A connectivity map-based drug repurposing study and integrative analysis of transcriptomic profiling of SARS-CoV-2 infection

Seyedeh Zahra Mousavi\textsuperscript{a}, Mojdeh Rahmanian\textsuperscript{b}, Ashkan Sami\textsuperscript{b}

\textsuperscript{a} Shiraz University of Medical Sciences, Shiraz, Iran
\textsuperscript{b} Shiraz University, Shiraz, Iran

Corresponding author

\textsuperscript{*}Ashkan Sami. Department of Computer Science and Engineering and IT, Shiraz University, Shiraz, Iran.

Email: sami@shirazu.ac.ir

Abstract

Aims: The recent outbreak of COVID-19 has become a global health concern. There are currently no effective treatment strategies and vaccines for the treatment or prevention of this fatal disease. The current study aims to determine promising treatment options for the COVID-19 through a computational drug repurposing approach.

Materials and methods: In this study, we focus on differentially expressed genes (DEGs), detected in SARS-CoV-2 infected cell lines including “the primary human lung epithelial cell line NHBE” and “the transformed lung alveolar cell line A549”. Next, the identified DEGs are used in the connectivity map (CMap) analysis to identify similarly acting therapeutic candidates. Furthermore, to interpret lists of DEGs, pathway enrichment and protein network analysis are performed. Genes are categorized into easily interpretable
pathways based on their biological functions, and overrepresentations of each pathway are tested in comparison to what is expected randomly.

**Key findings:** The results suggest the effectiveness of Saquinavir, lansoprazole, folic acid, ebselen, aminocaproic acid, simvastatin, surfactant stimulant drugs, heat shock protein 90 (HSP90) inhibitors, histone deacetylase (HDAC) inhibitors, metronidazole, inhaled corticosteroids (ICS) and many other clinically approved drugs and investigational compounds as potent drugs against COVID-19 outbreak.

**Significance:** Making new drugs remain a lengthy process, so the drug repurposing approach provides an insight into the therapeutics that might be helpful in this pandemic. In this study, pathway enrichment and protein network analysis are also performed, and the effectiveness of some drugs obtained from the CMap analysis has been investigated according to previous research.

Keywords: COVID-19; SARS-CoV-2; Drug Repurposing; Connectivity Map; Transcriptomic Profiling.

1. **Introduction**

Coronavirus disease 2019 (COVID-19) is an emerging global health concern. It initially appeared in December 2019 in China with cases of unknown origin pneumonia. Subsequently, a novel betacoronavirus was isolated from the throat sample of a patient, furtherly named 2019 novel coronavirus (2019-nCoV) and formally Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) with 79 and 52 percent sequence similarity with Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Middle East Respiratory Syndrome coronavirus (MERS-CoV), two other members of betacoronavirus family and leading causes of global outbreaks in past two decades, respectively [1-3].
As of 9 June 2020, there have been more than 7 million definite cases of COVID-19 with approximately 404,000 deaths worldwide [4]. Although clinical trials have been launched to determine potential treatments, no specific therapeutic agent is available at present. Considering the urgency of identifying effective treatments for COVID-19 outbreak, repositioning of available approved drugs for other indications, is an alternative to de novo drug development. Different approaches are available for drug repurposing and one of the most commonly used strategies is signature matching. It is based on comparing the transcriptomic signature of a specific disease with the unique signature of a drug resulting in the creation of a connectivity map (CMap) [5].

The CMap concept was initially presented by Lamb et al. [6, 7] in 2006. They generated a large-scale genomic signatures database to translate gene functions, diseases, and drug actions into the same language. It begins with identifying gene expression signatures of a specific disease that contains a list of differentially expressed genes (DEGs), genes that are significantly up or down-regulated in a specific disorder in comparison to its control sample. This signature is then compared with a collection of gene expression signatures of cell lines treated with various small molecules. The result of signature matching for each pair is shown with a connectivity score ranging from -1 to +1. A negative connectivity score reflects the potential effect of that small molecule to reverse the signature of a specific disorder. Therefore, it can be used as a therapeutic option.

Since the CMap introduction, this method has been applied in pharmacological research to define drug-disease connections [8, 9]. In 2016, So et al. [10] used the CMap concept to identify available approved drugs for seven common psychiatric disorders associated with promising results. Several other researchers have applied this approach to find potential novel therapies for different cancers (e.g., small cell lung carcinoma, gastric cancer, etc.) as well as rare disorders such as Hirschsprung disorder (HD) [11-13].
In this study, we set out to use a drug repurposing approach based on the CMap concept to identify possible treatments for COVID-19. Furthermore, to help interpret lists of DEGs, pathway enrichment and protein network analysis are performed. Genes are categorized into easily interpretable pathways based on their biological functions and overrepresentation of each pathway in the selected gene list is tested in comparison to what is expected randomly.

2. Material and Methods

2.1. Dataset selection and Analysis

The NCBI (National Center for Biotechnology Information) Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) was used to identify microarray datasets [14]. The GEO database was manually searched using the terms “SARS-CoV-2” or “COVID-19”. The collected datasets were further selected if they met the following inclusion criteria: (a) whole-genome transcriptome profiling, and (b) species of origin were “Homo sapiens”. Finally, GSE147507 [15] dataset was selected.

