Integrative transcriptome analyses empower the anti-COVID-19 drug arsenal

Nehme El-Hachem (nehme.el-hachem@umontreal.ca)
University of Montreal; American University of Beirut https://orcid.org/0000-0002-5778-5342
Edward Eid
American University of Beirut https://orcid.org/0000-0002-9355-2474
Georges Nemer
Hamad Bin Khalifa University, Doha, Qatar https://orcid.org/0000-0003-2157-5279
Ghassan Dbaibo
American University of Beirut,Lebanon
Ossama Abbas
American University of Beirut
Nelly Rubeiz
American University of Beirut
Salah Zeineldine
American University of Beirut
Ghassan M. Matar
American University of Beirut,Lebanon https://orcid.org/0000-0002-4691-8998
Jean-Pierre Bikorimana
University of Montreal, Montreal, QC, Canada
Riam Shammaa
University of Toronto, ON, Canada
Benjamin Haibe-Kains
University of Toronto, ON, Canada https://orcid.org/0000-0002-7684-0079
Mazen Kurban (mk104@aub.edu.lb)
American University of Beirut
Moutih Rafei (moutih.rafei.1@umontreal.ca)
University of Montreal, Montreal, QC, Canada

Research Article

Keywords: SARS-CoV-2, COVID-19, transcriptomics, drug repurposing, ACE2

Posted Date: July 15th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-37528/v2

Version of Record: A version of this preprint was published at iScience on November 1st, 2020. See the published version at https://doi.org/10.1016/j.isci.2020.101697.

Abstract

The beginning of the twenty-first century has been marked by three distinct waves of zoonotic coronavirus outbreaks into the human population. The current pandemic COVID-19 (Coronavirus disease 2019) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). With a rapid infection rate, it is a global threat endangering the livelihoods of millions worldwide. Currently, and despite the collaborative efforts of governments, researchers, and the pharmaceutical industries, there are no substantially significant treatment protocols for the disease. To address the need for such an immediate call of action, we leveraged the largest dataset of drug-induced transcriptomic perturbations, public SARS-CoV-2 transcriptomic datasets, and expression profiles from normal lung transcriptomes. Our unbiased systems biology approach not only shed light on previously unexplored molecular details of SARS-CoV-2 infection (e.g., interferon signaling, inflammation and ACE2 co-expression hallmarks in normal and infected lungs) but most importantly prioritized more than 50 repurposable drug candidates (e.g., Corticosteroids, Janus kinase and Bruton kinase inhibitors). Further clinical investigation of these FDA approved candidates as monotherapy or in combination with an antiviral regimen (e.g., Remdesivir) could lead to promising outcomes in COVID-19 patients.

Introduction

On December 31st, 2019, the world held its breath as the first cases of pneumonia of unknown etiology detected in the Chinese city of Wuhan were brought to light by the WHO China Country Office. Ever since, this new pandemic caused by the single strand positive RNA virus, SARS-CoV-2 spread quickly and caused a worldwide global panic. By May 22nd, 2020, the number of infected individuals has exponentially increased to 4,993,470 cases with the global death toll reaching 327,738 deaths 1. This dramatic rise in cases owes itself on one hand to the nonspecific flu-like manifestations that afflict the vast majority of COVID-19 patients and on the other to the heavy-travel profile characteristic of most populations in the current era of globalization. COVID-19 can clinically manifest on a spectrum ranging from mild non-specific flu-like symptoms to near-fatal Acute Respiratory Distress Syndrome (ARDS) leading to inflammation, pneumonia, acute lung injury (ALI), and sepsis especially in the elderly and persons with underlying comorbidities 2,3,4.

The current understanding of COVID-19 pathogenesis, etiology, and clinical features has been first extrapolated from knowledge gathered from other similar zoonotic infections associated with upper respiratory illness, such as the severe acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East respiratory syndrome coronavirus (MERS-CoV) 5,6,7,8. It is now established that SARS-CoV-2 specifically recognizes the human angiotensin-converting enzyme 2 protein (ACE2) as its main target on the cell surface 9,10. Following virus entry, the cell triggers the physiological response to the infection through several defense arms that are also thought to be implicated in the pathogenesis of the infection Thus, characterizing the downstream transcriptional hallmarks of COVID-19 is crucial to illuminating the mechanisms of this deadly respiratory infection and to guide potent therapeutic countermeasures. However, with the urgency of the current pandemic situation, developing new drugs capable of combating

COVID-19 is deemed difficult. To that end, drug repurposing represents a rapid approach to expediting the process of finding a therapeutic candidate[11,12]. Cases of drug repurposing have been serendipitous rather than data-driven discoveries and this extrapolates to COVID-19 research. The hydroxychloroquine-azithromycin combination has been proposed as a panacea for COVID-19 but has not been backed by a proven scientific rationale. Despite initial fanfare generated from favorable outcomes noted in an open-label, non-randomized clinical trial [13], the azithromycin-hydroxychloroquine combination fell short of expectations in later trials with therapeutic inefficacy overshadowed by cardiovascular side effects associated with its use [14–17]. This setback highlights the dire need for an unbiased methodology to further identify drug-target associations that can be efficiently translated to the clinic.

Motivated by both an unmatched global pandemic and a lack of plausible drug candidates to treat COVID-19, we have conducted a unique data-driven approach (Figure 1) that integrates three types of data: 1) A public resource of SARS-CoV-2 transcriptomics [18] 2) Drug-induced transcriptome profiles from the Library of Integrated Network-based Cellular Signatures (LINCS) Connectivity Map project [19] and 3) transcriptome-wide profiling in normal lung tissue from the Genotype-Tissue Expression (GTEx) project [20]. Briefly, our analyses highlighted several targetable hallmarks associated with the SARS-CoV-2 pathobiology and underscored plausible drug-target associations among which top-scoring hits have been acknowledged in ongoing COVID-19 clinical trials. Furthermore, our approach provides the largest set of pre-clinical candidates and FDA approved drugs with known safety profiles to be further investigated for COVID-19 treatment. In comparison to other studies which have used the old version of Connectivity Map [21,22] or the limited L1000 web interface [23], the novelty of our work is twofold: First, it provides the largest repertoire of drug repurposing candidates across available SARS-CoV-2, *in vivo,* and *in vitro* settings [R Shiny WebApp]. Second, it has identified drug candidates that interfere with ACE2 co-expressed signatures in SARS-CoV-2 infected settings.

Methods

SARS-CoV-2 datasets

We collected transcriptomic profiles from:

1) Cell lines infected with SARS-Cov-2 (Multiplicity of infection: 2) and COVID-19 patients from GEO: GSE147507 dataset. The in vitro setting includes A549 cells supplemented with a vector expressing ACE2, Calu-3 adenocarcinoma cells, and Human Bronchial Epithelial Cells (NHBE). Mock-treated cells were provided for each of the in vitro groups (N=3 per group). For the in vivo setting, two lung samples derived from COVID-19 patients were compared against two biopsied healthy lung tissues. The Ruxolitinib-treated group consisted of A549 cells overexpressing the ACE2 receptor and pre-treated with Ruxolitinib, a Janus kinase 2 inhibitor (JAK2i). A complete description of the dataset can be found in Blanco-Melo et al.[18]

2) Bronchoalveolar lavage fluid (BALF) samples from two COVID-19 patients 24. BALF healthy control samples corresponding to non-obese, non-asthmatic patients were downloaded from the SRA database with accession numbers: SRR10571724, SRR10571730, and SRR10571732 25

Differential expression analyses

We conducted a differential expression analysis separately for each of the collected RNA-seq datasets using Deseq2 26. Genes with zero counts were removed and shrunken values of the fold change were computed for later use with the gene set enrichment tool (GSEA). A gene is considered differentially expressed between the SARS-CoV-2 setting and its corresponding control/mock group if its adjusted p-value (FDR) is below 0.05.

In vivo comparisons are referred to as SARS2_BALF_WUHAN1-2 and SARS2_LUNG whereas in vitro contrasts appear in the manuscript as SARS2_NHBE, SARS2_A549_ACE2, SARS2_Calu3, and SARS2_A549_ACE2_RUXO.

