A Computational Framework of Host-Based Drug Repositioning for Broad-Spectrum Antivirals against RNA Viruses

Zexu Li¹,²,#, Yingjia Yao¹,²,#, Xiaolong Cheng³,⁴,#, Qing Chen³,⁴, Wenchang Zhao¹,², Shixin Ma¹,², Zihan Li¹,², Hu Zhou⁵,⁶, Wei Li³,⁴,* Teng Fei¹,²,*

¹College of Life and Health Sciences, Northeastern University, Shenyang 110819, People’s Republic of China; ²Key Laboratory of Data Analytics and Optimization for Smart Industry (Northeastern University), Ministry of Education, People’s Republic of China ³Center for Genetic Medicine Research, Children’s National Hospital, 111 Michigan Ave NW, Washington, DC, 20010, USA ⁴Department of Genomics and Precision Medicine, George Washington University, 111 Michigan Ave NW, Washington, DC, 20010, USA ⁵School of Pharmaceutical Sciences, Fujian Provincial Key Laboratory of Innovative Drug Target Research, Xiamen University, Xiamen, Fujian, 361102, China ⁶High Throughput Drug Screening Platform, Xiamen University, Xiamen, Fujian, 361102, China

#These authors contributed equally  *Correspondence: wli2@childrensnational.org; feiteng@mail.neu.edu.cn

Keywords: RNA virus; drug repositioning; drug repurposing; SARS-CoV-2; COVID-19

Summary
RNA viruses are responsible for many types of zoonotic diseases that post great challenges for public health system. Effective therapeutics against these viral infections remains limited. Here we deployed a computational framework for host-based drug repositioning to predict potential antiviral drug candidates from 2352 approved drugs and 1062 natural compounds embedded in Traditional Chinese Medicine herbs. By systematically interrogating public genetic screening data, we comprehensively catalogued human-specific host dependency genes that are indispensable for the successful viral infection corresponding to 10 families and 29 species of RNA viruses. In addition, we utilized these host dependency genes as potential drug targets, and interrogated extensive drug-target interactions through multiple ways such as database retrieval, literature mining and de novo prediction using artificial intelligence-based algorithms. Repurposed drugs or natural compounds were proposed for combating many viral pathogens such as coronaviruses (e.g., SARS-CoV-2), flaviviruses (e.g., Zika virus) and influenza viruses. This study helps to prioritize promising drug candidates for further therapeutic evaluation against these viral-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 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 caused 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 towards novel 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 anti-viral 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 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 unsatisfactory response during following clinical trials (Wang et al., 2020a; Wang et al., 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 novel 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 anti-viral 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 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 was illustrated in Figure 1. Firstly, we sought to systematically catalogue 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 work 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. Screening data from 56 independent studies spanning 10 families and 29 species of RNA viruses were collected (Figure 1A; Table S1). Considering the limited high-throughput screening data for Coronaviridae, we performed in depth literature mining to include individual HDGs identified from 27 Coronaviridae-focused studies (Table S1). Notably, we only considered studies using human-derived cells or tissues as host system 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 a ranked list for marketed drugs or natural compounds were recommended as potential antiviral solutions (Figure 1D).

**Cataloguing 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 27 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 61% (34 out of 56) 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, 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; 3) batch effect across independent studies.

To further examine the variations of HDG calling across different studies, we re-analyzed CRISPR screening data where raw sequencing or count data is 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 for a “β score”, the more positive selection for the corresponding gene, and the more likely for the gene to be a HDG hit in viral resistance screen. 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 as 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 interaction 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 source or biological difference per se (Figure S3A). To minimize the analytic bias due to data insufficiency and fluctuation, we primarily focused on *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). In addition, we also included *Coronaviridae* family for the following analysis in response to COVID-19 and earlier severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS). For HDGs in *Flaviviridae* or *Orthomyxoviridae* families, 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. On the other hand, all the HDGs for *Coronaviridae* family were considered due to data insufficiency. 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). For COVID-19 pathogen SARS-CoV-2, we have not found systematic study yet to identify HDGs in human cells despite one recent CRISPR screen performed in African green monkey Vero-E6 cells (Wei et al., 2020). Considering the potential discrepancy of host response between human and monkey, we did not include Vero-E6 data into our analysis. Nevertheless, we reasoned that SARS-CoV-2 may share some consensus HDGs with other pathogenic coronaviruses considering that ACE2 and TMPRSS2, the two critical host genes for SARS-CoV-2 cell entry, are also shared by SARS-CoV or MERS-CoV (Figure 3A). These results support the possibility to develop broad-spectrum host-based antivirals for multiple coronaviruses.

