Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Original Article

Transcriptome network analyses in human coronavirus infections suggest a rational use of immunomodulatory drugs for COVID-19 therapy

Henry Sung-Ching Wong a, Chin-Lin Guo b, Gan-Hong Lin c, Kang-Yun Lee d,1, Yukinori Okada a,d,g,h, Wei-Chiao Chang a,c,i,h,k,*

a Department of Clinical Pharmacy, School of Pharmacy, Taipei Medical University, Taipei, Taiwan
b Institute of Physics, Academia Sinica, Taipei, Taiwan
c Master Program for Clinical Pharmacogenomics and Pharmacoproteomics, School of Pharmacy, Taipei Medical University, Taipei, Taiwan
d Division of Pulmonary Medicine, Department of Internal Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
e Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
f Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan
g Laboratory of Statistical Immunology, WPI Immunology Frontier Research Center, Osaka University, Osaka, Japan
h Laboratory of Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
i Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Osaka, Japan
j Integrative Research Center for Critical Care, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan
k Division of Pulmonary Medicine, Department of Internal Medicine, Taipei Medical University-Shuang Ho Hospital, New Taipei City, Taiwan
l Department of Pharmacology, National Defense Medical Center, Taipei, Taiwan
m Department of Pharmacology, National Defense Medical Center, Taipei, Taiwan
n Division of Pulmonary Medicine, Department of Internal Medicine, Taipei Medical University-Shuang Ho Hospital, New Taipei City, Taiwan

A R T I C L E  I N F O

Keywords:
SARS-CoV-2
COVID-19
Transcriptomic and network analysis
Drug repositioning

A B S T R A C T

The recent outbreak of coronavirus disease 2019 (COVID-19) by SARS-CoV-2 has led to update 24.3 M cases and 0.8 M deaths. It is thus in urgent need to rationalize potential therapeutic targets against the progression of diseases. An effective, feasible way is to use the pre-existing ΔORF6 mutant of SARS-CoV as a surrogate for SARS-CoV-2, since both lack the moiety responsible for interferon antagonistic effects. By analyzing temporal profiles of upregulated genes in ΔORF6-infected Calu-3 cells, we prioritized 55 genes and 238 ligands to reposition currently available medications for COVID-19 therapy. Eight of them are already in clinical trials, including dexamethasone, ritonavir, baricitinib, tocilizumab, naproxen, budesonide, ciclesonide and formoterol. We also pinpointed 16 drug groups from the Anatomical Therapeutic Chemical classification system, with the potential to mitigate symptoms of SARS-CoV-2 infection and thus to be repositioned for COVID-19 therapy.

1. Introduction

The COVID-19 pandemic, which has led to huge economic losses by the Great Lockdown, is an ongoing global threat caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1]. As of August 2020, 24.3 M cases with ∼828 k deaths have been reported across 188 countries and territories, leading to an ∼3.4% mortality rate, 34 times higher than that of H1N1. Unfortunately, no drug or vaccine has been approved to treat human coronaviruses, and new interventions are likely to take months to years to develop. In this regard, rationalizing semi-empirical or empirical strategies to protect currently hospitalized victims (∼8.4 M up to date) from death is urgent and necessary. The SARS-CoV-2 is linked to the severe acute respiratory syndrome coronavirus (SARS-CoV, which led to 774 deaths out of 8096 infected individuals ~16 years ago) and the Middle East respiratory syndrome coronavirus (MERS-CoV, which led to 858 deaths out of 2494 cases ~8 years ago) in that they share a high degree of genomic similarity (~80% between SARS-CoV and SARS-CoV-2, e.g.) [2]. As such, one plausible way to rationalize empirical strategies in a timely manner is to use available SARS-CoV/MERS-CoV data and perform fast, effective in silico analyses, thereby identifying candidate remedies from currently available medications.

In addition to genomics, SARS-CoV-2 and SARS-CoV share a certain degree of similarity. It has been shown that SARS-CoV-2 shares similar transmission mechanisms and pathogenic sequelae with SARS-CoV. First, both viruses are group B coronaviruses utilizing the same receptor, angiotensin-converting enzyme 2 (ACE2), to bind to the host cell surface [3]. Second, the pathology of SARS-CoV-infected lung shows...
features of acute injury and fibroelastic organizing pneumonia patterns [4], which were also found in SARS-CoV-2 infected lung, including diffusive acute alveolar injury with edema, hyaline membrane formation, proteinaceous exudate, focal reactive hyperplasia, immune cell infiltration, and intra-alveolar fibroelastic plug formation [5]. The consequence is the respiratory failure, the leading cause for the mortality, as no gas exchange can be performed in such foamy, fibroelastic alveolar space. Intervening in the disease progression requires an explicit molecular cellular understanding of the pathogenesis, which is currently limited. What has been noted is the involvement of highly inflammatory responses that produce enormous chemokines implicated in cytokine storm [6,7]. Futhermore, a key difference between SARS-CoV and SARS-CoV-2 is a truncation of the open reading frame (ORF) 3b, ORF6, in SARS-CoV-2, which affects the viral antagonistic effect toward interferon response [6]. These lines of evidence suggest that it is plausible to use available SARS-CoV data as a surrogate for devising the empirical strategy and that immune modulation might be involved in such strategy. Fortunately, a mutant of SARS-CoV with the truncation in ORF6 has been found and used to infect a human lung cell line, Calu-3, for empirical study. Here, we took advantage of these preexisting data and applied bioinformatic analysis on SARS-CoV-infected cells (as the reference) and SARS-CoV-ΔORF6-infected cells (as the surrogate to mimic SARS-CoV-2 infection) to identify medications that can be repurposed for COVID-19 therapy.