The Genotype-Tissue Expression (GTEx) lung dataset

RNA-seq gene counts for 374 lung samples were downloaded from GTEx through the recount2 interface 27. Raw data were normalized using the variance stabilizing transformation (VST) technique in DESeq2. Gene co-expression matrix (gene-gene correlation) was constructed using the Pearson correlation distance measure. Genes co-expressed with ACE2 in lung samples were ranked from −1 (opposite gene expression patterns) to +1 (identical expression patterns). This is referred to as ACE2_GTEX in the text.

Gene set enrichment analysis

Gene sets were collected from several sources: MsigDB hallmark set 28, WikiPathways 29 , Reactome 30, and a custom set associated with SARS-CoV and SARS-CoV-2 from the literature. This custom gene collection, manually curated, is referred here as geneshot_SARS_CoV downloaded from Geneshot 31, respectively. We run GSEA (gene set enrichment analysis)32 with default settings against all lists of differentially expressed genes ranked by the shrunken log2 fold change, from *in vivo/vitro* SARS-CoV-2 and GTEX-ACE2 settings.

Drug-induced gene expression profiles and similarity search

Combined z-scores by biological replicates from LINCS L1000 Phase I (GSE92742) & Phase II (GSE70138) datasets (only small molecule perturbagen) were downloaded from 33. The pipeline processed raw data from L1000 based on the Bayesian approach and a probability-based z-score inference method that showed improved performance over the original L1000 data. We kept L1000 core

cell lines (A375, A549, HA1E, HCC515, HT29, MCF7, PC3, VCAP) for further downstream analysis. Signatures corresponding to the same perturbagen but different conditions (cell line, duration, and dose), were averaged using the MODZ method to create a consensus signature per chemical perturbagen 19. We applied the cosine similarity 34 to quantify the relationship between the drug-induced gene expression signature from L1000 and the ranked gene signatures from the SARS-CoV-2 groups to identify repurposing candidates. We assessed the significance of the cosine score from 10000 random permutations.

Drug-target databases

We collected and manually inspected drug-target interactions from several sources: DrugBank 35, IUPHAR/BPS Guide to PHARMACOLOGY 35,36, and the repurposing hub (https://clue.io/repurposing). We conducted a drug set enrichment analysis (drug-target associations are used instead of genesets) to assess the enrichment for known drug-target interactions among the ranked list of drugs from the cosine similarity analysis. Protein targets with a minimum of three representative drugs were kept for the analysis. Drug families enriched for negative associations constitute potential repurposing candidates for COVID-19.

Results

To fill the unmet need for effective COVID-10 treatments, we leveraged the Library of Integrated Network-based Cellular Signatures (LINCS) - the largest body of drug-induced transcriptional perturbations profiles available to date, the healthy lung transcriptome from the Genotype-Tissue Expression (GTEx) project and the recently published SARS-CoV-2 induced transcriptomic profiles from *in vivo* and i*n vitro* settings (All settings are described in the method section and illustrated in table 1).

Hallmarks of SARS-CoV-2 infection across *in vivo* and *in vitro* settings

In order to address pharmacological gaps in SARS-CoV-2 research, it was of paramount importance to understand the molecular mechanisms underlying the SARS-CoV-2 infection. To this end, we have collected public SARS-CoV-2 transcriptome-wide datasets and normal lung transcriptomics from GTEx. Subsequently, we conducted a differential expression analysis to compare SARS-CoV-2 conditions with their respective control groups. Next, all genes from the different comparisons were merged by the shrunken log2 fold change values (no-cutoff) to yield a full matrix corresponding to 7433 genes x 7 contrasts. These are referred to as SARS2_BALF_WUHAN1-2, SARS2_LUNG, SARS2_NHBE, SARS2_A549_ACE2, SARS2_Calu3 and SARS2_A549_ACE2_RUXO (ruxolitinib-treated cells). For each of these seven contrasts, the fraction of significant up- and down-regulated genes (|Log2FoldChange| > 1 and p-adjusted-value < 0.05) and the corresponding overlap between contrasts is given in Supplementary

Figure 1. As for the transcriptomic data from the Genotype-Tissue Expression (GTEx) project (N=374 samples), gene-wise Pearson correlation of the *ACE2* gene and all other expressed genes is referred to as ACE2_GTEX, (Pearson correlation $\in$ [-0.47, 0.59]), out of which 7427 genes were shared with the seven SARS-CoV-2 contrasts (Supplementary Table S2) and genes were ranked accordingly.

The ranked lists of genes across all SARS2 contrasts as well as genes ranked by their correlation coefficient from ACE2_GTEX were investigated at the pathway level using gene set enrichment analysis (GSEA). We assessed enrichment against a collection of hallmark genesets and a curated set of genes associated with SARS_CoV infection from the literature (See methods section and Table S3). To assess the similarity between settings, we first computed all pairwise correlations of pathway enrichment scores from all GSEA analysis (Figure 2A). Interestingly, both SARS2_A549_ACE2_RUXO and ACE2_GTEX were strongly concordant (rho= 0.63, $p < 0.01$). This suggests that ruxolitinib treatment in ACE2- expressing A549 cells and ACE2 baseline expression in the lungs engage a similar set of molecular pathways. Indeed, similarly to ACE2_GTEX, Ruxolitinib-induced pathways were highly anti-correlated with those from all the SARS2 settings (Fig 2A). Although SARS2_LUNG was correlated with BALF settings, this correlation was not significant (p-value > 0.05, Fig 2A) suggesting a functional diversity in response to SARS-CoV-2.

We further investigated the enriched hallmarks/pathways in both the normal lung and SARS-CoV-2 infected *in vivo* and *in vitro* settings. By keeping all gene sets with a normalized enrichment score |NES| greater or equal to 1 and a p-adjusted-value threshold of 5%, our analysis identified a highly conserved hallmark process, the TNF Alpha (Tumor necrosis factor-alpha) signaling via NF-κB. This hallmark was significantly upregulated by the SARS-CoV-2 infection but downregulated in both ruxolitinib-treated and ACE2_GTEX groups (Figure 2B). Moreover, to investigate the pathogenic cellular response of SARS-CoV-2, we proceeded as follow: (1) We extracted the union of all leading-edge genes from the conserved hallmark and (2) We extracted leading-edge genes belonging to the following enriched hallmarks: INTERFERON GAMMA RESPONSE, INTERFERON ALPHA RESPONSE, INFLAMMATORY RESPONSE,

IL6-JAK-STAT3 signaling and the curated set of SARS-CoV associated genes from the literature (geneshot_SARS_CoV). It is noteworthy to be mentioned that enriched hallmarks from (2) were negatively scoring genesets in ACE2_GTEX and positively enriched in all SARS2 settings but the WUHAN_BALF samples (Figure 2B). Across all settings, the resulting genes were assigned either the shrunken log2 fold change from their corresponding DEG analysis or the correlation value for ACE2_GTEX (Figure 2C). From this collection of relevant genes of innate immunity, we identified a set of conserved upregulated genes in SARS2 settings (group1, Figure 2C). These same upregulated genes were downregulated in ruxolitinib-treated cells and highly anti-correlated with reference to ACE2_GTEX. Most of these genes (e.g. *CXCL10, CXCL11, OAS1, TNFSF10, MX1, DDX58, ISG15, IFIT2, IFIT3, EIF2AK2*) are induced in response to interferons. Notably, the induction of these genes was much more pronounced in SARS2_LUNG, SARS2_Calu3, and SARS2_A549_ACE2 (Figure 2C). For instance, the pro-inflammatory chemokine CXCL10/IP-10 had a 32 fold-increase in SARS2_LUNG compared with a 3 fold increase in the WUHAN_BALF samples, and it was almost absent in NHBE infected cells. Chemokine's expression in the

postmortem lung sample (SARS2_LUNG) was similar to what is being observed in A549_ACE2 and Calu3 cells infected with high loads of virus (MOI:2). Although MOI of 2 was also used for NHBE cells, RNA-seq profiling 24 hours post-infection showed that these cells efficiently cleared the virus (no change observed for CXCL10-11). In contrast to groups 1 and 2 that were predominantly represented by interferon-induced chemokines and genes regulated by NF-κB in response to TNF alpha, group 3 seems to represent a set of genes belonging to the IL6-JAK-STAT3 pathway (*CSF1, IFNGR2, IL4R, IFNGR1, HMOX1, TYK2, TGFB1, TNFRSF1A*) and inflammatory responses (*PTGER4, IRS2, MAP2K3, AXL, C5AR1, PTAFR*); these genes were antagonized by SARS-CoV-2 specifically in BALF samples and not the cell lines. Lastly, group 4 showed a similar pattern to group 1 and its gene members are also related to the Interferon alpha/beta/gamma, and complement signaling (e.g., *TRIM25, UBE2L6, STAT1, JAK2, IL1B, IL2RG, GBP2, PLAUR*). For instance, STAT1, a known master regulator of interferon genes, has more than a 5 fold-induction in SARS2_LUNG, A549, and Calu3 but its expression was not tangible in the other in vivo settings or SARS2_NHBE. In summary, expression dynamics from the deceased patients are highly similar to those from SARS2_A549_ACE2 and SARS2_Calu3 but dissimilar to WUHAN_BALF and NHBE cells. The former elicits marked pro-inflammatory events which could explain the development of severe cases of COVID-19 symptoms and lung injury as seen in the postmortem samples.

