Transcriptomics-based drug repositioning pipeline identifies therapeutic candidates for COVID-19

Authors: Brian L. Le[1,2+], Gaia Andreoletti[1,2+], Tomiko Oskotsky[1,2+], Albert Vallejo-Gracia[3], Romel Rosales[4,5], Katharine Yu[1,2,6], Idit Kosti[1,2], Kristoffer E. Leon[3], Daniel G. Bunis[1,2,6], Christine Li[1,2,7], G. Renuka Kumar[3], Kris M. White[4,5], Adolfo García-Sastre[4,5,8,9], Melanie Ott[3,10], Marina Sirota[1,2*]

Affiliations:
[1] Department of Pediatrics, UCSF, SF, CA, USA
[2] Bakar Computational Health Sciences Institute, UCSF, SF, CA, USA
[3] Gladstone Institute of Virology, Gladstone Institutes, SF, CA, USA
[4] Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[5] Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[6] Biomedical Sciences Graduate Program, UCSF, SF, CA, USA
[7] Shanghai American School, Shanghai, China
[8] Department of Medicine, Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[9] The Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[10] Department of Medicine, UCSF, SF, CA, USA

[+] These authors contributed equally to this manuscript.

*Correspondence should be addressed to: marina.sirota@ucsf.edu

Abstract

The novel SARS-CoV-2 virus emerged in December 2019 and has few effective treatments. We applied a computational drug repositioning pipeline to SARS-CoV-2 differential gene expression signatures derived from publicly available data. We utilized three independent published studies to acquire or generate lists of differentially expressed genes between control and SARS-CoV-2-infected samples. Using a rank-based pattern matching strategy based on the Kolmogorov-Smirnov Statistic, the signatures were queried against drug profiles from Connectivity Map (CMap). We validated sixteen of our top predicted hits in live SARS-CoV-2 antiviral assays in either Calu-3 or 293T-ACE2 cells. Validation experiments in human cell lines showed that 11 of the 16 compounds tested to date (including clofazimine, haloperidol and others) had measurable antiviral activity against SARS-CoV-2. These initial results are encouraging as we continue to work towards a further analysis of these predicted drugs as potential therapeutics for the treatment of COVID-19.

Introduction

SARS-CoV-2 has already claimed at least a million lives, has been detected in at least 40 million people, and has likely infected at least another 200 million. The spectrum of disease caused by the virus can be broad ranging from silent infection to lethal disease, with an estimated infection-fatality ratio around 1%[1]. SARS-CoV-2 infection has been shown to affect many organs of the body in addition to the lungs[2]. Three epidemiological factors increase the risk of disease severity: increasing age, decade-by-decade, after the age of 50 years; being male; and various underlying medical conditions[1]. However, even taking these factors into account, there is immense interindividual clinical variability in each demographic category considered[3]. Recently, researchers found that more than 10% of people who develop severe COVID-19 have misguided antibodies—autoantibodies—that attack the innate immune system. Another 3.5% or more of people who develop severe COVID-19 carry specific genetic mutations that impact innate immunity. Consequently, both groups lack effective innate immune responses that depend on type I interferon, demonstrating a crucial role for type I interferon in protecting cells and the body from COVID-19. Whether the type I interferon has been neutralized by autoantibodies or—because of a faulty gene—is produced in insufficient amounts or induced an inadequate antiviral response, the absence of type I IFN-mediated immune response appears to be a commonality among a subgroup of people who suffer from life-threatening COVID-19 pneumonia[3].

While numerous efforts are underway to identify potential therapies targeting various aspects of the disease, there is a paucity of clinically proven treatments for COVID-19. There have been efforts to therapeutically target the hyperinflammation

associated with severe COVID-19[4], as well as to utilize previously identified antiviral medications[5,6]. One of these antivirals, remdesivir, an intravenously administered RNA-dependent RNA polymerase inhibitor, showed positive preliminary results in patients with severe COVID-19[7]. In October 2020, the FDA approved remdesivir for the treatment of COVID-19[8]. Dexamethasone has also been shown to reduce the mortality rate in cases of severe COVID-19[9].

Nevertheless, the lack of treatments and the severity of the current health pandemic warrant the exploration of rapid identification methods of preventive and therapeutic strategies from every angle. The traditional paradigm of drug discovery is generally regarded as protracted and costly, taking approximately 15 years and over $1 billion to develop and bring a novel drug to market[10]. The repositioning of drugs already approved for human use mitigates the costs and risks associated with early stages of drug development, and offers shorter routes to approval for therapeutic indications. Successful examples of drug repositioning include the indication of thalidomide for severe erythema nodosum leprosum and retinoic acid for acute promyelocytic leukemia[11]. The development and availability of large-scale genomic, transcriptomic, and other molecular profiling technologies and publicly available databases, in combination with the deployment of the network concept of drug targets and the power of phenotypic screening, provide an unprecedented opportunity to advance rational drug design.

Drug repositioning is being extensively explored for COVID-19. High-throughput screening pipelines have been implemented in order to quickly test drug candidates as they are identified[12–15]. In the past, our group has successfully applied a transcriptomics-based computational drug repositioning pipeline to identify novel therapeutic uses for

existing drugs[16]. This pipeline leverages transcriptomic data to perform a pattern-matching search between diseases and drugs. The underlying hypothesis is that for a given disease signature consisting of a set of up and down-regulated genes, if there is a drug profile where those same sets of genes are instead down-regulated and up-regulated, respectively, then that drug could be therapeutic for the disease. This method based on the Kolmogorov-Smirnov (KS) test statistic has shown promising results for a variety of different indications, including inflammatory bowel disease[17], dermatomyositis[18], cancer[19–21], and preterm birth[22].

In existing work from Xing et al.[23], this pipeline has been used to identify potential drug hits from multiple input disease signatures derived from SARS-CoV or MERS-CoV data. The results were aggregated to obtain a consensus ranking, with 10 drugs selected for *in vitro* testing against SARS-CoV-2 in Vero E6 cell lines, with four drugs (bortezomib, dactolisib, alvocidib and methotrexate) showing viral inhibition[23]. However, this pipeline has not yet been applied specifically to SARS-CoV-2 infection.

A variety of different transcriptomic datasets related to SARS-CoV-2 were published in the spring of 2020. In May 2020, Blanco-Melo et al. studied the transcriptomic signature of SARS-CoV-2 in a variety of different systems, including human cell lines and a ferret model[24]. By infecting human adenocarcinomic alveolar basal epithelial cells with SARS-CoV-2 and comparing to controls, the authors generated a list of 120 differentially expressed genes. They observed two enriched pathways: one composed primarily of type-I interferon-stimulated genes (ISGs) involved in the cellular response to viral infection; and a second composed of chemokines, cytokines, and complement proteins involved in the humoral response. After infecting the cell lines, Blanco-Melo et al. did not

detect either ACE2 or TMPRSS2, which are the SARS-CoV-2 receptor and SARS-CoV-2 protease, respectively[25]. However, supported viral replication was observed, thereby allowing the capture of some of the biological responses to SARS-CoV-2.

In May 2020, another study by Lamers et al. examined SARS-CoV-2 infection in human small intestinal organoids grown from primary gut epithelial stem cells[26]. The organoids were exposed to SARS-CoV-2 and grown in various conditions, including Wnt-high expansion media. Enterocytes were readily infected by the virus, and RNA sequencing revealed upregulation of cytokines and genes related to type I and III interferon responses.

A limited amount of transcriptomic data from human samples has also been published. One study detailed the transcriptional signature of bronchoalveolar lavage fluid (of which responding immune cells are often a primary component) of COVID-19 patients compared to controls[27]. Despite a limited number of samples, the results revealed inflammatory cytokine profiles in the COVID-19 cases, along with enrichments in the activation of apoptosis and the P53 signaling pathways.

On the drug side, data are available in the form of differential gene expression profiles from testing on human cells. Publicly-available versions include the Connectivity Map (CMap)[28], which contains genome-wide testing on approximately 1,300 drugs, wherein the differential profile for a drug was generated by comparing cultured cells treated with the drug to untreated control cultures.

Here, we applied our existing computational drug repositioning pipeline to identify drug profiles with significantly reversed differential gene expression compared to several diverse input signatures for SARS-CoV-2 effects on human cells. By taking into account

a broader view of differentially expressed gene sets from both cell line and organoid disease models and human samples, the predictions are complementary to other drug discovery approaches. We identified 102 unique drug hits, from which 25 were identified in at least two of the signatures, several of which have been already investigated in clinical trials. We furthermore explore our findings in the context of other computational drug repurposing efforts for COVID-19. Finally, we tested 16 of our top predicted hits in live SARS-CoV-2 antiviral assays. Four of the top predicted inhibitors were tested for virus inhibition in a human lung cell line, Calu-3, infected with SARS-CoV-2 with quantitation of the secreted virus assessed by RT-qPCR assay. Thirteen predicted inhibitors (including one tested in Calu-3) were incubated with SARS-CoV-2 infected human embryonic kidney 293T cells overexpressing ACE2 (293T-ACE2) with viral replication determined using an immunofluorescence-based assay.