2. Materials and method

2.1. Expression data

Transcriptomes of Calu-3 cells infected by infectious clone-derived (ic)SARS CoV and icSARS ΔORF6 mutant (GSE33267) were queried using GEOquery. The downloaded data had already been normalized. Probe-level expression values (n = 30,772) were log2-transformed and aggregated to become gene-level expression values (n = 18,113) by maximum value. Gene symbols were further checked and corrected using the HGNCchelper package.

2.2. Identification of differentially expressed genes

Differentially expressed tests were conducted on the time-series microarray data (ΔORF6-infected cells vs. control cells or SARS-infected cells vs. control cells) based on the following post-treatment time-points: 0, 3, 7, 12, 24, 30, 36, 48, 54, 60 and 72 h. To ensure the reliability of identified DEGs, we conducted three different methods, and defined differential expressed genes (DEGs) as genes that showed consensus (both on significance and effect direction) across methods. The three methods used different ways of normalization (normexp method and quantile normalization for background correction and robust multichip average (RMA) for method 1 and 3; normexp method and quantile normalization for background correction, RMA and between-sample normalization for method 2), different ways of gene summarization (maximum value aggregation for methods 1 and 3; Tukey Biweight estimator for method 2), and different ways of identifying significant genes across groups, as following: Method (1): Limma method. By assuming smooth expression changes over time, we fitted a regression spline with effective degrees of freedom of five to capture the temporal trends, and then compared the differences in curves between groups (ΔORF6 vs. mock and wild-type vs. mock). Then for each gene, a linear model that implemented in limma [8] was fitted and an empirical Bayes moderated F-statistic was calculated for testing the change in expression level over the time points. Five coefficients corresponded to differences in two curves were retrieved from the fitted model and then averaged to produce a mean coefficient value. To control the false positive rate, raw p-values were adjusted using the Benjamini-Hochberg (BH) method. Significant genes (out of 18,113 genes) of the method (1) were defined as genes with BH-adjusted p-value < 0.05. Up or downregulated DEGs were defined by the average coefficient of > 0 or < 0, respectively. Method (2): Rnits method. In this method, probe expression values were further undergone between sample normalization and collapsed into gene-level using a robust Tukey Biweight estimator, which performs penalization on outlier values. Next, significant genes were identified by testing whether each gene expression trajectory from the treatment group (ΔORF6 or wild-type) and mock (control) group are the realizations of the same underlying basis B-spline curve [10]. The significant genes (out of 18,109 genes) of the method (2) were defined as genes that passed a BH-adjusted p-value threshold of 0.05. Up or downregulated significant genes were defined by ratio statistic of > 0 or < 0, respectively. Method (3): timecourse (tc) method. This method was implemented in timecourse package, which calculated a T² statistic for each gene using multivariate empirical Bayes approach. This method incorporated the correlation coefficient information across genes to diminish false-positivity and false-negativity. The significance of each gene was calculated by F-statistic that derived from the T² statistic. The significant genes (from total 18,113 genes) of the method (3) were defined as genes that passed a Bonferroni-adjusted p-value threshold of 0.05. Since this method didn’t output information regarding the up or downregulation of significant genes, we thus calculated the logarithmic-transformed multiple of change (logFC) as:

\[
\text{logFC} = \frac{\sum_{i}^{n} \Delta F_i}{\sum_{i}^{N} \Delta F_i}, \quad \forall i \in \Delta ORF6 or wild-type-infected cells and \forall i \in control cells, of each gene from normalized expression data. Finally, we defined DEGs as genes that were identified as significant genes in the above three methods and with the same effect direction.

2.3. Construction of co-expression clusters

ΔORF6-infected cell expression data were summarized to Gencode v33 [11] and the median-centering before network analysis. Gene clusters were identified using weighted correlation network analysis (WGCNA) [12]. A weighted signed network was constructed using a thresholding power of 10, which is the smallest value to get a scale-free r² fit of 0.85. The pairwise gene-to-gene topological overlap (TO) matrix was then calculated. The pre-clusters were identified by applying a top-down dynamic tree cut method (module gene size > 200 and deepSplit parameter 2), and pre-clusters were defined as branches of the hierarchical cluster tree. Pre-clusters with pairwise cluster eigengene (PC1) correlation values (r) > 0.85 were further merged into clusters.

2.4. Cluster enrichment test for DEGs

Enrichment test for DEGs across multiple time points in co-expressed clusters was conducted using two-tailed Fisher’s exact test. The odds ratios (ORs) were calculated. We defined OR > 1 and p-value < 0.05 as significant overrepresentation (enrichment).

2.5. Transcription factor binding site (TFBS) analysis

Depending on the initial data (either PWM information or target genes of a TF) retrieved from databases, two different strategies were adopted as follows: (1) sequence-based enrichment test. Given a set a gene with size Ns, we found several candidate transcription factors (TFs) with known position weight matrix (PWM) or motif, we want to identify whether the TFs’ binding motif was enriched in the promoter sequence of genes in the given gene set. To do this, we scanned the corresponding region (2000 basepair (bp) upstream from the transcription start site (TSS)) for the Ns genes. We retrieved TFs’ human PWM information from MotifDB package. For each PWM, a threshold-free log-normal distribution was calculated by fitting 500 bp chunks of 2000 bp promoters from all human genes (hg19). The PWM enrichment test of each
TF was performed using PWMEnrich package. We further conducted Bonferroni correction to reduce false positivity, and (2) gene-based permutation test. Given a set of gene with size \( N_G \) and several candidate TFs with known target genes from database(s) (such as ChIP-seq results from the ENCODE), we conducted gene-based permutation test by counting the overlapped genes of TF-regulated genes to the given gene set and performed permutation test to assess the enrichment of overlapped count number as following: For each TF, we counted the overlapped target number \( (n) \) in the given gene set with size \( N \). Next, we randomly selected \( N \) genes from the background gene set and then counted the overlapped number of targets of these TFs. This process was repeated for 99,999 times to produce a null distribution \( n_1, \ldots, n_{99999} \) of target number for the given TFs. We then calculated the single-tailed \( p \)-value as \( \frac{(\text{no. of } n, n_1, \ldots, n_{99999} \geq n)}{\text{(size of } n, n_1, \ldots, n_{99999})} \).

