

Interactome of SARS-CoV-2 / nCoV19 modulated host proteins with computationally predicted PPIs

Kalyani B. Karunakaran¹, N. Balakrishnan¹ and Madhavi K. Ganapathiraju^{2,3*}

¹Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore, 560 012, India

²Department of Biomedical Informatics, School of Medicine, and ³Intelligent Systems Program, School of Computing and Information, University of Pittsburgh, Pittsburgh, USA

Correspondence should be sent to madhavi@pitt.edu

Corresponding author:

Dr. Madhavi K. Ganapathiraju
Department of Biomedical Informatics,
University of Pittsburgh
5607 Baum Blvd,
Suite 501
Pittsburgh, PA 15206, USA
Email: madhavi@pitt.edu
Phone: 412-648-9552
Fax: 412-624-5310

Running title: Computationally predicted PPIs of SARS-CoV-2 modulated host proteins

Highlights

- 1,941 novel interactions of proteins targeted by SARS-CoV-2 ('host proteins') are predicted with HiPPIP, our computational model.
- The interactome is made available via a web-server [Wiki-Corona](http://hagrid.dbmi.pitt.edu/corona) (<http://hagrid.dbmi.pitt.edu/corona>).
 - It is searchable by protein IDs and annotations.
 - More interestingly, it is searchable with queries like “[Show PPIs where one protein has to do with 'virus' and the other protein has to do with 'pulmonary'](#).”
- The interactome is made available as downloadable network files to facilitate future systems biology studies.
- Analysis of the interactome embedded with novel predicted PPIs resolved the apparent disconnect between two recent studies (transcriptional with 120 genes and proteomic with 332 genes) by showing that although they shared only two common genes, there are many direct PPIs between them and several shared common interactors.
- The interactome also showed connections between host-proteins of SARS-1 and SARS-2 coronaviruses, and the proteins common to both coronaviruses were enriched for mitochondrial proteins.
- We hypothesize that the novel interaction of the virus-induced NUP98 with CHMP5 may activate an antiviral mechanism mediated by the interferon-stimulated gene ISG15 leading to disruption of viral budding.
- The interactome had highly connected network modules which were enriched significantly for *cilium organization*, *cell-cell signaling by Wnt*, *epigenetic regulation of gene expression*, *nuclear transport*, *ribonucleoprotein complex biogenesis*, and *vesicle-mediated transport between endosomal compartments*, *cell cycle G2/M phase transition*, and *DNA replication*
- Midbody proteins are enriched in the HoP interactome. Viruses are known to recruit protein complexes that also localize to the midbody, a microtubule-rich structure that marks the site of abscission during cytokinesis, to the host cell membrane, to promote scission and thereby the release of viral particles. 11 host proteins targeted by SARS-CoV-2 (RHOA, CENPF, CIT, RAB8A, NUP62, SCCPDH, SPART, RDX, ARF6, CNTRL and RALA), 63 known interactors and 9 novel interactors (KIF4A, BIRC5, INCENP, ALKHB4, DNM2, DDX11, ARL2BP, ABRAXAS2 and WIS) are midbody localized proteins.
- The interactome demonstrated significant association with 35 diseases, including hypertension, diabetes and cancers which are found to be comorbidities in critical COVID-19 cases. Additionally, enrichment for neurological disorders such as Alzheimer's disease and schizophrenia were also found, warranting investigations into potential comorbidities.

