detailed information concerning to these literatures and HDGs was summarized in Table S1. For *Coronaviridae*, *Flaviridae* and *Orthomyxoviridae* viruses, we only took a subset of HDGs that occurred more than once within its corresponding family as high confidence HDGs for further analysis. In general, around one hundred HDGs for each group of the above three virus families were used for molecular characterization and drug repurposing analysis (Table S6).

Phylogenetic tree construction

The sequences of nucleic acid and protein corresponding to viral RNA-dependent RNA polymerase (RdRp) gene for indicated RNA viruses were downloaded from online sources (https://www.ncbi.nlm.nih.gov) and were used for phylogenetic tree analysis (Table S2). The nucleic acid and protein sequences were analyzed by Multiple Sequence Alignment in Muscle calculation using MEGA X software. The phylogenetic tree was subsequently constructed based on neighbor-joining (NJ) method or maximum parsimony (MP) method using pairwise phylogenetic distance with 1000 bootstrap replicates.

Re-analysis of CRISPR screening data

Among the 25 CRISPR screening studies, we downloaded the raw sequencing or read count data from 7 studies wherever these raw data were available. We re-analyzed these CRISPR screening data to re-call the HDGs using the same

MAGeCK-VISPR pipeline (Li et al., 2015). In total, 36 samples across the 9 viruses are included in the analysis. The beta scores of each screening, generated by MAGeCK-VISPR, were combined together and normalized using quantile normalization. Next, we filtered the data using the following two thresholds: First, the maximum of the beta score of a gene across all the samples must be greater than 3. Second, the average beta score of a gene across all the samples must be greater than 1. After filtering, 261 genes were retained as positively selected HDG hits. Then hierarchical clustering and protein-protein interaction network was performed using StringDB.

KEGG and GO enrichment analysis

The high confidence HDGs for *Coronaviridae*, *Flaviridae* and *Orthomyxoviridae* viruses (166, 81 and 63, respectively) were used for this analysis (Table S6). KEGG and GO enrichment analysis were performed using clusterProfiler R package with a strict cutoff of p-value < 0.001 and false discovery rate (FDR) < 0.05 (Yu et al., 2012). Enrichment analyses were visualized using the R package clusterProfiler with default settings.

Network analysis

The input HDGs were uploaded to the STRING database (version 11.0, https://string-db.org) and high confidence protein-protein interactions (PPIs) were extracted with a minimum required interaction score ≥ 0.7. Next, the interactions were imported into Cytoscape 3.2.1 software to visualize PPI Network. The druggable HDG-encoding proteins with predicted drug candidates in this study and proteins classified into certain functional protein complexes or biological processes are highlighted.

Drug candidate selection for repurposing

FDA-approved drug information was extracted from DrugBank database (version 5.1.7, released 2020-07-02; https://www.drugbank.ca) corresponding to 2352 marketed drugs with InChI (the IUPAC International Chemical Identifier) key information. Natural compound information is downloaded from Traditional Chinese Medicine Systems Pharmacology (TCMSP) online database (version 2.3, released 2014-05-31; https://tcmspw.com/tcmsp.php) which is a unique systems pharmacology platform of Chinese herbal medicines (Ru et al., 2014). To select the most favorable compound candidates, we filtered the pool of 1455 natural compounds by requiring each candidate passing the criteria of oral bioavailability (OB) ≥ 30.0 %, drug-likeness (DL) ≥ 0.18 and blood-brain barrier (BBB) ≥ -0.30, and finally ended up with 1062 selected natural compounds for the downstream DTI analysis.

DTI retrieval from related databases

Known drug-target interactions were extracted according to annotated information associated with related drugs, compounds or target genes from multiple databases including BindingDB (updated 2020-03-01), DGIdb3.0 (version 3.0.2), DrugCentral (version 10.12) and Stitch (version 5.0) (Cotto et al., 2018; Gilson et al., 2016; Kuhn et al., 2010; Ursu et al., 2019). The high confidence HDGs for *Coronaviridae*, *Flaviridae* and *Orthomyxoviridae* viruses were used for the DTI analysis (Table S6). One HDG may be associated with multiple drugs or compounds. Only FDA-approved drugs and selected natural compounds were considered for compiling these known DTI information for drug repurposing.

DTI prediction by DeepCPI

The source code of DeepCPI can be downloaded from https://github.com/FangpingWan/DeepCPI. The binding activity score for each drug-target pair was predicted by providing the InChl key information of a drug or compound and the amino acid sequence of a protein target from UniProt database. We applied DeepCPI on 4,563 high confidence DTIs out of 7,444,710 putative pairs (3,030 druggable proteins and 2,457 FDA-approved drugs) extracted from DGIdb3.0 database (version 3.0.2) as a benchmark analysis and determined an optimal threshold with a normalized z-score ≥ 0.641 (sensitivity: 73%; specificity: 51.9%) by receiver operating characteristics (ROC) analysis. We then used this cutoff to filter confident DTI in our analysis for virus-related HDGs and FDA-approved drugs as well as selected natural compounds.