Results

In this study, we applied our drug repositioning pipeline to SARS-CoV-2 differential gene expression signatures derived from publicly available RNA-seq data (Figure 1). The transcriptomic data were generated from distinct types of tissues, so rather than aggregating them together, we predicted therapeutics for each signature and then combined the results. We utilized three independent gene expression signatures (labelled "ALV", "EXP", and "BALF"), each of which consisted of lists of differentially expressed genes between SARS-CoV-2 samples and their respective controls. The ALV signature was generated from human adenocarcinomic alveolar basal epithelial cells by comparing SARS-CoV-2 infection to mock-infection conditions[24]. The EXP signature originated from a study where organoids, grown from human intestinal cells expanded in Wnt-high expansion media, were infected with SARS-CoV-2 and then compared to controls[26]. The BALF signature was from a contrast of primary human BALF samples from two COVID-19 patients versus three controls[27]. Each of these signatures was contrasted with drug profiles of differential gene expression from CMap.

For each of the input signatures, we applied a significance threshold false discovery rate (FDR) < 0.05. We further applied minimum fold change thresholds in order to identify the driving genes. The ALV signature had only 120 genes, with 109 genes shared with the drug profiles; in order to maintain at least 100 genes for the pattern-matching algorithm to work with, we applied no fold-change threshold. For the EXP signature, we applied a $|\log_2 FC| > 2$ cutoff, resulting in 125 genes for the expansion signature (108 shared with the drug profiles). For the BALF signature, we processed the raw read count data to calculate differential gene expression values. We applied a

199 $|\log_2 FC| > 4$ cutoff, with the BALF data yielding 1,349 protein-coding genes for the lavage
200 fluid signature (941 shared with the drug profiles). As the sample types across the three
201 datasets were very different, we used different fold change thresholds to identify the
202 appropriate gene signatures to be used for drug repurposing. The gene lists for each of
203 these signatures can be found in the supplement (Tables S1, S2, S3).

204 We used GSEA (Gene Set Enrichment Analysis)[29,30] to annotate enriched (FDR
205 0.05) Hallmark pathways from each of the input signatures (Figure 2A). A number of
206 pathways common to at least two signatures were found. Interferon alpha response and
207 interferon beta response were upregulated in the ALV and EXP signatures. Adipogenesis
208 and cholesterol homeostasis pathways were downregulated in the EXP and BALF
209 signatures. KRAS signaling, and mTORC1 (mammalian target of rapamycin complex 1)
210 signaling were enriched in all three signatures, but not in the same direction, showing the
211 diversity of effects SARS-CoV-2 may have on human cells, and highlighting a need for
212 utilization of diverse profiles as we do in the present study. When we look at the
213 contributing genes within the three signatures (Figure 2B), we found one overlapping
214 upregulated gene - Dickkopf WNT Signaling Pathway Inhibitor 1 (DKK1). We used the
215 publicly available single-cell RNAseq dataset GSE128033[31] composed of 13 patients (4
216 healthy, 3 presenting with mild COVID-19 symptoms, and 6 presenting with severe
217 COVID-19 symptoms) to further characterize the expression of DKK1 (Figure S1). Data
218 were re-analyzed following the standard Seurat pipeline. From the analyses of the single-
219 cell data, DKK1 is highly expressed in COVID-19 patients compared to controls,
220 specifically in severe patients and it is expressed by epithelial cells.

After analyzing the input SARS-CoV-2 signatures, we utilized our repositioning pipeline to identify drugs with reversed profiles from CMap (Figure 1). Significantly reversed drug profiles were identified for each of the signatures using a permutation approach: 30 hits from the ALV signature (Table S4), 15 hits from the EXP signature (Table S5), and 86 hits from the BALF signature (Table S6). When visualizing the gene regulation of the input signatures and their respective top 15 drug hits, the overall reversal pattern can be observed (Figure 2C-E). Interestingly, we found several drugs shared across datasets that significantly reversed the disease signature. For example, haloperidol, highlighted in purple in Figures 2C-E, was shown to reverse the disease gene signature from three datasets, whereas levopropoxyphene, shown in green in Figures 2D-E, was observed to reverse the disease gene signature from two of the datasets. In total, our analysis identified 102 unique drug hits (Table S7). Twenty-five drug hits reversed at least two signatures ($p = 0.0334$, random sampling), and four drug hits reversed three signatures ($p = 0.0599$, random sampling) (Table 1, Figure 3A).

We further characterized the common hits by examining their interactions with proteins in humans. We used known drug targets from DrugBank[32] and predicted additional targets using the similarity ensemble approach (SEA)[33]. We visualized the known interactions from DrugBank in a network. Figure 3B shows the connectivity across compounds highlighting both single drug genes (such as *SIGMAR1* for haloperidol) and genes shared across drugs, such as *ADRA2A* and *DRD1* for haloperidol and co-dergocrine mesilate. The proteins with the most known interactions with our list of 25 drugs included adrenergic receptors (particularly α2 adrenoreceptors), dopamine receptors, and serotonin receptors.

To confirm the validity of our approach, the inhibitory effects of 16 of our drug hits which significantly reversed multiple SARS-CoV-2 profiles were assessed in live **pCC50** antiviral assays. Next, we wanted to test the potential of the predicted compounds to inhibit viral activity robustly using different human cell lines - Calu-3 and 293T-ACE2. The respective selection of 16 and 13 compounds for testing was based on side effect profiles and compound availability.

The inhibitory effects of haloperidol, clofazimine, valproic acid, and fluticasone were evaluated in SARS-CoV-2 infected Calu-3 cells (human lung epithelial cell line), with remdesivir also tested as a positive control. From these five, remdesivir and haloperidol inhibited viral replication (Figure 4A), and the inhibitory effect was also observed by microscopy (Figure 4B). Toxicity assessments for haloperidol, clofazimine, valproic acid, and fluticasone were evaluated using viability assays (Alamarblue) in Calu-3 cells treated with each compound for 72h (n=1, with 2 technical replicates). No significant differences between controls and the biological replicates were detected using a non-parametric test (Kruskal-Wallis) (Figure S6). Fluticasone and bacampicillin showed some toxicity in a dose-dependent manner at the highest doses tested. Haloperidol, clofazimine and valproic acid did not show significantly reduced cell viability (Figure S6).

Additionally, 13 drugs (bacampicillin, ciclopirox, ciclosporin, clofazimine, dicycloverine, fludrocortisone, isoxicam, lansoprazole, metixene, myricetin, pentoxifylline, sirolimus, tretinoin) were independently assessed in a live SARS-CoV-2 antiviral assay. Remdesivir was again used as a positive control. This testing involved six serial dilutions of each drug to inhibit the replication of SARS-CoV-2 in 293T-ACE2 cells using an immunofluorescence-based antiviral assay[34]. All antiviral assays were paired with

cytotoxicity assays using identical drug concentrations in uninfected human 293T-ACE2 cells. Positive control remdesivir and 10 of our predicted drugs (bacampicillin, ciclopirox, ciclosporin, clofazimine, dicycloverine, isoxicam, metixene, pentoxifylline, sirolimus, and tretinoin) showed antiviral efficacy against SARS-CoV-2, reducing viral infection by at least 50%, that was distinguishable from their cytotoxicity profile when tested in this cell line (Figure 5). Several inhibitors showed micromolar to sub-micromolar antiviral efficacy, including clofazimine, ciclosporin, ciclopirox, and metixene. These results not only confirm our predictive methods, but have also identified several clinically-approved drugs with potential for repurposing for the treatment of COVID-19.

Discussion

Here, we used a transcriptomics-based drug repositioning pipeline to predict therapeutic drug hits for three different input SARS-CoV-2 signatures, each of which came from distinct human cell or tissue origins. We found significant overlap of the therapeutic predictions for these signatures. From 102 total drug hits, 25 drugs reversed at least two signatures ($p = 0.0334$) and 4 drugs reversed three signatures ($p = 0.0599$). The diversity of such signatures yet overlap of highlighted drugs underscores the utility of the current pipeline for identification of drugs which might be therapeutic for the diverse effects of SARS-CoV-2 infection.

Twenty-five of our drug hits reversed at least two of the three input signatures (Figure 3). Notably, 14 of the 15 hits from the EXP signature were also hits for the BALF signature, despite being generated from different types of tissue. The EXP signature was generated from intestinal tissue, whereas the BALF signature was generated from constituents of the respiratory tract. Among the common hits reversing at least two of the signatures were two immunosuppressants (ciclosporin and sirolimus), an anti-inflammatory medication (isoxicam), and two steroids (fludrocortisone and fluticasone). Sirolimus (or rapamycin), an immunosuppressant and an mTOR inhibitor, is currently undergoing investigation in several clinical trials in COVID-19 patients (NCT04371640, NCT04341675, NCT04461340). Other hits currently in clinical trials for COVID-19 treatment include ciclosporin (NCT04412785, NCT04392531), niclosamide in combination with diltiazem (NCT04558021), and clofazimine in combination with interferon beta-1b (NCT04465695).

