Reversal of Infected Host Gene Expression Identifies Repurposed Drug Candidates for COVID-19

Jing Xing[1#], Rama Shankar[1#], Aleksandra Drelich[2#], Shreya Paithankar[1], Eugene Chekalin[1], Thomas Dexheimer[3], Surender Rajasekaran[1,4], Chien-Te Kent Tseng[2,5*] and Bin Chen[1,3*].

[1] Department of Pediatrics and Human Development, Michigan State University, Grand Rapids, Michigan, USA.

[2] Departments of Microbiology and Immunology, University of Texas Medical Branch, Galveston, Texas, USA

[3] Department of Pharmacology and Toxicology, Michigan State University, Grand Rapids, Michigan, USA.

[4] Helen Devos Children Hospital, Grand Rapids, Michigan, USA

[5] Center of Biodefense and Emerging Disease, University of Texas Medical Branch, Galveston, Texas, USA

[#] These authors contributed equally

[*] Correspondence to Bin Chen: chenbi12@msu.edu, Kent Chien-Te, Tseng: sktseng@UTMB.EDU

Abstract

Repurposing existing drugs is a timely option to cope with COVID-19. We predicted therapeutic candidates that could reverse the gene expression of coronavirus-infected host cells. Thirteen expression signatures computed from various experimental conditions and preclinical models could be reversed by those compounds known to be effective against SARS- or MERS-CoV, as well as the drug candidates recently shown to be effective against SARS-CoV-2. We selected ten novel candidates to further evaluate their *in vitro* efficacy against SARS-CoV-2 infection. Four compounds bortezomib, dactolisib, alvocidib and methotrexate inhibited the formation of virus infection-induced cytopathic effect in Vero E6 cells at 1 μM, yet such a concentration seems toxic to the cells as well. While the evaluation in other permissive cells and

the prediction of toxicity are needed to optimize and minimize their antiviral activity and cytotoxicity, respectively, this computational approach has the potential to rapidly and rationally identify drug candidates against COVID-19.

Main

Since early December 2019, the newly emerged SARS-CoV-2 has infected almost 1 million people globally[1]. In the United States, confirmed cases increased from a few dozen to almost 1000 within two weeks, and this number rapidly grew to over 80,000 within one month[2]. Among the patients, 15% suffered from severe acute respiratory distress syndrome (ARDS) and approximately 3% died from acute respiratory failure, acute cardiac injury, secondary infection, and other serious complications[1,3,4]. The World Health Organization (WHO) declared this rapidly spreading and highly pathogenic COVID-19 first a global public health emergency and then a pandemic. As SARS-CoV-2 infection continues to endanger lives, effective therapeutics are urgently needed. Repurposing existing drugs could be an efficient and timely means of identifying drugs that have activity against coronavirus. There are a few repurposed drugs such as lopinavir/ritonavir, baricitinib, remdesivir and chloroquine currently under clinical investigation[5,6]. These drugs are expected to target key steps of viral entry, or specific proteins involved in viral replication, including viral proteases[7]. In addition to viral replication, the viral pathogen associated molecular pattern (PAMP) (e.g., immune dysfunction and endoplasmic reticulum stress, Figure 1A) could be targeted to improve the clinical outcome [8]. PAMPs-mediated signaling pathways are attractive drug targets to alleviate diseases caused by human pathogens. Therefore, effectively targeting these pathways to stop the progression to ARDS caused by SARS-CoV-2 might save lives. Independent of SARS-CoV-2 infection, in aging adult populations, ARDS is associated with mortality rates of 30-50%[9]. Thus, a methodical and unbiased search for new drug candidates from a large drug library could uncover agents that have potential to arrest the infection and ameliorate its effect. To accomplish this, we sought to target infection-induced genes in the host cells, hoping to mitigate disease progression and alleviate symptoms.

We and others[10–16] have utilized a systems-based approach that employs gene expression profiles of disease samples and drug-induced gene expression profiles from cell lines to discover new therapeutic candidates for diseases. The essential idea is to identify drugs that reverse the disease gene expression signature by suppressing the over-expressed disease genes and activating the repressed genes (Figure 1A). A disease signature is defined as a list of differentially expressed genes between disease samples and normal control samples. We recently found that the reversal of gene expression (namely sRGES) correlates to drug efficacy in cancers[17], demonstrating the feasibility of applying this approach to predict drug candidates for other diseases, including viral infection.

To utilize this approach for drug discovery against SARS-CoV-2, we first need to collect virus-related host gene expression profiles, which were not available at the time of writing. Given the high genomic similarity between SARS-CoV, MERS-CoV, and SARS-CoV-2, we reasoned that existing host gene expression profiles of the samples infected by SARS- or MERS-CoV could approximate to those infected by SARS-CoV-2. To verify this assumption, we compiled 331 virus-induced signatures from enrichR and GEO (Table S1) and used an established pipeline to score 1740 drugs in our repurposing library regarding their reversal of signature gene expression. Clustering of these signatures based on their drug prediction scores suggests that signatures derived from the same virus or the virus family under the similar experimental model tend to cluster together (Figure S1). An example cluster includes one signature derived from primary human microvascular endothelial cells (MMVE001) after 48h of MERS-CoV infection (study id: GSE79218) and another derived from melanoma cells in mice after seven days of SARS-CoV infection (study id: GSE68820). In addition, Spearman correlation coefficient of the *in vitro* drug efficacy data (EC_{50}: Half maximal effective concentration) of SARS-CoV and MERS-CoV is up to 0.6 (Figure 1B). The clustering and correlation results suggested that drugs predicted based on the signatures related to SARS-CoV and MERS-CoV could also be applied for SARS-CoV-2. Therefore, we developed a pipeline to repurpose existing drugs against MERS-CoV and SARS-CoV, and then experimentally evaluate these drugs in SARS-CoV-2 (Figure 1C).

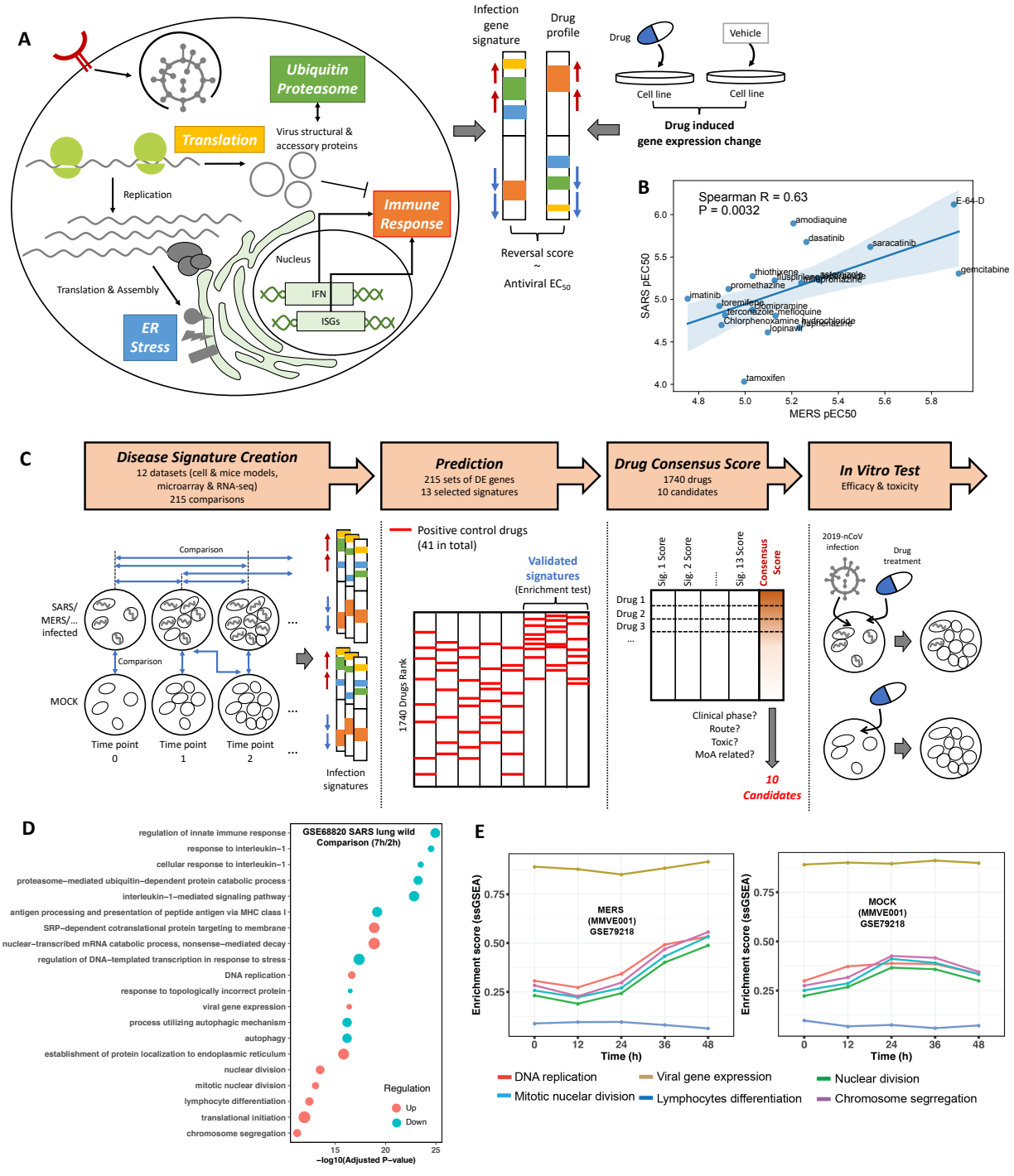

Figure 1. Study workflow and biological processes dysregulated by viral infection. **A**, An illustration of reversing the expression of host genes comprising multiple biological processes (highlighted with colors) induced by coronavirus infection. Drug-induced gene expression profiles are taken from the LINCS