Briefly, the GSE147507 dataset included independent biological triplicates of “primary human lung epithelial cell line NHBE”, “transformed lung alveolar cell line A549” and some other cell lines which were mock-treated or infected with SARS-CoV-2 for 24hrs. For the generation of this dataset, total RNA from infected and mock-treated cells were extracted. Subsequently, mRNA enriched libraries were prepared from total RNA using TruSeq Stranded mRNA LP. cDNA libraries were sequenced using an Illumina NextSeq 500 platform. Raw sequencing reads were then aligned to the human genome (hg19) using the RNA-Seq Alignment App on Basespace (Illumina, CA, USA). For the differential expression
analysis of the GSE147507 dataset, DESeq2 was used. Genes with an adjusted p value< .05 were identified as DEGs and selected for further analysis [14].

2.2. CMap Analysis

The basic idea of this work was taken from Tinka Vidovik [16]. She conducted a CMap-based drug repurposing approach on available icSARS-CoV data and suggested valproic acid as a therapeutic option. However, considering the genomic diversity of SARS-CoV-2 from icSARS-CoV and unique features of COVID-19 infection makes the study results less reliable. The CMap is a collection of gene-expression profiles of drug-treated human cancer cells, which has been widely used for investigation of polypharmacology and drug repurposing [6, 7]. The current version (build 02; https://portals.broadinstitute.org/cmap/) of CMap collects more than 7000 gene-expression profiles representing 1309 compounds. Because most CMap compounds are the United States Food and Drug Administration (FDA)-approved drugs, this database has become a powerful tool for drug repurposing. The CMap provides a useful tool for screening associations between compounds and diseases and identifying highly correlated gene expression patterns [17]. In this study, PharmacoGx [18], an R package for analysis of large pharmacogenomic datasets, was used. Gene-expression signatures of SARS-CoV-2 infected cell lines, lists of significantly up and downregulated genes, resulting from the GSE147507 dataset were used as the input data and compared with gene-expression signatures of available CMap drugs. Connectivity scores were calculated for each drug-disease pair to investigate the correlation between the drug and the COVID-19 signature. Drugs with significant negative connectivity scores can potentially reverse the gene-expression profile of SARS-CoV-2 infection and can be used as therapeutic options. Since the publication of transcriptomic profiling of SARS-CoV-2, some researchers have analyzed this data to better understand SARS-CoV-2 pathogenesis, host responses, and
evaluate the efficacy of possible treatments. Fagone et al. [19] have analyzed and compared this data with the transcriptome profile of SARS-CoV from the 2003 pandemic as well as predicting potential drugs. They proposed Mitogen-activated protein kinase (MEK), serine-threonine kinase (AKT), mammalian target of rapamycin (mTOR), and I kappa B Kinase (IKK) inhibitors as candidate drugs.

Taguchi et al. [20] have applied the tensor decomposition (TD)-based unsupervised feature extraction (FE) method to gene expression profiles of COVID-19 infected cells summarized in this data to identify important genes and furtherly identify drug candidates that altered the expression of the genes selected by TD-based unsupervised FE.

However, our work differs from these studies as we have used a different tool, PharmacoGX, for our CMap analysis leading to propose many different novel potential drugs mentioned in the following. PharmacoGx, an R package, represented a unified framework for obtaining and analyzing large-scale pharmacogenomic datasets [18]. Since the introduction of the CMap concept, perturbation datasets such as CMap and LINCS L1000 were created. However, lacking standard frameworks for annotation, storage, access, and analysis challenged the use of these datasets. Therefore, unifying integrative platforms could remove biases of different sources, profiling platforms, and cell-specific differences. PharmacoGx is one of the few or even only integrative platforms developed for this reason. It stores essential data provided by pharmacogenomics datasets and removes the biases and performs CMap analysis [9].

The pre-processed CMap data (build 2), available in the PharmacoGX package, was used for further analysis. This version of the CMap dataset consists of 1309 compounds and 7000 gene expression profiles. Results of previous similar studies have shown the effectiveness of some small investigational molecules [19]. However, considering the urgency of identifying COVID-19 treatment, already available approved drugs should be repurposed for this
pandemic. Considering this, the second version of the CMap dataset (build 2) was selected since available approved drugs account for a higher percentage of this dataset than investigational molecules, compatible with our results.

2.3. **Enrichment analysis**

Pathway enrichment analysis was performed with the Metascape web tool (http://metascape.org) to determine whether the identified DEGs are enriched for specific biological functions. Metascape is a web-based portal integrating over 40 bioinformatic knowledgebases within one portal, implementing an automated analysis workflow including conversion, gene annotation, membership, and enrichment analysis [21]. Pathway enrichment analysis was carried out for upregulated and downregulated genes of NHBE and A549 cell lines separately using Metascape with the following ontology sources: KEGG Pathway, GO Biological Processes, Reactome Gene Sets, Canonical Pathways and CORUM.

2.4. **Protein-protein interaction**

The protein-protein interaction (PPI) network was constructed, visualized, and analyzed with the Metascape tool. Analysing DEGs in the context of protein interaction networks provide a framework for a better understanding of transcriptome and functional mechanisms underlying disease. For each gene list, the PPI network was constructed using Metascape with the following protein interaction databases: BioGrid, InWeb.IM, and OmniPath.
3. Results

3.1. Drug prediction analysis using CMap

We aimed to study the genes differentially expressed upon SARS-CoV-2 infection, in NHBE cells and A549 cells separately. First, all genes with adjusted p value < .05, irrespective of the fold-change were considered as DEGs. For NHBE, a total of 548 genes were found to be modulated by SARS-CoV-2 (378 upregulated and 170 downregulated genes) and for A549, a total of 119 genes were found to be modulated by SARS-CoV-2 (100 upregulated and 19 downregulated genes) (suppl file 1). Then, a more stringent selection of the DEGs was applied (i.e., adj. p value < .05 and |fold-change| > 2). Applying these two criteria, a set of 72 genes (64 genes were upregulated and 8 genes were downregulated) from NHBE cells, and 46 genes (44 genes were upregulated and 2 genes were downregulated) from A549 cells were identified that showed evidence for differential expression. (suppl file 2). These identified lists of DEGs were further analyzed through CMap study separately.