2.6. Gene ontology and REACTOME pathway analysis

Genes in co-expressed clusters were characterized using Gene Ontology (GO) biological processes (BPs) and REACTOME pathways using clusterProfiler [13] and ReactomePA [14], respectively. Q-values were calculated for multiple testing corrections. The top 20 enriched terms were selected for visualization.

2.7. Identification of targetable genes

Targetable genes were defined as genes with its (protein) products that could be targeted by specific ligands (drugs) from the IUPHAR/BPS Guide to PHARMACOLOGY [15]. Further data sanitization steps including (i) limited only drugs with target species of human; (ii) ligands that annotated as non-endogenous agents; (iii) ligands with clear and explicit “action” and “type” annotations; and (iii) ligands with explicit target gene (symbol). Finally, the drug-gene pairs were collapsed across time points and further visualized using circlize, VennDiagram, RColorBrewer, and igraph packages.

2.8. Permutation test for the ATC groups of drugs

The Anatomical Therapeutic Chemical (ATC) classification of each ligand was annotated from the DrugBank database [16]. Given the size of \( N_G \) for the final targetable genes, we randomly selected \( N_G \) genes from the targetable gene list and mapped the ligands that may target the genes. Next, we summarized the number of drugs that belong to a specific ATC group. We repeated this process for \( n = 99,999 \) times and calculated the probability of obtaining a number of drugs that equal to or greater than original data in the particular ATC group. We further computing BH-adjusted \( p \)-values for multiple testing corrections and selected a BH-adjusted \( p \)-value of 0.1 as a threshold of significance.

2.9. Statistical analyses

In this study, we used R (http://www.r-project.org/ and http://cran.r-project.org/) or Bioconductor (http://www.bioconductor.org/) to perform all statistical association tests, enrichment analyses, and visualization tasks.

3. Results

3.1. Characterization of icSARS \( \Delta ORF6 \) infection-related genes

In contrast to SARS-CoV, SARS-CoV-2 shows ablated antagonism to interferon (IFN) due to an amino acid truncation in the viral ORF6 gene [8]. We therefore leveraged a previously published dataset (GSE33267) [17] that contains time-dependent transcriptomic profiling of infectious clone-derived (ic)SARS-CoV \( \Delta ORF6 \) (henceforth called “\( \Delta ORF6 \)”) in human Calu-3 cells. Since \( \Delta ORF6 \) is a mutant form of SARS-CoV that cannot express ORF6, it is reasonable to use \( \Delta ORF6 \) as a surrogate for SARS-CoV-2 (which also contains truncated ORF6) to reveal its pathophysiology and explore drug repositioning (Fig. 1).

To identify genes that respond to \( \Delta ORF6 \) infection in Calu-3 cells (\( \Delta ORF6 \) infection-related genes), we analyzed the differential expression of genes by comparing \( \Delta ORF6 \)-infected cells (\( n = 33; 3 \) replicates for 11 different time points including \( 0, 3, 7, 12, 24, 30, 36, 48, 54, 60 \) and 72 h) to control cells (\( n = 33; 3 \) replicates for 11 different time points) by applying three complementary methods (i.e. method (1), (2) and (3) as described in “Materials and Methods” section) that used different strategies to identify differentially expressed genes (DEGs) of time-series expression data. In parallel, we conducted similar tests on wild-type icSARS-CoV-infected Calu-3 cells (henceforth called “wild-type” or “SARS”) for comparison (Fig. 2a). We revealed a wide range of percentages of up or downregulated DEGs (0% to 77.74%), suggesting different sensitivity and specificity across methods.

For significant genes detected by each method, we defined DEGs as genes identified by all three methods and with the same effect of direction (upregulated and downregulated) across methods (Fig. 2b). Therefore, we detected no down-regulated DEGs in this study. Especially, we detected 3373 and 848 upregulated DEGs in \( \Delta ORF6 \)-infected cells and SARS-infected cells, respectively. However, none of the \( \Delta ORF6 \) group or SARS group shows consensus results regarding downregulated genes.

Significance values of genes were further transformed into 0–1, and the outcome similarity of three methods was quantified using Pearson’s product-moment correlation coefficients \( (r) \). We observed moderate-to-high correlations \( (r_p = 0.7 \) and 0.69 in \( \Delta ORF6 \)-infected cells and SARS-infected cells, respectively; Fig. S1) between the results of methods (1) and (3); while the results from the method (2) deviated from that of methods (1) and (3) \( (r_p \) ranged from 0.25–0.60; Figs. 2c and S2).

To investigate the effect of \( \Delta ORF6 \) or SARS infection on cytokine profiles, we assessed the significance of 121 cytokines/chemokines (ligands) and 144 corresponding receptor genes (in total 265 cytokines/chemokines-related genes, a.k.a. cytokine genes; Fig. S3) from the results of the differentially expressed test across multiple time points. We quantified the number (and percentage) of ligands or receptors. The \( p \)-values were calculated by using a one-tailed binomial test (with a “greater” alternative hypothesis and confidence level of 95%) by comparing it to the background gene list (excluding cytokine genes). False positivity was controlled by using Bonferroni correction. As a result, we found that the numbers of ligand genes (cytokines/chemokines) were significantly higher than the background genes (Fig. S4). However, we didn’t identify any significance in receptor genes. These data suggested that \( \Delta ORF6 \) or SARS infection has a great effect on the cytokine (and chemokine) levels.