Among our four drug hits that reversed all three signatures, three drugs demonstrated in vitro antiviral efficacy - bacampicillin, clofazimine, and haloperidol with

no toxicity effects (Figure S6). Our group found haloperidol decreased viral growth in SARS-CoV-2 infected Calu-3 cells (Figure 4B) in a dose-dependent manner (Figure 4A). Haloperidol is a psychiatric medication that is indicated for the treatment of psychotic disorders including schizophrenia and acute psychosis. By blocking dopamine (mainly D2) receptors in the brain, haloperidol eliminates dopamine neurotransmission which leads to improvement of psychotic symptoms[35]. Haloperidol can also bind to the sigma-1 and sigma-2 receptors, which are implicated in lipid remodeling and cell stress response[12]. As reported by Gordon et al.[12], the SARS-CoV-2 proteins Nsp6 and ORF9c interact with the sigma-1 receptor and the sigma-2 receptor, respectively. Moreover, they found that haloperidol decreased viral replication in SARS-CoV-2-infected Vero E6 cells with, based on their reported pIC50 and pCC50 values, a calculated Selectivity Index (SI) of 53.7[12]. An SI greater than 10 is the generally accepted minimum cut-off for an antiviral worth pursuing[36,37]. In another more recent study, Gordon et al. found in their analysis of a national electronic medical record database that fewer hospitalized COVID-19 patients who were newly prescribed haloperidol and other sigma-binding typical antipsychotic medications progressed to requiring mechanical ventilation compared to those who were newly prescribed atypical antipsychotic medications that do not bind to sigma receptors[14].

Our testing of clofazimine demonstrated submicromolar antiviral effects of this drug in SARS-Co-V-2 infected 293T-ACE2 and Vero E6 cells (Figures 4 and S3). Clofazimine is an orally administered antimycobacterial drug used in the treatment of leprosy. By preferentially binding to mycobacterial DNA, clofazimine disrupts the cell cycle and eventually kills the bacterium[38]. In addition to being an antimycobacterial agent, clofazimine also possesses anti-inflammatory properties primarily by inhibiting T

lymphocyte activation and proliferation[39]. Yuan et al. found that clofazimine inhibits SARS-CoV-2 replication by interfering with spike-mediated viral entry and viral RNA replication[40]. Their work also demonstrated that clofazimine has antiviral efficacy against SARS-CoV-2 in human embryonic stem cell-derived cardiomyocytes and in an ex vivo human lung culture system, as well as antiviral synergy with remdesivir demonstrating the potential of clofazimine as part of a combination treatment regimen for COVID-19[40].

Our group found bacampicillin to have micromolar antiviral efficacy in SARS-Co-V-2 infected 293T-ACE2 cells. Bacampicillin is an orally administered prodrug of ampicillin typically prescribed for treating bacterial infections[41]. As identified by SPOKE[42], bacampicillin was found to downregulate the GDF15 gene and upregulate the NFKB2 (Nuclear Factor Kappa B Subunit 2) gene in studies by Cmap[28] and LINCS[43]. The GDF15 protein acts as a cytokine and is involved in stress response after cellular injury, and the NFKB2 is a central activator of genes involved with inflammation and immune function[44]. Circulating levels of GDF15 have been found to be significantly higher in COVID-19 patients who die[45]. Zhou et al.'s work revealed NF-kappa B signaling as one of the main pathways of coronavirus infections in humans. While the rapid conversion of bacampicillin to ampicillin in vivo makes this prodrug a less optimal therapeutic candidate for COVID-19, our findings nevertheless provide insights into the immunologic and inflammatory landscape from SARS-CoV-2 infection.

Overall, in testing of our drug hits across two human cell line assays, 11 of 16 exhibited inhibition of SARS-CoV-2 infection. In particular, three of our four consensus drug hits demonstrated antiviral efficacy, with haloperidol showing reproducible inhibition in Calu-3 cells, and bacampicillin and clofazimine inhibiting viral activity in 293T-ACE2

cells without cytotoxicity. Many of our tested drugs can be administered orally, and several are on the WHO Model List of Essential medications, including ciclosporin, clofazimine, and haloperidol[46]. These results suggest that our drug repositioning pipeline can rapidly identify readily available potential therapeutics in antiviral contexts.

There are several limitations of our approach that should be recognized. In general, drug repositioning pipelines are reliant upon data being encoded in a computable format. Considering variability of conditions under which experimental and clinical data are obtained, this also implies a particular set of limitations and biases to the ensuing results. Therefore, significant and concerted community efforts are necessary for efficient usage of the existing biomedical and clinical information and extraction of knowledge from this information, which may allow better repositioning of the current drugs[47].

In our work, we applied the KS-based similarity metric on the CMap database. Rank-based methods, such as the KS statistic, may suffer from high false positive rates, as genes not differentially expressed can be ranked high and contribute to the similarity measurement for drug and disease signatures[48,49]. Moreover, rank-based approaches can also miss many potential drugs as ranking captured just a small part of information in a gene expression profile. Alternative methods have bene proposed, such as EMUDRA[49] and XSum[50]; however, they have not been widely adopted by the community. Future work might include evaluating multiple similarity metrics on larger datasets[49]. The studies that we leveraged here are also limited because of their small sample size, which might explain the small gene overlap across the signatures. While the different sample types are able to capture the heterogeneity of the response to viral infection, because data generated from cell lines and organoids (the ALV and EXP signatures, respectively) might

not accurately represent the biological changes and responses in human infection, *in vivo* experiments are needed to better understand the biology of the disease and the effect of these drugs on it. Moreover, although the BALF signature was generated from fluid recovered from lavage of infected human tissues, this primary response data was aggregated from a very limited sample size (2 cases and 3 controls). As the sample types across the three datasets were very different, we used different fold change thresholds to identify the signature genes to be used for drug repurposing. Gathering samples from a larger number of patients should generate a more robust gene expression signature and better inform therapeutic predictions. Furthermore, the drug profiles from CMap were generated from cell line data; drug data generated from more relevant tissue cultures (e.g. lung tissue) may generate more appropriate comparisons. Finally, our validation approaches focus on in vitro studies, which are limited, and warrant further in-vivo testing of the proposed compounds.

The drug development response for SARS-CoV-2 / COVID-19 is rapidly developing. One drug, remdesivir, recently received FDA approval for the treatment of COVID-19, and numerous other drugs are being actively explored for possible therapeutic value in COVID-19 cases. Utilizing a diverse set of transcriptomic SARS-CoV-2 signatures, our drug repositioning pipeline identified 25 therapeutic candidates. Validation experiments revealed antiviral activity for 11 of 16 drug hits. Further clinical investigation into these drug hits, *in vivo* assays as well as potential combination therapies is warranted to further investigate both the anti-viral as well as side effect profile of the drugs.

Materials and Methods

Study design

We have previously developed and used a transcriptomics based bioinformatics approach for drug repositioning in various contexts including inflammatory bowel disease, dermatomyositis, and spontaneous preterm birth. For a list of differentially expressed genes, the computational pipeline compares the ranked differential expression of a disease signature with that of a profile[16,19,28]. A reversal score based on the Kolmogorov-Smirnov statistic is generated for each disease-drug pair, with the idea that if the drug profile significantly reverses the disease signature, then the drug could be potentially therapeutic for the disease.

SARS-CoV-2 gene expression signatures

Blanco-Melo et al. generated a differential gene expression signature using RNA-seq on human adenocarcinomic alveolar basal epithelial cells infected with SARS-CoV-2 propagated from Vero E6 cells (GSE147507, 67 samples)[24]. Due to the fast-moving nature of the research topic, we opted to use this cell line data in lieu of waiting for substantial patient-level data. This work identified 120 differentially expressed genes (DEGs) – 100 upregulated and 20 downregulated. We used these 120 genes as the ALV signature for our computational pipeline (Table S1).

Lamers et al. performed RNA-seq on their organoid samples, from which differentially expressed genes were calculated[26]. These samples were grown in a medium with a Wnt surrogate supplement and infected with SARS-CoV-2 propagated from Vero E6 cells (GSE149312, 22 samples). They detected 434 significant DEGs (FDR < 0.05). We additionally applied a fold-change cutoff ($|\log_2 FC| > 2$), resulting in 125 genes used as the EXP signature (Table S2).

Xiong et al. performed RNA-seq analysis of BALF samples from two COVID-19 patients (two samples per patient) and three healthy controls[27]. We processed their raw read counts in order to construct a differential signature (see below for details). FASTQ files were downloaded from the Genome Sequence Archive[51,52] under accession number CRA002390. Paired-end reads were mapped to the hg19 human reference genome using Salmon (v.1.2.0) and assigned Ensembl genes. After read quality control, we obtained quantifications for 55,640 genes in all samples. In order to identify genes differentially expressed between cases and controls for the BALF samples, we quantified gene expression as raw counts. Raw counts were used as inputs to DESeq2 (v.1.24.0 R package) to call differentially expressed genes (DEGs). After adjusting for the sequencing platform, the default settings of DESeq2 were used. Principal components were generated using the DESeq2 function (Figure S2), and heat maps were generated using the Bioconductor package pheatmap (v.1.0.12) using the rlog-transformed counts (Figure S3). Values shown are rlog-transformed and row-normalized. Volcano plots were generated using the Bioconductor package EnhancedVolcano (v.1.2.0) (Figure S4). Retaining only protein-coding genes and applying both a significance threshold and a fold-change cutoff (FDR < 0.05, $|\log_2 FC| > 4$), we obtained 1,349 genes to be used as the BALF signature (Table S3).