Pathway and functional gene category enrichment analysis with Kyoto Encyclopedia of Genes and Genomes (KEGG) and gene ontology (GO) tools showed that MAPK pathway and protein phosphorylation regulators are significantly enriched among HDGs for *Coronaviridae* viruses (Figures 3B, 3C and S3B; Table S4). In contrast, *Flaviridae* 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).
Mining known drug-target interaction 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, 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 2352 drugs approved by 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 1062 natural compounds that are active ingredients of traditional Chinese medicine (TCM) herbs 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 30~130 FDA-approved drugs targeting HDGs for respective Coronaviridae, Flaviridae and Orthomyxoviridae viral families (Figure 4A and Tables S5 and S6).

Predicting novel drug-target interactions

We employed two machine learning methods (DeepCPI and DTINet) to predict novel 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. 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 under stringent cutoffs for FDA-approved drug repurposing (Figure 4B and S4A). 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 (Figure 4A; Table S5). Importantly, there is a significant portion of repurposed FDA-approved drugs that are retrieved from both known and predicted DTIs (Figure 4C), further strengthening the confidence of DTIs and resultant drug repositioning.

Prioritizing candidate drugs and natural compounds

We prioritized potential drug candidates from known and predicted DTIs. Since one drug may target multiple HDG targets that may produce enhanced anti-viral response, we ranked these repurposed drugs according to the sum score
of predicted interactions (Tables S5). Top candidate drugs and natural compounds for anti-viral purpose against *Coronaviridae*, *Flaviridae* and *Orthomyxoviridae* viruses were listed in Tables 1, 2 and Tables S7-S10.

Top ten FDA-approved drugs repurposed for *Coronaviridae* viruses are predominantly kinase inhibitors marketed to treat either cancer or immune-related diseases (Table 1). Our analysis predicted many more drug targets compared to the original targets for each drug (Table 1). To further determine the quality of the predicted interactions, we took the top candidate drug Fostamatinib as an example and performed molecular docking analysis to simulate its interaction with several predicted target proteins. The *in silico* structure-based analysis showed high binding affinity between Fostamatinib and its predicted targets as shown by the docking scores, further supporting their potential interactions (Figures 4D and S4B; Table S11). Of note, among these top drug candidates for *Coronaviridae* viruses, Baricitinib, a JAK inhibitor approved for rheumatoid arthritis treatment, has been shown to lower the cytokine effect and reduce the viral load in COVID-19 patients by targeting 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). Another JAK inhibitor Tofacitinib among the top ten repurposed *Coronaviridae*-targeting drugs is also being evaluated for COVID-19 treatment in active clinical trial (NCT04415151). Moreover, Tofacitinib was previously shown to be a potent inhibitor for immunodeficiency virus type 1 (HIV-1) replication *in vitro*, further supporting its antiviral activity (Gavegnano et al., 2014). Another repurposed drug Everolimus approved for cancer therapeutics was reported to suppress cytomegalovirus infection (Tan et al., 2019). In addition, approved CDK4/6 inhibitor Palbociclib was shown to inhibit herpes simplex virus type 1 (HSV-1) replication in primary monocyte-derived macrophages (Badia et al., 2016). Although these drugs are not originally designed for antiviral uses, it is not surprising that they can act to inhibit viral infection by targeting host restriction factors for viruses. Taken together, our analysis provides encouraging repurposing candidates for antiviral application.

Candidate natural compounds were ranked primarily according to the sum of DeepCPI scores (Tables 2, S8 and S10). We also summarized the herbs that include the corresponding compound as part of their active ingredients. The top predicted natural compound against *Coronaviridae* viruses is an anticholinergic agent hyoscyamine that is often used to treat some gastrointestinal and bladder conditions (Table 2). This compound ranked the best because of its maximal multi-targeting potential for a plethora of HDGs. Other selected natural compounds such as sophocarpine and β-carboline derivatives have been reported to exhibit antiviral activities against enterovirus 71 (EV71) and HSV-1.
respectively (Gonzalez et al., 2018; Jin et al., 2017). Interestingly, Asari Radix et Rhizoma (Xi Xin) and Codonopsis Radix (Dang Shen), the TCM herbs that contain selected compounds 4,9-Dimethoxy-1-vinyl-beta-carboline 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, 8th edition (http://www.nhc.gov.cn/yzygj/). Representative molecular docking analysis was also performed for compound Lysergol and its predicted targets, and again high-affinity interaction modules can be generated between the compound and predicted targets (Figures 4D and S4B; Table S11). 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.

**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 a novel 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, 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, 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; 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 novel 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.

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; Rıfaioglu 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 covers FDA-approved drugs, but also include natural compounds especially present in TCM herbs. When the host targets are not enriched for kinase proteins, the repurposed natural compounds generally exhibit broader targeting capability than FDA-approved drugs as is the case for Flaviridae and Orthomyxoviridae (Tables S7-S10). This is not surprising because FDA-approved drug is often designed to specifically target one target and many drugs are kinase inhibitors. Therefore, taking natural compounds extracted from TCM herbs into consideration may help to compensate the targeting bias whiles maintain
appreciable level of safety and efficacy potential. Since our approach is primarily based on targeting HDGs, the viral families that share common druggable host targets will result in similar repurposed drug or compounds as exemplified by Flaviridae and Orthomyxoviridae viruses (Tables S7-S10). The fundamental difference of this study with previous drug repositioning work largely lies in target selection, DTI determination and final repurposed drug candidates.