database. A good candidate should activate the repressed biological processes and inhibit the upregulated processes. **B**, Correlation of the published antiviral activities of 30 drugs (pEC_{50}, $-\log_{10}$ transformed EC_{50} value in mol/L) against MERS- and SARS-CoV. **C**, Study workflow including creation of disease signatures, prediction of drug candidates, selection of a final drug list, and in vitro validation. One disease signature composed by the differentially expressed genes of each comparison led to one drug prediction list. Only the signature resulting in a prediction list where known positive drugs were enriched on the top was considered as a valid signature. **D**, Dysregulated pathways after SARS infection at 7h compared with 2h in lungs. **E**, The enrichment of top six dysregulated pathways in primary human microvascular endothelial (MMVE001) cells through 0h to 48h in MERS-CoV infection (left) and in mock (right). Only one study was selected for D and E, respectively, and the dysregulated pathways and their dynamics for other studies are available in supplementary materials (Figure S2 and S3, Extended Data 1).

In total, 430 samples infected by either MERS- or SARS-CoV under different models (e.g., cell line, mouse models) across multiple time points from public repositories were used for the identification of disease signatures (Table S1, 12 studies in total). Their expression profiles were created using either microarray or RNA-Sequencing. Depending on the profiling platform, data processing and signature creation methods varied (see Methods). The previous clusters are highly confounded by post-infection time points (Figure S1), meaning the disease signature and their predicted drugs are strikingly different under different time points. Therefore, we enumerated all the possible comparisons (Figure 1C), including (1) comparisons between infection and mock at each time point, (2) comparisons between different time points within the infection or the mock group (e.g., time point 1 vs. time point 0, time point 2 vs. time point 1), and (3) comparisons both between time points and between infection vs. mock. These comparisons revealed different virus-related biological processes and their dynamic regulation. For instance, analysis of SARS-CoV infected lung tissue data showed that various biological processes, including viral gene expression, DNA replication, nuclear division, lymphocytes differentiation and translation-related processes, were activated (Figure 1D, S2 and Extended Data 1). In contrast, interleukin and autophagy-related processes were repressed in infected samples (Figure 1D, S2 and Extended Data 1). Interestingly, some processes were only activated after a certain time point (e.g., 24 h in Fig 1E) in infected samples, while such patterns were not observed in mock samples (Figure 1E and S3), suggesting that only some of the comparisons captured the biology of critical viral infection processes.

For each comparison, we computed a disease signature to characterize the infection status, followed by the prediction of which drugs may have activity. As we could not directly evaluate the quality and pathologic relevance of each disease signature, we turned to validating their action using those drugs that are positive *in vitro* MERS/SARS testing (41 positive drugs in total, 30 with EC_{50} values, Table S2). Among 215 MERS-CoV or SARS-CoV infection signatures, 13 signatures led to the drug list, where positive drugs are highly enriched on the top (Extended Data 2). Moreover, EC_{50} of these drugs significantly correlated with sRGES (Figure 2A and Figure S4). In contrast, for the H1N1 infection signatures, we didn't observe any significant enrichment of anti-coronavirus positive drugs (Extended Data 2). The recent clinical data suggested that the combination of two drugs lopinavir/ritonavir present clinical benefits to patients with coronavirus[18,19]. Both drugs were profiled in LINCS[20] individually; however, none of them could induce substantial gene expression change (absolute z score < 2) under 10 μM concentration in cancer cell lines; therefore, neither was predicted as a hit. We next found another dataset where HepaRG cells were treated by ritonavir under 10 different concentrations (from 9 nM to 300 μM). We observed that ritonavir could reverse nearly all the 13 disease signatures under high concentration (i.e., 10 μM to 200 μM) (Figure S5). Together, we reasoned that those highly ranked drugs with greater negative sRGES in these valid comparisons could be new therapeutic candidates.

To obtain the final drug prediction, as it's not trivial to merge disease signatures derived from different platforms and biological conditions, instead of merging them, we calculated the consensus score of all the drug prediction lists derived from individual valid disease signatures,. We took the median of the ranks across multiple comparisons for each drug, and ranked them based on their median rank (Extended Data 3). Mechanism of action (MoA) enrichment analysis of the final ranked drug list revealed a few significant drug classes including CDK inhibitors, mTOR inhibitors and NF-κB pathway inhibitors (Figure 2B). Lately, Jeon et al.[21] tested the antiviral efficacy of 35 FDA-approved drugs against SARS-CoV-2, among which 14 positive drugs overlapped with our screening library. These drugs were also significantly enriched at the

top of our prediction (Figure 2C, p = 1.39E-4), further validating our predictions. Interestingly, three anti-parasite drugs pyrvinium, ivermectin, niclosamide were ranked among the top 30. Both ivermectin, and niclosamide could inhibit the replication of SARS-CoV-2 in vitro[21,22] and pyrvinium and niclosamide were effective in MERS and SARS[23,24].

We manually inspected the top candidates and selected ten representative candidates according to their reversal scores (Figure 2D), clinical applicability, MoA, administration routes and safety (Table S3). Finally, we evaluated their cytotoxicity and ability of the prevention of cytopathic effect (CPE) in the Vero E6 cell line (Table 1). Instead of measuring EC_{50}, we used more stringent criteria, which required a 4-day complete prevention of CPE observed under the microcopy. Six of the proposed drugs showed less than 10 μM minimum concentration that could completely prevent CPE. However, five of them presented unfavorable cytotoxicity at their effective concentration. In the following confirmatory experiment, we examined the efficacy and toxicity for the four most effective drugs (CPE preventing concentration <= 1 μM), bortezomib, methotrexate, nvp-bez235 and alvocidib as well as chloroquine, a drug being used as the first-line treatment for SARS-CoV-2. None of the predicted drugs could completely prevent CPE, though chloroquine could at the dose of 15 μM.