Pathway enrichment analysis

Functional enrichment gene-set analysis for GSEA (Gene Set Enrichment Analysis) was performed using fgsea (v.1.12.0 R package) and the input gene lists were ranked by log2 fold change. The 50 Hallmark Gene Sets used in the GSEA analysis were

downloaded from MSigDB Signatures database[29,53]. For GO (Gene Ontology) terms, identification of enriched biological themes was performed using the DAVID database[54].

Drug gene expression profiles

Drug gene expression profiles were sourced from Connectivity Map (CMap), a publicly-available database of drugs tested on cancer cell lines[28]. CMap contains a set of differential gene expression profiles generated from treating cultured human cells with a variety of different drugs and experimental compounds. These profiles were generated using DNA microarrays to assay mRNA expression. These drug profiles are ranked genome-wide profiles (~22,000 genes) of the effects of the drugs on various cell lines. 6,100 gene expression profiles are presented in CMap. A total of 1,309 compounds were tested in up to 5 different cell lines. The overlap between the gene lists of CMap and the SARS-CoV-2 signature is 109 genes.

Computational gene expression reversal scoring

To compute reversal scores, we used a non-parametric rank-based method similar to the Kolmogorov-Smirnov test statistic. This analysis was originally suggested by the creators of the CMap database and has since been implemented in a variety of different settings[16–19,22,28]. As also described by others, the drug signature is compared with the gene expression profiles. By splitting the gene signature into two lists containing only upregulated genes and downregulated genes, a so-called connectivity score is estimated via several auxiliary variables using a nonparametric rank-ordered Kolmogorov–Smirnov (KS) test)[48]. Similar to past works, we applied a pre-filtering step to the CMap profiles to

maintain only drug profiles which were significantly correlated with another profile of the same drug. Drugs were assigned reversal scores based on their ranked differential gene expression profile relative to the SARS-CoV-2 ranked differential gene expression signature. A negative reversal score indicated that the drug had a profile which reversed the SARS-CoV-2 signature; that is, up-regulated genes in the SARS-CoV-2 signature were down-regulated in the drug profile and vice versa.

Statistical analysis

P-values were adjusted using the false discovery rate (FDR; Benjamini-Hochberg) procedure. P-values for individual drug hits were obtained by comparing reversal scores to a distribution of random scores. Negative reversal scores were considered significant if they met the criterion FDR < 0.05. For drugs tested multiple times (e.g. different cell lines), we used the most reversed profile (lowest negative score). For significance values of the number of drugs reversing multiple signatures, we constructed distributions of the common reversal (reversing two of three signatures) and the consensus reversal (reversing three of three signatures) by randomly sampling the same number of drug profiles for each signature from CMap.

Single-cell data analysis

Quantification files were downloaded from GEO GSE145926. An individual Seurat object for each sample was generated using Seurat v.3. While the data has been filtered by 10x's algorithm, we still needed to ensure the remaining cells are clean and devoid of artifacts. We calculated three confounders for the dataset: mitochondrial percentage,

ribosomal percentage, and cell cycle state information. For each sample, cells were normalized for genes expressed per cell and per total expression, then multiplied by a scale factor of 10,000 and log-transformed. Low quality cells were excluded from our analyses— this was achieved by filtering out cells with greater than 5,000 and fewer than 300 genes and cells with high percentage of mitochondrial and ribosomal genes (greater than 10% for mitochondrial genes, and 50% for ribosomal genes). SCTransform is a relatively new technique that uses "Pearson Residuals" (PR) to normalize the data. PRs are independent of sequencing depth[55]. We "regress out" the effects of mitochondrial and ribosomal genes, and the cell cycling state of each cell, so they do not dominate the downstream signal used for clustering and differential expression. We then performed a lineage auto-update disabled r dimensional reduction (RunPCA function). Then, each sample was merged together into one Seurat object. Data were then re-normalized and dimensionality reduction and significant principal components were used for downstream graph-based, semi-unsupervised clustering into distinct populations (FindClusters function) and uniform manifold approximation and projection (UMAP) dimensionality reduction was used. For clustering, the resolution parameter was approximated based on the number of cells according to Seurat guidelines; a vector of resolution parameters was passed to the FindClusters function and the optimal resolution of 0.8 that established discernible clusters with distinct marker gene expression was selected. We obtained a total of 21 clusters representing the major immune and epithelial cell populations. To identify marker genes driving each cluster, the clusters were compared pairwise for differential gene expression (FindAllMarkers function) using the Likelihood ratio test assuming an underlying negative binomial distribution (negbinom). For visualization of

519 gene expression data between different samples a number of Seurat functions were used:
520 FeaturePlot, VlnPlot and DotPlot.
521
522 ***Cell Lines***
523 For studies at the Gladstone Institutes, Calu-3 cells, a human lung epithelial cell line
524 (American Type Culture Collection, ATCC HTB-55), were cultured in advanced MEM
525 supplemented with 2.5% fetal bovine serum (FBS) (Gibco, Life Technologies), 1% L-
526 GlutaMax (ThermoFisher), and 1% penicillin/streptomycin (Corning) at 37°C and 5% CO_2.
527 SARS-CoV-2 Isolate USA-WA1/2020 was purchased from BEI Resources and
528 propagated and titered in Vero E6 cells. For studies carried out at Mount Sinai, SARS-
529 CoV-2 was propagated in Vero E6 cells (ATCC CRL-1586) and 293T-ACE2 cells (ATCC
530 CRL-3216).
531
532 ***Compounds***
533 Selection of compounds for testing was based on side effect profiles and compound
534 availability. Bacampicillin (B0070000), ciclopirox (SML2011-50MG), ciclosporin
535 (C2163000), clofazimine (1138904-200MG), dicycloverine (D1060000), fludrocortisone
536 (1273003-200MG), fluticasone (1285873-100MG), haloperidol (H1512-5G), isoxicam
537 (I1762-1G), lansoprazole (1356916-150MG), metixene (M1808000), myricetin (M6760-
538 10MG), pentoxifylline (1508901-200MG), sirolimus (S-015-1ML), tretinoin (1674004-
539 5X30MG), and valproic acid (1708707-500MG) were purchased from Sigma-Aldrich.
540 Remdesivir (GS-5734) was purchased from Selleckchem.
541

Compounds were resuspended in DMSO according to manufacturer's instructions and serially diluted to the relevant concentrations for treatment of infected cells.

Infection Experiments

Work involving live SARS-CoV-2 was performed in the BSL3 facility at the Gladstone Institutes with appropriate approvals. Calu-3 cells were seeded in 96-well plates for 24h, infected with SARS-CoV-2 at a multiplicity of infection (MOI) of 0.05, and treated with compounds. 72 hours post infection, supernatant was collected for RNA extraction and the RNA was analyzed using RT-qPCR to quantify viral genomes present in the supernatant. SARS-CoV-2 specific primers targeting the E gene region: 5'-*ACAGGTACGTTAATAGTTAATAGCGT-3' (Forward)* and *5'-ATATTGCAGCAGTACGCACACA-3' (Reverse)* were used to quantify cDNA on the 7500 Fast Real-Time PCR system (Applied Biosystems). Cells were fixed with paraformaldehyde and used for immunofluorescence analysis with dsRNA antibody (SCICONS) and DAPI stain. Images were acquired and analyzed using ImageXpress Micro Confocal High-Content Imaging System.

In Vitro Microneutralization Assay for SARS-CoV-2 Serology and Drug Screening

For studies at Mount Sinai, SARS-CoV-2 was propagated in Vero E6 cells (ATCC CRL-1586) and 293T-ACE2 cells (ATCC CRL-3216), as previously described[12,34]. Two thousand cells were seeded into 96-well plates in DMEM (10% FBS) and incubated for 24 h at 37 °C,5% CO2. Then, 2 h before infection, the medium was replaced with 100 µl of DMEM (2% FBS) containing the compound of interest at concentrations 50% greater than those indicated, including a DMSO control. The Vero E6 cell line used in this study

is a kidney cell line; therefore, we cannot exclude that lung cells yield different results for some inhibitors. Plates were then transferred into the Biosafety Level 3 (BSL3) facility and 100 PFU (MOI = 0.025) was added in 50 µl of DMEM (2% FBS), bringing the final compound concentration to those indicated. Plates were then incubated for 48 h at 37 °C. After infection, supernatants were removed and cells were fixed with 4% formaldehyde for 24 h before being removed from the BSL3 facility. The cells were then immunostained for the viral NP protein (an in-house mAb 1C7, provided by Dr. Thomas Moran) with a DAPI counterstain. Infected cells (488 nM) and total cells (DAPI) were quantified using the Celigo (Nexcelcom) imaging cytometer. Infectivity is measured by the accumulation of viral NP protein in the nucleus of the Vero E6 cells and 293T-ACE2 cells (fluorescence accumulation). Percentage infection was quantified as ((infected cells/total cells) − background) × 100 and the DMSO control was then set to 100% infection for analysis. The IC50 and IC90 for each experiment were determined using the Prism (GraphPad) software. Cytotoxicity was also performed using the MTT assay (Roche), according to the manufacturer's instructions. Cytotoxicity was performed in uninfected VeroE6 cells with same compound dilutions and concurrent with viral replication assay. All assays were performed in biologically independent triplicates.

