Bayesian model infers drug repurposing candidates for treatment of COVID-19

Michael A. Kiebish¹*, Punit Shah¹¶, Rangaprasad Sarangarajan¹, Vivek K. Vishnudas¹, Stephane Gesta¹, Poornima K. Tekumalla¹, Chas Bountra², Elder Granger¹, Eric Schadt³,², Leonardo O. Rodrigues³, and Niven R. Narain¹*

**Affiliation:** ¹BERG, Framingham, MA USA.

²Center for Medicines Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, UK.

³ Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA.

⁴ Sema4, Stamford, CT, USA.

¶ *Co-first Authors

* Corresponding author:

E-mail: Niven.Narain@berghealth.com (NN)
Abstract

Background

The emergence of COVID-19 progressed into a global pandemic that has functionally put the world at a standstill and catapulted major healthcare systems into an overburdened state. The dire need for therapeutic strategies to mitigate and successfully treat COVID-19 is now a public health crisis with national security implications for many countries.

Methods

The current study employed Bayesian networks to a longitudinal proteomic dataset generated from Caco-2 cells transfected with SARS-CoV-2 (isolated from patients returning from Wuhan to Frankfurt) [1]. Two different approaches were employed to assess the Bayesian models, a titer-center topology analysis and a drug signature enrichment analysis.

Results

Topology analysis identified a set of proteins directly linked to the SAR-CoV2 titer, including ACE2, a SARS-CoV-2 binding receptor, MAOB and CHECK1. Aligning with the topology analysis, MAOB and CHECK1 were also identified within the enriched drug-signatures.

Conclusions

Taken together, the data output from this network has identified nodal host proteins that may be connected to 18 chemical compounds, some already marketed, which provides an immediate opportunity to rapidly triage these assets for safety and efficacy against COVID-19.

Keywords: COVID-19, Proteomics, Bayesian, Drug Repurposing, ACE2, CHEK1, MAO
Introduction

On March 11th, 2020, the World Health Organization officially declared coronavirus disease 2019 (COVID-19) as a pandemic. The rate of incidence and progression of cases has rapidly thrust major healthcare systems worldwide into an overburdened state with advanced cases requiring extended ICU care. At time of this manuscript submission, there are greater than 31.6 million cases confirmed across 208 countries and territories with incidence and mortality still rising in Africa, South America and most of the Western Hemisphere. The rapidity and intensity of the COVID-19 transmission and associated morbidity and mortality [2] have created a critically emergent need for the identification and characterization of biological drivers of infection and clinical progression.

The mean time to progression of observable symptoms is reported to be 9.1 to 12.5 days, thereby providing researchers with information on viral pathogenicity and a plausible therapeutic window of opportunity [3, 4]. Global efforts are underway to investigate the potential utility of anti-virals and other agents in mitigating the clinical course of SARS-CoV-2 induced complications.

In pharma and healthcare, when data is available but there are unmet needs for markers and therapeutic targets, causal inference has been successfully applied to generate knowledge from data, such as in the identification of novel biomarkers[5], disease regulators[6] and outcome predictors[7]. Similarly, the SARS-CoV-2 9 viral/host interaction necessitates a more comprehensive understanding which can be achieved by coupling protein-based interactome with a Bayesian network (BN) approach. Herein, we applied Bayesian networks[8], differential expression analysis and drug-enrichment analysis to a publicly available longitudinal dataset of proteomics and viral titer from Caco-2 cells infected with SARS-CoV-2[1]. Utilizing this BN-based
approach, a *de novo* map of host proteins was established as well as drug enrichment signatures were defined. Several of the host proteins presented as linked to the virus titer, providing new insights into the pathogenesis of this virus and defining potential targets for therapeutic intervention that could mitigate the viral load and associated complications. The approach described herein largely complements artificial intelligence (AI) tools and algorithms that have been mobilized during this emerging health crisis.