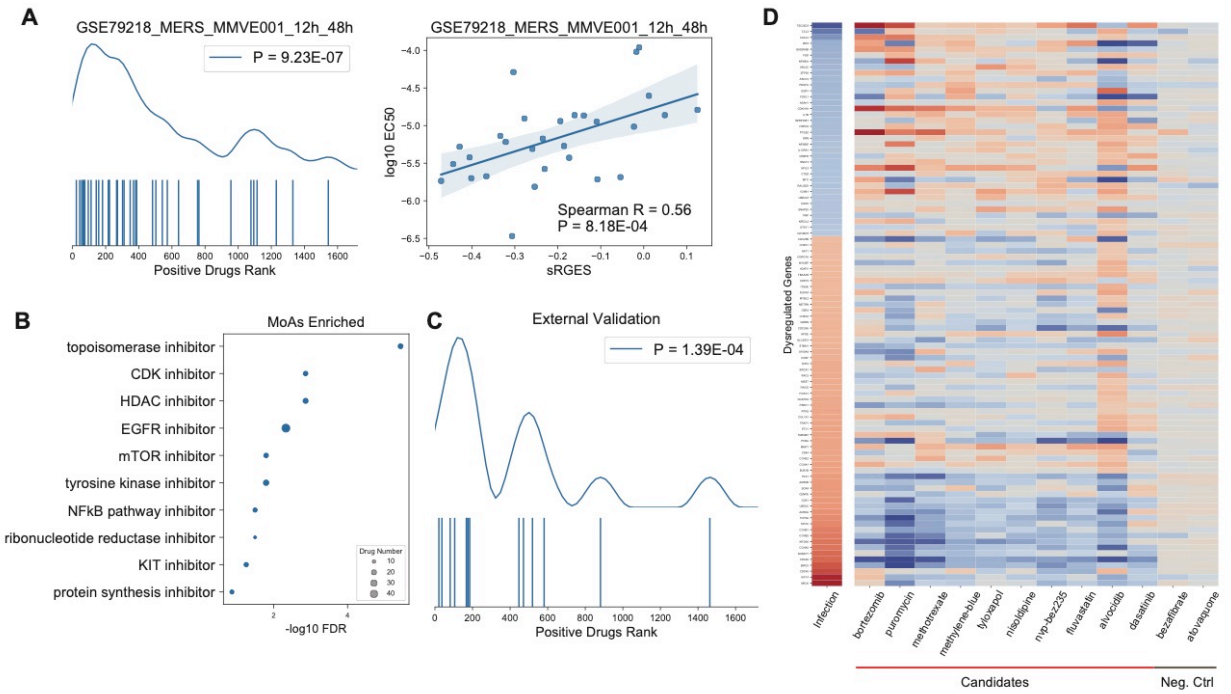

Figure 2. Meta drug analysis and candidate evaluation. A, an example of a validated signature derived from differentially expressed genes between 48h and 12h after MERS-CoV infection in the MMVE001 cell line (GSE79218 dataset). The left panel shows the enrichment of positive drugs (p value was computed by Wilcoxon rank sums test, see Methods). The curve shows enrichment density, and each bar under the curve represents the rank of a positive drug on the prediction list. The right panel shows the correlation between sRGES and EC_{50} (in mol/L, log_{10} transformed) of the positive controls. Each point indicates a positive drug. **B**, enriched MoA (blue dots) and their FDR based on meta drug ranking. Dot size corresponds to the number of drugs associated with the MoA. **C**, external validation using published anti-COVID-19 efficacy of FDA-approved drugs[21]. **D**, heatmap of the summarized infection signature and the summarized reversal effects of 10 selected candidates and two negative examples (see Methods). Red color indicates up-regulated genes and blue indicates down-regulated genes.

Table 1. Efficacy and toxicity of selected drugs in SARS-CoV-2

	MoA	Primary Indication	CPE Preventing[*1] (μM)	Toxic[*2] (μM)	Repeat[*3] CPE Preventing (μM)	Repeat Toxic (μM)
Bortezomib	Proteasome inhibitor	Multiple myeloma	0.05	0.002	> 30	> 30
Puromycin	Protein synthesis inhibitor	Antibiotic	20	0.002	-[*4]	-
Methotrexate	Dihydrofolate reductase inhibitor	Rheumatoid arthritis	0.78	> 25	> 30	30

	MoA	Primary Indication	CPE Preventing[*1] (μM)	Toxic[*2] (μM)	Repeat[*3] CPE Preventing (μM)	Repeat Toxic (μM)
Methylene-blue	Guanylyl cyclase inhibitor	Methemoglobinemia	20	5	-	-
Tyloxapol	NF-kB pathway inhibitor	Bronchopulmonary secretions with mucus and pus	> 100	20	-	-
Nisoldipine	Calcium channel blocker	Hypertension	6.25	0.2	-	-
Nvp-bez235	PI3K/MTOR inhibitor	Advanced Solid Malignancies	1	1	> 30	7.5
Fluvastatin	HMGCR inhibitor	Hypercholesterolemia	3.125	0.2	-	-
Alvocidib	CDK inhibitor	Acute myeloid leukemia	0.05	0.002	> 30	3.75
Dasatinib	SRC/ABL kinase inhibitor	Leukemia	3.125	1.6	-	-
Chloroquine[*5]	Endosomal acidification inhibitor	Malaria	-	-	15	> 30

[*1], the lowest concentration of a drug to prevent CPE.
[*2], the lowest concentration of a drug showing cytotoxicity.
[*3], when repeated, drug treatment time was shorter (2 hours) than the previous test (3 hours).
[*4], not tested.
[*5], the positive control of this assay.

Discussion

In this work, we investigated host gene expression change after coronavirus infection to characterize the infection signatures for drug discovery against SARS-CoV-2. Thirteen out of 215 disease signatures led to the prediction where published SARS or MERS drug candidates could be recovered. Those uninformative signatures might be due to variation of models or experimental conditions. For example, samples using virus strains (dORF6, BatSRBD) in GSE4796 might be different from the wild SARS type. Samples using ferret as the model for immune response investigation in study GSE22581 might require additional

deconvolution of profiles to dissect the response in host cells. In study GSE79218, as viral infection pathways were activated in human microvascular endothelial cells only after 24 h post-infection of MERS (Figure 1E), the signature from the comparison between 24 h and 0h could not capture the biology of viral infection, thus failed to recover known drugs. Those findings also reiterate the challenge of choosing appropriate experimental models and the right time points for drug treatment in COVID-19[25].

We then developed a consensus score for each drug based on their potency to reverse these thirteen signatures. The prediction was externally validated by a recent drug screening against SARS-CoV-2[21]. The validation justified the approach of reversal of host cell gene expression derived from SARS-CoV or MERS-CoV infection. Ten drugs were proposed and their antiviral efficacy and cytotoxicity in the Vero E6 cell line were further examined. We observed that four drugs prevented CPE under 1 μM, but also were toxic at this concentration. We initially only excluded those highly toxic compounds such as topoisomerase inhibitors, but didn't attempt to evaluate their cytotoxicity since all those candidates are FDA approved drugs with established safety profiles.

Genes related to mitotic G2/M cell cycle such as *MELK*, *KIF14*, *BIRC5* were enriched (Figure 2D), thus reversal of the expression of these genes might induce apoptosis and cytotoxicity. On the other side, several immune signaling related genes such as *CCL2*, *CXCL2* and *NFKBIA* were down-regulated by the virus, thus reversal of the expression of these genes may augment the host response to fight viral infection. Our initial efficacy study neither checked their ability of inhibiting viral replications nor used the model that has necessary immune environment (e.g., Vero cells are interferon-deficient). For example, the candidate methotrexate, a chemotherapeutic at high doses and, an immune suppressor at lower doses, being used to treat rheumatoid arthritis (RA), showed considerable *in vitro* antiviral effect. Methotrexate could down regulate genes related to G2/M cell cycle, and up regulate genes related to cytokine signaling. This is not surprising as studies have shown some of the interferon related pathways that govern inflammation overlap with cancer[26]. Recently, another RA drug, tocilizumab, an IL-6 antibody, was applied to treat severe patients

infected by SARS-CoV-2 to protect them from life-threatening lung damage caused by cytokine storm. Further analysis of drug-induced *in vivo* profiles revealed that methotrexate and tocilizumab could stimulate several interferon-stimulated genes for virus blocking (Figure S6), but chloroquine could stimulate less, suggesting a different MoA (e.g. pH dependent control of viral entry). We are interested in testing methotrexate on other interferon sufficient cell lines in the future.

Notably, all four effective drugs (proteasome inhibitor bortezomib, PI3K-mTOR inhibitor nvp-bez235 (dactolisib), CDK inhibitor alvocidib and antimetabolite methotrexate) are being used for cancer treatment. In several independent screening efforts, similar drugs (AXL receptor tyrosine kinase inhibitor Gilteritinib, CDK inhibitor Abemaciclib) are identified[21,27]. Although those anti-cancer drugs might reverse the host expression and prevent CPE, because they are often toxic or have severe side effects, they are not appropriate to treat patients weakened by SARS-CoV-2.

For this work, in addition to the known limitations of the LINCS datasets (e.g., a limited coverage of the transcriptome), another limitation lies in the use of drug-induced gene expression profiles derived from cancer cells that are different from those infected cells. This partially explains why anti-viral drugs were not predicted as hits. The limited resources and the immeasurable damage of this pandemic call for the urgent needs of studying existing drugs in infectious disease models.

In conclusion, thanks to the open science initiatives, we were able to leverage the open resources to rationally predict drug candidates that might reverse coronavirus-induced transcriptomic change, and validated *in silico* and *in vitro*. The unbiased search of drug candidates based on reversal of gene expression could offer an effective and rapid means to propose candidates for further experimental testing, yet more layers of information such as toxicity, validation experimental setting and clinical applicability could be incorporated to find better therapeutics. The prediction list would be of some value to those labs who have the capability to investigate a few compounds, especially during this urgent time.

Methods

Disease signatures

We obtained a total of 430 samples for "SARS" or "MERS" related data from ArrayExpress, Gene Expression Omnibus (GEO) and Sequence Read Archive (SRA). The meta information of each sample was manually annotated, including virus strain, model, organism, and time point. The expression matrix for each microarray data was downloaded via the GEOquery R package. The matrix was further filtered by removing the probes with expression only in half of the samples. Expression values were normalized using quantile normalization and $\log_2$ transformation was applied for each matrix. The probe values were collapsed based on Entrez Gene ID. The Significance Analysis of Microarrays (SAM) method was used to compute differentially expressed (DE) genes with criteria fold change > 1 and false discovery rate (FDR) < 0.05). Gene symbols of other organisms were converted to HUGO gene symbols. For RNA-Seq datasets, raw sequence data were downloaded from SRA and processed with the TOIL pipeline[28,29]. EdgeR was used to compute DE genes using the same criteria as used for microarray data. Gene ontology enrichment analysis of DE genes for each comparison was performed using the clusterprofiler R package. Further, gene set enrichment (ssGSEA) for each biological process was performed using ssGSEA method in the GSVA R package. For the infection group, we enumerated all the comparisons across all time points, and the corresponding comparison was performed in the mock group as well. The DE genes that were uniquely present in the infection group were selected for further analysis. We also compared DE genes between infection and mock at each time point, together with consistently dysregulated genes from time point 0 to the end.