3.2. Time-dependent co-expression networks of icSARS \( \Delta ORF6 \)-induced genes

To depict the co-expression dynamics of \( \Delta ORF6 \) infection-related genes in human lung cells after viral infection, we constructed time-dependent co-expression networks from the GSE33267. Post-infection transcriptomic profiles of \( \Delta ORF6 \)-infected Calu-3 cells (from 0 to 72 h, across 11 time-points, each with 3 replicates) were found to be partitioned into 6 clusters based on pairwise correlations between gene expression values; the number of genes in each cluster ranged from 232 to 4517 (Fig. 3a). In the following steps, we categorized the significantly upregulated DEGs in \( \Delta ORF6 \)-infected cells into two groups: (i) specifically in \( \Delta ORF6 \)-infected cells (denoted by “\( \Delta ORF6 \)-specific” or “specific”; which containing 2613 genes) and shared in both \( \Delta ORF6 \) and wild-type-infected cells (denoted by “shared”; which containing 760 genes).

We posited that additional mapping of \( \Delta ORF6 \) infection-related DEGs in an identified cluster should be observed if the particular cluster reflects a pathophysiological process in \( \Delta ORF6 \)-infected Calu-3 cells. To determine the convergence of identified co-expressed clusters of
ΔORF6 infection-related genes, we conducted enrichment tests for ΔORF6-specific and shared DEGs using an over-representation analysis (ORA). Among up-regulated DEGs (both ΔORF6-specific and shared genes) in response to ΔORF6 infection, significant enrichment was detected for clusters 1 (Fig. 3b). We next depicted the temporal trajectory of cluster 1 by calculating the first principal component (PC1; i.e., cluster eigengene) and assessing the value of eigengene across time. Eigengene can be viewed as a summarized value of gene expression in the specific cluster. In particular, genes in cluster 1 were up-regulated and sustained at approximately 30 to 72 h upon ΔORF6 infection (Fig. 3c). In details, the eigengene of cluster 1 from ΔORF6-infected cells gradually increased from 0 to 54 h post-infection, and then slightly dropped from 60 to 72 h.

3.3. Biological meanings of cluster enriched for upregulated ΔORF6 infection-related DEGs

To characterize the key factor(s) in up-regulated DEGs, we hypothesized that if there exists a key factor that causes the up-regulation and sustainability of genes from approximately 30 to 72 h after ΔORF6 infection, the key factor gene(s) should be also continuously expressed, and exert its/their regulatory functions. Since key factors should be able...
to regulate—turn on and off—genes' expression (upregulation), we thence focused on transcription factor(s) (TFs). Therefore, we derived a sequence-based enrichment test (please see “Materials and Methods” section) to identify TFs that may regulate the expression of ΔORF6-specific or shared (with wild-type/SARS) DEGs. As a result, we identified 57 and 18 TFs that passed the significance threshold in ΔORF6-specific and shared groups, respectively (Fig. S5). To validate our findings in ΔORF6-specific up-regulated DEGs, we leveraged the chromatin immunoprecipitation sequencing (ChIP-seq) data from the Encyclopedia of DNA Elements (ENCODEx) database. Of the 10 available TFs data in ENCODE, only the FOSL1 was overlapped with the 57 identified TFs. We, therefore, adopted the FOSL1 ChIP-seq results for validation. Given the information on regulatory interactions regarding FOSL1, we conducted a gene-based permutation test (please see “Materials and Methods” section) to assess the enrichment of (ΔORF6-specific) up-regulated DEGs in FOSL1 targets. As a result, we successfully replicated the significance of enrichment regarding FOSL1 using ENCODE ChIP-seq evidence (p-value < 0.00001; Fig. S6).

We next sought to determine whether correlated genes in cluster 1 share biological functions or regulatory mechanisms. The top 20 most significantly enriched Reactome pathways and Gene Ontology (GO) biological process (BP) annotation terms suggested that the genes in cluster 1 were implicated in cellular defensive response to the virus, which includes interferon (especially, type I) signaling pathways.
3.4. Discovery of candidate pharmacological ligands (drugs) that potentially oppose SARS-CoV-2 infection

Based on the fact that SARS-CoV-2 possesses a truncated ORF6 gene, which is critical for viral sensitivity to type I IFN and may influence host immune response [17], we utilized ΔORF6 as a model for SARS-CoV. Specific and shared genes induced in ΔORF6-infected cells were of interest for our analysis, as both groups may provide clues for drug repositioning. By restricting our study to upregulated genes responsive to ΔORF6 infection, we were able to pinpoint candidate drugs that may inhibit the gene products. DEGs upregulated at different time-points were classified into two cluster categories (i.e., “Cluster 1” and compare specific and/or shared biological functions of DEGs in human lung cancer cell line infected with wild-type or ΔORF6 SARS-CoV. Specific and shared genes induced in ΔORF6-infected cells were of interest for our analysis, as both groups may provide clues for drug repositioning. By restricting our study to upregulated genes responsive to ΔORF6 infection, we were able to pinpoint candidate drugs that may inhibit the gene products. DEGs upregulated at different time-points were classified into two cluster categories (i.e., “Cluster 1” and

(Fig. 3d).
“Other”). Together, DEGs that were specifically induced by ΔORF6 (denoted as “Specific”) or up-regulated in both ΔORF6 and SARS treatment (denoted as “Shared”) were subjected to drug repositioning analysis using the drug (ligand)-gene pairing information from the IUPHAR PHARMACOLOGICAL database. In total, 2396 drug-gene pairs (305 unique targetable genes and 1931 unique ligands) were identified, with 1365 (56.97%) pairs from the “Specific” group and 1031 (43.03%) from the “Shared” group (Fig. 4a). Among them, 1811 (75.58%) pairs were from “Cluster 1” and the remaining 585 (24.42%) pairs were from “Other”. Therefore, we identified a substantial number of candidate drug-gene pairs in “Cluster 1” and “Other” groups. The identified 3373 up-regulated genes were considered “targetable” or “druggable” and may be potentially useful to guide repositioning of drugs for the treatment of SARS-CoV-2 (Fig. 4b).