For the CMap analysis, PharmacoGx was used. Lists of DEGs identified in the previous step were applied as the input data and compared with CMap compounds library. Compounds with p-value < .05 and negative connectivity scores were selected and analyzed. First, based on genes with adjusted p-value < .05, 57 and 25 compounds were shown to significantly reverse genomic signatures of SARS-CoV-2 infected NHBE and A549 cells, respectively (Table 1). After applying a more stringent selection of DEGs, results were limited to 8 and 6 effective compounds against infected NHBE and A549 cells, respectively (Table 2).
Table 1. Top-scoring CMap study results based on DEGs with adjusted p-value<0.05. Compounds are ranked based on connectivity score, ranging from -1 (indicating the exactly reverse correlation of signatures between the compound and the disease) to +1.

| Rank | Compound       | Connectivity score | P-value   | Rank | Compound       | Connectivity score | P-value   |
|------|----------------|--------------------|-----------|------|----------------|--------------------|-----------|
| 1    | LY-294002      | -0.39943           | 0.00802929| 1    | STOCK1N-35696  | -0.89139           | 0.011560885 |
| 2    | Prestwick-1080 | -0.369915          | 0.00333631| 2    | 5182598        | -0.64096           | 0.003296745 |
| 3    | SC-560         | -0.351215          | 0.032486753| 3    | saquinavir     | -0.56532           | 0.007435467 |
| 4    | molindone      | -0.346195          | 0.003667894| 4    | lomustine      | -0.55598           | 0.019594195 |
| 5    | benzethonium chloride monorden | -0.344755 | 0.038870062 | 5    | ebselen        | -0.50165           | 0.027388767 |
| 6    | 0317956-0000   | -0.33464           | 0.013722457| 6    | 2,6-dimethylpiperidine lansoprazole | -0.48938 | 0.029438883 |
| 7    | 0198306-0000   | -0.33455           | 0.010090282| 7    | lansoprazole   | -0.47769           | 0.013600831 |
| 8    | sodium phenylbutyrate cinchonine | -0.32522 | 0.019984598 | 8    | aminocaproic acid | -0.47111          | 0.012364051 |
| 9    | alpha-ergocryptine ambroxol | -0.31816 | 0.013801023 | 9    | propylthiouracil | -0.46929           | 0.020724117 |
| 10   | oxolinic acid  | -0.31205           | 0.003082633| 10   | 6-azathymine   | -0.46812           | 0.012742531 |
| 11   | Budesonide     | -0.31004           | 0.003831323| 11   | iodixanol      | -0.4673            | 0.014734297 |
| 12   | Benzamid      | -0.307535          | 0.003182858| 12   | procaine       | -0.46708           | 0.021523513 |
| 13   | metronidazole  | -0.30554           | 0.003302104| 13   | propantheline bromide | -0.45344          | 0.012814118 |
| 14   | Pretanide      | -0.299785          | 0.016762662| 14   | canavanine     | -0.45256           | 0.013249204 |
| 15   | phenelzine     | -0.29952           | 0.006201045| 15   | simvastatin    | -0.43633           | 0.027487493 |
| 16   | Prestwick-864  | -0.297415          | 0.006801292| 16   | arecoline      | -0.43464           | 0.036092982 |
| 17   | Levobunolol    | -0.294205          | 0.015858995| 17   | vitexin        | -0.43402           | 0.040829469 |
| 18   | hyoscyamine    | -0.293515          | 0.006778838| 18   | trimethadione  | -0.4314            | 0.023400373 |
| 19   | Meptazinol     | -0.2913            | 0.01690587 | 19   | halcinonide    | -0.42312           | 0.038250087 |
| 20   | folic acid     | -0.61697           | 0.0151563  | 20   | atovaquone     | -0.42082           | 0.034635937 |

Table 2. Top-scoring CMap study results based on DEGs with adjusted p-value<0.05 and fold change≥2. Compounds are ranked based on connectivity score, ranging from -1 (indicating the exactly reverse correlation of signatures between the compound and the disease) to +1.

| Rank | Compound       | Connectivity score | P-value   | Rank | Compound       | Connectivity score | P-value   |
|------|----------------|--------------------|-----------|------|----------------|--------------------|-----------|
| 1    | Lansoprazole   | -0.61956           | 0.030081  | 1    | lansoprazole   | -0.61956           | 0.016883  |
| 2    | folic acid     | -0.61697           | 0.0151563 | 2    | Folic acid     | -0.61697           | 0.024286  |
| 3    | sulfonamethoxine | -0.61222          | 0.008615  | 3    | sulfonamethoxine | -0.61222         | 0.026554  |
| 4    | Tolnaftate     | -0.5996            | 0.0289097 | 4    | tolnaftate     | -0.5996           | 0.032699  |
| 5    | diclofenamide  | -0.59354           | 0.042645  | 5    | halcinonide    | -0.56766         | 0.037322  |
| 6    | metronidazole  | -0.54316           | 0.039469  | 6    | benzocaine     | -0.49907         | 0.043365  |
3.2. *Enrichment analysis*

To analyze the biological process of identified DEGs in SARS-CoV-2 infected cells, the Metascape tool was used. Pathway enrichment analysis was carried out for each gene list. Terms with a p-value < 0.01, a minimum count of 3, and an enrichment factor > 1.5 were grouped into clusters based on their membership similarities. Each cluster was represented by the most statistically significant term. Figures 1 and 2 show top clusters with their representative enriched terms for each gene list.