Code Availability Statement

The data used for the repositioning pipeline are all publicly available. The code for the drug repositioning pipeline was adapted from reference 19 and is available at https://github.com/brianlle/sirota_lab_covid_drug_repositioning.

Acknowledgements

This work was funded in part by the Program for Breakthrough Biomedical Research (PBBR) Grant. This research was also partly funded by the Defense Advanced Research Projects Agency (HR0011-19-2-0020); by CRIP (Center for Research for Influenza Pathogenesis), a NIAID supported Center of Excellence for Influenza Research and Surveillance (CEIRS, contract # HHSN272201400008C); by supplements to NIAID grant U19AI135972 and DoD grant W81XWH-20-1-0270; by the generous support of the JPB Foundation and the Open Philanthropy Project (research grant 2020-215611 (5384)); and by anonymous donors to A. García-Sastre. M. Ott acknowledges support through a gift from the Roddenberry Foundation and by NIH 5DP1DA038043. We would like to thank members of the Sirota Lab for useful discussion, Dr. Boris Oskotsky for technical support, Edna Rodas for administrative support, Randy Albrecht for support with the BSL3 facility and procedures at the ISMMS, and Richard Cadagan for technical assistance.

Author contributions

B.L., T.O. and M.S. designed and coordinated the study. B.L. led the drug repurposing efforts. G.A., K.Y., I.K., and C.L. helped with data analyses. SARS-CoV-2 virus assays were led by A.V.G., G.R.K., K.L., R.R., K.W., A.G.S., and M.O. All the authors contributed to making figures, writing and editing the manuscript.

Competing interests

M.S. is on the advisory board of twoXAR. The García-Sastre Laboratory has received research support from Pfizer, Senhwa Biosciences and 7Hills Pharma. A.G.S.

has consulting agreements for the following companies involving cash and/or stock: Vivaldi Biosciences, Contrafect, 7Hills Pharma, Avimex, Vaxalto, Accurius and Esperovax. Other authors declare no competing financial interests.

References

1. Ghisolfi S, Almås I, Sandefur JC, Carnap T von, Heitner J, Bold T. Predicted COVID-19 fatality rates based on age, sex, comorbidities and health system capacity. *BMJ Glob Health*. 2020;5(9):e003094. doi:10.1136/bmjgh-2020-003094
2. Gupta A, Madhavan MV, Sehgal K, et al. Extrapulmonary manifestations of COVID-19. *Nat Med*. 2020;26(7):1017-1032. doi:10.1038/s41591-020-0968-3
3. Zhang Q, Bastard P, Liu Z, et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science*. Published online September 24, 2020. doi:10.1126/science.abd4570
4. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ. COVID-19: consider cytokine storm syndromes and immunosuppression. *The Lancet*. 2020;395(10229):1033-1034. doi:10.1016/S0140-6736(20)30628-0
5. Cava C, Bertoli G, Castiglioni I. A protein interaction map identifies existing drugs targeting SARS-CoV-2. *BMC Pharmacol Toxicol*. 2020;21(1):65. doi:10.1186/s40360-020-00444-z
6. Cava C, Bertoli G, Castiglioni I. In Silico Discovery of Candidate Drugs against Covid-19. *Viruses*. 2020;12(4):404. doi:10.3390/v12040404
7. Beigel JH, Tomashek KM, Dodd LE, et al. Remdesivir for the Treatment of Covid-19 — Preliminary Report. *N Engl J Med*. 2020;0(0):null. doi:10.1056/NEJMoa2007764

8. Commissioner O of the. FDA Approves First Treatment for COVID-19. FDA. Published October 22, 2020. Accessed October 22, 2020. https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-covid-19
9. RECOVERY Collaborative Group, Horby P, Lim WS, et al. Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report. *N Engl J Med*. Published online July 17, 2020. doi:10.1056/NEJMoa2021436
10. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. *J Health Econ*. 2003;22(2):151-185. doi:10.1016/S0167-6296(02)00126-1
11. Aronson JK. Old drugs – new uses. *Br J Clin Pharmacol*. 2007;64(5):563-565. doi:10.1111/j.1365-2125.2007.03058.x
12. Gordon DE, Jang GM, Bouhaddou M, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. Published online April 30, 2020:1-13. doi:10.1038/s41586-020-2286-9
13. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov*. 2020;6:1268-1281. doi:10.1038/s41421-020-0153-3
14. Gordon DE, Hiatt J, Bouhaddou M, et al. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science*. Published online October 15, 2020. doi:10.1126/science.abe9403
15. Shah B, Modi P, Sagar SR. In silico studies on therapeutic agents for COVID-19: Drug repurposing approach. *Life Sci*. 2020;252:117652. doi:10.1016/j.lfs.2020.117652
16. Sirota M, Dudley JT, Kim J, et al. Discovery and Preclinical Validation of Drug Indications Using Compendia of Public Gene Expression Data. *Sci Transl Med*. 2011;3(96):96ra77. doi:10.1126/scitranslmed.3001318

17. Dudley JT, Sirota M, Shenoy M, et al. Computational Repositioning of the Anticonvulsant Topiramate for Inflammatory Bowel Disease. *Sci Transl Med*. 2011;3(96):96ra76. doi:10.1126/scitranslmed.3002648

18. Cho HG, Fiorentino D, Lewis M, Sirota M, Sarin KY. Identification of alpha-adrenergic agonists as potential therapeutic agents for dermatomyositis through drug-repurposing using public expression datasets. *J Invest Dermatol*. 2016;136(7):1517-1520. doi:10.1016/j.jid.2016.03.001

19. Chen B, Wei W, Ma L, 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*. 2017;152(8):2022-2036. doi:10.1053/j.gastro.2017.02.039

20. Chen B, Ma L, Paik H, et al. Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. *Nat Commun*. 2017;8:16022. doi:10.1038/ncomms16022

21. Jahchan NS, Dudley JT, Mazur PK, et al. A Drug Repositioning Approach Identifies Tricyclic Antidepressants as Inhibitors of Small Cell Lung Cancer and Other Neuroendocrine Tumors. *Cancer Discov*. 2013;3(12):1364-1377. doi:10.1158/2159-8290.CD-13-0183

22. Le BL, Iwatani S, Wong RJ, Stevenson DK, Sirota M. Computational discovery of therapeutic candidates for preventing preterm birth. *JCI Insight*. 5(3). doi:10.1172/jci.insight.133761

23. Xing J, Shankar R, Drelich A, et al. Reversal of Infected Host Gene Expression Identifies Repurposed Drug Candidates for COVID-19. *bioRxiv*. Published online April 9, 2020:2020.04.07.030734. doi:10.1101/2020.04.07.030734

24. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*. Published online May 13, 2020. doi:10.1016/j.cell.2020.04.026

685 25. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 Cell Entry Depends on
686 ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*.
687 2020;181(2):271-280.e8. doi:10.1016/j.cell.2020.02.052
688 26. Lamers MM, Beumer J, Vaart J van der, et al. SARS-CoV-2 productively infects human gut
689 enterocytes. *Science*. 2020;369(6499):50-54. doi:10.1126/science.abc1669
690 27. Xiong Y, Liu Y, Cao L, et al. Transcriptomic characteristics of bronchoalveolar lavage fluid
691 and peripheral blood mononuclear cells in COVID-19 patients. *Emerg Microbes Infect*.
692 2020;9(1):761-770. doi:10.1080/22221751.2020.1747363
693 28. Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene-expression
694 signatures to connect small molecules, genes, and disease. *Science*.
695 2006;313(5795):1929-1935. doi:10.1126/science.1132939
696 29. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-
697 based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S*
698 *A*. 2005;102(43):15545-15550. doi:10.1073/pnas.0506580102
699 30. Mootha VK, Lindgren CM, Eriksson K-F, et al. PGC-1α-responsive genes involved in
700 oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*.
701 2003;34(3):267-273. doi:10.1038/ng1180
702 31. Liao M, Liu Y, Yuan J, et al. Single-cell landscape of bronchoalveolar immune cells in
703 patients with COVID-19. *Nat Med*. 2020;26(6):842-844. doi:10.1038/s41591-020-0901-9
704 32. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank
705 database for 2018. *Nucleic Acids Res*. 2018;46(D1):D1074-D1082.
706 doi:10.1093/nar/gkx1037
707 33. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein
708 pharmacology by ligand chemistry. *Nat Biotechnol*. 2007;25(2):197-206.
709 doi:10.1038/nbt1284

34. Amanat F, White KM, Miorin L, et al. An In Vitro Microneutralization Assay for SARS-CoV-2 Serology and Drug Screening. *Curr Protoc Microbiol*. 2020;58(1):e108. doi:10.1002/cpmc.108