DTI prediction by DTINet

The source code of DTINet can be downloaded from https://github.com/luoyunan/DTINet. The drug-protein interactions and protein-protein interactions were extracted from UniProt database. The drug-disease associations and protein-disease associations were extracted from the Therapeutic Target Database (Wang et al., 2020). The drug-drug interactions were extracted from the BioSNAP Network database (http://snap.stanford.edu/biodata/). Then the Jaccard similarity for these interactions/associations was calculated to further augment the heterogeneity. A heterogeneous network (including three types of nodes and five types of edges) are constructed using these diverse drug-related and protein-related information for the prediction task. The informative, but low-dimensional feature vector was obtained by integrating the diverse information from the heterogeneous network by combining the network diffusion algorithm (random walk with restart, RWR) with a dimensionality reduction scheme (diffusion component analysis, DCA). The restart probability is set to 0.50 and the maximum number of iterations is set to 20. Intuitively, the low-dimensional feature vector is used to encode the relational properties (e.g., similarity), association information and topological context of each drug (or protein) node

in the heterogeneous network. Finally, the score for each drug-protein pair was calculated based on the feature vectors by DTINet default parameters. Similar to DeepCPI analysis, we also applied DTINet on the benchmark datasets and determined an optimal threshold with a normalized z-score ≥ 0.973 (sensitivity: 88.9%; specificity: 63.8%) by ROC analysis. We then use this cutoff to filter confident DTI in our analysis for virus-related HDGs and FDA-approved drugs. Due to the insufficient prior data for proper modeling, DTINet was not applied for natural compound DTI analysis.

Prioritizing repurposed drug candidates

The repurposed FDA-approved drugs were prioritized by both known DTI and predicted DTI with high confidence. The candidate drugs were ranked by predicted DTI scores with known DTI annotation accompanied to the drug if any. We adopted two ranking methods to prioritize these candidates. The first ranking method only considers the HDG target-associated DTIs. For FDA-approved drugs with both DeepCPI and DTINet DTI prediction, we extracted mutual confident DTIs by both prediction algorithms and the mean of normalized z-score by each prediction tool was calculated as a positive score (P-score). A joint P-score by the sum of DeepCPI and DTINet P-score was employed to rank the drug candidates. The second ranking method not only considers HDG targets, but also incorporates non-HDG targets and common essential gene targets to evaluate drug promiscuousness and cytotoxicity effects. In addition to P-score, we introduced a negative score for DTIs between a given drug and non-HDG (among 3,030 druggable proteins in DGIdb3.0 database) or essential gene targets (676 core essential gene-encoded proteins) (Wang et al., 2019). An arbitrary weight was set for positive score (1) and negative score (-0.333) for multiplexing to generate a PN-score. For FDA-approved drugs, a joint PN-score was reported by adding the DeepCPI and DTINet PN-score together, and used for ranking the drugs. For natural compounds, we also employed these two ranking methods using either DeepCPI P-score or DeepCPI PN-score.

The detailed formula was as follows:

For a given drug-target pair, we calculated the DTI score t_{CPI} and t_{Net} by DeepCPI and DTINet, respectively. By collecting all the DTI scores, two score matrices T_{CPI} and T_{Net} were defined to quantify the confidence of predicted DTIs:

$$\begin{cases} T_{CPI} \in \mathbb{R}^{l \times k} \\ T_{Net} \in \mathbb{R}^{l \times k} \end{cases} \quad (1)$$

Where, l refers to the length of drug list and k refers to the length of target list.

To ensure them comparable, the score matrices T_{CPI} and T_{Net} were normalized by Z-Score measurement:

$$\begin{cases} Z_{CPI} = \frac{x_{CPI} - \mu_{CPI}}{\sigma_{CPI}}, x_{CPI} \in T_{CPI} \\ Z_{Net} = \frac{x_{Net} - \mu_{Net}}{\sigma_{Net}}, x_{Net} \in T_{Net} \end{cases} \quad (2)$$

Where, μ is mean value of the scores and σ is standard deviation of the scores.