From the gene perspective, for the initial 3373 up-regulated DEGs, we consequently identified 305 unique targetable (druggable) genes by intersecting the DEGs with druggable genes from the database, with 200 (65.57%) genes were “Specific” and 105 (34.43%) genes were “Shared” (Fig. 4c). Moreover, 236 (77.38%) genes were from “Cluster 1” and 69 (22.62%) genes were from “Other”. Therefore, we ascertained in total 1931 unique candidate ligands (Fig. 4d), with some duplicated ligands been identified in different categories due to multiple target genes (Fig. 4e). We thus investigated the number of multiple gene targets of each ligand and found that most of the ligands (1557 out of 374 (19.37%) candidate ligands have multiple targets.

Since McCreary et al. suggested several putative pharmacotherapeutic options for COVID-19 based on current clinical evidence (up to March 28, 2020) [18], we sought to verify our primary findings by examining the candidate drugs identified and discussed in the study. Among previously identified drug candidates, our analysis successfully identified baricitinib (targeting JAK1 and JAK2), sarilumab (targeting IL6R) and ritonavir (targeting CYP3A4) as potential drugs for the treatment of SARS-CoV-2 infection (Fig. 5). Since the upstream (IL6R) and downstream (JAK1/2) genes in IL6 signaling pathway were up-regulated, we thus checked the significance of IL6 gene. Indeed, IL6 has successfully been identified as up-regulated DEGs (i.e., passed the significant threshold of three methods, Fig. S7). After prioritization, we found two drugs (siltuximab and sirukumab) as candidate ligands for the treatment COVID-19. Furthermore, we also detected dexamethasone (targeting gene product of NC3R1 and NC3R2) as a candidate ligands for SARS-CoV-2, which was proved to reduce mortality in COVID-19 patients who received respiratory support [19]. These results suggested that our pipeline could be empirically supported by current therapeutic options.

We reasoned that our transcriptomic-driven drug discovery approach may provide a novel perspective regarding the pharmacological mechanisms of drugs that could be considered for repositioning to SARS-CoV-2 infection. To assess the pharmacological actions of the candidate ligands, pharmacological and chemical properties of drugs were annotated using the Anatomical Therapeutic Chemical (ATC) classification code. As a result, we identified 345 (17.87% of 1931) ligands with available ATC code which may target 92 druggable genes. A permutation test was conducted to assess the enrichment of specific ATC
groups for the 92 unique targetable genes and revealed significant enrichment in 16 out of 57 (28.07%) ATC groups (FDR < 0.1; Fig. 6a). Especially, the top 3 most significantly enriched ATC groups with FDR < 0.05 were “ophthalmologicals” (FDR = 0.00057), “immunosuppressants” (FDR = 0.00086), and “drugs for obstructive airway diseases” (FDR = 0.0268).

Since “immunosuppressants” was identified as an enriched ATC term, implying the roles of immune-related functions for drug repositioning to COVID-19. Therefore, we investigated whether other immune-relevant molecules could be exploited for drug repositioning in SARS-CoV-2 infection, which exemplified by immune checkpoint markers, such as PDCD1 and HAVCR2. However, in this study, differentially expressed tests and following drug repositioning analyses were derived from virus-infected human Calu-3 cells. Since T cell data was not being included in our research, we could not analyze the differential expression pattern of PDCD1 and HAVCR2, which mainly be expressed on T cells. We thus investigated the DEG results of CD274 and PDCD1LG2 genes due to the following reasons: (1) CD274 and PDCD1LG2 were druggable by immune checkpoint inhibitors (ICIs) and (2) CD274 and PDCD1LG2 could bind to PDCD1 gene product to trigger T-cell exhaustion. We found a significant up-regulation of CD274 (but not for PDCD1LG2) as DEG, and thus identified atezolizumab, avelumab, and durvalumab (targeting gene product of CD274) as candidate ligands.

### 3.5. Prioritization of candidate ligand-gene pairs for drug repositioning to SARS-CoV-2 infection

Given the substantial number of identified drugs with the potential to be repositioned for treatment of SARS-CoV-2 infection, we utilized several drug-based (a.k.a. ligand-based) and gene-based criteria to prioritize our candidate drug-gene pairs. For six identified clusters, we first restricted the results to “Cluster 1” genes based on following reasons (Fig. S8): (1) known candidate targetable genes (JAK1, JAK2, CYP3A4, IL6R, NC3R1 and NC3R2) were restricted in Cluster 1; (2) In compared to Cluster 1 (18% ~ 61%, approximately), Clusters 2–6 showed a lower level of intersection to the ∆ORF6-infected DEGs (0% ~ 11%, approximately); (3) For Cluster 1, a number of 236 druggable genes were found. However, for Clusters 2–6, only 0–1 druggable gene was identified; (4) Similar results were inspected while examining the number of drug-gene pairs and number of candidate ligands. In light of the underrepresentation of ∆ORF6-infected DEGs and candidate druggable genes in
Clusters 2–6, Cluster 1 was focused for drug repositioning, and resulting in 236 (77.38%) unique genes and 1467 (75.97%) unique ligands. We then leveraged a network (cluster) property, so-called within-cluster connectivity (WCC; i.e. the sum of weights connecting to the gene) to perform gene-based prioritization. We compared the distributions of WCC scores between druggable genes and the remaining genes and found that WCC scores of druggable genes were significantly lower than that of the remaining genes in Cluster 1 (asymptotic general independence test $Z$ value = 2.85; $p$-value = 0.0043; Fig. 6 b). This result indicated that genes with lower connectivity (non-hub) tend to be developed as a pharmacotherapeutic target. We therefore selected the genes with a WCC score of $\leq 2177.606$, which was the WCC score of NR3C1 (targeted by dexamethasone), and resulting in 188 druggable genes (Fig. 6 c). These 188 genes corresponded to 1165 candidate ligands. Notably, JAK2 and IL6R were filtered in this step due to higher WCC scores.