Drug signatures

Drug gene expression profiles have been widely used in our previous studies. Briefly, a full matrix comprising 476,251 signatures and 22,268 genes including 978 landmark genes was downloaded from the

LINCS website[30] as of September 2013. The meta-information of the signatures (for example, cell line, treatment duration, treatment concentration) was retrieved via LINCS Application Program Interfaces. The matrix and metadata are now available via GSE92742 in GEO. The signature derived from the comparison of expressions between the samples treated with the perturbagen of interest and vehicle control, represents gene expression change upon treatment. We further downloaded the LINCS drug information from the Drug Repurposing Hub. Only small-molecule with high-quality gene expression profiles (is_gold=1, annotated in the meta-information) and listed in the drug repurposing hub were further analyzed.

Reversal correlation

The computation of RGES and the summarization RGES were detailed elsewhere and recently implemented as a standalone R package[31]. Briefly, we quantified the reversal of disease gene expression as RGES (Reversal Gene Expression Score), a measure modified from the connectivity score developed in other studies[20]. To compute RGES scores, we first rank genes based on their expression values in each drug signature. An enrichment score/s for each set of up- and down-regulated disease genes were computed separately using a Kolmogorov–Smirnov-like statistic, followed by the merge of scores from both sides (up/down). The score is based on the amount to which the genes (up or down-regulated) at either the top or bottom of a drug-gene list ranked by expression change after drug treatment. One compound may have multiple available expression profiles because they have been tested in various cell lines, drug concentrations, treatment durations, or even different replicates, resulting in multiple RGES for one drug-disease prediction. We termed this score summarized RGES (sRGES). We set a reference condition (i.e., concentration of 10 μM, treatment duration of 24 hours) and used a model to estimate a new RGES if the drug profile under the reference condition was not available. We did not weight the LINCS cell lines. Those comparisons, where maximum of the absolute sRGES is less than 0.25 were considered as insignificant predictions.

Signatures validation and selection

Drugs with known *in vitro* activity against two coronaviruses (i.e., SARS-CoV and MERS-CoV) served as positive controls for signature validation. Qualified signatures should meet the following criteria: (1) derived from SARS-CoV or MERS-CoV infection experiments; (2) the number of differentially expressed genes was greater than 50 (mapped to LINCS); (3) the maximum absolute sRGES prediction was greater than 0.25; (4) the sRGES of positive drugs was enriched at the top (one side Wilcoxon rank-sum test $p < 0.05$, FDR < 0.25); (5) the sRGES and the average EC_{50} value of positive drugs were highly correlated (Spearman $r >= 0.4$, $p < 0.05$).

Clustering of virus predictions

We downloaded the compiled virus-perturbed signatures from EnrichR (323 in total). Since the EnrichR dataset did not include any MERS signatures, we manually added the signatures of two MERS datasets (GSE79218, GSE79172, 8 in total) computed from the comparison between the infection group and the mock group, separately. Only the virus signature containing more than 50 LINCS landmark genes was selected. Each virus signature was queried against the LINCS library using the established pipeline. Only those signatures where the maximum of the absolute sRGES was greater than 0.25 were chosen for the following analysis. Viruses were clustered based on the sRGES scores using pvclust[32] (distance method: spearman correlation, nboot = 100).

Ritonavir correlation analysis

We found one RNA-Seq dataset (SRA: SRX4939022) for HepaRG cells treated with multiple compounds including ritonavir under multiple concentrations (ranging from 9 nM to 300 μM). We processed 90 profiles in the plate (2D_RG_PLATE2) consisting of 10 concentrations (each concentration has 9 profiles). The $\log_2$ TPM of each profile was subtracted by the median of $\log_2$ TPM of all DMSO-treated samples in this plate, resulting in one drug-induced gene expression profile. The spearman correlation between the drug-induced gene expression and the disease gene expression used for the LINCS prediction was computed. The negative correlation means a reversal relation. Ritonavir was originally developed as an inhibitor of

HIV protease and now often used at a low dose with other protease inhibitors to boost the antiviral effect. Our analysis suggests that ritonavir could only reverse the disease signature at a very high dose.

Reversal infection signature visualization

For better visualization of the selected drug candidates' reversal of coronavirus induced gene expression changes, we combined validated infection signatures into one meta signature, and summarized drug profiles from different experiments into one profile. Dysregulated genes were included into the meta infection signature if 25% quantile of $\log_2$ fold changes was less than -1 or 75% quantile was greater than 2. For each drug, all profiles (z-scores, level 5) in L1000 measured at 10 μM were extracted (including different cell lines and treatment times). The value of each gene in the summarized profile was defined as the median of the head or tail 25% (depending on which absolute value is larger) if this quantile absolute value was greater than 1, we defined it as the median of the z-scores of this gene across all the profiles extracted. The matrix was composed of the meta signature genes and signatures of selected drug candidates. We also included the profiles of two drugs predicted as negative hits, and ordered the rows by the fold change of infection signature genes. A heatmap was used to visualize the effect of selected drugs in reversing virus-induced genes.

When visualizing the transcriptome datasets (for Figure S6) from GEO (https://www.ncbi.nlm.nih.gov/geo/), we processed the gene expression matrix the same as the disease signature creation aforementioned. Then all gene log2 fold change values were converted to ranking percentages. Finally, a clustermap was computed for genes of our interest, colored in red (up-regulation) and blue (down-regulation).

Gene Ontology (GO) Enrichment Analysis

The processed compound transcriptome profile was categorized into up- and down-regulated genes, with a threshold of log2 fold change greater than 1 or less than -1, respectively. Then each group of genes was submitted to Enrichr[33] (https://amp.pharm.mssm.edu/Enrichr/) to compute the Gene Ontology enrichment. GO terms with p-value less than 0.05 and adjusted p-value less than 0.2 were taken as significant.

Cell culture, virus infection and drug evaluation

Tissue Cultures and Virus

Vero E6 cells [CRL:1586, ATCC] were grown in Eagle's minimal essential medium (EMEM) supplemented with penicillin (100 units/ml), streptomycin (100 µg/ml), and 10% fetal bovine serum (FBS). SARS-CoV-2 (US_WA-1 isolate), the 3rd passage in Vero E6 cells from the original CDC (Atlanta) material and sequence confirmed, was used throughout the study. The titer of the viral stock was 7.5×10^7 50% tissue culture infectious doses ($TCID_{50}$)/ml. All experiments involving using infectious virus were conducted at the University of Texas Medical Branch in an approved biosafety level 3 laboratory.

Drug screening by cell-based assay

A slightly modified Vero E6-based standard micro-neutralization assay was used to rapidly evaluate the drug efficacy against SARS-CoV-2 infection. Briefly, confluent Vero E6 cells grown in 96-wells microtiter plates were pre-treated with serially 2-folds diluted individual drugs for two hours before infection with 100 infectious SARS-CoV-2 particles in 100 µl EMEM supplemented with 2% FBS. Vero E6 cells treated with parallelly diluted dimethyl sulfoxide (DMSO) with or without virus were included as positive and negative controls, respectively. After cultivation at 37 ℃ for 4 days, individual wells were observed under the microcopy for the status of virus-induced formation of CPE. The efficacy of individual drugs was calculated and expressed as the lowest concentration capable of completely preventing virus-induced CPE in 100% of the wells. The toxicity to the treated cells was assessed by observing floating cells and altered morphology of adhered Vero E6 cells in wells under the microcopy. All compounds were ordered from Selleckchem or Cayman Chemical. All compounds were dissolved in 100% DMSO as 10 mM stock solutions and diluted in culture media.

Software tools and statistical methods

All the analysis was conducted in R (v3.5.1) or Python (v3.7) programming language. The ggplot2, pheatmap and seaborn packages were used for data visualization. Student's t-test was performed for normally distributed data and Wilcoxon rank-sum test was used for other types of data to compute the p-value.

Data and code availability

Authors declare that all data used in this study are available within the article and its supplementary information files. Any other specific files can be provided from the corresponding author upon reasonable request. The code is available at GitHub (https://github.com/Bin-Chen-Lab/wars).

Contributions

B.C. conceived the study. J.X., R.S. and B.C. performed computational analyses with the input from S.P. and E.C.. A.D. performed biological experiments supervised by C.T.K.T.. S.R. provided clinical insights, and T.D. prepared reagents. J.X., R.S., B.C., A.D., C.T.K.T wrote the manuscript with the input from all coauthors. B.C. supervised the project.

Acknowledgements

The research is supported by R01GM134307 and K01 ES028047 and the MSU Global Impact Initiative. The content is solely the responsibility of the authors and does not necessarily represent the official views of sponsors. The authors would like to thank all researchers who shared their data publicly that made this project possible.