Identification of new COVID-19 repurposing opportunities through mining L1000 connectivity map

In the previous section, we highlighted molecular vulnerabilities providing a rationale for targeting hyperinflammatory response in COVID-19. Although ruxolitinib provided a proof of concept that abrogating inflammation is an appealing therapeutic strategy, we conducted a full data-driven computational approach to identify potential drug candidates that, in addition to ruxolitinib, could perturb the pathological mechanism induced by the new SARS-CoV. To this end, we have interrogated the L1000, the largest repository of chemical-induced gene perturbation profiles. The curation and pre-filtering of the L1000 dataset used in this study are described in the methods section. Briefly, we have assembled expression perturbations of 978 landmark genes across 4487 chemical compounds from LINCS L1000. 727 out of 978 genes (~75%) were shared with our SARS2 and ACE2_GTEX settings. To identify potential drug repurposing candidates, we used the cosine similarity distance between vectors of gene expression signatures. For each of the drug-induced expression profiles, a cosine score was calculated against all SARS2 and ACE2_GTEX settings (shinyApp table provided upon request). A positive cosine score corresponds to similar expression patterns (useful to identify drugs inducing similar modes of action/molecular mechanisms) while a negative score would suggest an opposite expression pattern (likely beneficial in the context of a disease, specifically COVID-19).

We first validated our approach on ruxolitinib signature in SARS-CoV2 infected cells (SARS2_A549_ACE2_RUXO). Our pipeline could accurately and unbiasedly identify Ruxolitinib from L1000 as the second-best matching hit out of 4,487 chemical compounds (cosine score= 0.18, p-value from 10,000 random permutations = 2e-04).

To further ascertain that our approach can identify not only ruxolitinib but its drug target class, we performed a drug set enrichment analysis. In this setting, a drug set is a target gene/protein that is associated with at least 3 drugs. We first ranked drugs by the cosine value and ran the enrichment analysis against 295 target genes (see method section for drug-target preparation and database). With respect to SARS2_A549_ACE2_RUXO, we could confirm that indeed JAK2 inhibitors were significantly overrepresented at the top of the ranked list of cosine scores (ES= 0.825, adjusted p-value=0.08) (Figure 3C).

Next, we investigated drug-target classes that display a negative enrichment score (their drug members likely reverse the phenotype) against a computed consensus cosine score which is the median value of the cosine spanning 4,487 drugs across all SARS2 settings. We postulate that the consensus score would emphasize drug classes that target generic processes induced by SARS-CoV-2 across diverse settings and infectivity phases/severity levels. Furthermore, drug-target enrichment analyses were conducted separately for each of the SARS2 settings (excluding RUXO group). (Figure 3A).

Importantly, when considering an enrichment drug score (ES) less or equal to -0.5 and adjusted-p-value < 0.25 (Figure 3A), our consensus signature exhibited the strongest enrichment for JAK2 inhibitors (filgotinib, baricitinib, fostamatinib), Bruton kinase-BTK inhibitors (GDC-0834, ibrutinib, dasatinib; Figure 3D), CACNA1C blockers (mibefradil, diltiazem, verapamil), HCRTR1 (orexin) antagonists (SB-334867/almorexant/suvorexant), PTGER2 agonists (carbacyclin, treprostinil), KCNA7 Channel blockers (amiodarone, flecainide), ASIC3 Channel blocker (amiloride, nafamostat, diclofenac), PDE5A inhibitors (milrinone, zaprinast, vardenafil, ibudilast, sildenafil) and ADRA1B antagonists (prazosin, tamsulosin, alfuzosin, doxazosin, mianserin, ritanserin, phentolamine, indoramin). (Figure 3A). When leveraging the drug enrichment analysis for each of SARS2 settings, we observed that SARS2_LUNG exhibited very few enriched drug classes. Among these enriched classes, BRD4 inhibitors (BI-2536, colchicine) and NR1I2 agonists (hyperforin, mifepristone, nifedipine, clotrimazole, paclitaxel). Interestingly, corticosteroid agonists (methylprednisolone, triamcinolone, budesonide, mometasone, beclomethasone, dexamethasone, fluocinonide, betamethasone, fluorometholone) were enriched in both WUHAN_BALF, consensus, and NHBE settings.

Given the important role of ACE2 in both SARS-CoV-2 infectivity and normal lung physiology, we next sought to identify drugs that could counteract the SARS-CoV-2 invasion by restoring the baseline co-expression network of ACE2 in the healthy lung.

In this particular analysis and similar to SARS2_A549_ACE2_RUXO, a positive enrichment score is the preferred outcome (Figure 3B). A drug-target class with an enrichment score equal or greater than 0.5 is considered a putative candidate to maintain a normal ACE2 expression whereas a drug-target with a negative enrichment score is likely to be beneficial for the virus. The search for such modulators identified drug-targets highly enriched for inhibitors of cereblon protein (lenalidomide, pomalidomide, thalidomide; ES=0.93, adjusted-p value=0.05) and Bruton Kinase BTK (fostamatinib, ibrutinib, dasatinib, GDC-0834; ES=0.86, adjusted-p value=0.023). Strikingly, Angiotensin-converting enzyme inhibitors (enalapril,

fosinopril, captopril, perindopril) also showed positive connectivity with ACE2_GTEX (ES=0.61, adjusted-p value=0.22). This could be explained that the Angiotensin-converting enzyme (ACE) expression from GTEx was highly anti-correlated with ACE2 (rho = -0.47). Last, and similar to Bruton kinase inhibitors, we identified drug-target classes such as NR3C1 agonists (corticosteroids), PDE5A inhibitors (milrinone, zaprinast, vardenafil, ibudilast, sildenafil), ADRA2B antagonists (phenoxybenzamine, prazosin, bromocriptine, apomorphine, terguride, ARC-239, chlorpromazine, mirtazapine, phentolamine), HTR2A agonists (MK-212, pergolide, pindolol, terguride, donitriptan, quipazine, lorcaserin, aripiprazole), and KCNA10 channel blockers (verapamil, pimozide). These classes might have a dual effect by modulating the inflammatory response, protecting the lung, and restoring ACE2-induced downregulation by the virus (Figure 3E). We are aware that only a limited fraction of the drugs is assigned to a target class (1030 out of 4487 drugs; 23%). As such, all results from our analyses will be made publicly available to accelerate the search for COVID-19 treatments.

Pathways/molecular hallmarks associated with potential modulators of COVID-19

To understand the mechanism by which the candidate drug candidates block the molecular events underlying COVID-19, we focused on significant drug-target lists associated with the previously described consensus score. This involved 25 drug-target families and an ensemble of 106 drugs (Supplementary Figure S2). Given the small number of genes from the L1000 dataset (727 out of the 978 landmark genes are shared with our data), we opted out to choose gene set enrichment analysis but instead, we conducted a row-wise Welch's t-test to identify the landmark genes that show significant changes between the 106 drugs and the SARS2 groups. This analysis yielded a set of 124 genes that displayed extreme values from the two-tailed t-test

(False discovery rate < 5%, 73 genes with positive t-stat values genes, 51 genes with negative t-stat values)(Figure 4A).