**Methods**

**Data description**

Data used in this study was extracted from a publicly available source from ProteomeXchange Consortium via the PRIDE39 partner repository with the dataset identifier PXD017710[1].

Based on Bojkova et al, data was generated using SARS-CoV-2 isolated from 2 travelers from Wuhan to Frankfurt. Caco-2 cells were infected with SARS-CoV-2. Samples were harvested at 2, 6, 10 and 24 hours post infection. Media was changed to DMEM containing 84 mg/L L-Arginine (13C615N4; Cambridge Isotope Laboratories, Cat#CNLM–539-H) and 146 mg/L L-Lysine (13C615N2, Cambridge Isotope Laboratories, Cat#CNLM–291-H) 2 hours before harvesting. Cells were then lysed, proteins were precipitated, digested using LysC and trypsin, TMT-labeled and the samples combined. Peptides were then fractionated using HpH RP Fractionation Kit. Easy nLC 1200 coupled to a Q Exactive HF mass spectrometer collected LC-MS/MS data[1].
Data extraction

Data source for the current study was as described by Bojkova et al [1] where the concentrations of SARS-CoV-2 titer in the supernatant at 2h, 6h, 10h and 24h were reported. The high quality of unprocessed proteomics was accessed and confirmed. The available processed proteomics was used in this analysis with no further modification to maintain the optimal processing-parameters defined in Bojkova et al[1].

Model building

Bayesian networks (BN)[9] are graphical models that characterize the probabilistic dependencies over a set of random variables in a directed acyclic graph. The final models contain a set of nodes connected through directed edges, that denote the inferred causal dependencies within the data. As part of the Interrogative Biology® platform BERG developed bAlcis®, a scalable score-based BN-learning software which uses the Bayesian information criteria[10] for local and global model selection and Bayesian methods to estimate the parameters[11] as previously described[7, 8]. From the combined longitudinal proteomics and titer data, 300 networks were generated by bAlcis® and compiled into a unified ensemble network, which was filtered for the edges present in at least 65% of the 300 models. A subnetwork containing the titer-node, and all nodes with a distance of 1 or 2 degrees from the titer-node, was extracted for further analysis (Fig 1).
Proteomics and virus titer information were extracted from Bojkova et al., 2020, merged, and analyzed by the platform’s Bayesian network tool, bAlcis®. From the ensemble model, titer-first and second degree nodes were selected for a literature evaluation, and a differential-expression enriched sub-network was used in a drug enrichment analysis.

**Drug enrichment analysis**

Differential expression analysis from samples with the lowest/highest viral titer was performed using the Bioconductor package LIMMA[12]. Proteins with a significant differential expression (adjusted p-value ≤ 0.001) were projected onto the ensemble network, and the projection further expanded to the 1-degree neighbor nodes, generating a sub-network that
overlays differential expression and Bayesian network analysis. This subnetwork contained the titer subnetwork comprised of nodes within a path length of two from the titer node.

Drug signatures (DsigDB) and drug-target coregulated gene sets (ARCHS4) were extracted from the Enrichr database[13]. Subnetworks around each node in the full network were defined both in an undirected and downstream fashion. For each, node enrichment for signatures in the tier subnetwork from a path length 1 through 6 were computed with the R based function fisher.test and adjusted for multi-testing using the FDR setting of the p.adjust function. The proportion of nodes with significant enrichment was defined at each path length as the proportion of nodes with existing perturbation and FDR<0.05 divided by the total number of nodes with existing perturbation.

Results

Titer-centered sub-network

Bayesian network modeling can infer potential causal associations on integrated phenotypic, molecular, biochemical, and omic datasets from disparate sources. Herein, we engaged a Bayesian network approach which identified 1869 causally associated nodes (Fig 2A). To evaluate specific associations with titer we extracted a subnetwork for 1st degree and 2nd degree nodes from the network (Fig 2B).
Fig 2: Ensemble Model

A probabilistic ensemble model from proteomics and viral titer was generated by bAlcis®, with nodes representing proteins and titer and edges the conditional dependence between nodes (A). The titer 2-degree sub-network was extracted for further evaluation (B).