References

1. COVID-19 Report from WHO. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/.
2. CDC Cases and Updates. https://www.cdc.gov/coronavirus/2019-ncov/cases-in-us.html.
3. COVID-19 Report from National Health Commission of China. http://www.nhc.gov.cn/xcs/xxgzbd/gzbd_index.shtml.
4. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* **395**, 497–506 (2020).
5. Peeri, N. C. *et al.* The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? *International Journal of Epidemiology* (2020) doi:10.1093/ije/dyaa033.
6. Wang, M. *et al.* Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Research* (2020) doi:10.1038/s41422-020-0282-0.
7. Zumla, A., Chan, J. F. W., Azhar, E. I., Hui, D. S. C. & Yuen, K.-Y. Coronaviruses — drug discovery and therapeutic options. *Nature Reviews Drug Discovery* **15**, 327–347 (2016).
8. de Wit, E., van Doremalen, N., Falzarano, D. & Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nature Reviews Microbiology* **14**, 523–534 (2016).
9. Bellani, G. *et al.* Epidemiology, Patterns of Care, and Mortality for Patients With Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries. *JAMA* **315**, 788–800 (2016).
10. Chen, B. *et al.* Computational Discovery of Niclosamide Ethanolamine, a Repurposed Drug Candidate That Reduces Growth of Hepatocellular Carcinoma Cells In Vitro and in Mice by Inhibiting Cell Division Cycle 37 Signaling. *Gastroenterology* **152**, 2022–2036 (2017).
11. Qu, X. A. & Rajpal, D. K. Applications of Connectivity Map in drug discovery and development. *Drug Discovery Today* **17**, 1289–1298 (2012).

12. Sirota, M. *et al.* Discovery and Preclinical Validation of Drug Indications Using Compendia of Public Gene Expression Data. *Science Translational Medicine* **3**, 96ra77 (2011).

13. van Noort, V. *et al.* Novel Drug Candidates for the Treatment of Metastatic Colorectal Cancer through Global Inverse Gene-Expression Profiling. *Cancer Res* **74**, 5690 (2014).

14. Pessetto, Z. Y. *et al.* In silico and in vitro drug screening identifies new therapeutic approaches for Ewing sarcoma. *Oncotarget* **8**, (2016).

15. Wu, H., Huang, J., Zhong, Y. & Huang, Q. DrugSig: A resource for computational drug repositioning utilizing gene expression signatures. *PLOS ONE* **12**, e0177743 (2017).

16. Lee, B. K. B. *et al.* DeSigN: connecting gene expression with therapeutics for drug repurposing and development. *BMC Genomics* **18**, 934 (2017).

17. Chen, B. *et al.* Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. *Nature Communications* **8**, 16022 (2017).

18. Chu, C. M. *et al.* Role of lopinavir/ritonavir in the treatment of SARS: initial virological and clinical findings. *Thorax* **59**, 252 (2004).

19. Lim, J. *et al.* Case of the Index Patient Who Caused Tertiary Transmission of Coronavirus Disease 2019 in Korea: the Application of Lopinavir/Ritonavir for the Treatment of COVID-19 Pneumonia Monitored by Quantitative RT-PCR. *J Korean Med Sci* **35**, (2020).

20. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452.e17 (2017).

21. Jeon, S. *et al.* Identification of antiviral drug candidates against SARS-CoV-2 from FDA-approved drugs. *bioRxiv* 2020.03.20.999730 (2020) doi:10.1101/2020.03.20.999730.

22. Caly, L., Druce, J. D., Catton, M. G., Jans, D. A. & Wagstaff, K. M. The FDA-approved Drug Ivermectin inhibits the replication of SARS-CoV-2 in vitro. *Antiviral Research* 104787 (2020) doi:10.1016/j.antiviral.2020.104787.

23. Wu, C.-J. *et al.* Inhibition of Severe Acute Respiratory Syndrome Coronavirus Replication by Niclosamide. *Antimicrob. Agents Chemother.* **48**, 2693 (2004).

24. Shen, L. *et al.* High-Throughput Screening and Identification of Potent Broad-Spectrum Inhibitors of Coronaviruses. *J. Virol.* **93**, e00023-19 (2019).

25. Kupferschmidt, K. & Cohen, J. Race to find COVID-19 treatments accelerates. *Science* **367**, 1412 (2020).

26. Musella, M., Manic, G., De Maria, R., Vitale, I. & Sistigu, A. Type-I-interferons in infection and cancer: Unanticipated dynamics with therapeutic implications. *OncoImmunology* **6**, e1314424 (2017).

27. Weston, S., Haupt, R., Logue, J., Matthews, K. & Frieman, M. B. FDA approved drugs with broad anti-coronaviral activity inhibit SARS-CoV-2 in vitro. *bioRxiv* 2020.03.25.008482 (2020) doi:10.1101/2020.03.25.008482.

28. Vivian, J. *et al.* Toil enables reproducible, open source, big biomedical data analyses. *Nature Biotechnology* **35**, 314–316 (2017).

29. Liu, K. *et al.* Evaluating cell lines as models for metastatic breast cancer through integrative analysis of genomic data. *Nature Communications* **10**, 2138 (2019).

30. lincscloud. http://www.lincscloud.org/.

31. Zeng, B. *et al.* OCTAD: an open workplace for virtually screening therapeutics targeting precise cancer patient groups using gene expression features. *bioRxiv* 821546 (2019) doi:10.1101/821546.

32. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).

33. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

Supplementary Information

Reversal of Infected Host Gene Expression Identifies Repurposed Drug Candidates for COVID-19

Jing Xing[1#], Rama Shankar[1#], Aleksandra Drelich[2#], Shreya Paithankar[1], Eugene Chekalin[1], Thomas Dexheimer[3], Surender Rajasekaran[1,4], Chien-Te Kent Tseng[2,5*] and Bin Chen[1,3*].

[1] Department of Pediatrics and Human Development, Michigan State University, Grand Rapids, Michigan, USA.

[2] Departments of Microbiology and Immunology, University of Texas Medical Branch, Galveston, Texas, USA

[3] Department of Pharmacology and Toxicology, Michigan State University, Grand Rapids, Michigan, USA.

[4] Helen Devos Children Hospital, Grand Rapids, Michigan, USA

[5] Center of Biodefense and Emerging Disease, University of Texas Medical Branch, Galveston, Texas, USA

[#] These authors contributed equally

[*] Correspondence to Bin Chen: chenbi12@msu.edu, Kent Chien-Te, Tseng: sktseng@UTMB.EDU

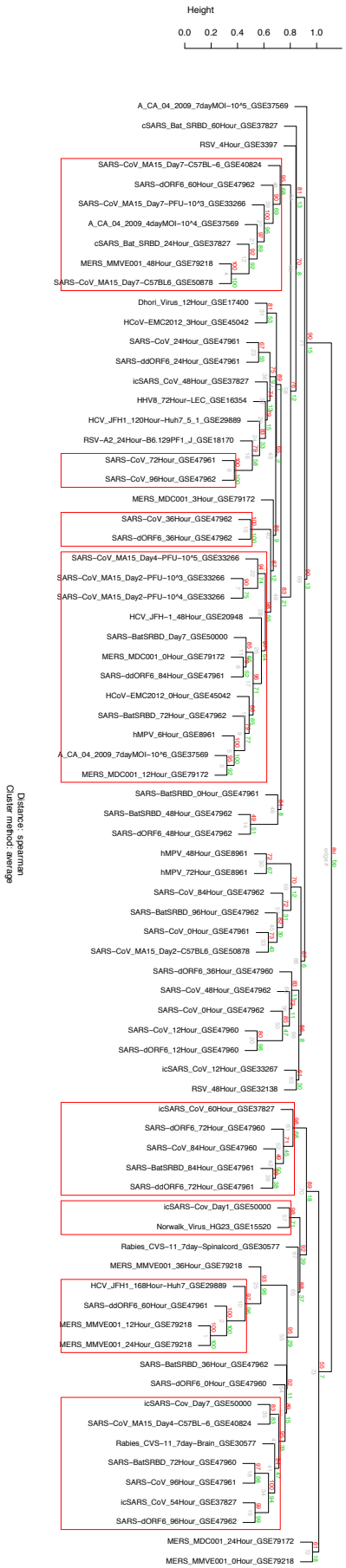

Figure S1. Hierarchical clustering of virus infection signatures based on their drug prediction results. Significant SARS/MERS clusters with approximately unbiased (AU) p-values no less than 95% were highlighted with red rectangles. Red and green numbers indicate AU and bootstrap probability (BP) p-values, respectively. Edge numbers are shown in grey. Among the 291 virus-induced signatures, only those that mapped to at least 50 LINCS landmark genes were used in this analysis.

Table S1. Data sets used in this study for disease signature creation.