Interestingly, upregulated genes in SARS-CoV-2 infected NHBE cells were mainly enriched in “cytokine-mediate signaling pathway”, “response to virus” and “regulation of cytokine production”. Furthermore, the most enriched pathways in upregulated genes of infected A549 cells were “defense response to virus”, “response to interferon-gamma” and “antiviral mechanisms by IFN-stimulated genes” as well as “response to interferon-beta” and “response to interferon-alpha”.

However, downregulated genes of infected NHBE cells were enriched in “kidney development”, “response to growth factor” and “inner ear development”. Downregulated genes of A549 cells were mainly enriched in the “macroautophagy” pathway.
Figure 1. Bar graph of top clusters with their representative enriched terms (one per cluster) across input gene lists, colored by p-values. "Log10(P)" is the p-value in log base 10. (A) top enriched functional pathways based on upregulated genes in SARS-CoV-2 infected NHBE cells. (B) same as (A) for downregulated genes in SARS-CoV-2 infected NHBE cells.
Figure 2. Bar graph of top clusters with their representative enriched terms (one per cluster) across input gene lists, colored by p-values. "Log10(P)" is the p-value in log base 10. (A) top enriched functional pathways based on upregulated genes in SARS-CoV-2 infected A549 cells. (B) same as (A) for downregulated genes in SARS-CoV-2 infected A549 cells.

3.3. Protein-protein interaction

PPI enrichment network was constructed for each gene list with Metascape online tool. After constructing the network, the Molecular Complex Detection (MCODE) method was applied to identify closely connected network components. Furthermore, pathway enrichment analysis was applied to each MCODE component independently, and the three most significant terms by p-value were shown as the functional description of the corresponding components.
Figures 3, 4 and tables 3, 4 show the PPI network and detailed proteins involved in each MCODE subcluster with their corresponding biological pathways based on upregulated genes of infected NHBE and A549 cell lines separately. Downregulated genes did not show a significant PPI network.
Figure 3. PPI network and MCODE components analysis identified for upregulated genes of infected NHBE cells. (A) The resultant protein interaction network for the given gene list containing the subset of proteins that form physical interactions with at least one other member in the list. (B) three most densely connected network components based on MCODE clustering.
Table 3. Pathway enrichment analysis applied to MCODE 1-3 components represented in figure 3 independently, and the three most significant terms by p-value represent the functional description of the corresponding components.

| Color | MCODE  | GO | Description | Log10(P) |
|-------|--------|----|-------------|----------|
| MCODE_1 | R-HSA-380108 | Chemokine receptors bind chemokines | -12.9 |
| MCODE_1 | R-HSA-375276 | Peptide ligand-binding receptors | -12.7 |
| MCODE_1 | R-HSA-373076 | Class A/1 (Rhodopsin-like receptors) | -12.4 |
| MCODE_2 | R-HSA-69275 | G2/M Transition | -10.9 |
| MCODE_2 | R-HSA-453274 | Mitotic G2-G2/M phases | -10.9 |
| MCODE_2 | R-HSA-380320 | Recruitment of NuMA to mitotic centrosomes | -10.8 |
| MCODE_3 | R-HSA-168928 | DDX58/IFIH1-mediated induction of interferon-alpha/beta | -18.4 |
| MCODE_3 | R-HSA-918233 | TRAF3-dependent IRF activation pathway | -17.6 |
| MCODE_3 | GO:0032479 | regulation of type I interferon production | -16.7 |

Figure 4. PPI network and MCODE components analysis identified for upregulated genes of infected A549 cells. (A) The resultant protein interaction network for the given gene list containing the subset of proteins that form
physical interactions with at least one other member in the list. (B) densely connected network components based on MCODE clustering.

Table 4. Pathway enrichment analysis applied to MCODE 1-3 components presented in figure 4 independently, and the three most significant terms by p-value represent the functional description of the corresponding components.

| Color | MCODE   | GO                  | description                                                                 | Log10(P) |
|-------|---------|---------------------|-----------------------------------------------------------------------------|----------|
| Red   | MCODE_1 | R-HSA-380108        | Chemokine receptors bind chemokines                                         | -14.9    |
| Red   | MCODE_1 | R-HSA-375276        | Peptide ligand-binding receptors                                             | -14.6    |
| Red   | MCODE_1 | GO:0070098          | chemokine-mediated signaling pathway                                         | -13.9    |
| Blue  | MCODE_2 | R-HSA-1169408       | ISG15 antiviral mechanism                                                    | -12.7    |
| Blue  | MCODE_2 | R-HSA-1169410       | Antiviral mechanism by IFN-stimulated genes                                 | -12.4    |
| Blue  | MCODE_2 | GO:0032479          | regulation of type I interferon production                                  | -11.4    |
| Blue  | MCODE_3 | R-HSA-918233        | TRAF3-dependent IRF activation pathway                                       | -9.8     |
| Blue  | MCODE_3 | R-HSA-933541        | TRAF6 mediated IRF7 activation                                               | -8.8     |
| Blue  | MCODE_3 | GO:0039528          | cytoplasmic pattern recognition receptor signaling pathway in response to virus | -8.8     |