When performing a hypergeometric test, separately for the negatively and positively modulated genes considering the set of significant genes, we showed that several down-regulated processes were enriched for Cytokine Signaling, Jak-STAT signaling, Complement activation, TNF-alpha effects on cytokine activity, lung fibrosis and inflammatory response. Gene members include transcripts such as *RALA, CD40, FRS2, CRKL, SOCS2, UBE2L6, STX4, SMAD3, CCL2, IFNAR1, SPRED2, PIK3CA,*

IKBKB(Figure 4B). In contrast, up-regulated processes were enriched for metabolic processes, mainly glucose metabolism, cell cycle, and mRNA processing. Such transcripts include *CTD, LBR, PGM1, NUP85, PFKL, PLCB3, PSMD10, SLC5A6, CAT, NUP133, HK1, HSD17B10, ETFB, NUP62, ADH5, NUP88, NNT, RPN1, GNAI2, VAPB,*

SACM1L (Figure 4C). Finally, although L1000 contains a limited set of genes and lots of transcriptome-wide information is missing, we could confirm that the identified drugs interfere with the hallmarks of the

SARS-CoV-2 infection described earlier. More importantly, most of these drugs are FDA approved and patients would benefit from them before reaching severe courses of COVID-19.

Discussion

In this study, we focus on (*i*) defining aspects of molecular pathogenesis with regards to SARS-CoV-2 and (*ii*) validating potential therapeutic targets using integrative data-driven computational analyses. This was carried on large transcriptomic datasets encompassing both in vitro and in vivo SARS-CoV-2 infected samples as well as normal human lung biopsy samples in order to provide a global solid basis to validate our results.

GSEA was initially carried out on all of the above datasets to ascertain the molecular pathways involved in SARS-CoV-2 pathogenesis. TNF alpha signaling via NFkB was found to be a highly conserved hallmark process across all SARS-CoV-2 models. This aligns with published research on the pathogenesis of beta-coronaviruses (SARS-CoV in MERS-CoV)37,38,3939,40,41. Our analysis provides further validation of the central role of TNFα signaling via NF-κB in SARS-CoV-2 pathogenesis. It has been reported that serum levels of TNFα along with many other pro-inflammatory cytokines and chemokines were higher in SARS-CoV-2 ICU patients relative to their non-ICU counterparts 41,42. TNFα plays an instrumental role in orchestrating the clinical outcomes seen in severe SARS-CoV and SARS-CoV-2 infections (reviewed in 43). Besides its role in the ominous cytokine storm, TNF-α strongly induces hyaluronan-synthase-2 (HAS2) in pulmonary epithelial and fibroblasts potentially leading to ARDS 44. Downregulation of ACE2 expression by SARS-CoV-1 S protein has been incriminated in the acute lung injury characteristic of severe SARS-CoV and possibly SARS-CoV-2 45,10. In one study, it was shown that SARS-CoV S protein downregulates ACE2 by upregulating TNFα production which in turn acts in an autocrine fashion to induce the TNF-α-converting enzyme (TACE)-dependent shedding of the ACE2 ectodomain 46.

Our analyses highlighted a strong transcriptional response and induction of a subset of pro-inflammatory cytokines and interferon-induced genes (ISGs). Subsequently, the in vivo dataset also allowed us to dichotomize the lung biopsy and BALF samples into two degrees of clinical severity. Lung biopsy samples are likely to represent severe SARS-CoV-2 groups, as they were derived from SARS-CoV-2 deceased victims, while milder SARS-CoV-2 infection was represented by BALF samples. Stronger estrogen enrichment in BALF samples as opposed to the SARS-CoV-2 samples alludes to a possible protective role for estrogen in SARS-CoV-2 infection. This notion is further strengthened by epidemiologic data showing a preponderance of males in patients with severe SARS-CoV-2 infection 47–49. Furthermore, estrogen (E2) was found to decrease ACE2 expression in vitro cell models and thus potentially regulate SARS-CoV-2 infection 47,48. Interferon-related genes such as STAT1 and CXCL10 as well as inflammatory response and IL-6-JAK-STAT pathway genes were significantly upregulated in the lung biopsy samples and minimally expressed in BALF tissue samples. The latter result possibly stems from the impressive arsenal of anti-IFN molecular tricks that beta coronaviruses have evolved to evade the impending immune response. The SARS-CoV nsp1 protein has demonstrated the ability to interfere with IFN-signaling pathways by decreasing the levels of phosphorylated STAT1 50 while MERS-CoV

membrane protein (M protein) and SARS-CoV ORF3b and -6 were found to prevent nuclear IRF3 translocation and subsequent interferon-stimulated genes expression 51,52. A testament to its pathogenic prowess, SARS-CoV-2 was found to trigger significantly lower levels of IFNs than SARS-CoV while boasting a superior infection and replication rate in the lungs 53. In case of progression into severe SARS-CoV-2, the immune milieu adopts the transcriptomic profile of the lung sample dataset. Unabated viral replication eventually induces the delayed release of IFN-α/β and the subsequent influx of inflammatory macrophages. Consequently, proinflammatory cytokines (TNFalpha, IL-6, and IL1-β) are released to devastating effect. Inevitably, the aforementioned “cytokine storm” hinders viral clearance via T cell apoptosis and damages pulmonary microvascular and alveolar epithelial cell barriers via endothelial and epithelial cell apoptosis. The resultant vascular leakage and alveolar edema induce ARDS and lung damage 54. Interestingly and in stark contrast to the other in vitro SARS-CoV-2 datasets, the NHBE SARS-CoV-2 set failed to register any tangible hallmark-associated gene expression. This could be explained by the cell tropism of SARS-CoV-2 in the human lungs that favors type I pneumocytes, type II pneumocytes, and alveolar macrophages over bronchial epithelia 53. Alternatively, NHBE might be innately capable of clearing the virus. Elucidation of the exact cause might be addressed in future cellular studies.

To identify potential repurposing candidates, we mined the L1000 connectivity map for drug inducing SARS-CoV-2-relevant gene perturbations. The ranked list of drugs was then subjected to drug enrichment analysis targeting 2 aspects of SARS-CoV-2 treatment: the reversal of the inflammatory SARS-CoV-2 phenotype and the maintenance of normal ACE2 expression. The output of these aforementioned analyses yielded a panoply of repurposing candidates that conformed to either one of the mentioned aspects of treatment or both. This data achieves 2 major goals depending on the drug in question. This data provides novel therapeutic targets for SARS-CoV-2 treatment. Investigators are then encouraged to seek further validation, biological or otherwise, en route to clinical consideration. For output drugs that are currently being trialed or under trial consideration such as Colchicine (NCT04375202), Chlorpromazine (ReCoVery trial; NCT04366739), Corticosteroids (NCT04381936), Prazosin (PREVENT trial; NCT04365257), lenalidomide (GETAFE trial; NCT04361643), Thalidomide (NCT04273581), Ibrutinib (NCT04375397) and Sildenafil (NCT04304313) among others, our results provide an added layer of validation to imbue the concerned investigator with added confidence moving forward. As the drug class with the highest enrichment in our analysis, currently trialed JAK2 inhibitors such as ruxolitinib (NCT04348071) and baricitinib (BARI-COVID trial; NCT04320277) present promising therapeutic repurposing targets for SARS-CoV-2 treatment. Blockade can target IL-6 signaling via JAK-STAT, a process that was enriched in our hallmark analysis and a powerful driver of immune dysregulation in SARS-CoV-2 vis-à-vis its lymphopenic effect 53,55 as well as the pro-inflammatory angiotensin 2-AT1R signaling via JAK-STAT.

Other enriched targets interfere with viral entry. During SARS-CoV and SARS-CoV-2 entry, cell surface-associated transmembrane protease serine 2 (TMPRSS2) cleaves the S protein to permit viral and cellular membrane fusion 56. Nafamostat (RACONA trial; NCT04352400) and to a lesser extent, camostat (NCT04353284), both enriched in our analysis, have been shown to prevent SARS-CoV-2 entry by

inhibiting TMPRSS2 10,56,57. Clathrin-mediated endocytosis then ensues with regulation afforded by numb-associated kinases (NAK), such as AP2-associated protein kinase 1 (AAK1) and cyclin G-associated kinase (GAK) 58. In addition to the JAK-STAT blockade discussed above, the JAK inhibitor baricitinib might also disrupt viral entry via AAK1 and GAK inhibition 59. Interestingly, chloroquine and hydroxychloroquine were not significantly enriched in our drug analysis. This pertinent negative finding is in line with the failure of hydroxychloroquine to demonstrate clinical efficacy. In effect, this finding lends credence to our data-driven analysis as it outperformed extrapolated biological validation of chloroquine and hydroxychloroquine, albeit on non-SARS-CoV-2 cell models.