Additionally, in consideration of some candidate ligands from IUPHAR PHARMACOLOGY database were under investigation and thus may not be suitable for drug repositioning, we performed ligand-based prioritization by limiting our candidate drugs to those with ATC annotation. Finally, 55 druggable genes and 238 candidate ligands were discovered for drug repositioning to SARS-CoV-2 infection (Fig. S9).

To validate our findings, we next asked whether some of the candidate drugs for repositioning to SARS-CoV-2 infection identified from our study are in ongoing clinical trials for COVID-19. Indeed, our analysis identified seven drugs (including ritonavir, baricitinib, tofacitinib, naproxen, budesonide, ciclesonide and formoterol) in ongoing clinical trials. These seven drugs target 7 genes found in our study, including CYP3A4, JAK1, PTGS1, PTGS2, NR3C1, NR3C2 and ADRB2. It is noteworthy that dexamethasone was also been identified as candidate ligands, which targets PTGS1 and PTGS2 genes (Fig. 7).

4. Discussion

With the pandemic threat of COVID-19 as it currently stands and the lack of time to develop new medical intervention for over millions of hospitalized victims, devising an empirical yet rationally effective therapeutic strategy is in urgent need. Unfortunately, no preclinical omic-level data from SARS-CoV-2-infected human tissues, especially for lung transcriptomics, are currently not available, posing substantial challenges for drug development or repositioning. A feasible, appealing approach is to use pre-existing data that closely resembled SARS-CoV-2-infection for bioinformatics analyses, thereby identifying candidates of currently available medications to be repositioned for SARS-CoV-2 infection.
therapy. In this study, we performed such task by analyzing the gene expression profile of a human lung cell line, Calu-3, infected with SARS-CoV-ΔORF6, a mutant closely resembling SARS-CoV-2. The rationale for using the mutant as a surrogate is justified by the fact that both the mutant and SARS-CoV-2 lack the antagonistic effect toward interferon response due to a truncation in ORF6. Our method of transcriptomics-driven drug discovery ultimately provided a candidate list of marketed drugs (approved for other indications) to be used in empirical clinical therapy for the urgent need of SARS-CoV-2 infection.

Specifically, we prioritized 55 targetable genes that respond to ΔORF6 infection. These genes can be modulated by 238 candidate drugs that are currently available and hence can be repositioned for COVID-19 therapy. The reliability of our analyses were justified by the following facts. First, we predicted several drugs including ritonavir, baricitinib, tofacitinib, naproxen, budesonide, ciclesonide and formoterol that are currently under clinical trials. Second, despite of an intensive discussion on the empirical use of several drugs including hydroxychloroquine/chloroquine, remdesivir, lopinavir and ritonavir, interleukin (IL)-1 and – 6 inhibitors, and Janus kinase (JAK) inhibitors, no consistent or conclusive evidence for the safety and efficacy is available, except for dexamethasone, which remarkably was on our list of candidates. As such, the analyses we provided support the idea that analyzing transcriptomic data from virus-infected cells can yield valuable information for potential treatments, which can further be leveraged to identify potential candidates for drug repositioning.

Without any preclinical data or clinical experiments, current therapeutic options on COVID-19 infection are lacking in solid evidence and empirically limited to antiviral agents and host modifiers. In light of recent studies and reviews on potential drugs for COVID-19 therapy [18,20–23], our list of candidates contains unexpected drugs. These candidate drug categories can expand the scope of COVID-19 drug targets, open new avenues for future study, and include agents with solid track records in clinical treatment of infectious disease for empirical tests. For instance, we identified several repositioned candidates (e.g., ophthalmological agents, immunosuppressive agents, drugs for obstructive airway diseases, etc.), which have proven useful in treating other viral infections. Also identified were drugs for obstructive airway diseases (e.g. formoterol and ciclesonide) as candidate for repositioning to COVID-19. Notably, a previous in vitro study has shown that combination use of formoterol with budesonide (both were identified as...
candidate ligands) and glycopyrronium can inhibit the viral replication and cytokine production of HCoV-229E in human respiratory epithelial cells [24,25]. Besides, several studies have suggested that ciclosporine may block the replication of SARS-CoV-2 ribonucleic acid and thus reduce the virus-induced cytopathogenic effect in COVID-19 patients [26,27]. Thus, it appears reasonable to reposition and include drugs from other categories for COVID-19 treatment.

One intriguing feature in our prediction was the emergence of drugs for immune modulation. Yet several lines of molecular clinical evidence suggest that taking into account immune modulation might be necessary for COVID-19 therapy. First, the pathology of the SARS-CoV-2 infected lungs showed mixed features of acute alveolar injury and diffusive immune responses including immune infiltration, focal reactive hyperplasia, and intra-alveolar fibrillar plug formation [5]. Second, the clinical severity of COVID-19 infection is linked to highly inflammatory responses that produce enormous chemokines implicated in a cytokine storm [6,7]. Third, a key difference between SARS-CoV and SARS-CoV-2 is a truncation of the open reading frame (ORF) 3b, ORF6, in SARS-CoV-2, which affects the viral antagonistic effect toward interferon response [5]. On our list, we found a significant enrichment in the category of “anti-inflammatory and antirheumatic products”, which may alleviate immune-inflammatory injury that is known to cause increased oxidative stress and organ failure in virus-infected patients [28,29]. Also identified was β2-adrenergic signaling (elicited by some drugs for obstructive airway diseases such as formoterol, salbutamol, etc.) to be critical in norepinephrine-mediated processes and in the serine of TNFα and IL-10, which may relieve tissue-damaging inflammation (and thus exerts a protective effect to patients) during viral infection [30].