As for candidate drugs for COVID-19, we may resort to the repurposed drug list for Coronaviridae viral family. In addition to those promising hits currently under clinical trials, here we also put forward several novel options either from marketed drugs or from natural products for further investigation. As host-based drugs do not directly act on viruses, rather they target host function that may orchestrate in complicated molecular, cellular, and/or even more advanced systematic levels such as immune-related processes. In this regard, we expect that appropriate evolution of these host-based drugs would require human-specific integrative systems including but not limited to in vitro human cell infection model, human induced pluripotent stem (iPS) cell model, human organoid and explant model, primate animal model and human clinical trials. The importance of appropriate testing model is exemplified by the recent reports on the therapeutic evaluation of chloroquine and hydroxychloroquine against SARS-CoV-2. Although these two compounds exhibit significant antiviral activity in SARS-CoV-2 challenged Vero-E6 cells and has been applied for COVID-19 treatment as a repurposed drug without rigid clinical trials, two recent reports denied their roles using human lung cell and macaque models (Hoffmann et al., 2020; Maisonnasse et al., 2020). Despite the efforts to test drug efficacy using SARS-CoV-2 challenged monkey kidney Vero-E6 cells (Riva et al., 2020), it is necessary to fully evaluate their antiviral function in more advanced human-specific models.

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.

In summary, our study presents a novel 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 repositioning scheme towards 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.
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. Under this criteria, data from 56 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 but insufficient screen data for Coronaviridae virus family, we collected additional 27 individual gene-focused non-screening studies to include as many human-specific Coronaviridae HDGs as possible. The detailed information concerning to these literatures and HDGs was summarized in Table S1. For Flaviridae and Orthomyxoviridae viruses, we only took a subset of HDGs that occurred more than once across within its corresponding family as high confidence HDGs for further analysis. For Coronaviridae viruses, we considered all the HDGs for downstream analysis because of scarce data available. 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 18 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 Flaviviridae and Orthomyxoviridae viruses (69 and 63, respectively) and all the 107 HDGs for Coronaviridae viruses 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) < 0.01 (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), DrugCentra (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 Flaviviridae and Orthomyxoviridae viruses and all the HDGs for Coronaviridae 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 InChI key information of a drug or compound and the amino acid sequence of a protein target from UniProt database. For DeepCPI analysis, the top (4% for Coronaviridae; 5% for Flaviviridae; 3.23% for Orthomyxoviridae) predicted drug-target pairs with a DeepCPI score ≥ 0.99 were selected. For DTINet analysis, different cutoffs were used to maintain similar percentage and number of DTIs for indicated virus families (the top 1.35% DTIs with DTINet score ≥ 0.4 for Coronaviridae; the top 1.8% DTIs with DTINet score ≥ 0.1 for Flaviviridae; the top 1.44% DTIs with DTINet score ≥ 0.2 for Orthomyxoviridae).

**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., 2020c). 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.

**Prioritizing repurposed drug candidates**
The repurposed FDA-approved drugs were prioritized by both known DTI and predicted DTI with high confidence. The high confidence DTI prediction was defined as the one passing through both DeepCPI and DTINet top score cutoff. The top drug candidates that meet this criteria is further ranked by the sum of
both DeepCPI and DTINet scores. Due to lack of known DTIs, the repurposed natural compounds were ranked just according to the sum of DeepCPI score.

**Molecular Docking**

The structures of target protein were downloaded from PDB database ([http://www.rcsb.org](http://www.rcsb.org)). The drug or compound structures were downloaded from TCMSP and PubChem database ([https://pubchem.ncbi.nlm.nih.gov](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.
Acknowledgements
This work was supported by the National Natural Science Foundation of China (31871344), the Fundamental Research Funds for the Central Universities (N2020001, N182005005), the 111 Project (B16009), XingLiao Talents Program (XLYC1807212) and China Postdoctoral Science Foundation (2020T130012ZX).

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

Declaration of Interests
The authors declare no conflict of interest.
Figure Titles and Legends

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. 2352 FDA-approved drugs and 1062 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.

Figure 2. Systematically Cataloguing 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.

Figure 3. Characterization of HDGs for corresponding RNA virus families
(A) Subsets of shared HDG 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) KEGG 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.

Figure 4. Drug repositioning using multiple prediction models
(A) Summary and statistics of repurposed drug candidates for corresponding RNA virus families.
(B) The heatmap showing DTI prediction by DeepCPI or DTINet methods. Each row represents a targetable HDG and each column represents a FDA-approved drug. The top predicted DTI is color-coded according to the color legend.

(C) The venn diagram of FDA-approved drugs repurposed from known and de novo prediction sources for the three indicated virus families.