We further applied an optimal threshold (as discussed above, 0.641 and 0.973 were used for Z_{CPI} and Z_{Net}, respectively) to filter the non-significant scores and only keep the confident DTI scores:

$$Z_{CPI\_sig} = \begin{cases} z, if\ z \geq 0.641 \\ 0, if\ z < 0.641 \end{cases} \quad z \in Z_{CPI} \quad (3)$$

$$Z_{Net\_sig} = \begin{cases} z, if\ z \geq 0.973 \\ 0, if\ z < 0.973 \end{cases} \quad z \in Z_{Net} \quad (4)$$

For each FDA-approved drug, the mean value of the normalized z-scores was defined as its positive score:

$$\begin{cases} P\_score_{CPI} = \sum_{i=1}^{k} z_i^{CPI} / k \\ P\_score_{Net} = \sum_{i=1}^{k} z_i^{Net} / k \end{cases} \quad (5)$$

Similar as above, we defined negative scores $N\_score_{druggable}$ and $N\_score_{essentialome}$ for non-HDG and essential gene targets, respectively. The final negative was the sum of $N\_score_{druggable}$ and $N\_score_{essentialome}$:

$$\begin{cases} N\_score_{CPI} = N\_score_{Druggable_{CPI}} + N\_score_{Essentialome\_CPI} \\ N\_score_{Net} = N\_score_{Druggable_{Net}} + N\_score_{Essentialome\_Net} \end{cases} \quad (6)$$

The PN-score was the sum of weighted positive score and negative score:

$$\begin{cases} PN\_score_{CPI} = 1 * P\_score_{CPI} + (-0.333) * N\_score_{CPI} \\ PN\_score_{Net} = 1 * P\_score_{Net} + (-0.333) * N\_score_{Net} \end{cases} \quad (7)$$

Here, we defined a joint P-score by the sum of P_score_{CPI} and P_score_{Net} for each drug:

$$Joint\_P\_score = P\_score_{CPI} + P\_score_{Net} \quad (8)$$

The joint PN-score was the sum of PN_score_{CPI} and PN_score_{Net} for each drug:

$$Joint\_PN\_score = PN\_score_{CPI} + PN\_score_{Net} \quad (9)$$

Molecular Docking

The structures of target protein were downloaded from PDB database (http://www.rcsb.org). The drug or compound structures were downloaded from TCMSP and PubChem database (https://pubchem.ncbi.nlm.nih.gov). The structures of proteins and compounds were imported into prime tool of Maestro

(version 11.8.012) suite of Schrödinger software (released 2018-4). Next the preprocessing step was performed by adding hydrogens and missing atoms as well as removing water molecules for the proteins using the Protein Preparation tool. Ligand preprocessing was performed using default settings with Ligprep tool of Maestro software. Then, the top-ranked potential binding site was defined using Receptor Grid Generation tool. Glide tool was used to detect the interactions between ligands and proteins. The docking score ≤ -6 was considered as a high confidence binding event between tested ligand and protein. The Glide energy for each docking pair was also shown in Table S15. The 2D structures of the top candidate drugs were presented in Table S16.

Supplemental References

Cotto, K.C., Wagner, A.H., Feng, Y.Y., Kiwala, S., Coffman, A.C., Spies, G., Wollam, A., Spies, N.C., Griffith, O.L., and Griffith, M. (2018). DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. Nucleic acids research *46*, D1068-D1073.

Gilson, M.K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L., and Chong, J. (2016). BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic acids research *44*, D1045-1053.

Kuhn, M., Szklarczyk, D., Franceschini, A., Campillos, M., von Mering, C., Jensen, L.J., Beyer, A., and Bork, P. (2010). STITCH 2: an interaction network database for small molecules and proteins. Nucleic acids research *38*, D552-556.

Li, W., Koster, J., Xu, H., Chen, C.H., Xiao, T., Liu, J.S., Brown, M., and Liu, X.S. (2015). Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. Genome biology *16*, 281.

Ru, J., Li, P., Wang, J., Zhou, W., Li, B., Huang, C., Li, P., Guo, Z., Tao, W., Yang, Y.*, et al.* (2014). TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. Journal of cheminformatics *6*, 13.

Ursu, O., Holmes, J., Bologa, C.G., Yang, J.J., Mathias, S.L., Stathias, V., Nguyen, D.T., Schurer, S., and Oprea, T. (2019). DrugCentral 2018: an update. Nucleic acids research *47*, D963-D970.

Wang, B., Wang, M., Zhang, W., Xiao, T., Chen, C.H., Wu, A., Wu, F., Traugh, N., Wang, X., Li, Z.*, et al.* (2019). Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. Nature protocols *14*, 756-780.

Wang, Y., Zhang, S., Li, F., Zhou, Y., Zhang, Y., Wang, Z., Zhang, R., Zhu, J., Ren, Y., Tan, Y.*, et al.* (2020). Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. Nucleic acids research *48*, D1031-D1041.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics : a journal of integrative biology *16*, 284-287.