S. No.	Accession	Platform	Disease/ Infection	Organism	Model	Time points	Number of samples	PMID
1	GSE17400	GPL570	MOCK, DOHV and SARS	Homo sapiens	Calu-3	3	27	20090954
2	GSE30589	GPL570	SARS and MOCK	Homo sapiens	Vero E6, Vero E6 DeltaE, MA-104 and MA-104 DeltaE	4	33	22028656
3	GSE45042	GPL6480	MOCK and EMC	Homo sapiens	Calu-3	6	33	23631916 24846384
4	GSE47960	GPL6480	MOCK, SARS, H1N1, SARS-BatSRBD and SARS-dORF6	Homo sapiens	HAE	11	163	23935999
5	GSE79218	GPL13497	MOCK and MERS	Homo sapiens	MMVE001	5	49	
6	GSE79172	GPL13497	MOCK and MERS	Homo sapiens	MDC001	5	29	28830941
7	GSE22581	GPL3738	MOCK and SARS	Canis lupus	Ferret Lung	3	9	21035159
8	GSE36016	GPL7202	MOCK and SARS	Mus musculus	lung WT, lung IFNAR1 and lung STAT	3	36	
9	GSE68820	GPL7202	MOCK and SARS	Mus musculus	lung wt and lung TLR3	3	52	26015500
10	GSE30351	GPL6480	JFH1	Homo sapiens	Huh7	2	7	
11	GSE71063	GPL570	HIV	Homo sapiens	patients	2	40	26935044
12	SRP166108	Illumina HiSeq		Homo sapiens	HepaRG		50	

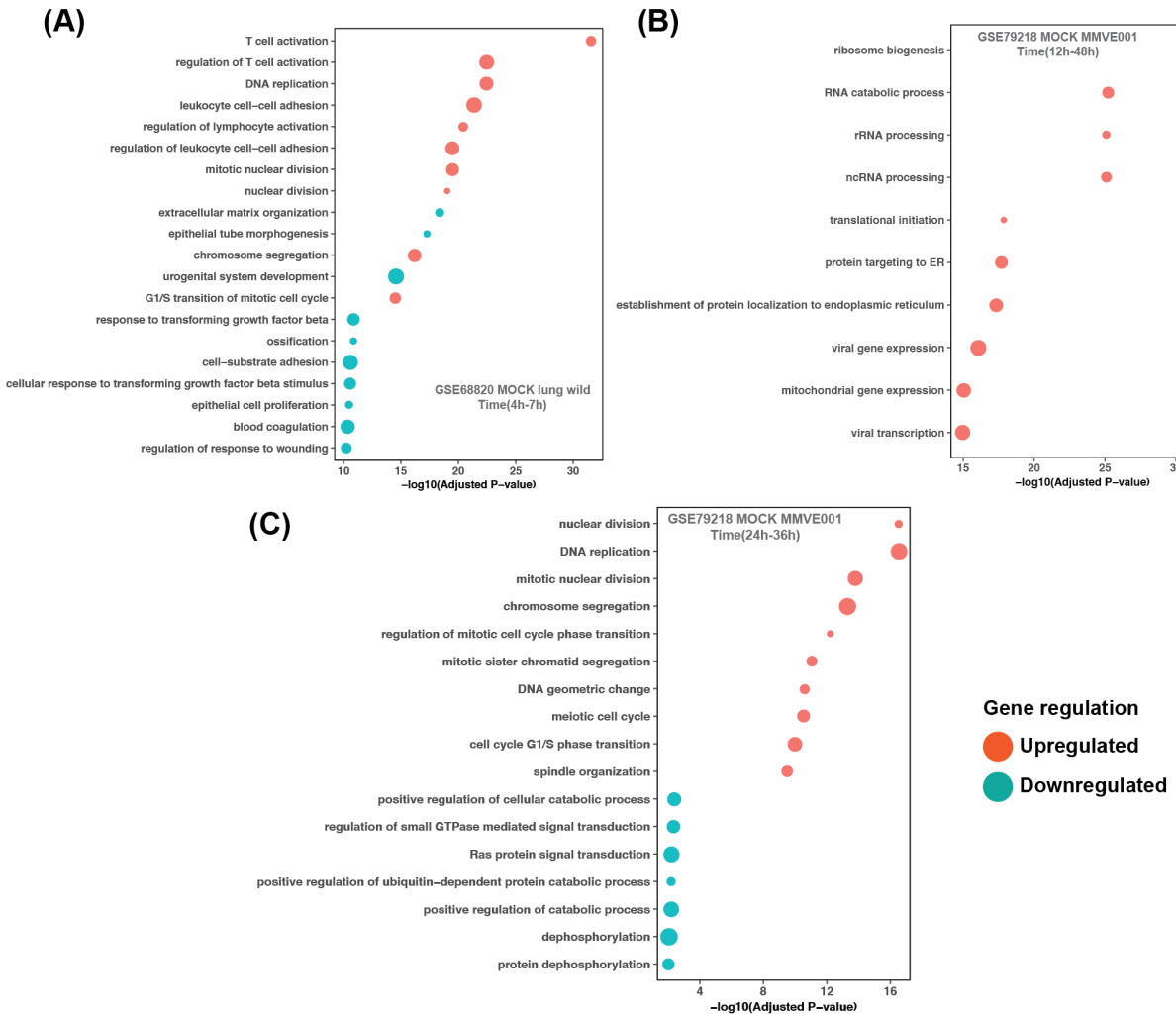

Figure S2. Gene ontology enrichment of differentially expressed genes obtained among different time points in MOCK group of samples. Gene ontology terms enriched in MOCK infected samples in **(A)** lung tissues without any knockout and **(B)** lung tissues with TLR3 knockout in GSE68820, and **(C)** MMVE001 cell lines in GSE79172. Terms enriched in MOCK infected samples are mostly cell cycle related. The size of the circles represents the ratio of genes in each process and color of the circle represents the regulation (Upregulated- red and downregulated- green) of genes.

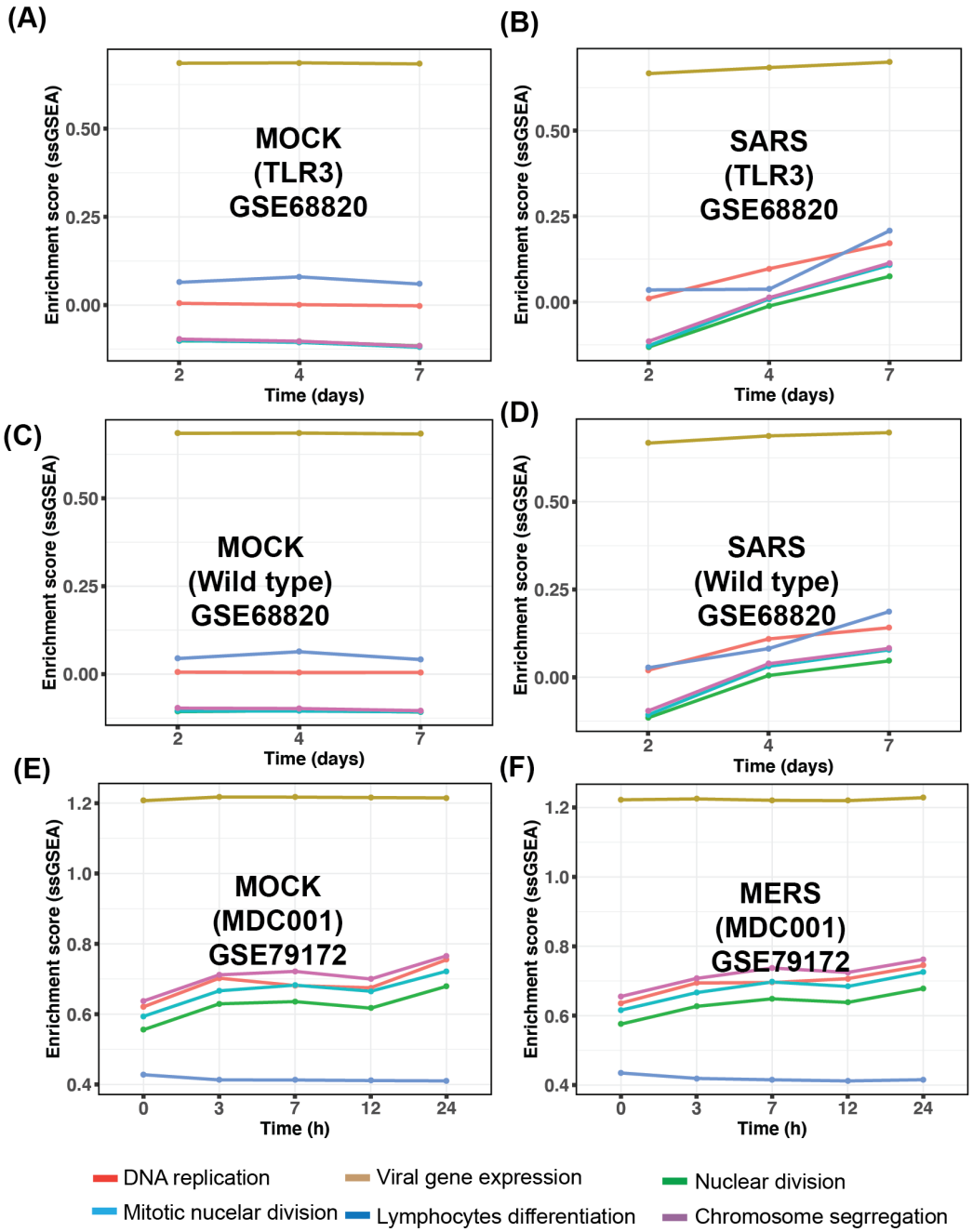

Figure S3. Gene set enrichment of the biological biological processes at different time points. Variations of gene set enrichment (ssGSEA) of biological processes obtained at different time points in MOCK **(A)** and SARS infected **(B)** lung tissue with TLR3 mutant. Similarly, variation of ssGSEA of biological processes obtained at different time points in MOCK **(C)** and SARS infected **(D)** lung tissue in GSE68820, and MOCK **(E)** and MERS **(F)** infected MDC001 cell lines in GSE79172. The enrichment of expression of genes involved in these processes was increasing in SARS/MERS infected tissue but found to be consistent in MOCK samples with time.