(D) Molecular docking analysis showing the potential binding pocket of the repurposed drug Fostamatinib and natural compound Lysergol with targeted host factors CHEK2 and CLK1, respectively.
Figure 1.

A Compendium of Host Dependency Genes for RNA Viruses

- Genome-wide screen data
  - CRISPR screen (18)
  - RNAi screen (34)
  - Gene-trap screen (4)
- Literature mining (Coronaviridae)
  - Host genes from 27 individual studies

B Virus - Host Gene Association Analysis

C Host Gene Enrichment Analysis

D Drug Repositioning Strategy

- Host factors
- Drug-Target Interaction
- Drugs
  - FDA-approved drugs
  - Natural compounds

Virus Family
- Arenaviridae
- Bunyaviridae
- Coronaviridae
- Filoviridae
- Flaviviridae
- Orthomyxoviridae
- Picomaviridae
- Retroviridae
- Rhabdoviridae
- Togaviridae

Drug Candidates
- Known
  - BindingDB
  - DGIdb3.0
  - DrugCentral
  - Stitch
- Drug ranking
- Literature Support
- Molecular Docking

Prediction
- DeepCPI
- DTINet
Figure 2.

A

B

C

| Family                | CRISPR | RNAi | Gene-trap | Individual hit | Total unique |
|-----------------------|--------|------|-----------|----------------|--------------|
| Arenaviridae          | 58     | 27   | -         | -              | 85           |
| Bunyaviridae          | 8      | 33   | 7         | -              | 48           |
| Coronaviridae         | -      | 90   | -         | 17             | 107          |
| Filoviridae           | 41     | 158  | 16        | -              | 211          |
| Flaviviridae          | 250    | 563  | -         | -              | 610          |
| Orthomyxoviridae      | 169    | 457  | 2         | -              | 604          |
| Picoxviridae          | 48     | 342  | -         | -              | 385          |
| Retroviridae          | 100    | 309  | -         | -              | 405          |
| Rhabdoviridae         | -      | 124  | -         | -              | 124          |
| Togaviridae           | -      | 57   | -         | -              | 57           |
| Total unique          | 503    | 1823 | 50        | 17             | 2316         |
Figure 3.

A

Frequency of gene occurrence

Coronaviridae

TMPRSS2
ACE2
PPA1
CTSL
DPP4
FURIN

Flaviviridae

UBC2J1
DDNL2
SLC16A1
WDR7

Orthomyxoviridae

SANS
PARK2
COPA

B

KEGG Enrichment Pathways

Coronaviridae

Count

MAPK signaling pathway
Vaccinia infection
Human immunodeficiency virus 1 infection
PD-L1 expression and PD-1 checkpoint pathway in cancer
Autoimmune disease
Embryonic development

Flaviviridae

Count

Protein processing in endoplasmic reticulum
Human papillomavirus infection
Influenza A
NOD-like receptor signaling pathway
Viral cholesterol infection
Programmed cell death
JAK-STAT signaling pathway
Synaptic vesicle cycle
Protein export

Orthomyxoviridae

Count

Human papillomavirus infection
Viral cholesterol infection
Rheumatoid arthritis
Caspase activation
Programmed cell death
Tuberculosis
Collecting duct acid secretion
Lysosome
RNA transport

C

Coronaviridae

Cell cycle

Activin receptor

MAPK signaling

Splicosomal complex

Caspase

Cdkk

Druggable host protein
Host protein
Figure 4.

A

Statistics of repurposed drug candidates for corresponding RNA virus families

| Virus Family  | FDA-approved drugs (2352) | Natural compounds (1062) |
|---------------|---------------------------|--------------------------|
|               | Predicted | Known* | Predicted | Known* |
| Coronaviridae | 180       | 130    | 172       | 5      |
| Flaviviridae  | 206       | 30     | 289       | 0      |
| Orthomyxoviridae | 88       | 80     | 62        | 0      |

*Interaction Databases: BindingDB, DGIdb3.0, DrugCentral, Stich

B

Drug - Target Interaction Prediction

Approved Drugs - Coronaviridae

C

FDA-approved drugs

Coronaviridae

Flaviviridae

Orthomyxoviridae

D

Fostamatinib - CHEK2 (PDB ID: 2CN8)

Lysergol - CLK1 (PDB ID: 5X8I)
Table 1. The top ten repurposed FDA-approved drugs against Coronaviridae viruses