Finally, and given the limited set of genes profiled in LINCS L1000, we investigated whether the observed drug mechanism of action is associated with molecular events triggered by SARS-CoV-2 and not a ubiquitous modulation of pathways. Notably, the enriched drug candidates down-regulated proinflammatory processes such as complement activation, JAK-STAT pathways, and lung fibrosis. Conversely, metabolic processes such as transcription, glucose metabolism, and cell cycle processes were upregulated. The latter resumption of normal function attests to SARS-CoV-2's disruptive influence on cellular resources during infection. This is demonstrated by the nsp1 protein expressed by SARS-CoV during infection. SARS-CoV nsp1 achieves translational inhibition either via direct binding to the 40S subunit and cleavage of cellular mRNA or via inhibition of the 48S initiation complex formation and its conversion to the 80S initiation 60–62. This results in the shutoff mechanism whereby viral RNA is preferentially transcribed and translated over host mRNA. Moreover, the massive metabolic suppression observed concur with recent findings from proteomic and metabolomic characterization from a severe case of COVID-19 63.

In conclusion, through a rigorous data-driven computational analysis, this study validates aspects of a complex pathogenetic framework for SARS-CoV-2 infection; however, targeted biological validation is still a requisite for a comprehensive understanding of the pathogenesis. A significant number of therapeutic repurposing candidates, many of which are FDA-approved, safe, and cheap, are presented to the scientific community for further validation and clinical consideration. The different mechanisms of action of the various validated candidates serve to emphasize the myriad of pathways involved in the pathogenesis and the implicit complexity of the virus. With that in mind and as time is of the essence, the notion of combining drug candidates with different mechanisms of actions might provide a synergistic effect while simultaneously decreasing the risk of associated adverse drug reactions and expediting the process of testing each therapeutic candidate based on the severity of COVID-19 symptoms.

Declarations

Contributions

N.EH conceived the study, performed all bioinformatic analyses and wrote the manuscript with the significant input of E.E, G.N., B.HK, M.K and M.R. G.D.,O.A.,S.Z.,G.M.,JP.B and R.S. provided clinical insights, immunology expertise and contributed to the critical review of the manuscript.

Competing interests

R.S. is founder of IntelliStem Technologies Inc., Toronto, ON, Canada. The other authors have no competing interests.

Acknowledgements

The authors would like to thank all researchers who shared their data publicly and made this project possible.This work is partially supported by a Discovery Grant from the National Sciences and Engineering Research Council of Canada (RGPIN/06101-2014) and an operating grant from the Cancer Research Society (OG24054). M.R holds a Fonds de la Recherche en Santé du Québec Junior II Award.

All authors declare no conflict of interest

References

1. Coronavirus disease (COVID-19).
2. Chen, N. *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive *The Lancet* vol. 395 507–513 (2020).
3. Wang, , Zhou, Y., Yang, Z., Xia, D. & Geng, S. Clinical Characteristics of Patients with Severe Pneumonia Caused by the 2019 Novel Coronavirus in Wuhan, China. doi:10.1101/2020.03.02.20029306.
4. Guan, -J. *et al.* Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. *Eur. Respir. J.* **55**, (2020).
5. Menter, T. *et al.* Post-mortem examination of COVID19 patients reveals diffuse alveolar damage with severe capillary congestion and variegated findings of lungs and other organs suggesting vascular dysfunction. *Histopathology* (2020) doi:1111/his.14134.
6. Rota, A. Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *Science* vol. 300 1394–1399 (2003).
7. Gu, & Korteweg, C. Pathology and Pathogenesis of Severe Acute Respiratory Syndrome. *The American Journal of Pathology* vol. 170 1136–1147 (2007).
8. Ng, D. L. *et al.* Clinicopathologic, Immunohistochemical, and Ultrastructural Findings of a Fatal Case of Middle East Respiratory Syndrome Coronavirus Infection in the United Arab Emirates, April 2014. *J. Pathol.* **186**, 652–658 (2016).
9. Gheblawi, M. *et al.* Angiotensin-Converting Enzyme 2: SARS-CoV-2 Receptor and Regulator of the Renin-Angiotensin System: Celebrating the 20th Anniversary of the Discovery of ACE2. *Res.* **126**, 1456–1474 (2020).

10. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271–280.e8 (2020).
11. Harrison, Coronavirus puts drug repurposing on the fast track. *Nat. Biotechnol.* **38**, 379–381 (2020).
12. Gordon, E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* (2020) doi:10.1038/s41586-020-2286-9.
13. Gautret, *et al.* Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int. J. Antimicrob. Agents* 105949 (2020).
14. Geleris, *et al.* Observational Study of Hydroxychloroquine in Hospitalized Patients with Covid-19. *N. Engl. J. Med.* (2020) doi:10.1056/NEJMoa2012410.
15. Bessière, F. *et al.* Assessment of QT Intervals in a Case Series of Patients With Coronavirus Disease 2019 (COVID-19) Infection Treated With Hydroxychloroquine Alone or in Combination With Azithromycin in an Intensive Care Unit. *JAMA Cardiology* (2020) doi:1001/jamacardio.2020.1787.
16. Rosenberg, S. *et al.* Association of Treatment With Hydroxychloroquine or Azithromycin With In-Hospital Mortality in Patients With COVID-19 in New York State. *JAMA* (2020) doi:10.1001/jama.2020.8630.
17. Tang, *et al.* Hydroxychloroquine in patients with mainly mild to moderate coronavirus disease 2019: open label, randomised controlled trial. *BMJ* **369**, m1849 (2020).
18. Blanco-Melo, *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* (2020) doi:10.1016/j.cell.2020.04.026.
19. Subramanian, A., Narayan, R., Corsello, S. M. & Peck, D. D. A Next Generation Connectivity Map: L1000 Platform And The First 1,000,000 *bioRxiv* (2017).
20. GTEx The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
21. Zhou, *et al.* Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov* **6**, 14 (2020).
22. Li, , Bai, T., Yang, L. & Hou, X. Discovery of potential drugs for COVID-19 based on the connectivity map. doi:10.21203/rs.2.24684/v1.
23. Duarte, R. R. *et al.* Repurposing FDA-Approved Drugs for COVID-19 Using a Data-Driven Approach. doi:10.26434/chemrxiv.12148764.
24. Xiong, *et al.* Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg. Microbes Infect.* **9**, 761–770 (2020).
25. Michalovich, *et al.* Obesity and disease severity magnify disturbed microbiome-immune interactions in asthma patients. *Nat. Commun.* **10**, 5711 (2019).
26. Love, , Anders, S. & Huber, W. Differential analysis of count data–the DESeq2 package. *Genome Biol.* **15**, 550 (2014).
27. Collado-Torres, *et al.* Reproducible RNA-seq analysis using recount2. *Nature Biotechnology* vol. 35 319–321 (2017).