In the titer subnetwork, 14 nodes were inferred to be associated with the viral titer. In this sub-network, ACE2, a receptor for the SARS-CoV-2 virus, was directly connected to the titer node with increasing levels of the viral titer potentially repressing the expression of ACE2 (Fig 3A). In addition, ACE2 was inferred as potentially repressed by COX6A1 and induced by SLC6A3.

The expression CHEK1 was predicted as potentially induced by the viral infection and UBAP2L but repressed by CAPN7 and, in a downstream event, CHEK1 levels had a reverse effect on BAZ1B levels (Fig 3B). Contrary to ACE2, the increasing titer concentration might induce the expression of MAO-B (Fig 3C), whose expression is also induced by PRMT7, but repressed by
TMEM238. Both MAO-B and CHEK1 were also identified as component of drug signatures enriched in the full network.

**Fig 3: Model nodes**

Selected nodes were highlighted to demonstrate model validation such as ACE2 (A), or as potential new therapeutic targets, such as CHEK1 (B) and MAOB (C). Arrowed edges indicate a positive causal connection (E.g. A → B for A inducing B), and T-shaped arrows indicate a negative causal connection (E.g. A–| B for A repressing B).

**Identified drug signatures through network integration and associated with titer**

In order to assess possible drug signatures associated with the viral titer, an enriched titer sub-network was generated, and compared to known drug signatures and to drug-target co-regulated gene sets extracted from the Enrichr database[13]. This analysis identified kinase signatures for the protein kinase DNA-Activated catalytic subunit (PRKDC)[OR:7.9; p-value: 1.3e-11], serine/threonine protein kinase (mTOR)[OR:7.1; p-value: 2.6e-7], serine/threonine protein
kinase 24 (STK24) [OR: 4.9; p-value: 1.7e-5], ribosomal protein S6 kinase alpha-3 (RPS6KA3) [OR: 4.4; p-value: 9.4e-5], and lysine deficient protein kinase 1 (WNK1) [OR: 4.4; p-value: 9.4e-5] that were all enriched in the titer subnetwork at the indicated odds ratios/pvalues. In addition to the kinase signatures, signatures for known drug compounds were also significantly enriched in the titer subnetwork Table 1. The prediction that results from these enrichments is that these compounds through their impact on the titer subnetwork have the potential influence of modifying the titer.

Table 1: List of drug signature based on enrichment analysis to the titer

<table>
<thead>
<tr>
<th>Drug Signature</th>
<th>Fold Enrichment</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>emetine HL60 DOWN DOWN</td>
<td>5.14</td>
<td>2.46E-07</td>
</tr>
<tr>
<td>ambroxol PC3 DOWN DOWN</td>
<td>3.99</td>
<td>1.30E-06</td>
</tr>
<tr>
<td>Vitinoin CTD 00007069</td>
<td>3.36</td>
<td>3.85E-06</td>
</tr>
<tr>
<td>diclofenac CTD 00005804</td>
<td>4.09</td>
<td>1.91E-05</td>
</tr>
<tr>
<td>sertraline CTD 00007358</td>
<td>8.42</td>
<td>2.03E-05</td>
</tr>
<tr>
<td>ellagic acid BOSS</td>
<td>5.84</td>
<td>7.23E-05</td>
</tr>
<tr>
<td>ellagic acid CTD 00005891</td>
<td>6.36</td>
<td>0.00012</td>
</tr>
<tr>
<td>ascorbic acid CTD 00005445</td>
<td>4.73</td>
<td>0.00013</td>
</tr>
<tr>
<td>piperine TTD 00010245</td>
<td>21.65</td>
<td>0.00034</td>
</tr>
<tr>
<td>oxygen CTD 00006454</td>
<td>3.25</td>
<td>0.00035</td>
</tr>
<tr>
<td>Fulvestrant CTD 00002740</td>
<td>3.91</td>
<td>0.00053</td>
</tr>
<tr>
<td>carbamazepine CTD 00005574</td>
<td>3.03</td>
<td>0.00183</td>
</tr>
<tr>
<td>menadione PC3 DOWN DOWN</td>
<td>4.64</td>
<td>0.00193</td>
</tr>
<tr>
<td>indomethacin CTD 00006147</td>
<td>3.27</td>
<td>0.00328</td>
</tr>
</tbody>
</table>