| Drug candidate | Structure | Approved Indication | PubChem CID | Predicted Host Target | Sum of DeepCPI Score | Sum of DTNet Score | Known Interaction |
|---------------|-----------|---------------------|-------------|-----------------------|---------------------|------------------|------------------|
| Fostamatinib  | Chronic immune thrombocytopenia | 11671467 | ACE2, AURK2, CAMK1D, CD4, CDEK2, CHEK2, CLK1, CLG3, CLK4, CSNK2A1, DAPK3, DPP4, DYRK1A, EPHA4, FLTR1, FAK, JAK3, MAP2K3, MAP3K10, MAP3K9, MAPK9, MKK2, MYC, NEK4, NUK1, SIK3, STK3 | 26.94 | 33.14 | SYK |
| Bosutinib     | Chronic Myelogenous Leukemia | 5328940 | CD4, CD63, CHEK2, DOK, DDX5, DPP4, EPHA2, ERBB3, FLTR4, JAK1, JAK2, JAK3, MAP2K1, MAPK9, MST1R, PLAG2A, SIRT1 | 15.00 | 10.22 | SRC, ABL1 |
| Palbociclib   | Breast cancer | 5330285 | ABCC10, ALK, CD4, CDS5, DOK, DPP4, FGFR2, FLTR4, JAK1, JAK2, JAK3, LYN, MAP2K1, PDGFR, PLAG2A | 14.95 | 9.62 | CDK4, CDK6 |
| Crizotinib    | Lung cancer | 1162550 | CD4, CHEK2, DOK, DDX5, DPP4, ERBB3, FLTR4, JAK3, MAP2K1, MAPK9, MST1R, PLAG2A | 12.60 | 8.75 | ALK, ROS1 |
| Temsirolimus  | Kidney cancer | 6918289 | CD4, CHEK2, DOK, DDX5, DPP4, ERBB3, FLTR4, JAK3, MAP2K1, MAPK9, MST1R, PLAG2A | 11.99 | 8.45 | FKB12 |
| Vandetanib    | Thyroid Cancer | 3091361 | CD4, CHEK2, DOK, DPP4, ERBB3, FLTR4, JAK3, MAP2K1, MAPK9, MST1R, PLAG2A, TLR4 | 11.98 | 7.55 | VEGFR2-3, EGFR, RET |
| Everolimus    | Breast cancer | 6442177 | CD4, CHEK2, DOK, DDX5, DPP4, FLTR4, JAK3, MAP2K1, MAPK9, MST1R, PLAG2A | 10.99 | 7.70 | FKB12 |
| Bafacetinib   | Rheumatoid arthritis | 44205240 | CD4, CHEK2, DPP4, EPHA2, FLTR4, JAK3, MAPK9, MST1R, PTGS2 | 8.99 | 8.83 | JAK1, JAK2, JAK3, TYK2 |
| Lenvatinib    | Kidney cancer | 9823820 | CD4, CHEK2, DOK, DPP4, ERBB3, FLTR4, MAP2K1, MAPK9, MST1R, PLAG2A | 10.00 | 7.56 | VEGFR2-3, PDGFR, KIT |
| Tolacetinib   | Rheumatoid arthritis | 9920791 | ACE2, CD4, DOK, DPP4, EPHA3, FLTR4, JAK3, MAPK9 | 8.90 | 8.98 | JAK3, JAK2, JAK1 |
Table 2. The top ten repurposed natural compounds against *Coronaviridae* viruses

| Drug candidate | Structure | TCMS* MOL ID | PubChem CID | Herb | Predicted Host Target | Sum of DeepCP score |
|----------------|----------|-------------|-------------|------|-----------------------|---------------------|
| Hyoscyamine    | ![Image](image1.png) | MOL001552   | 154417      | Lysil Cortex, Hyoscyami Semen | ACE2, ACVR1B, ALK, CD4, CKS1B, CLK1, CLK3, CLK4, CSNK2A1, DAPK1, DDIT3, DPI4, DYRK1A, EF1A, E2F2AK3, EPAA3, FLJ25006, FLT4, MKRN2, PLA2G4A, THNS1, TLRI, TLR4, TMPRSS11, D | 18.94 |
| Lysergol       | ![Image](image2.png) | MOL005261   | 14067       | Pharbitidis Semen | ACE2, ADK, CD4, CDXL3, CDKN3, CKS1B, CLK1, CLK3, CLK4, DPI4, DYRK1A, MAP2K1, MKRN2, RDN, ROCK1, THNS1, TLRI | 18.92 |
| Solanocapine   | ![Image](image3.png) | MOL007305   | 73419       | Solanum Nigra | CD4, CLK1, CLK3, CLK4, CSNK2A1, DPI4, MKRN2, THNS1, TLRI, TLRI | 9.95 |
| 4,9-Dimethoxy-1-vinyl-benz-carboline | ![Image](image4.png) | MOL012140   | 5316876     | Asari Radix et Rhizoma | ALRKB, CD4, CLK1, CLK3, CLK4, CSNK2A1, MAP2K1, MAP2K7, MAP3K10, MKRN2 | 9.95 |
| Sophocarpine   | ![Image](image5.png) | MOL006066   | 115269      | Cnidopsis Radix | ACE2, CD4, CLK1, CLK3, CSNK2A1, DPI4, TLRI | 7.97 |
| Guggulsterol-VI | ![Image](image6.png) | MOL001000   | 6922406     | Myrrha | CD4, CLK1, CLK3, CLK4, DAPK1, MAP2K1, MKRN2, THNS1 | 7.97 |
| Stelphae- hydroxysoforonine | ![Image](image7.png) | MOL006570   | 50056119    | Sophora Flavescentis Radix | CD4, CLK1, CLK3, CLK4, CSNK2A1, DPI4, MKRN2 | 6.98 |
| Lehmannanine   | ![Image](image8.png) | MOL006566   | 3041752     | Sophora Flavescentis Radix | CD4, CLK1, CLK3, CSNK2A1, DPI4, MAP2K1, MKRN2 | 6.97 |
| o-Dichroine    | ![Image](image9.png) | MOL005190   | 9851693     | Dichroa Radix | ALK, DDIT3, CDXL3, CD4, CLK4, ACVR1, CLK3, CLK1 | 6.96 |
| (+)-16alpha- Hydroxyatrema   | ![Image](image10.png) | MOL005563   | 15385683    | Sophora Flavescentis Radix | CD4, CLK1, CLK3, CLK4, DPI4, MKRN2 | 5.98 |