28. Liberzon, *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems* vol. 1 417–425 (2015).
29. WikiPathways : An Experiment in the Community-based Curation of *SciVee* (2008) doi:10.4016/6670.01.
30. Joshi-Tope, *et al.* Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* **33**, D428–32 (2005).
31. Lachmann, *et al.* Geneshot: search engine for ranking genes from arbitrary text queries. *Nucleic Acids Res.* **47**, W571–W577 (2019).
32. Subramanian, *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545–15550 (2005).
33. Qiu, , Lu, T., Lim, H. & Xie, L. A Bayesian approach to accurate and robust signature detection on LINCS L1000 data. *Bioinformatics* **36**, 2787–2795 (2020).
34. Leydesdorff, L. Similarity measures, author cocitation analysis, and information *Journal of the American Society for Information Science and Technology* vol. 56 769–772(2005).
35. Wishart, S. *et al.* DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **36**, D901–6 (2008).
36. Southan, C. *et al.* The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 *Nucleic Acids Res.***44**, D1054–68 (2016).
37. DeDiego, L. *et al.* Inhibition of NF-κB-mediated inflammation in severe acute respiratory syndrome coronavirus-infected mice increases survival. *J. Virol.* **88**, 913–924 (2014).
38. Zhang, *et al.* Nucleocapsid protein of SARS-CoV activates interleukin-6 expression through cellular transcription factor NF-kappaB. *Virology* **365**, 324–335 (2007).
39. Kanzawa, N. *et al.* Augmentation of chemokine production by severe acute respiratory syndrome coronavirus 3a/X1 and 7a/X4 proteins through NF-kappaB *FEBS Lett.***580**, 6807–6812 (2006).
40. de Wit, , van Doremalen, N., Falzarano, D. & Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **14**, 523–534 (2016).
41. Siu, -L. *et al.* Severe acute respiratory syndrome coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC. *FASEB J.***33**, 8865–8877 (2019).
42. Huang, *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
43. Tay, Z., Poh, C. M., Rénia, L., MacAry, P. A. & Ng, L. F. P. The trinity of COVID-19: immunity, inflammation and intervention. *Nat. Rev. Immunol.* (2020) doi:10.1038/s41577-020-0311-8.

44. Xu, *et al.* Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir Med* **8**, 420–422 (2020).
45. Kuba, *et al.* A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat. Med.* **11**, 875–879 (2005).
46. Haga, S. *et al.* Modulation of TNF-alpha-converting enzyme by the spike protein of SARS-CoV and ACE2 induces TNF-alpha production and facilitates viral entry. *Proc. Natl.Acad. Sci. U. S. A.* **105**, 7809–7814 (2008).
47. Stelzig, E. *et al.* Estrogen regulates the expression of SARS-CoV-2 receptor ACE2 in differentiated airway epithelial cells. *Am. J. Physiol. Lung Cell. Mol. Physiol.* (2020) doi:10.1152/ajplung.00153.2020.
48. Bi, *et al.* Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases* (2020) doi:10.1016/s1473-3099(20)30287-5.
49. Channappanavar, *et al.* Sex-Based Differences in Susceptibility to Severe Acute Respiratory Syndrome Coronavirus Infection. *J. Immunol.* **198**, 4046–4053 (2017).
50. Wathelet, G., Orr, M., Frieman, M. B. & Baric, R. S. Severe acute respiratory syndrome coronavirus evades antiviral signaling: role of nsp1 and rational design of an attenuated strain. *J. Virol.* **81**, 11620–11633 (2007).
51. Yang, *et al.* The structural and accessory proteins M, ORF 4a, ORF 4b, and ORF 5 of Middle East respiratory syndrome coronavirus (MERS-CoV) are potent interferon antagonists. *Protein Cell* **4**, 951–961 (2013).
52. Kopecky-Bromberg, S. A., Martínez-Sobrido, L., Frieman, M., Baric, R. A. & Palese, P. Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. *Virol.* **81**, 548–557 (2007).
53. Chu, *et al.* Comparative replication and immune activation profiles of SARS-CoV-2 and SARS-CoV in human lungs: an ex vivo study with implications for the pathogenesis of COVID-19. *Clin. Infect. Dis.* (2020) doi:10.1093/cid/ciaa410.
54. Ye, , Wang, B. & Mao, J. The pathogenesis and treatment of the `Cytokine Storm' in COVID-19. *J. Infect.* **80**, 607–613 (2020).
55. Giamarellos-Bourboulis, E. J. *et al.* Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory *Cell Host Microbe* (2020) doi:10.1016/j.chom.2020.04.009.
56. Glowacka, I. *et al.* Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *Virol.* **85**, 4122–4134 (2011).
57. Hoffmann, M. *et al.* Nafamostat Mesylate Blocks Activation of SARS-CoV-2: New Treatment Option for COVID-19. *Antimicrobial agents and chemotherapy* 64 (2020).

58. Sorrell, F. J., Szklarz, M., Abdul Azeez, K. R., Elkins, J. M. & Knapp, S. Family-wide Structural Analysis of Human Numb-Associated Protein *Structure* **24**, 401–411 (2016).
59. Richardson, *et al.* Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *Lancet* **395**, e30–e31 (2020).
60. Lokugamage, K. G., Narayanan, K., Huang, C. & Makino, S. Severe acute respiratory syndrome coronavirus protein nsp1 is a novel eukaryotic translation inhibitor that represses multiple steps of translation initiation. *Virol.* **86**, 13598–13608 (2012).
61. Kamitani, *et al.* Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. *Proc. Natl. Acad. Sci. U. S. A.***103**, 12885–12890 (2006).
62. Kamitani, W., Huang, C., Narayanan, K., Lokugamage, K. G. & Makino, S. A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 *Nat. Struct. Mol. Biol.* **16**, 1134–1140 (2009).
63. Shen, B. *et al.* Proteomic and Metabolomic Characterization of COVID-19 Patient *SSRN Electronic Journal* doi:10.2139/ssrn.3570565.

Supplemental File Legends

Supplementary Figure 1.

(A) Stacked barplot depicting fractions of up- and down-regulated genes in each of the contrasts. (B) UpSet plots to summarize the ratio of key differentially expressed (DE) genes. This panel shows the overlap of up-regulated genes across all comparisons. (C) Same as for (B) but shows the overlap of down-regulated genes across all comparisons. For all 3 panels, the FDR value of DE significance was set to 0.05 and a |log2FC| > 1.

Table1.

Summary of public datasets used in the analytical pipeline

Supplementary Figure S2.

Heatplot depicting top enriched drug-target associations against the ranked cosine consensus drug score (putative candidates to reverse SARS2-induced transcriptome perturbations in lung tissue)

Table S2.

Co-expression relationships of ACE2 from GTEX normal lung tissue. Pearson correlation coefficients were determined for ACE2 versus all other genes using data from GTEx. 7433 overlapping genes with all SARS-

CoV-2 contrasts were included in the table

Table S3.

Edge List of hallmark collections from MsigDB and custom gene sets used to test enrichment for immune, inflammatory and SARS-CoV related biological imprinting in all settings