Table S2. Positive control drugs with known activity against MERS-CoV/SARS-CoV/SARS-CoV-2.

Name	MERS EC50 µM	SARS EC50 µM	COVID-19 EC50 µM	MoA	PMID
amodiaquine	6.21	1.27	NA*	Antiparasitic agent	32006468
astemizole	4.88	5.59	NA	Neurotransmitter inhibitor	32006468
bisindolylmaleimide-ix	inhibition 74% @ 10uM	NA	NA	PKC inhibitor	25653449

Name	MERS EC50 μM	SARS EC50 μM	COVID-19 EC50 μM	MoA	PMID
bufalin	inhibition > 90% @ 10 nM	NA	NA	inhibit MERS-CoV entry by blocking clathrin-mediated endocytosis	25653449
chloroquine	3	NA	1.13	Endosomal acidification inhibitor	32006468
chlorpromazine	4.9	NA	NA	Neurotransmitter inhibitor	32006468
clomipramine	9.33	13.23	NA	Neurotransmitter inhibitor	32006468
dasatinib	5.46	2.1	NA	ABL1 inhibitor	32006468
disulfiram	NA	NA	NA	MERS-CoV PL-pro inhibitor	32006468
emetine	0.34	NA	NA	Inhibits RNA, DNA and protein synthesis	32006468
fluphenazine	5.86	21.43	NA	Neurotransmitter inhibitor	32006468
fluspirilene	7.47	5.96	NA	Neurotransmitter inhibitor	32006468
gemcitabine	1.21	4.95	NA	DNA metabolism inhibitor	32006468
GW-5074	inhibition 52% @ 10uM	NA	NA	Raf inhibitor	25653449
imatinib	17.68	9.82	NA	ABL1 inhibitor	32006468
loperamide	4.8	5.9	NA	antidiarrheal opioid receptor agonist	32006468
lopinavir	8	24.4	NA	HIV-1 inhibitor	32006468
mefloquine	7.41	15.55	NA	Antiparasitic agent	32006468
mg-132	NA	NA	NA	cys-protease m-calpain inhibition	22787216
monensin	3.27	NA	NA	Antibacterial	32006468
mycophenolate-mofetil	1.54	NA	NA	Immune suppressant, antineoplastic, antiviral	32006468
niclosamide	NA	2	NA	Antiparasitic agent	15215127
nitazoxanide	NA	NA	2.12	antiprotozoal, type I IFN inducer	32020029
ouabain	inhibition 70% @ 50 nM	NA	NA	inhibit MERS-CoV entry by blocking clathrin-mediated endocytosis	25653449
penciclovir	NA	NA	95.96	Inhibition of virus DNA synthesis	32020029
phenazopyridine	1.93	NA	NA	Analgesic	32006468
promethazine	11.8	7.54	NA	Neurotransmitter inhibitor	32006468
pyrvinium-pamoate	1.84	NA	NA	Anthelmintic	32006468
ribavirin	NA	NA	109.5	ribonucleic analog	32020029
ritonavir	24.9	NA	NA	HIV protease inhibitor	31924756
saracatinib	2.9	2.4	NA	Src family of tyrosine kinases inhibitor	32006468
SB-203580	inhibition 45% @ 10uM	NA	NA	p38 MAPK inhibitor	25653449
selumetinib	inhibition >= 95% @ 10uM	NA	NA	MEK1, ERK1/2 inhibitor	25653449
sirolimus	inhibition 61% @ 10uM	NA	NA	MTOR inhibitor	25653449
tamoxifen	10.11	92.88	NA	Estrogen receptor inhibitor	32006468
terconazole	12.2	15.32	NA	Sterol metabolism inhibitor	32006468
thiothixene	9.29	5.31	NA	Neurotransmitter inhibitor	32006468
toremifene	12.91	11.96	NA	Estrogen receptor inhibitor	32006468
trametinib	inhibition >= 95% @ 10uM	NA	NA	MEK1/2 inhibitor	25653449
trifluprom azine	5.75	6.39	NA	Neurotransmitter inhibitor	32006468
U-0126	inhibition 51% @ 10uM	NA	NA	MEK1/2 inhibitor	25653449

* “NA” indicates not available.

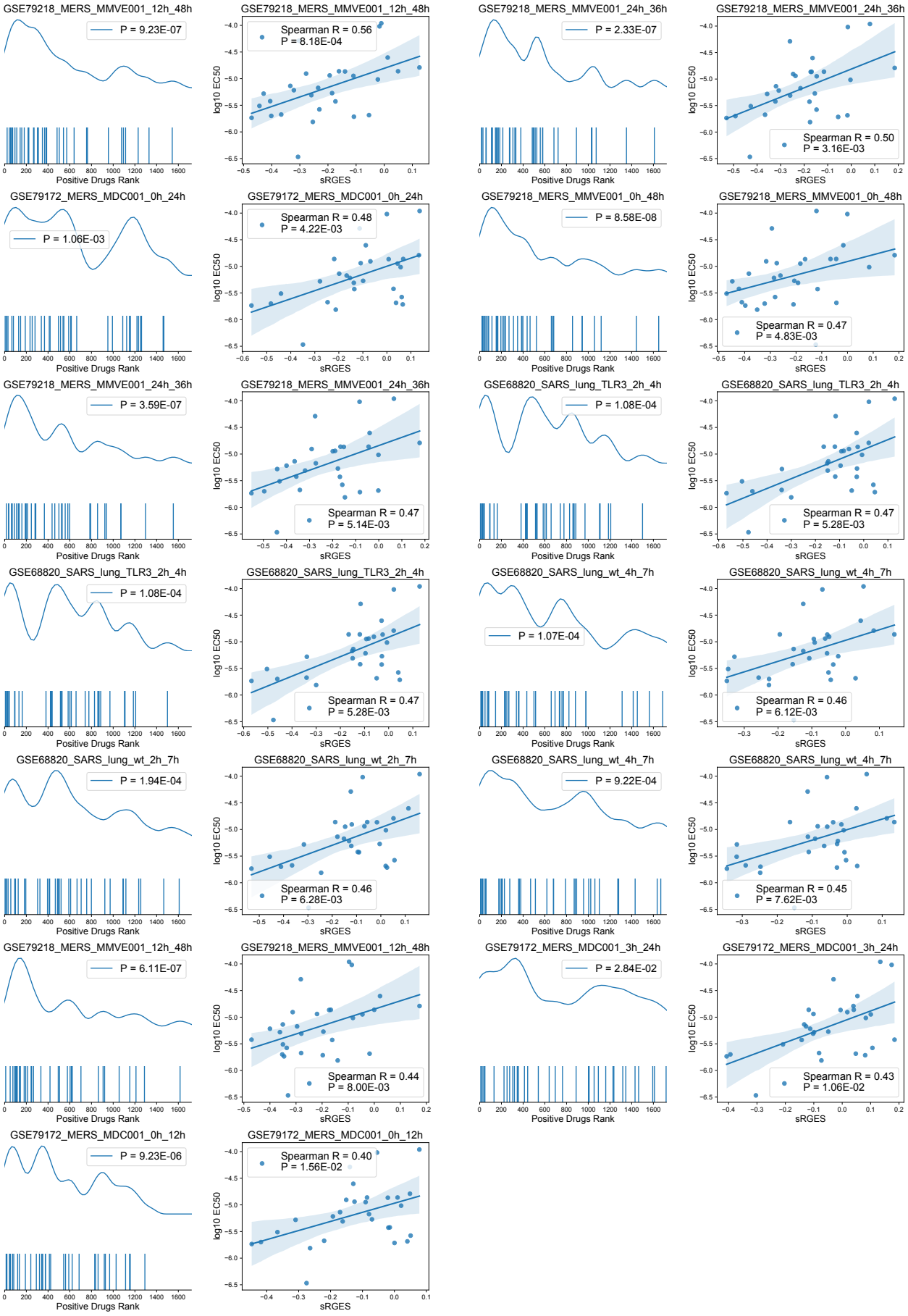

Figure S4. SARS/MERS signatures validation using known active drugs (positive controls). The first and third columns show enrichment density and barcode, with p values shown at upper right. The curve shows

enrichment density, and each bar under the curve represents the rank of a positive drug among all the drugs in one prediction. The second and last columns list the correlation between sRGES and EC50 (in M, log10 transformed) of the positive controls, with Spearman R and p value shown at lower right. Each point indicates a positive control drug.