*TCMS database: Traditional Chinese Medicine Systems Pharmacology Database
Supplementary Figure 1.

Figure S1. HDG collection for different RNA viruses

(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.
Supplementary Figure 2.

Figure S2. Re-analysis of CRISPR screening data for HDGs

(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.
Supplementary Figure 3.

Figure S3. Comparative analysis and characterization of HDGs for indicated RNA virus families

(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) GO term 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.
**Supplementary Figure 4.**

Figure S4. DTI prediction for indicated virus families and docking analysis for indicated drug-target pairs

(A) The heatmap showing DTI prediction by DeepCPI or DTINet methods for repurposed FDA-approved drugs corresponding to *Flaviviridae* and *Orthomyxoviridae* virus families. Each row represents a targetable HDG and each column represents a FDA-approved drug. The top predicted DTI is color-coded according to the color legend.

(B) Molecular docking analysis showing the potential binding pockets between the repurposed drug Fostamatinib and natural compound Lysergol with their corresponding targeted host factors.
Supplementary Tables

Supplementary Table 1. Compendium of host dependency genes for multiple RNA viruses (attached dataset)

Supplementary Table 2. Sequence sources for phylogenetic analysis (attached dataset)

Supplementary Table 3. Re-analysis of CRISPR screening data (attached dataset)

Supplementary Table 4. Functional gene enrichment analysis of host dependency genes (attached dataset)

Supplementary Table 5. List of drug-target interactions and repurposed drug candidates (attached dataset)

Supplementary Table 6. Summary of host dependency genes with repurposed drugs or natural compounds

| Virus Family       | Druggable host dependence gene | High confidence host dependence gene |
|--------------------|--------------------------------|--------------------------------------|
| Coronaviridae      | Predicted 42                   | Known* 67                            |
|                   |                                 |                                      |
| Flaviviridae       | Predicted 9                     | Known* 7                             |
|                   |                                 |                                      |
| Orthomyxoviridae   | Predicted 10                    | Known* 13                            |

* Interaction Database: BindingDB, DGIdb3.0, DrugCentral, Stitch
† All HDGs for Coronaviridae are included.

Supplementary Table 7. The top five repurposed FDA-approved drugs against Flaviviridae viruses

| Drug candidate | Structure | Approved Indication | PubChem CID | Predicted Host gene | Sum of DeepCI score | Sum of DTINet score | Known interaction |
|----------------|-----------|---------------------|-------------|---------------------|---------------------|---------------------|--------------------|
| Bexarotene     | Rheumatoid arthritis | 442052400 | JAK1 TYK2 NUAK2 | 4.99                | 3.04                | JAK1 JAK2           |
| Flaviviridae   | Chronic immune thrombocytopenia | 11671467 | JAK1 TYK2 TGFBR1 NUAK2 | 3.99                | 4.01                | SYK                |
| Tipifarnib     | Rheumatoid arthritis | 9926791 | JAK1 TYK2 CHUK AXL TGFBR1 NUAK2 | 3.99                | 3.47                | JAK1 JAK2 JAK3      |
| Bosutinib      | Chronic Myelogenous Leukemia | 5328940 | JAK1 TGFBR1 NUAK2 | 5.00                | 0.82                | SRC ABL1            |
| Simvastatin    | Hypercholesterolemia | 54454 | JAK1 TYK2 TGFBR1 | 3.99                | 1.00                | HMGCR               |
**Supplementary Table 8.** The top five repurposed natural compounds against *Flaviviridae* viruses