The results of PPI were compatible with the results of the pathway enrichment analysis mentioned above. In the PPI network constructed based on upregulated genes of infected NHBE cells, 24 densely connected proteins were grouped as MCODE1 mainly associated with the “chemokine receptors bind chemokines” pathway. PPI network of A549 upregulated genes also showed similar results; seven proteins including CXCL1, CXCL2, CXCL3, CXCL5, CXCL8, CCL20, and C3 were grouped as MCODE1. Pathway and process analysis showed that these seven proteins mainly belong to chemokine related pathways. Other MCODE subclusters were mainly associated with the regulation of interferon (IFN) production, IFN signaling pathways, and antiviral mechanisms by IFN-stimulated genes (ISG).
4. Discussion

The newly emerged COVID-19 has become a major global public health concern. However, complete life cycle details of SARS-CoV-2 and host response are not properly known yet. As long as our knowledge of this virus is limited, designing effective novel treatments are almost impossible. Transcriptome analysis of SARS-CoV-2 infected cells can aid our understanding of this virus and accelerate the discovery of novel therapies. In this study, we conducted a drug repurposing strategy based on CMap to identify therapeutic candidates for COVID-19. Transcriptional profiling of NHBE and A549 cell lines infected with SARS-CoV-2 were prepared. Drugs tending to reverse transcriptional response to SARS-CoV-2 infection were considered as therapeutic options. Interestingly, many of the top-scoring drugs are justified by the available evidence and can be proposed as available therapeutic candidates for COVID-19.

An approved drug with one of the highest connectivity scores is saquinavir, an HIV protease inhibitor. Virtual screening of clinically approved drugs has shown saquinavir as a potential inhibitor of SARS-CoV-2 main protease, RNA-dep RNA polymerase, and other non-structural proteins (nsp) [22, 23]. Saquinavir can also suppress SARS-CoV-2 replication based on in vitro models in less than 10μM concentration [24]. Other drugs of this group such as lopinavir and ritonavir are currently used as COVID-19 treatment.

Lansoprazole, a proton pump inhibitor (PPI), is another candidate. Watanabe et al. [25] have published an article in 2020, confirming potential antiviral properties of prazoles against different viruses such as HIV, Ebola virus (EBOV) and Epstein–Barr virus (EBV). Bojkova et al. [26] have conducted in vitro experiments to identify novel inhibitors of SARS-CoV-2 replication and they showed the potential of omeprazole, another PPI, to increase the antiviral
activity of remdesivir. It is known that inhibiting ATPase proton pumps with medications such as PPI may interfere with the acidification of endosomal pathways that seems an essential step for coronavirus infectivity. Chloroquine and hydroxychloroquine also are shown to have the same effects as prazoles to increase lysosomal PH [27]. Besides this, other mechanisms should also be considered.

Folic acid is another option. Some previous studies have suggested folic acid as an adjunctive therapy. Besides immune-boosting properties, folic acid can be used as an inhibitor of furin activity, an enzyme responsible for SARS-CoV-2 spike protein cleavage [28]. Therefore, it can be considered as a prophylactic or therapeutic candidate for COVID-19, especially in initial phases. In silico studies have also proposed folic acid as a potential inhibitor of SARS-CoV-2 main protease [29].

Another top-scoring drug is ebselen. Ebselen is a small molecule with antioxidant, anti-inflammatory, and cytoprotective features. Previous computer-aided studies have revealed ebselen as a strong inhibitor of SARS-CoV-2 main protease, furtherly was approved through cell-based studies [30]. It is also known that ebselen can covalently bind to the Zn-bound/catalytic cysteines in SARS-CoV-2 papain-like protease (PLpro) and nsp10, two critical proteins for viral replication, to inhibit them [31]. In the last two decades, different analogs of ebselen were synthesized as antiviral agents [32]. These data strongly emphasize the clinical potential of ebselen for treating COVID-19 patients.

Another drug is aminocaproic acid, a plasminogen activator inhibitor (PAI). It is known that PAIs can inhibit other membrane-anchored serine proteases such as transmembrane serine protease 2 (TMPRSS2), a critical factor for cell entry of SARS-CoV-2 [33]. Therefore, they can be used as a therapeutic or even prophylactic agent for COVID-19. Shen et al. have shown the potential effect of aminocaproic acid for inhibiting influenza virus through suppressing cleavage of hemagglutinin (HA) precursor, a critical glycoprotein for viral
binding and entry [34]. Besides, our results would suggest other unknown antiviral mechanisms for aminocaproic acid.

One of the other candidates is simvastatin, a lipid-lowering agent with anti-inflammatory and immunomodulatory properties. Previous studies have reported statins as potential adjunctive therapy for enveloped viruses such as EBOV by inhibiting cholesterol/isoprenoid pathway. They can inhibit viral glycoprotein processing and suppress virus infectivity [35]. In silico models have also shown potential inhibitory effects of statins on SARS-CoV-2 main protease [36].

Two other top-scoring compounds are ambroxol, a secretolytic agent used in respiratory diseases and stimulant of surfactant synthesis, and benzethonium chloride with surfactant, antiseptic and broad-spectrum antimicrobial properties. Angiotensin-converting enzyme 2 (ACE2), SARS-CoV-2 receptor, is expressed in type 2 alveolar pneumocytes, cells producing surfactant, and infection of these cells with SARS-CoV-2 leads to downregulation of surfactant synthesis and release, contributing to acute respiratory distress syndrome (ARDS) [37]. Therefore, these medications and other drugs with the same mechanisms would be a potent adjunctive therapy for COVID-19.