Figures

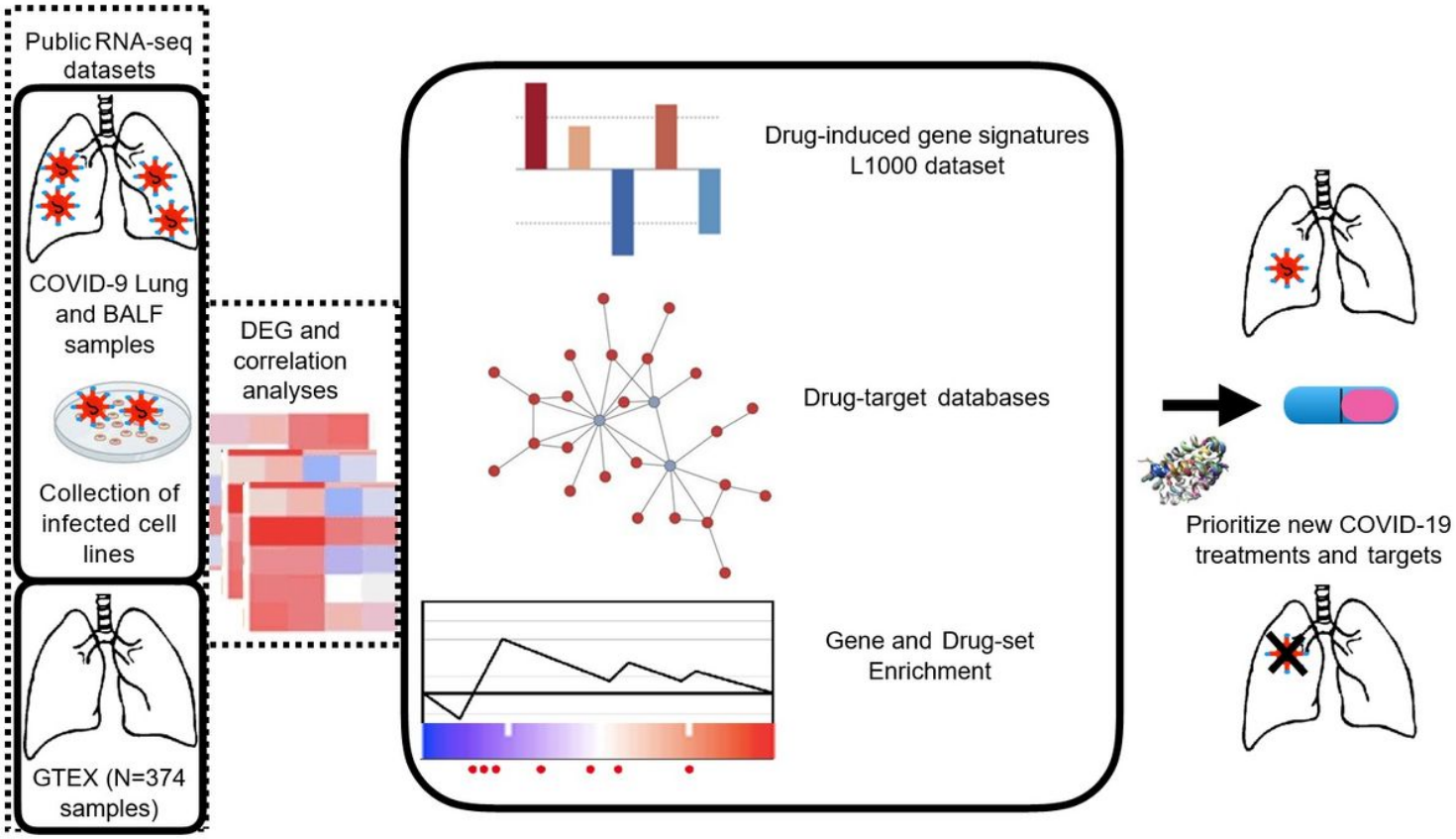

Figure 1

Schematic overview of the computational approach. Molecular signatures induced by SARS-CoV-2, chemical-induced gene expression profiles, gene, and drug-set enrichment analyses, drug-target databases have been integrated into the analytical framework to understand the molecular pathophysiology of SARS-CoV2 and to identify novel therapeutic options.

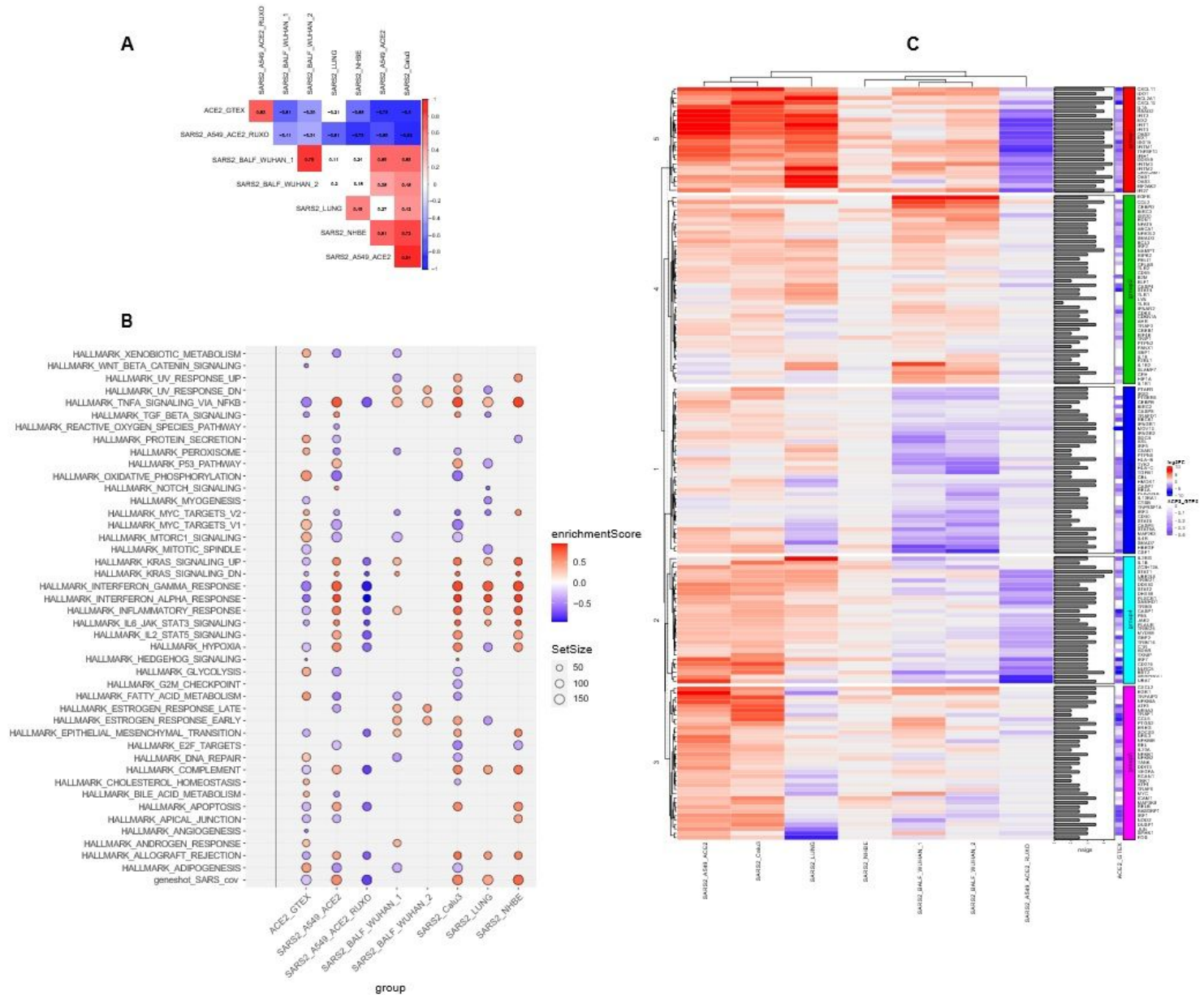

Figure 2

(A) Pairwise Spearman rank correlations from the Hallmark enrichment analyses. Red and blue colors represent positive and negative correlations respectively. While blank cells show non significant correlations at a given p-value threshold of 0.05. (B) Heatmap depicting the expression levels of 169 genes linked to the innate immune system across all SARS2 settings. These genes were extracted from the leading edge of the gene set enrichment analyses and belong to the following hallmarks/sets: TNFA SIGNALING VIA NFKB, INTERFERON GAMMA RESPONSE, INTERFERON ALPHA RESPONSE, IL6-JAK-STAT3 SIGNALING, INFLAMMATORY RESPONSE, and the geneshot_SARS_CoV gene list. The graph illustrates the shrunken log2 Fold change from all SARS2 contrasts. The sidebar plots show the number of settings for which the genes have a p-adjusted value of less than 0.05. The heatmap shows 5 distinct groups from k-mean clustering. A side heatmap depicts the correlation values for these same genes against ACE2 expression from GTEx (all genes have negative correlation values since all mentioned

hallmarks were significantly downregulated in ACE2_GTEX). (C) Dot plot visualization of enriched terms in all SARS2 and ACE2_GTEX settings. Gene set enrichment analyses (GSEA) were performed against the Hallmark dataset from MsigDB and a collection of genes related to SARS-CoV infection from the literature (gene shot, 224 genes). The color of the dots represents the enrichment score and the FDR value of significance was set to 0.05 for all sets. Size represents the total number of genes in the gene set.