| Drug candidate | Structure | TCMS* MOL ID | PubChem CID | Herb | Predicted Host gene | Sum of DeepCPI score |
|----------------|-----------|--------------|-------------|------|---------------------|---------------------|
| Lysargol       | ![Lysargol](image1.png) | MOL005261    | 14987       | Semen Pharthitis | B3GALT6 CD81 DAD1 EMCA6 EMCA7 FNAR1 KNC96 NQ63 O5TC3 LCL3RB2 SPC51 SPC3 SRR2 SRR3 STT3A STT3B TU3C3 | 16.98 |
| Costacavine    | ![Costacavine](image2.png) | MOL0008145   | 821486      | Ricini Semen    | B3GALT6 CD81 DAD1 EMCA6 EMCA7 KNC96 O5TC3 LCL3RB2 SPC51 SRR1 SRR3 STT3A STT3B TU3C3 | 13.98 |
| Penrizidin     | ![Penrizidin](image3.png) | MOL005257    | 115247      | Semen Pharthitis | B3GALT6 CD81 DAD1 EMCA6 EMCA7 KNC96 NQ63 O5TC3 LCL3RB2 SPC51 SRR1 SRR3 STT3A STT3B TU3C3 | 12.97 |
| Hygoxamrine    | ![Hygoxamrine](image4.png) | MOL001552    | 154417      | Lyci Cortex, Hygoxam Semi | EMCA3 EMCA7 EXTI1 SRR3 SRR4 STT3A STT3B TGBR1 TY92 UBE3J1 | 11.95 |
| Andronase      | ![Andronase](image5.png) | MOL003790    | 94144       | Styrox | B3GALT6 DAD1 EMCA6 EMCA7 KNC96 O5TC3 LCL3RB2 SPC51 SRR3 STT3A STT3B | 10.97 |

*TCMS* database: Traditional Chinese Medicine Systems Pharmacology Database

**Supplementary Table 9.** The top five repurposed FDA-approved drugs against *Orthomyxoviridae* viruses

| Drug candidate | Structure | Approved Indication | PubChem CID | Predicted Host gene | Sum of DeepCPI score | Sum of DTI score | Known Interaction |
|----------------|-----------|---------------------|-------------|---------------------|---------------------|-----------------|------------------|
| Fostamatinib   | ![Fostamatinib](image6.png) | Chronic immune thrombocytopenia | 11671467 | ABC10 TOP2A FGFR2 JAK2 LYN PDGRA | 4.99 | 4.93 | SYK |
| Barcitinib     | ![Barcitinib](image7.png) | Rheumatoid arthritis | 44205240 | ABC10 TOP2A FGFR2 LYN JAK2 PDGRA | 7.00 | 1.96 | JAK1 JAK2 |
| Tofacitinib    | ![Tofacitinib](image8.png) | Rheumatoid arthritis | 9926791 | ABC10 PDGRA FGFR2 JAK2 LYN JAK2 | 5.00 | 1.75 | JAK1 JAK2 JAK3 |
| Etoricoxib     | ![Etoricoxib](image9.png) | Rheumatoid arthritis | 123619 | ABC10 PDGRA FGFR2 JAK2 LYN JAK2 | 4.99 | 1.19 | COX2 |
| Ziprasidone    | ![Ziprasidone](image10.png) | schizophrenia | 60854 | ABC10 PDGRA FGFR2 JAK2 LYN JAK2 | 2.99 | 0.80 | DRD2 HTR2A |

**Supplementary Table 10.** The top five repurposed natural compounds against *Orthomyxoviridae* viruses

| Drug candidate | Structure | TCMS* MOL ID | PubChem CID | Herb | Predicted Host gene | Sum of DeepCPI score |
|----------------|-----------|--------------|-------------|------|---------------------|---------------------|
| Lysargol       | ![Lysargol](image11.png) | MOL005261    | 14987       | Semen Pharthitis | B3GALT6 CD81 CLK3 GSK3A MAPK13 POLD3 RNAE SPC51 SPC3 SRR2 SRR3 STT3A STT3B TMEM199 | 10.96 |
| Costacavine    | ![Costacavine](image12.png) | MOL0008145   | 821486      | Ricini Semen | ATRPAP1 B3GAT1 CD81 CLK3 RNAE SPC51 SPC3 SRR2 SRR3 STT3A STT3B TMEM199 | 7.97 |
| Hygoxamrine    | ![Hygoxamrine](image13.png) | MOL001552    | 154417      | Lyci Cortex, Hygoxam Semi | ATRPAP1 CLK3 CSE16 FGFR2 GABBR1 POLD3 SRR2 SRR3 STT3A STT3B TMEM199 | 7.06 |
| Penrizidin     | ![Penrizidin](image14.png) | MOL005257    | 115247      | Semen Pharthitis | CD81 GSK3A POLD3 RNAE SPC51 SPC3 SRR2 SRR3 STT3A STT3B TMEM199 | 5.06 |
| 8-hydroxy-2-methoxy-1,6-dimethyl-5-ethenyl-10,11-dihydrophenanthroline | ![8-hydroxy-2-methoxy-1,6-dimethyl-5-ethenyl-10,11-dihydrophenanthroline](image15.png) | MOL007917    | N/A | Juncti Medulla | ATRPAP1 CLK3 POLD3 SRR2 | 3.98 |