35. Seeman P, Kapur S. Schizophrenia: more dopamine, more D2 receptors. *Proc Natl Acad Sci U S A*. 2000;97(14):7673-7675. doi:10.1073/pnas.97.14.7673

36. Hafidh RR, Abdulamir AS, Abu Bakar F, Sekawi Z, Jahansheri F, Jalilian FA. Novel antiviral activity of mung bean sprouts against respiratory syncytial virus and herpes simplex virus −1: an in vitro study on virally infected Vero and MRC-5 cell lines. *BMC Complement Altern Med*. 2015;15. doi:10.1186/s12906-015-0688-2

37. Berezin V, Abdukhakimova D, Trenozhnikova L, et al. Antiviral activities of extremophilic actinomycetes extracts from Kazakhstan's unique ecosystems against influenza viruses and paramyxoviruses. *Virol J*. 2019;16. doi:10.1186/s12985-019-1254-1

38. PubChem. Clofazimine. Accessed May 21, 2020. https://pubchem.ncbi.nlm.nih.gov/compound/2794

39. Cholo MC, Steel HC, Fourie PB, Germishuizen WA, Anderson R. Clofazimine: current status and future prospects. *J Antimicrob Chemother*. 2012;67(2):290-298. doi:10.1093/jac/dkr444

40. Yuan S, Yin X, Meng X, et al. Clofazimine is a broad-spectrum coronavirus inhibitor that antagonizes SARS-CoV-2 replication in primary human cell culture and hamsters. *Res Sq*. Published online October 7, 2020. doi:10.21203/rs.3.rs-86169/v1

41. PubChem. Bacampicillin. Accessed May 21, 2020. https://pubchem.ncbi.nlm.nih.gov/compound/441397

42. Himmelstein DS, Lizee A, Hessler C, et al. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. Valencia A, ed. *eLife*. 2017;6:e26726. doi:10.7554/eLife.26726

43. Subramanian A, Narayan R, Corsello SM, et al. A Next Generation Connectivity Map: L1000 platform and the first 1,000,000 profiles. *Cell*. 2017;171(6):1437-1452.e17. doi:10.1016/j.cell.2017.10.049

44. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res*. 2005;33(Database Issue):D54-D58. doi:10.1093/nar/gki031

45. Luis García de Guadiana Romualdo, Mulero MDR, Olivo MH, et al. Circulating levels of GDF-15 and calprotectin for prediction of in-hospital mortality in COVID-19 patients: A case series. *J Infect*. Published online August 12, 2020. doi:10.1016/j.jinf.2020.08.010

46. WHO | WHO Model Lists of Essential Medicines. WHO. Accessed October 19, 2020. http://www.who.int/medicines/publications/essentialmedicines/en/

47. Poroikov V, Druzhilovskiy D. Chapter 1 - Drug Repositioning: New Opportunities for Older Drugs. In: Roy K, ed. *In Silico Drug Design*. Academic Press; 2019:3-17. doi:10.1016/B978-0-12-816125-8.00001-8

48. Musa A, Ghoraie LS, Zhang S-D, et al. A review of connectivity map and computational approaches in pharmacogenomics. *Brief Bioinform*. 2018;19(3):506-523. doi:10.1093/bib/bbw112

49. Zhou X, Wang M, Katsyv I, Irie H, Zhang B. EMUDRA: Ensemble of Multiple Drug Repositioning Approaches to improve prediction accuracy. *Bioinformatics*. 2018;34(18):3151-3159. doi:10.1093/bioinformatics/bty325

50. Cheng J, Yang L, Kumar V, Agarwal P. Systematic evaluation of connectivity map for disease indications. *Genome Med*. 2014;6:95. doi:10.1186/s13073-014-0095-1

51. National Genomics Data Center Members and Partners. Database Resources of the National Genomics Data Center in 2020. *Nucleic Acids Res*. 2020;48(D1):D24-D33. doi:10.1093/nar/gkz913

52. Wang Y, Song F, Zhu J, et al. GSA: Genome Sequence Archive<sup/>. *Genomics Proteomics Bioinformatics*. 2017;15(1):14-18. doi:10.1016/j.gpb.2017.01.001

53. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-1740. doi:10.1093/bioinformatics/btr260

54. Dennis G, Sherman BT, Hosack DA, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*. 2003;4(9):R60.

55. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*. 2019;20. doi:10.1186/s13059-019-1874-1

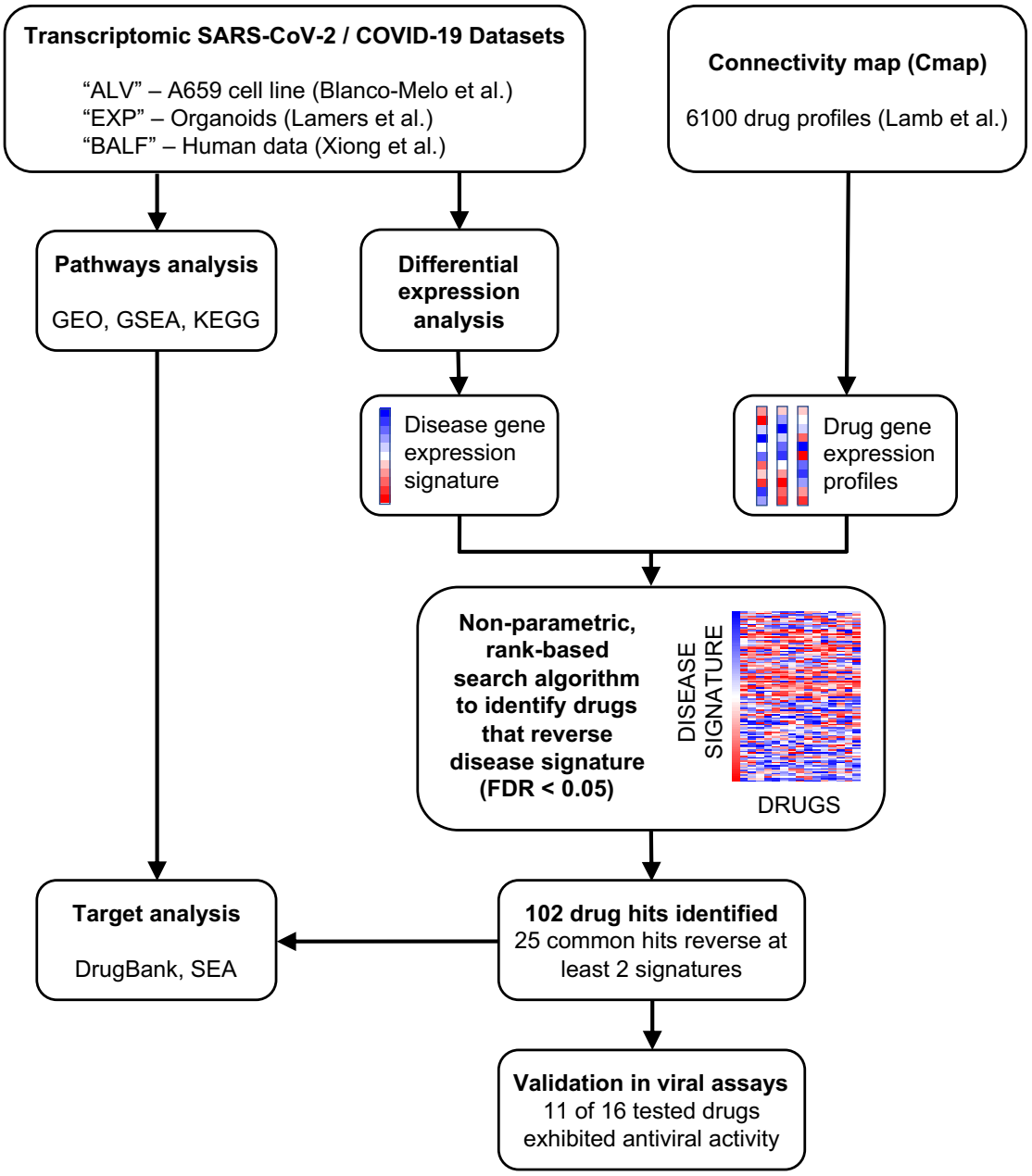


Figure 1. **COVID-19 transcriptomics-based bioinformatics approach for drug repositioning.** We generated lists of statistically significant differentially expressed genes from the analysis of three published studies of SARS-CoV-2 and COVID-19. The drug repositioning computational pipeline compares the ranked differential expression of the COVID-19 disease signature with that of drug profiles from CMap. A reversal score based on the Kolmogorov-Smirnov statistic is generated for each disease-drug pair. If a drug profile significantly (FDR < 0.05) reverses the disease signature, then the drug could be therapeutic for the disease. Across all datasets, a total of 102 drugs have been identified as potentially therapeutic for COVID-19. Twenty-five drugs were identified in analyses of at least two of the three datasets. We further conducted pathways analyses and targeted analyses on the results, focusing on the 25 shared hits. Finally, we validated sixteen of our top predicted hits in live SARS-CoV-2 antiviral assays.