Monorden or radicicol is a heat shock protein 90 (HSP90) inhibitor. Connor et al. [38] in 2007 showed that HSP90 is essential for viral replication and its inhibitors such as radicicol or geldanamycin could block the replication of RNA viruses. A recent study conducted by Wyler et al. [39] has examined the effect of HSP90 inhibitors on SARS-CoV-2 based on in vitro models and proposed this group as potent antiviral and anti-inflammatory agents. Sodium phenylbutyrate is a histone deacetylase (HDAC) inhibitor. They act as epigenetic modulators and regulate expression of genes involved in the cell cycle, proliferation, differentiation, and also inflammation [40]. A number of these compounds have been suggested for SARS-CoV-2. In vitro experiments have shown an inhibitory effect of
valproate, an HDAC inhibitor, on the NF-KB pathway, and subsequently suppressing TNF-α and IL 6 [41]. In addition, they can be used as antifibrotic agents in pulmonary fibrosis [42].

Another important drug is metronidazole, a commonly used antibiotic. Some recent studies have proposed metronidazole as an addition to the COVID-19 treatment regimen due to its immunomodulatory effects. Metronidazole can decrease the level of inflammatory cytokines which are increased during COVID-19 infection [43]. Chakraborty et al. [44] have proposed adding anaerobe-specific antibiotics such as metronidazole to COVID-19 treatment due to evidence of overexpression of anaerobic bacteria in COVID-19 patients.

Another choice is budesonide, an inhaled corticosteroid (ICS). Although some paradoxes about their effects on viral infections exist, some studies have shown a significant inhibitory effect of ICS on different members of the coronavirus family based on in vitro models [45-47]. Considering COVID-19 patient’s comorbidities such as asthma or COPD, this evidence suggests continuing ICS in course of coronavirus infection.

Some previous studies have also shown the effectiveness of many of the other resulting drugs. These mentioned drugs can be utilized for further in vitro and in vivo experiments to better examine their efficacy. In addition, some investigational compounds have resulted in high connectivity scores that can be chosen as candidate compounds for further analysis.

Furthermore, analyzing the transcriptome of SARS-CoV-2 infected cells identified significant upregulation of proinflammatory cytokines dominantly chemokines that are critical for immune cells chemotaxis. This revealed a chemokine-dominant hyperinflammatory response. However, the inflammatory response is a double-edged sword. On the one hand, it is critical to limit viral replication. On the other hand, exaggerated responses can lead to multiorgan failure including ARDS. Previous studies have shown dysregulated immune over activation in COVID-19 patients known as “cytokine storm”,

20
responsible for worsening of lung injury and COVID-19 exacerbations. Therefore, early immunomodulatory interventions might prevent severe cases and the need for invasive ventilation. To date, different immune targets are considered potentially effective [48-50].

Moreover, the results of SARS-CoV-2 infected cells confirm a robust IFN response with significant expression of ISGs pathway with known antiviral, inflammatory, and immune regulatory functions. However, previous analysis has shown impaired type 1 IFN production in severe COVID-19 cases and proposed different hypotheses whether IFN production is exhausted after the initial overexpression or SARS-CoV-2 can evade immune pathways [51-53]. However, COVID-19 immunopathology is not properly known and more comprehensive studies are needed to clarify dynamic immune response to SARS-CoV-2.

It is promising that the CMap strategy has identified a number of therapeutic candidates that could be repurposed for COVID-19. Since SARS-CoV-2 infection can change host transcriptome in the way optimized for viral replication, transcriptome profiling and further CMap analysis can help better understand virus pathophysiology and may discover new unknown genes and pathways. However, data used in this study was obtained 24 hours post-infection with SARS-CoV-2 and further studies should be conducted in variable time points to assess the dynamics of virus pathogenesis and host response. Furthermore, although CMap libraries have been improved since its introduction in 2006, all available small molecules are much more than current CMap libraries.

5. Conclusion

The newly emerged COVID-19 pandemic is the major global health crisis of our time affecting more than 217 countries and territories around the world. Several studies are ongoing to find effective therapies in particular through drug repurposing approaches. In the
In the present study, we suggested potential antiviral candidates based on the CMap concept. Available gene-expression profiles of SARS-CoV-2 infected NHBE and A549 cell lines were analyzed to identify DEGs. Subsequently, these signatures were compared with genomic signatures of cell lines treated with variable compounds available in CMap databases. Among these compounds, Saquinavir, lansoprazole, folic acid, ebselen, aminocaproic acid, simvastatin, surfactant stimulant drugs, HSP90 inhibitors, HDAC inhibitors, metronidazole, ICS and many other clinically approved drugs and investigational compounds showed promising results to reverse gene-expression profiles of SARS-CoV-2 infected cells. Further in vitro and in vivo studies should be conducted to test potential antiviral effects of these suggested compounds against SARS-CoV-2.

**Conflict of interest statement**

The authors declare that there are no conflicts of interest. All authors disclose any financial and personal relationships with other people or organizations that could inappropriately influence (bias) this work.

**Acknowledgments**

This research was partly funded through a research grant from Iran’s National Elites Foundation called Ahmadi Rowshan grant. Authors declare the role of study sponsors in the study design, in the collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.
References

[1] Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. The Lancet. 2020;395(10223):470-3.