*TCMS* database: Traditional Chinese Medicine Systems Pharmacology Database
N/A: Not applicable
### Supplementary Table 11. Key parameters of molecular docking analysis

| PubChem CID | Drug or natural compound                      | PDB ID | Target  | Docking score |
|-------------|------------------------------------------------|--------|---------|---------------|
| 11671467    | Fostamatinib                                   | 2CN8   | CHEK2   | -8.770        |
| 11671467    | Fostamatinib                                   | 3E7O   | MAPK9   | -7.766        |
| 11671467    | Fostamatinib                                   | 5LWM   | JAK3    | -6.669        |
| 11671467    | Fostamatinib                                   | 4GK3   | EPHA3   | -6.540        |
| 14987       | Lyaergol                                       | 5X8I   | CLK1    | -9.174        |
| 14987       | Lyaergol                                       | 6Z53   | CLK3    | -8.237        |
| 5316876     | 4,9-Dimethoxy-1-vinyl-beta-carboline           | 5X8I   | CLK1    | -8.141        |
| 154417      | Hyoscyamine                                    | 5X8I   | CLK1    | -7.774        |
| 14987       | Lyaergol                                       | 6FYV   | CLK4    | -7.132        |
| 5316876     | 4,9-Dimethoxy-1-vinyl-beta-carboline           | 6FYV   | CLK4    | -6.939        |
| 5316876     | 4,9-Dimethoxy-1-vinyl-beta-carboline           | 6Z53   | CLK3    | -6.826        |
| 154417      | Hyoscyamine                                    | 6Z53   | CLK3    | -6.514        |
| 154417      | Hyoscyamine                                    | 6FYV   | CLK4    | -6.279        |
| 14987       | Lyaergol                                       | 5T4E   | DPP4    | -6.196        |
| 154417      | Hyoscyamine                                    | 5T4E   | DPP4    | -6.064        |
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.

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.

Badia, R., Angulo, G., Riveira-Munoz, E., Pujantell, M., Puig, T., Ramirez, C., Torres-Torronteras, J., Marti, R., Pauls, E., Clotet, B., et al. (2016). Inhibition of herpes simplex virus type 1 by the CDK6 inhibitor PD-0332991 (palbociclib) through the control of SAMHD1. The Journal of antimicrobial chemotherapy 71, 387-394.

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. Frontiers in immunology 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 research 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 discovery 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. Clinical microbiology reviews 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. European journal of medicinal chemistry 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. Antimicrobial agents and chemotherapy 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 research 44, D1045-1053.
Gonzalez, M.M., Cabrerizo, F.M., Baiker, A., Erra-Balsells, R., Osterman, A., Nitschko, H., and Vizoso-Pinto, M.G. (2018). Beta-carboline derivatives as novel antivirals for herpes simplex virus. International journal of antimicrobial agents 52, 459-468.

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.

Hoffmann, M., Mösbauer, K., Hofmann-Winkler, H., Kaul, A., Kleine-Weber, H., Krüger, N., Gassen, N.C., Müller, M.A., Drosten, C., and Pöhlmann, S. (2020). Chloroquine does not inhibit infection of human lung cells with SARS-CoV-2. Nature.

Jin, Z., Yang, L., Ding, G., Yang, G., Han, Y., Zhang, X., and Li, W. (2017). Sophocarpine against enterovirus 71 in vitro. Experimental and therapeutic medicine 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. Nature structural & molecular biology 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 research 38, D552-556.

Li, C.C., Wang, X.J., and Wang, H.R. (2019). Repurposing host-based therapeutics to control coronavirus and influenza virus. Drug discovery 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. The New England journal of medicine 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 biology 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. Nature communications 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.

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 discovery 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. The Lancet Infectious diseases.

Pierson, T.C., and Diamond, M.S. (2020). The continued threat of emerging flaviviruses. Nature microbiology 5, 796-812.

Riﬁaioglu, 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.

Riva, L., Yuan, S., Yin, X., Martin-Sancho, L., Matsunaga, N., Pache, L., Burgstaller-Muehlbacher, S., De Jesus, P.D., Teriete, P., Hull, M.V., et al. (2020). Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. Nature.

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.

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.

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 molecular medicine, e12697.

Tan, L., Sato, N., Shiraki, A., Yanagita, M., Yoshida, Y., Takemura, Y., and Shiraki, K. (2019). Everolimus delayed and suppressed cytomegalovirus DNA synthesis, spread of the infection, and alleviated cytomegalovirus infection. Antiviral research 162, 30-38.

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.

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 research 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.

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

Wei, J., Alfajaro, M.M., Hanna, R.E., DeWeirdt, P.C., Strine, M.S., Lu-Culligan, W.J., Zhang, S.-M., Graziano, V.R., Schmitz, C.O., Chen, J.S., et al. (2020). Genome-wide CRISPR screen reveals host genes that regulate SARS-CoV-2 infection. bioRxiv.

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 pharmaceutica Sinica B.

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.

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. Current medicinal chemistry 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.

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 discovery 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. The New England journal of medicine 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. Nature reviews Drug discovery 15, 327-347.