Most striking was CD274 (PD-L1) identified as druggable gene in this study, suggesting the plausibility of repurposing ICIs (immune checkpoint inhibitors) for COVID-19 therapy. We attribute this finding to the cell lines applied in current analysis, namely, Calu-3, a human lung cancer cell line. These drugs include atezolizumab, durvalumab. The idea of implementing ICIs for COVID-19 therapy is supported by a previous study showing that T-cell exhaustion markers (PDCD1 and HAVCR2) were overexpressed in severe cases of COVID-19 patients [31]. Two recent commentary articles also addressed the possibility of using ICIs to reinvigorate exhausted T-cell in the early phase of COVID-19 infection, yet the discussion was mainly focused on cancer patients receiving ICI treatment [32,33]. Should ICIs be given during pandemic conditions, especially for cancer patients? We note that drugs targeting immune checkpoints may lead to immune checkpoint inhibitor-related pneumonitis (ICI-P) at a relatively low occurrence rate (4–13%) [34–37]. Given that immunostimulants such as ICIs may exacerbate the clinical symptoms (especially pneumonitis) of SARS-CoV-2 infection, repositioning of these drugs to COVID-19 should be evaluated with great caution. Moreover, it may be useful to consider combining ICIs with other drugs that mitigate unwanted IC-I-induced inflammation, such as IL6 and JAK2 inhibitors [38–40]. Consistently, the potential of targeting IL6 signaling pathway for COVID-19 therapy has already been considered [18,41]. Of note, our analyses pinpointed several ligands targeting IL6, IL6R, JAK1 and JAK2, suggesting the possibility of using ICIs with IL6 signaling pathway blockers in a combination therapy.

For COVID-19, concerns have been raised on enhancing SARS-CoV-2 virulence by treatment of widely used drugs, such as angiotensin-converting enzyme 2 (ACE2) inhibitors and angiotensin-receptor blockers (ARBs) [42–45]. Consistent with this concern, we did not detect significant enrichment for the category “agents acting on the renin-angiotensin system”. To maximize the accuracy of our drug repositioning predictions, we prioritized the initially identified 2396 drug-gene pairs into a list of 296 (12.35%) drug-gene pairs. The final list was restricted to (i) drugs with gene target(s) in “Cluster 1”, (ii) drugs with target gene(s) with lower WCC compared to dexamethasone, and (iii) drugs with an available ATC code. Excluding candidate drug-gene pairs that lack ATC annotations, we identified high-priority drug-gene pairs with likely disease relevance. Note that our analysis may be undermined by a lack of data from preclinical cases of SARS-CoV-2 infection and a lack of clinical specimens from COVID-19 patients. It is possible that the DEGs identified from the surrogate system may not fully reflect genes that respond to SARS-CoV-2 and promote disease progression in COVID-19 infection. Another potential caveat is that we did not take into account physical binding affinities in drug-gene interactions. Moreover, since our analyses were based on data from infected cells, we might not be able to identify candidate drugs that act on viral proteins or human gene products involved in the process of viral entry. For example, ACE2 and TMEM16S were identified as the entry factors for SARS-CoV-2 [46]. Among those genes we analyzed, TMEM16S was identified as a differentially-expressed gene, and TMEM16S-regulating variants were more frequently found in European and American populations [47]. However, we failed to detect the corresponding candidate ligand targeting TMEM16S. Neither was ACE2 identified as a DEG in our study. Based on the ATC enrichment test, we indeed failed to identify enrichment of “Agents acting on the renin-angiotensin system” in the candidate drug list. Thus, it is likely that the renin-angiotensin system may play an important role in the early stage of infection but not associate with the ongoing pathogenesis after viral infection. Although our initial attempt was to predict drugs for empirical clinical use, we should point out that the clinical feasibility of using candidate drugs for repositioning was not accessed in our study. As such, comorbidities and potential drug-drug interactions arising from polypharmacy in COVID-19 patients should be carefully considered and investigated before repositioning specific drugs and regimens for COVID-19 therapy. Likewise, preclinical experiments are needed to demonstrate the potential efficacy of identified drugs, and preclinical evidence should be gathered before moving to clinical application.

5. Conclusions

In summary, our results reveal potential candidates for drug repositioning to SARS-CoV-2 therapy. The candidates we found include drugs from categories other than “antivirals for systemic use” (based on ATC code), suggesting that drugs with various indications should be considered in future studies. Given the scarcity of omics-level data from SARS-CoV-2 infected lung tissue, our surrogate transcriptomic-driven drug discovery approach has prioritized the list of potential candidate drugs to be experimentally examined in preclinical studies, to be considered for empirical clinical use when it is needed, and to open novel therapeutic avenues for the treatment of COVID-19.

Authors’ Contributions

H.W.S.C., G.H.L., Y.O. and W.C.C. designed the study. H.W.S.C. and G.H.L. performed the bioinformatics. H.W.S.C., C.L.G., G.H.L., Y.O. and W.C.C. analyzed and discussed the data. H.W.S.C., G.H.L., K.Y.L., Y.O. and W.C.C. interpreted the data. H.W.S.C., C.L.G., G.H.L. and W.C.C. wrote the paper.

Funding

This work was supported by grants from Ministry of Science and Technology, Taiwan (MOST109-2314-B-038-151 and MOST109-2628-B-038-012), and Taipei Medical University, Taiwan (12310-106079; Yusuke Nakamura Chair Professorship).