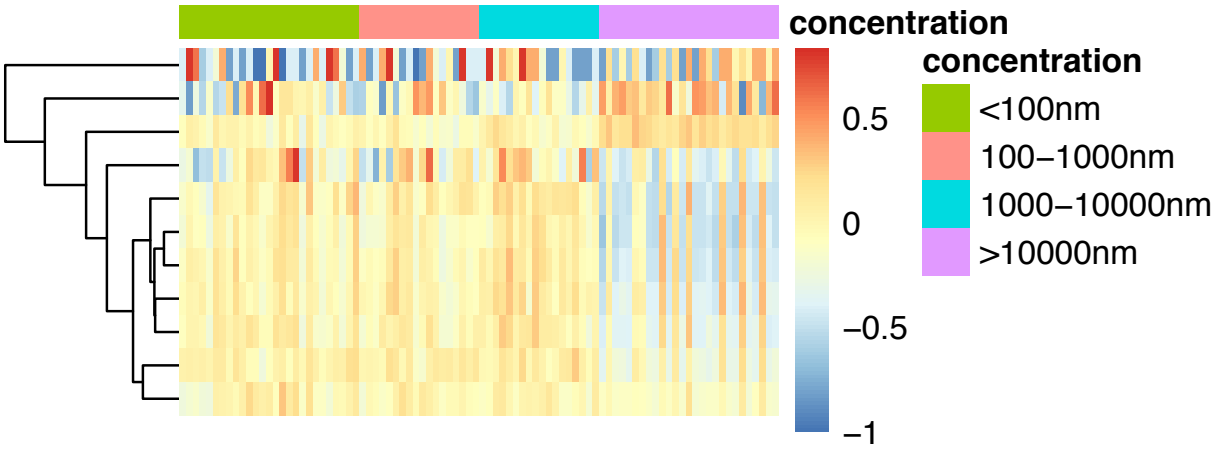

Figure S5. Ritonavir profiles measured under different concentrations. Each row indicates an infection signature, each column indicates a ritonavir profile under a different concentration annotated as the colorbar on the right side. The heatmap shows the spearman correlation coefficients between drug signatures and disease signatures. Blue means a better reversal effect.

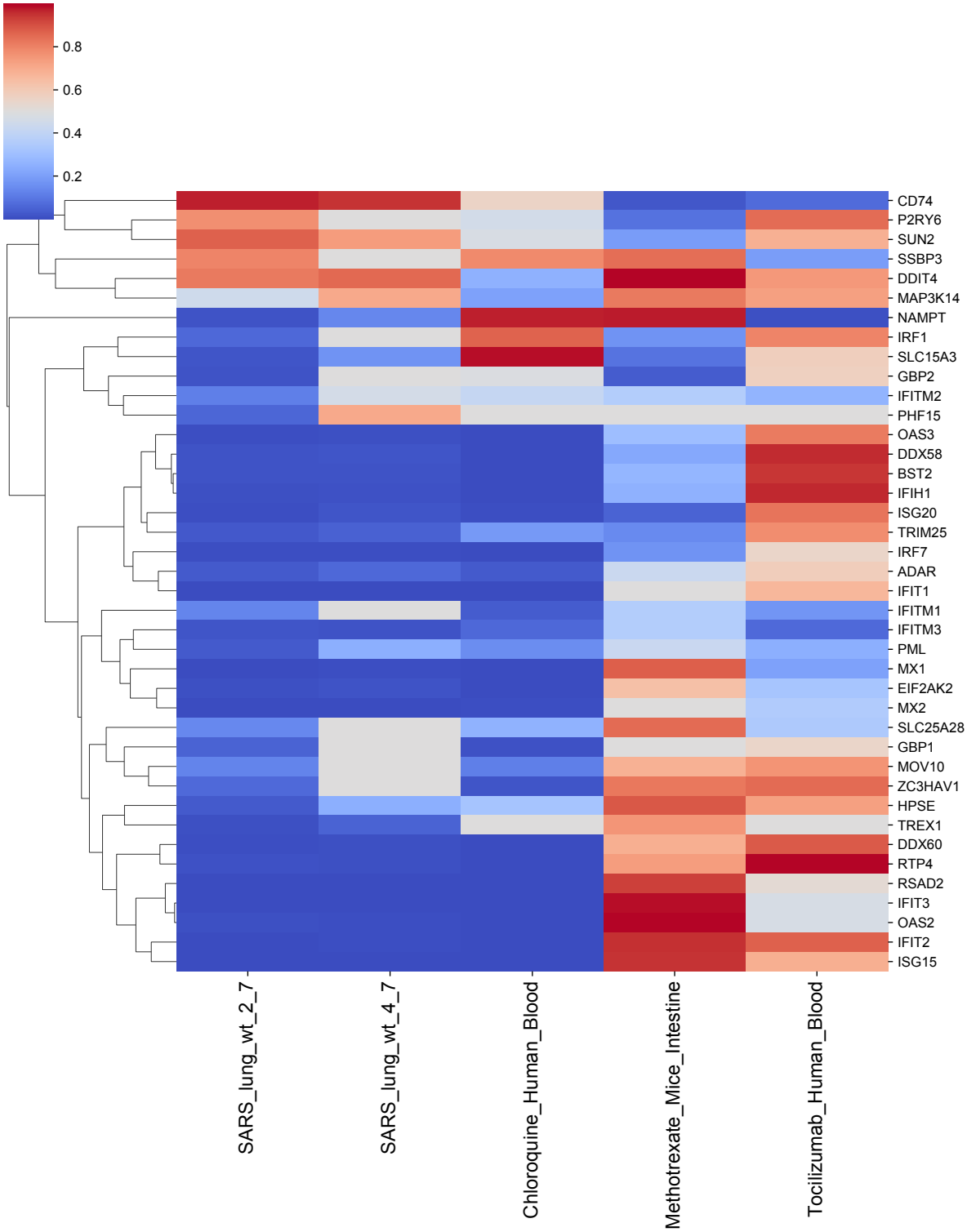

Figure S6. Heatmap showing the interferon-stimulated genes expression (rank percentage in the whole transcriptome) in infection (GSE68820), chloroquine (GSE71063), methotrexate (GSE56426) and tocilizumab (GSE25160) treated mice or patients. Red color indicates up-regulated genes and blue indicates down-regulated genes.

Table S3. Antiviral effect and other information of the selected candidates.

Name	Median Rank (CoV)	Median Rank (MOCK)	Clinical Phase	Route of Administration	Targets	MoA	Indication
Bortezomib	3	23	Launched	Intravenous	PSMA1-8 PSMB1-11 PSMD1-2 RELA	NFkB pathway inhibitor Proteasome inhibitor	Multiple myeloma, mantle cell lymphoma
Puromycin	18	128	Launched	Oral	60S ribosomal proteins	Protein synthesis inhibitor	
Methotrexate	37	356	Launched	Oral	DHFR	Dihydrofolate reductase inhibitor	Gestational choriocarcinoma, hydatidiform mole, acute lymphoblastic leukemia, psoriasis, rheumatoid arthritis
Methylene blue	57	153	Launched	Intravenous	ACHE	Guanylyl cyclase inhibitor nitric oxide production inhibitor	Methemoglobinemia
Tyloxapol	106	293	Launched	Inhalant	LPL NFKB2	NFkB pathway inhibitor	
Nisoldipine	108	245	Launched	Oral	Voltage-dependent L-type calcium channels	Calcium channel blocker	Hypertension
Nvp-bez235	178	173	Phase 2	Oral	ATR MTOR PI3Ks	mTOR inhibitor PI3K inhibitor	
Fluvastatin	283	362	Launched	Oral	HMGCR	HMGCR inhibitor	Hypercholesterolemia, congenital heart defects
Alvocidib	290	823	Phase 2	Intravenous	CDK1-2 CDK4-9 EGFR PYGM	CDK inhibitor	
Dasatinib	378	1103	Launched	Oral	ABL1/2 Src family kinases EPHA2 KIT PDGFRB SRMS STAT5B	Bcr-Abl kinase inhibitor Ephrin inhibitor KIT inhibitor PDGFR tyrosine kinase receptor inhibitor Src family inhibitor	Chronic myeloid leukemia, acute lymphoblastic leukemia