[2] Ren L-L, Wang Y-M, Wu Z-Q, Xiang Z-C, Guo L, Xu T, et al. Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. Chinese Medical Journal. 2020;133(9):1015-1024.

[3] Paules CI, Marston HD, Fauci AS. Coronavirus infections—more than just the common cold. Jama. 2020;323(8):707-8.

[4] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/

[5] Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nature Reviews Drug Discovery. 2019;18(1):41-58.

[6] Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006;313(5795):1929-35.

[7] Lamb J. The Connectivity Map: a new tool for biomedical research. Nature Reviews Cancer. 2007;7(1):54-60.

[8] Qu XA, Rajpal DK. Applications of Connectivity Map in drug discovery and development. Drug Discovery Today. 2012;17(23-24):1289-98.

[9] Musa A, Ghoraie LS, Zhang S-D, Glazko G, Yli-Harja O, Dehmer M, et al. A review of connectivity map and computational approaches in pharmacogenomics. Briefings in Bioinformatics. 2018;19(3):506-23.
[10] So H-C, Chau CK, Chiu W-T, Ho K-S, Lo C-P, Yim SH-Y, et al. When GWAS meets the Connectivity Map: drug repositioning for seven psychiatric disorders. bioRxiv. 2016:096503.

[11] Jahchan NS, Dudley JT, Mazur PK, Flores N, Yang D, Palmerton A, et al. A drug repositioning approach identifies tricyclic antidepressants as inhibitors of small cell lung cancer and other neuroendocrine tumors. Cancer Discovery. 2013;3(12):1364-77.

[12] Zhang L, Kang W, Lu X, Ma S, Dong L, Zou B. Weighted gene co-expression network analysis and connectivity map identifies lovastatin as a treatment option of gastric cancer by inhibiting HDAC2. Gene. 2019;681:15-25.

[13] Xiao S-j, Zhu X-c, Deng H, Zhou W-p, Yang W-y, Yuan L-k, et al. Gene expression profiling coupled with Connectivity Map database mining reveals potential therapeutic drugs for Hirschsprung disease. Journal of Pediatric Surgery. 2018;53(9):1716-21.

[14] Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, et al. NCBI GEO: mining tens of millions of expression profiles—database and tools update. Nucleic Acids Research. 2007;35(suppl_1):D760-D765.

[15] Blanco-Melo D, Nilsson-Payant B, Liu W-C, Møller R, Panis M, Sachs D, et al. SARS-CoV-2 launches a unique transcriptional signature from in vitro, ex vivo, and in vivo systems. bioRxiv. 2020.

[16] Tinka Vidovik. Using Connectivity map data and PharmacoGx R package to find drugs against coronaviruses. https://github.com/tinkavidovik/competition.

[17] Liu T-P, Hsieh Y-Y, Chou C-J, Yang P-M. Systematic polypharmacology and drug repurposing via an integrated L1000-based Connectivity Map database mining. Royal Society Open Science. 2018;5(11):181321.
[18] Smirnov P, Safikhani Z, El-Hachem N, Wang D, She A, Olsen C, et al. PharmacoGx: an R package for analysis of large pharmacogenomic datasets. Bioinformatics. 2016;32(8):1244-6.

[19] Fagone P, Ciurleo R, Lombardo SD, Iacobello C, Palermo CI, Shoenfeld Y, et al. Transcriptional landscape of SARS-CoV-2 infection dismantles pathogenic pathways activated by the virus, proposes unique sex-specific differences and predicts tailored therapeutic strategies. Autoimmunity Reviews. 2020;102571.

[20] Taguchi Y, Turki T. A New Advanced In Silico Drug Discovery Method for Novel Coronavirus (SARS-CoV-2) with Tensor Decomposition-Based Unsupervised Feature Extraction. Preprints. 2020.

[21] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature Communications. 2019;10(1):1-10.

[22] Hall Jr DC, Ji H-F. A search for medications to treat COVID-19 via in silico molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease. Travel Medicine and Infectious Disease. 2020;101646.

[23] Ruan Z, Liu C, Guo Y, He Z, Huang X, Jia X, et al. Potential inhibitors targeting RNA-dependent RNA polymerase activity (NSP12) of SARS-CoV-2. Preprints. 2020.

[24] Yamamoto N, Matsuyama S, Hoshino T, Yamamoto N. Nelfinavir inhibits replication of severe acute respiratory syndrome coronavirus 2 in vitro. bioRxiv. 2020.

[25] Watanabe SM, Ehrlich LS, Strickland M, Li X, Soloveva V, Goff AJ, et al. Selective targeting of Virus Replication by proton pump inhibitors. Scientific Reports. 2020;10(1):1-15.
[26] Bojkova D, McGreig JE, McLaughlin K-M, Masterson SG, Widera M, Kraehling V, et al. SARS-CoV-2 and SARS-CoV differ in their cell tropism and drug sensitivity profiles. bioRxiv. 2020.

[27] Touret F, Gilles M, Barral K, Nougairède A, Decroly E, de Lamballerie X, et al. In vitro screening of a FDA approved chemical library reveals potential inhibitors of SARS-CoV-2 replication. bioRxiv. 2020.

[28] Sheybani Z, Dokoohaki MH, Negahdaripour M, Dehdashti M, Zolghadr H, Moghadami M, et al. The role of folic acid in the management of respiratory disease caused by COVID-19. chemRxiv. 2020.