Declaration of Competing Interest

The authors declare that they have no known conflicts of interest associated with this work.
Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2020.12.041.

References

[1] C.I. Paules, H.D. Marston, A.S. Fauci, Coronavirus infections—more than just the common cold, JAMA 323 (2020) 707–708.

[2] A.H. de Wilde, E.J. Snijder, M. Kikkert, M.J. van Hemert, Host factors in coronavirus replication, Curr. Top. Microbiol. Immunol. 419 (2018) 1–42.

[3] M. Letko, A. Marzi, V. Munster, Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses, Nat. Microbiol. 5 (2020) 562–569.

[4] J. Gu, C. Korteweg, Pathology and pathogenesis of severe acute respiratory syndrome, Am. J. Pathol. 170 (2007) 1136–1147.

[5] Z. Xu, L. Shi, Y. Wang, J. Zhang, L. Huang, C. Zhang, S. Liu, P. Zhao, H. Liu, L. Zhu, et al., Pathological findings of COVID-19 associated with acute respiratory distress syndrome, Lancet Respir. Med. 8 (2020) 420–422.

[6] X. Li, M. Geng, Y. Peng, L. Meng, S. Lu, Molecular immune pathogenesis and diagnosis of COVID-19, J. Pharm. Anal. 10 (2020) 102–108.

[7] M. Liao, Y. Liu, J. Yuan, W. Wen, G. Xu, J. Zhao, L. Chen, J. Li, X. Wang, F. Wang, et al., Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19, Nature Medicine 26 (2020) 842–844.

[8] K.G. Lokugamage, A. Hage, M.d. Vries, A. M. Valero-Jimenez, C. Schindewolf, B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis, Stat. Appl. Genet. Mol. Biol. (2005) 4. Article17.

[9] G. Yu, L.G. Wang, H. Yan, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (2012) 284–297.

[10] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, G. Yu, Q.Y. He, ReactomePA: an R/Bioconductor package for reactome pathway enrichment analysis and visualization, Mol. BioSyst. 12 (2016) 477–479.

[11] B. Zhang, S. Horvath, A framework for weighted gene co-expression network analysis, Stat. Appl. Genet. Mol. Biol. 12 (2016) 477–479.

[12] S.D. Harding, J.J. Sharman, E. Facenda, C. Southan, A.J. Powson, S. Ireland, A.J. G. Gray, L. Bruce, S.P.H. Alexander, S. Anderson, et al., The IUPHAR/BPS guide to PHARMACOLOGY in 2018: updates and expansion to encompass the new guide to IMMUNOPHARMACOLOGY, Nucleic Acids Res. 46 (2018) D1091–D1106.

[13] S.K. Mulik, O. Hamid, J. Larkin, A. Covre, M. Altomonte, L. Calabro, S.A. Vardhana, C. Robert, R. Ibrahim, A. Anichini, et al., Immune checkpoint inhibitors for cancer therapy in the COVID-19 era, Clin. Cancer Res.: Off. J. Am. Assoc. Cancer Res. 26 (2020) 4201–4205.

[14] J. Domaga-Kulawik, Immune checkpoint inhibitors in non-small cell lung cancer – towards daily practice, Adv. Respir. Med. 86 (2018).

[15] K.R. Kalisz, N.H. Ramaiya, K.R. Lauthamp, A. Gupta, Immune checkpoint inhibitor therapy-related pneumonitis: patterns and management, Radiographics 39 (2019) 1923–1937.

[16] B.T. Larsen, J.M. Chae, A.S. Dixit, T.-E. Hartman, T. Peikert, A.C. Roden, Clinical and Histopathological features of immune checkpoint inhibitor-related pneumonitis, Am. J. Surg. Pathol. 43 (2019) 1331–1340.

[17] J.Y. Cho, J. Kim, J.S. Lee, Y.J. Kim, S.H. Kim, Y.J. Lee, Y.J. Cho, H.I. Yoon, J.H. Lee, C.T. Lee, J.S. Park, Characteristics, incidence, and risk factors of immune checkpoint inhibitor-related pneumonitis in patients with non-small cell lung cancer, Lung Cancer 125 (2018) 150–156.

[18] R.J. Sullivan, D.B. Johnson, B.I. Rini, T.G. Neilan, C.M. Lovly, J.J. Moslehi, L. Reynolds, COVID-19 and immune checkpoint inhibitors: initial considerations, J. Immunother. 8 (2020).

[19] S. Di Cosimo, A. Malfetone, J.M. Pérez-García, A. Llombart-Cussac, R. Miceli, G. Curigliano, J. Cortes, Immune checkpoint inhibitors: a physiology-driven approach to the treatment of coronavirus disease 2019, Eur. J. Cancer 135 (2020) 62–65.

[20] C.R. Straw, A. Hegdé, C. Churry, A.B. Nagabh, N. Sharma, S. Addepalli, S. Cherukuri, T. Parent, J. Hardin, P. Walker, Tocilizumab for the management of immune mediated adverse events secondary to PD-1 blockade, J. Oncol. Pharm. Pract. 25 (2019) 355–357.

[21] S. Atal, Z. Fatima, IL-6 inhibitors in the treatment of serious COVID-19: a promising therapy? Pharm. Med. (2020) 1–6.

[22] J.A. Wilkins, Preventing a covid-19 pandemic, BMJ 368 (2020) m810.

[23] M. Esler, D. Esler, Can angiotensin receptor-blocking drugs perhaps be harmful in COVID-19? BMJ 327 (2020) 1477–1478.

[24] D.M. Halpin, D. Singh, R.M. Hadfield, Inhaled corticosteroids and COVID-19: a systematic review and clinical perspective, Eur. Respir. J. 55 (2020), 2001099.