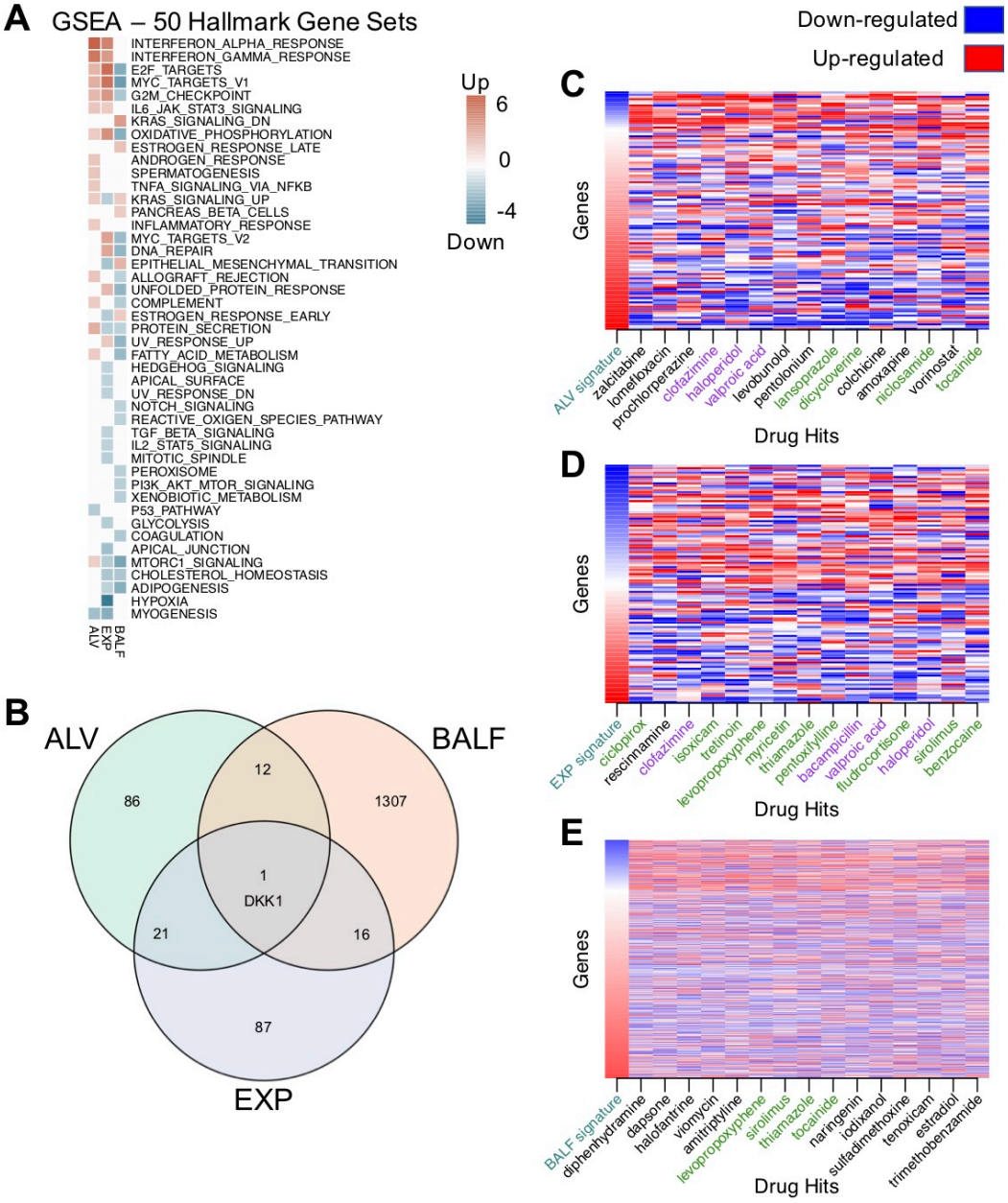


Figure 2. **SARS-CoV-2 differential gene expression signatures reversed by drug profiles from CMap. (A)** Enrichment analysis using GSEA reveals common pathways among input signatures. **(B)** DEG overlap from input signatures. Only 1 gene, DKK1, was shared by all 3 signatures. **(C)** Top 15 drug profiles reversing the ALV signature (109 genes). For each column, the gene expression values were ranked, with rank 1 being the most up-regulated gene (in red) and the maximum rank (109 for ALV) being the most down-regulated gene (in blue). Drug names highlighted in green were hits for a second signature, and drug hits highlighted in purple reversed all three signatures. **(D)** Top 15 drug profiles reversing the EXP signature (108 genes). **(E)** Top 15 drug profiles reversing the BALF signature (941 genes).