Further, *de novo*, we identified novel links between SARS-CoV-2 titer and host proteins expression that illuminated unique biological signature associated with COVID-19. Several of these associated proteins have therapeutic options already on the market or in clinical trials.
(Table 2) and serve as possible repurposing opportunities to impact the levels of the SARS-CoV-2 virus in patients.

**Table 2: List of targets with approved drugs or in advanced clinical development**

<table>
<thead>
<tr>
<th>Target</th>
<th>Drug Name</th>
<th>Indications</th>
<th>Development Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAOB</td>
<td>rasagiline</td>
<td>Parkinsons disease</td>
<td>Launched</td>
</tr>
<tr>
<td></td>
<td>ropinirole</td>
<td>Parkinsons disease; Restless legs syndrome</td>
<td>Launched</td>
</tr>
<tr>
<td></td>
<td>safinamide</td>
<td>Movement disorder; Muscle disease; Parkinsons disease</td>
<td>Launched</td>
</tr>
<tr>
<td></td>
<td>selegiline</td>
<td>Parkinsons disease</td>
<td>Launched</td>
</tr>
<tr>
<td></td>
<td>tolcapone</td>
<td>Parkinsons disease</td>
<td>Launched</td>
</tr>
<tr>
<td>CHECK1</td>
<td>SRA-737</td>
<td>Cancer; Colorectal tumor; Non-small-cell lung cancer; Ovary tumor; Small-cell lung cancer; Soft tissue sarcoma; Transitional cell carcinoma; Uterine cervix tumor</td>
<td>Phase 2 Clinical</td>
</tr>
</tbody>
</table>

**Discussion**

The COVID-19 pandemic has ignited a sense of urgency, purpose, and scientific/medical necessity to engage the entire armamentarium of technology, informatic capabilities, therapeutic tools and collaborative spirit to combat this accelerating foe of humanity. Due to the rapid growth kinetics of the disease and impact globally, utilizing emerging crowd sourced data and applying innovative bioinformatic modeling is an optimal approach to illuminate potential drug-targets to mitigate the SARS-CoV-2 infection[1]. Thus, we engaged a Bayesian network approach to identify potential therapeutic targets[7, 8] causally integrated with viral titer.

Despite the limited data size, which was underpowered to support the inferred causalities, the robustness of bAlcis® analysis was evidenced by the data-driven linkage of SARS-
CoV-2 titer to ACE2[14]. Several studies have confirmed that ACE2 serves as a receptor for binding of the SARS-CoV-2 spike proteins required for host cell entry and viral replication[3]. Overexpression of ACE2 from different species, including human (HeLa cells), pig and civet, was associated with SARS-CoV-2 infection and replication, further demonstrating its role as a receptor for entry into host cells[15]. Furthermore, SARS-CoV-2 entry into host cells is dependent on ACE2 and serine protease TMPRSS 2 for priming, an activity that can be blocked by a specific inhibitor[16]. Interestingly, the expression of ACE2 is found in multiple organs including lungs, kidneys and heart, and is associated with incidence and severity of complications such as acute respiratory distress syndrome (ARDS), acute cardiac injuries and kidney injuries[17-20]. In addition, the viral titer was also predicted as possibly inhibiting carboxypeptidase (CPD), that belongs to a class of enzymes that includes ACE and ACE2, supporting the robustness of the network model.