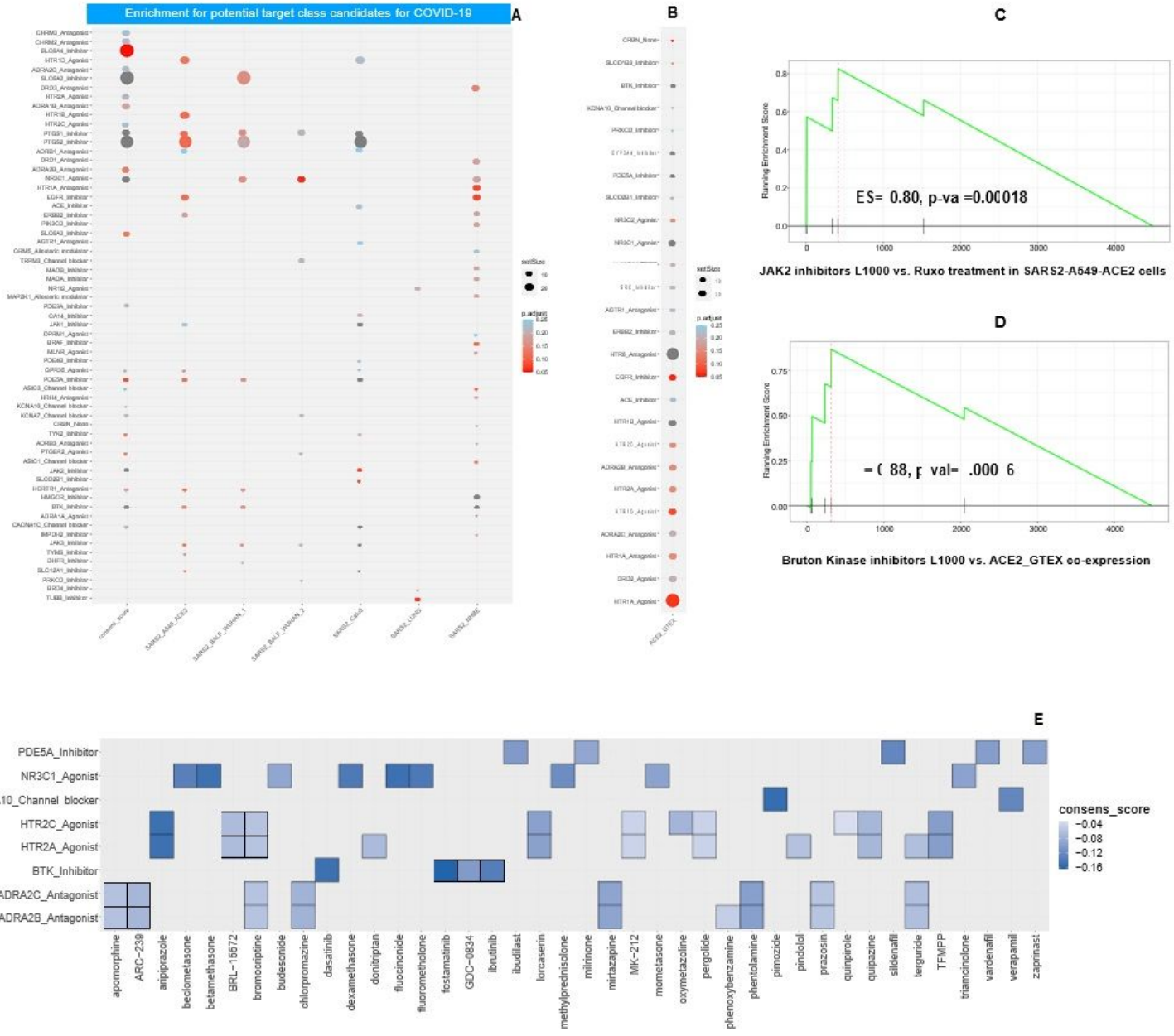

Figure 3

(A) Dotplot visualization of enriched drug-target associations in all SARS2 including the generated consensus cosine score Drug set enrichment analyses were performed against 295 drug-target families. Only those with absolute enriched score ES <= -0.5 and a FDR < 0.25 are shown (B) ACE2_GTEX settings respectively. Drug set enrichment analyses were performed against 295 drug-target families. Only those with absolute enriched score ES >= 0.5 and a FDR < 0.25 are shown. (C) GSEA plot of the cosine signature for Janus kinase 2 inhibitors in SARS2_A549_ACE2_RUXO drug set enrichment analysis and (D) GSEA

plot of the cosine signature for Bruton Kinase inhibitors in ACE2_GTEX drug set enrichment analysis. (E) Heatlpot showing all enriched drug-target families with putative dual activities: restoring both ACE2 co-expression partners in the lungs and pathway perturbations following SARS-CoV-2 infection.

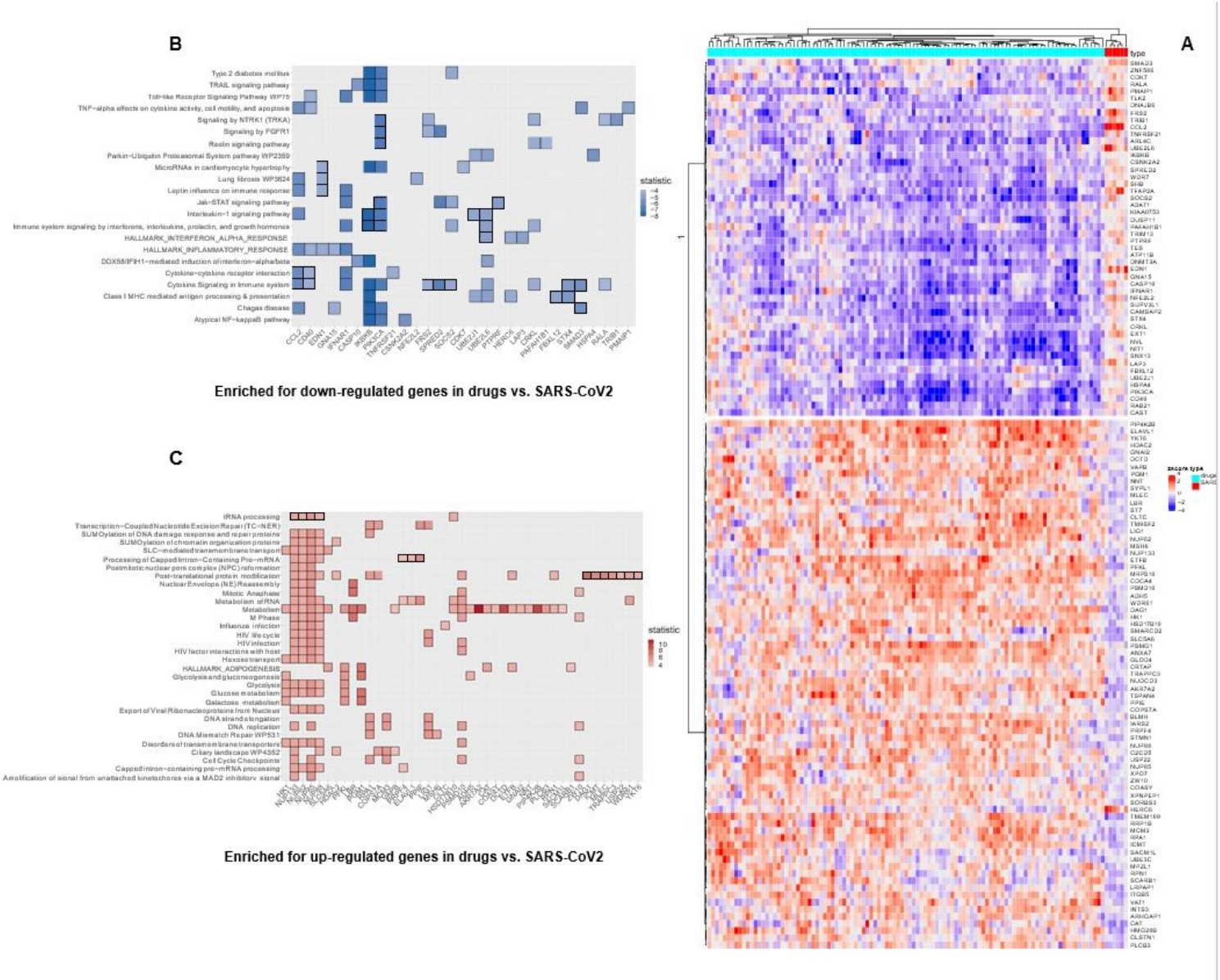

Figure 4

(A) Welch's t-test was performed between gene levels induced by 106 selected drugs from L1000 and SARS-CoV-2 expression settings. Heatmap illustrating the expression level of 124 genes that significantly separate the two groups (p-adjusted value from the t-test of less than 0.05). (B) Heatplot for core pathways in which gene members were significantly downregulated in response to L1000 drugs. (C) Heatplot for core pathways in which gene members were significantly upregulated in response to L1000 drugs, in contrast to SARS2 settings (p-value < 0.05).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.jpg
- SupplementaryFigure2.jpg
- Table1.pdf
- TableS2.csv
- TableS3.csv