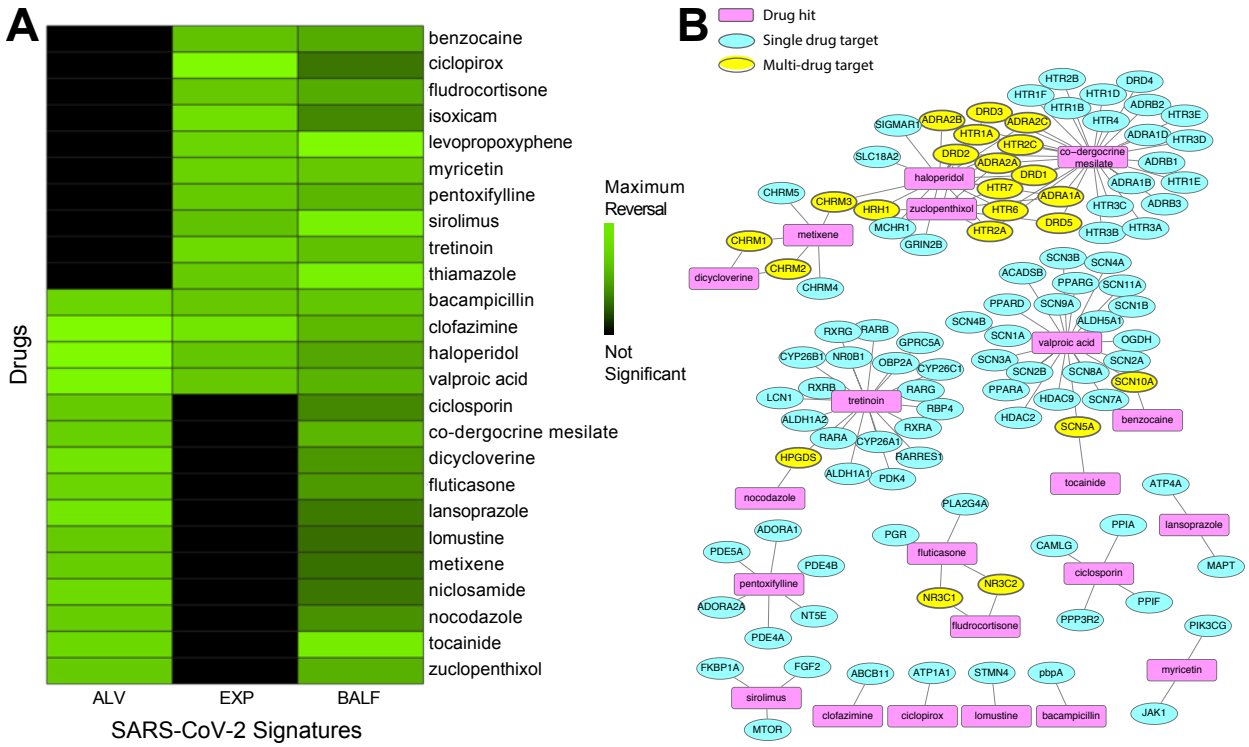
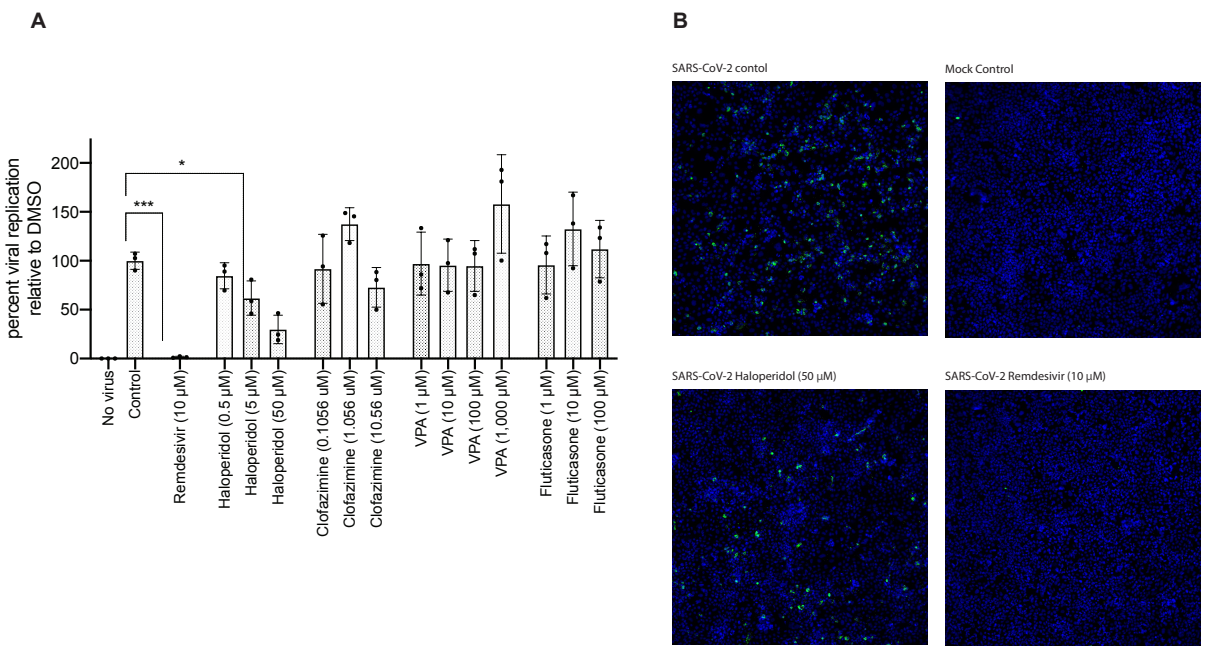
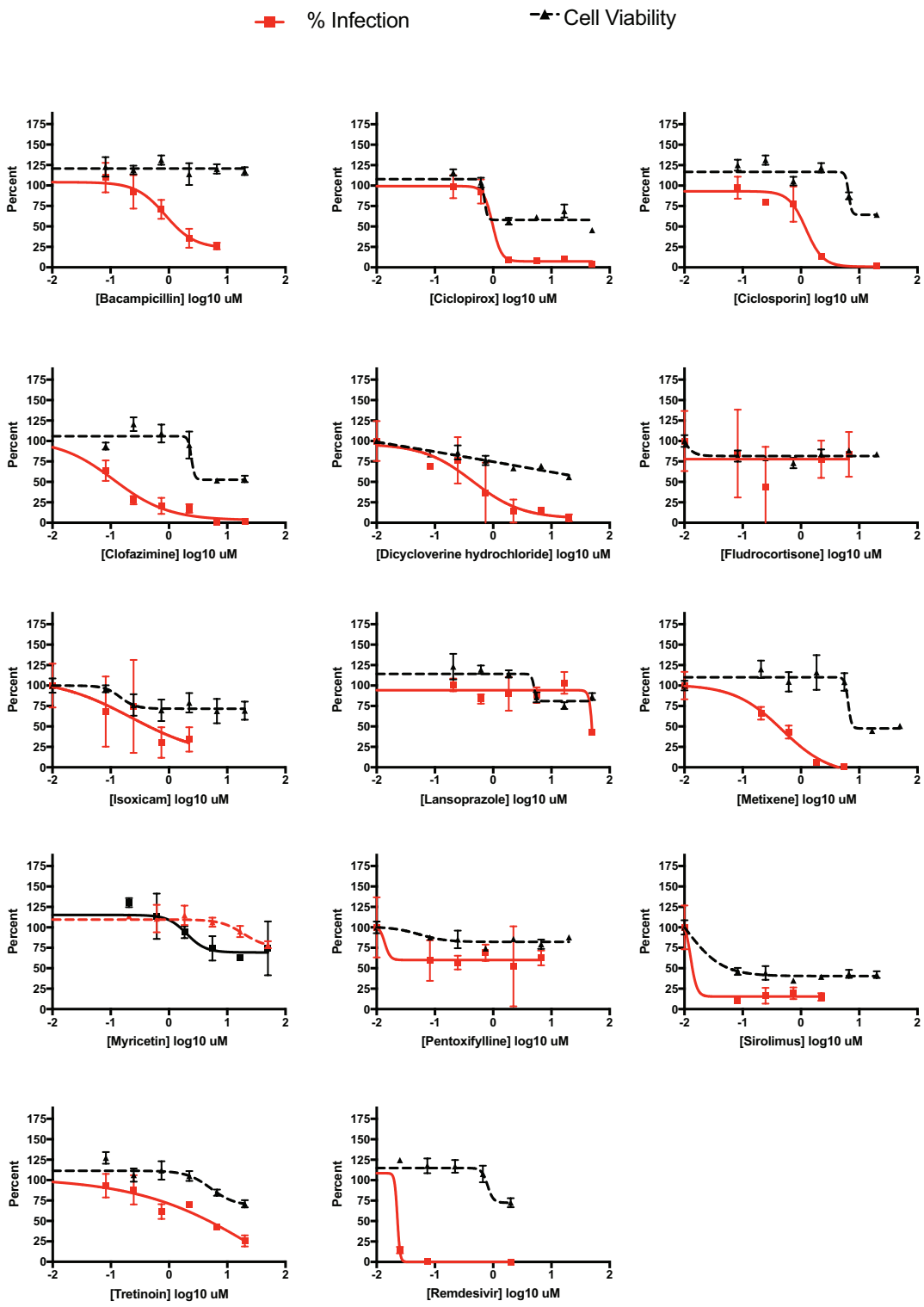


Figure 3. **Common therapeutic hits from drug repurposing pipeline applied to SARS-CoV-2 signatures. (A)** Drug profiles from CMap significantly reversed signatures from the ALV, BALF, and EXP signatures. 25 of the drugs were significant in at least 2 of the signatures. **(B)** Drug-protein target network. For the 25 drugs that reversed at least 2 of the signatures, target information was gathered from DrugBank to identify clusters of drugs from shared targets.



806

807 Figure 4. **Haloperidol inhibits viral replication of SARS-CoV-2 in the Calu-3 lung cell line.**
808 **(A)** Calu-3 cells were infected with SARS-CoV-2 at an MOI of 0.05 for 72h. Viral replication
809 levels were determined by RT-qPCR from supernatant RNA using specific primers for the E
810 gene. Viral RNA levels relative to DMSO are graphed. Error bars represent 3 or 4 independent
811 experiments. One-way ANOVA analysis was used to determine significance. **(B)** Microscopy:
812 Calu-3 cells were infected with SARS-CoV-2 at an MOI of 0.05 for 72h. Cells were fixed with
813 paraformaldehyde and used for immunofluorescence analysis with dsRNA antibody (SCICONS)
814 and DAPI stain. Images were acquired and analyzed using ImageXpress Micro Confocal High-
815 Content Imaging System.
816



817

818 Figure 5. **Viral inhibition and cell viability tests of 13 compounds in 293T-ACE2 cell**
819 **assays.** Several drugs inhibit viral infectivity. Red, viral infectivity (anti-NP); black, cell viability.
820 The lack of a dose response in cell viability probably reflects cytostatic and not cytotoxic effects.
821 Data are mean ± s.d.; *n* = 3 biologically independent samples for cell viability data.

Drug hit	Description (current uses)	ALV Reversal Score	EXP Reversal Score	BALF Reversal Score
Bacampicillin	Antibiotic	0.789	0.790	0.596
Benzocaine	Anesthetic	n.s.	0.766	0.546
Ciclopirox	Antifungal	n.s.	1	0.361
Ciclosporin	Immunosuppressant (RA, psoriasis, Crohn's)	0.756	n.s.	0.409
Clofazimine	Antimycobacterial (leprosy)	0.946	0.893	0.558
Co-dergocrine mesilate	Ergoid mesylate (dementia, Alzheimer's, stroke)	0.775	n.s.	0.553
Dicycloverine	Antispasmodic (IBS)	0.847	n.s.	0..461
Fludrocortisone	Corticosteroid	n.s.	0.782	0.519
Fluticasone	Steroid (asthma, COPD)	0.790	n.s.	0.463
Haloperidol	Antipsychotic (schizophrenia)	0.937	0.773	0.507
Isoxicam	NSAID	n.s.	0.873	0.410
Lansoprazole	Proton-pump inhibitor (acid reflux)	0.856	n.s.	0.370
Levopropoxyphene	Antitussive	n.s.	0.835	0.770
Lomustine	Antineoplastic (Hodgkin's disease, brain tumors)	0.748	n.s.	0.338
Metixene	Anticholinergic (Parkinson's)	0.759	n.s.	0.344
Myricetin	Flavonoid	n.s.	0.823	0.603
Niclosamide	Anthelmintic (tapeworms)	0.812	n.s.	0.360
Nocodazole	Antineoplastic	0.766	n.s.	0.439
Pentoxifylline	Vasodilatory and anti-inflammatory (claudication)	n.s.	0.791	0.552
Sirolimus	Immunosuppressive	n..s.	0.768	0.729

Thiamazole	Antithyroid agent (Graves disease)	n.s.	0.796	0.724
Tocainide	Antiarrhythmic	0.798	n.s.	0.714
Tretinoin	Vitamin A derivative (acne, acute promyelocytic leukemia)	n.s.	0.854	0.579
Valproic acid	Anticonvulsant (seizures, bipolar disorder)	0.917	0.786	0.546
Zuclopenthixol	Antipsychotic (schizophrenia)	0.754	n.s.	0.535

822 Table 1. Therapeutic hits reversing at least 2 of input SARS-CoV-2 signatures. A wide
823 variety of drugs were identified by the analysis of multiple signatures. Drug reversal
824 scores are normalized for each signature; drug entries marked “n.s.” were not significant
825 for reversing that signature.
826