Analysis of the ensemble network to identify enriched drug signatures ranked mTOR as a potential target with the ability to influence the viral titer. This is consistent with clinical reports suggesting Azithromycin, an mTOR inhibitor, as a therapeutic for SARS-CoV-2 infections[21]. In fact, the use of mTOR inhibitors to manage viral infections has been extensively investigated as well as their influence on immune systems, forming the basis for a potential use in SARS-CoV-2 infections[22, 23]. Additionally, enriched drug signature(s) also indicated an association of COVID-19 and the renin-angiotensin system e.g. cyclopenthiazide, a thiazide diuretic that inhibits angiotensin converting enzyme (ACE).

Further, a direct linkage of ACE2 to COX6A1 (a member of the mitochondrial Cytochrome C Oxidase complex associated with electron transport and critically engaged with oxidative
phosphorylation) and SLC26A3 (electroneutral Cl-/HCO₃⁻ antiporter) were identified. Changes in expression of SCL26A3 in the colon has been associated with incidence of infectious diarrhea[24]. Interestingly, in Caco-2 cells (model used in this study), the SLC26A3 transporter activity is regulated by the cystic fibrosis transmembrane conductance regulatory (CFTR)[25]. Given the importance of CFTR and Cl-/HCO₃⁻ balance in lung physiology[26] [21], the data from the proteomic network provides the basis for a potential lung-regulated acid-base imbalance in SARS-CoV-2 pulmonary manifestations.

The viral titer was also inferred as associated with MAOB and CHEK1. The role of MAO as virus receptors has been explored since the 1980’s and further supported by literature, demonstrating a relationship between MAO and viral infections[27]. Increases in MAO activity in the brain models of simian immunodeficiency virus (SIV) and human immunodeficiency virus (HIV) have been reported to be associated with central nervous system disorders[28]. An increase in MAOB mRNA was observed in macaques with severe SIV-associated lesion in the brain and correlated with viral loads[28]. Deprenyl, a MAOB inhibitor, improved verbal memory tests in subjects with mild HIV-associated cognitive impairment compared to patients not taking deprenyl[29]. In addition, given the wide-spread use of MAOB inhibitors (e.g. selegiline, rasagiline, safinamide) in clinical management of Parkinson’s Disease (PD)[30], determination of the prevalence of COVID-19 in cohort of PD patients could provide a rapid validation of a role of MOAB in the time-course of COVID-19. Furthermore, such validation would support the use of MOAB inhibitors as supportive therapy in patients with active SARS-CoV-2 infections to improve management and outcomes. Orthogonal support of the potential role of MAO-B in the regulation of SARS-CoV-2 viral titer is the enrichment of these linkages identified within the sertraline drug
signatures. Sertaline, a class of Selective Serotonin Reuptake Inhibitors (SSRIs), is used in the treatment of major depressive disorder (MDD), including patients with Parkinson’s Disease diagnosed with depression as comorbidity. Although increase in expression of MAO-A in the brain in MDD has been established, the status of MAO-B is unclear. However, inhibitors of MAO enzyme have the ability to inhibit activity of both enzymes and have known drug interactions with SSRIs, including Sertaline.