[29] Serseg T, Benarous K, Yousfi M. Hispidin and Lepidine E: two Natural Compounds and Folic acid as Potential Inhibitors of 2019-novel coronavirus Main Protease (2019-nCoVMpro), molecular docking and SAR study. arXiv preprint arXiv:200408920. 2020.

[30] Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, et al. Structure of Mpro from COVID-19 virus and discovery of its inhibitors. bioRxiv. 2020.

[31] Yang HS, Lim C. Multi-Targeting of Functional Cysteines in Multiple Conserved SARS-CoV-2 Domains by Clinically Safe Zn-ejectors. chemRxiv. 2020.

[32] Wójtowicz H, Kloc K, Maliszewska I, Młochowski J, Piętka M, Piasecki E. Azaanalogues of ebselen as antimicrobial and antiviral agents: synthesis and properties. Il Farmaco. 2004;59(11):863-8.

[33] Jankun J. COVID-19 pandemic; transmembrane protease serine 2 (TMPRSS2) inhibitors as potential drugs. Translation: The University of Toledo Journal of Medical Sciences. 2020;7:1-5.

[34] Shen X, Zhang X, Liu S. Novel hemagglutinin-based influenza virus inhibitors. Journal of Thoracic Disease. 2013;5(Suppl 2):S149.
[35] Shrivastava-Ranjan P, Flint M, Bergeron É, McElroy AK, Chatterjee P, Albariño CG, et al. Statins suppress Ebola virus infectivity by interfering with glycoprotein processing. MBio. 2018;9(3):e00660-18.

[36] Reiner Ž, Hatamipour M, Banach M, Pirro M, Al-Rasadi K, Jamialahmadi T, et al. Statins and the COVID-19 main protease: in silico evidence on direct interaction. Archives of Medical Science. 2020;16(3):490-6.

[37] Bombardini T, Picano E. Angiotensin converting enzyme 2 as the molecular bridge between epidemiologic and clinical features of COVID-19. Canadian Journal of Cardiology. 2020.

[38] Connor JH, McKenzie MO, Parks GD, Lyles DS. Antiviral activity and RNA polymerase degradation following Hsp90 inhibition in a range of negative strand viruses. Virology. 2007;362(1):109-19.

[39] Wyler E, Mösbauer K, Franke V, Diag A, Gottula LT,Arsie R, et al. Bulk and single-cell gene expression profiling of SARS-CoV-2 infected human cell lines identifies molecular targets for therapeutic intervention. bioRxiv. 2020.

[40] Seto E, Yoshida M. Erasers of histone acetylation: the histone deacetylase enzymes. Cold Spring Harbor Perspectives in Biology. 2014;6(4):a018713.

[41] Bhargava P, Panda P, Ostwal V, Ramaswamy A. Repurposing valproate to prevent acute respiratory distress syndrome/acute lung injury in COVID-19: A review of immunomodulatory action. Cancer Research, Statistics, and Treatment. 2020;3(5):65.

[42] Lyu X, Hu M, Peng J, Zhang X, Sanders YY. HDAC inhibitors as antifibrotic drugs in cardiac and pulmonary fibrosis. Therapeutic Advances in Chronic Disease. 2019;10:2040622319862697.
[43] Gharebaghi R, Heidary F, Moradi M, Parvizi M. Metronidazole; a Potential Novel Addition to the COVID-19 Treatment Regimen. Archives of Academic Emergency Medicine. 2020;8(1).

[44] Chakraborty S, Das G. Secondary infection by anaerobic bacteria possibly ensues a battle for oxygen in SARS-Cov2 infected patients: anaerobe-targeting antibiotics (like doxycycline/Metronidazole) to supplement Azithromycin in the treatment regimen of COVID19?. OSF Preprints. 2020.

[45] Halpin DM, Singh D, Hadfield RM. Inhaled corticosteroids and COVID-19: a systematic review and clinical perspective. Eur Respiratory Soc; 2020.

[46] Yamaya M, Nishimura H, Deng X, Sugawara M, Watanabe O, Nomura K, et al. Inhibitory effects of glycopyrronium, formoterol, and budesonide on coronavirus HCoV-229E replication and cytokine production by primary cultures of human nasal and tracheal epithelial cells. Respiratory Investigation. 2020.

[47] Matsuyama S, Kawase M, Nao N, Shirato K, Ujike M, Kamitani W, et al. The inhaled corticosteroid ciclesonide blocks coronavirus RNA replication by targeting viral NSP15. bioRxiv. 2020.

[48] Zhou Z, Ren L, Zhang L, Zhong J, Xiao Y, Jia Z, et al. Heightened innate immune responses in the respiratory tract of COVID-19 patients. Cell Host & Microbe. 2020.

[49] 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):1033-4.

[50] Merad M, Martin JC. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. Nature Reviews Immunology. 2020:1-8.

[51] Li G, Fan Y, Lai Y, Han T, Li Z, Zhou P, et al. Coronavirus infections and immune responses. Journal of Medical Virology. 2020;92(4):424-32.
[52] Hadjadj J, Yatim N, Barnabei L, Corneau A, Boussier J, Pere H, et al. Impaired type I interferon activity and exacerbated inflammatory responses in severe Covid-19 patients. medRxiv. 2020.

[53] Sallard E, Lescure F-X, Yazdanpanah Y, Mentre F, Peiffer-Smadja N, Florence A, et al. Type 1 interferons as a potential treatment against COVID-19. Antiviral Research. 2020:104791.