CHEK1, a highly evolutionary conserved serine/threonine-specific protein kinase, preserves the genome integrity by regulating DNA damage and cell cycle checkpoint responses [26]. CHEK1 activation is specifically controlled by ATR through phosphorylation, forming the ATR-CHEK1 pathway. This pathway is involved in a broad spectrum of DNA abnormalities including inhibition of DNA replication, UV-induced DNA damage, inter-strand DNA crosslinking and virus infection[31]. The inferred association between the viral titer and the expression level of CHEK1 in the host cell suggests that SARS-CoV-2 could use the ATR-CHEK1 pathway for its replication. Indeed, several studies have reported that activation of the ATR-CHEK1 pathway occurs during viral DNA synthesis in various viruses including coronavirus[32], hepatitis B virus[33], human papillomavirus[34], parvovirus B19[35]. Furthermore, utilization of CHEK1 inhibitors has been successfully used to reduce viral replication. For example, the CHEK1 inhibitor MK-8776 reduced viral DNA amplification by 90-99% in HPV infected cells[36]. Similarly, the CHEK1 inhibitor UCN-01 has been reported to decrease hepatitis B virus DNA yield of HL7702 cells by 80 to 90%, without significantly affecting cell survival[33]. The identification of a relationship between CHEK1 expression and a high SARS-CoV-2 titer associated with the literature on CHEK1 inhibitors and viral replication, supports the investigation of these molecules
as a potential therapeutic modality for the treatment of COVID-19. Due to the known functions of CHEK1 in cancer through the regulation of the DNA damage response and cell cycle checkpoints, numerous CHEK1 or CHEK1/CHEK2 inhibitors have been developed over the past two decades or are currently in clinical development, mainly for oncology applications. Evaluating the repurposing of these drugs for the treatment of COVID-19 presents an opportunity to rapidly bring therapeutic solutions to patients currently affected by this disease.

The drug signature for diclofenac was identified influencing SARS-CoV-2 viral titer to both CHEK1 and MAOB. Diclofenac is used for pain treatment and belongs to a class of non-steroidal anti-inflammatory drugs (NSAIDs). The ability of thiazolidinediones and ibuprofen to increase ACE2 expression was reported as a potential susceptibility factor for SARS-CoV-2 infection in patients with diabetes and hypertension. However, the World Health Organization (WHO) and the European Medicines Agency (EMA) provided follow up guidance on the lack of scientific evidence supporting role NSAIDs such as ibuprofen in SARS-CoV-2 infection. The identification of diclofenac drug signature and the association of CHEK1 and MAO-B with the viral titer network provide additional insights into an indirect influence of these agents in the viral infectivity. The targets identified in this study supporting underlying mechanism(s) of SARS-CoV-2 infectivity are consistent with targets/therapeutic profiles identified using physical protein interaction map[37].

In summary, we applied a Bayesian network approach to a publicly available dataset that modeled a cellular time course experiment integrating proteomics and SARS-CoV-2 titer. Utilization of Bayesian network demonstrates promises in identifying novel biological signatures in disease pathogenesis, of which, holds significant potential in influencing the clinical outcome of diseases, such as COVID-19. Taken together, the data herein and associated topology
knowledge derived from the model could serve as a foundational for an information database to triage and advance the repurposing of existing drugs or agents currently in development for the treatment, management, and mitigation of COVID-19.

**Declarations**

**Ethics approval and consent to participate**
Not Applicable

**Consent for publication**
Not Applicable

**Availability of data and materials**
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Competing Interests**
The authors declare that they have no competing interests.

**Funding**
This study was funded in its entirety by Mr. Carl Berg.

**Authors’ Contributions**
MK: Provided substantial contributions to the conception and design of the work, interpretation of data, and drafted the work or substantively revised it
PS: Drafted the work or substantively revised it, provided substantial contributions to the conception and design of the work and acquisition, analysis
RS: Drafted the work or substantively revised it, provided substantial contributions to the conception and design of the work and interpretation of data
VV: Drafted the work or substantively revised it and interpretation of data
SG: Drafted the work or substantively revised it and interpretation of data
PT: Drafted the work or substantively revised it
CB: Substantively revised the work and interpretation of data
EG: Substantively revised the work and interpretation of data
ES: Drafted the work or substantively revised it and interpretation of data
LR: Drafted the work or substantively revised it, provided substantial contributions to the conception and design of the work and acquisition, analysis
NN: Drafted the work or substantively revised it, provided substantial contributions to the conception and design of the work and interpretation of data

All authors have read and approved the final manuscript.
Acknowledgments
BERG would like to acknowledge the generous funding of Mr. Carl Berg for this entire study.

References