Interactome of SARS-CoV-2 / nCoV19 modulated host proteins with computationally predicted PPIs

Kalyani B. Karunakaran¹, N. Balakrishnan¹ and Madhavi K. Ganapathiraju^{2,3*}

¹Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore, 560 012, India

²Department of Biomedical Informatics, School of Medicine, and ³Intelligent Systems Program, School of Computing and Information, University of Pittsburgh, Pittsburgh, USA

Correspondence should be sent to madhavi@pitt.edu

Abstract

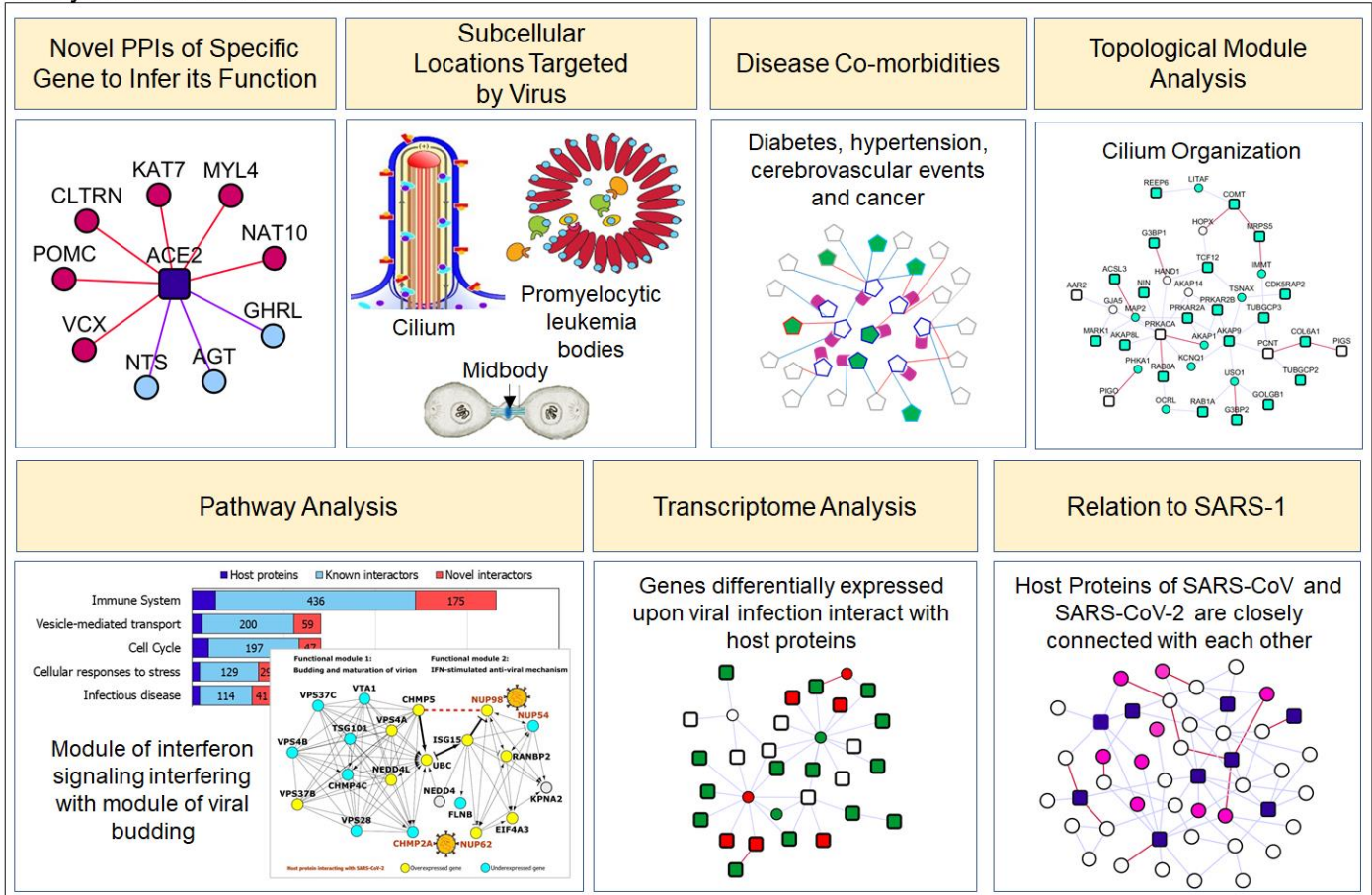
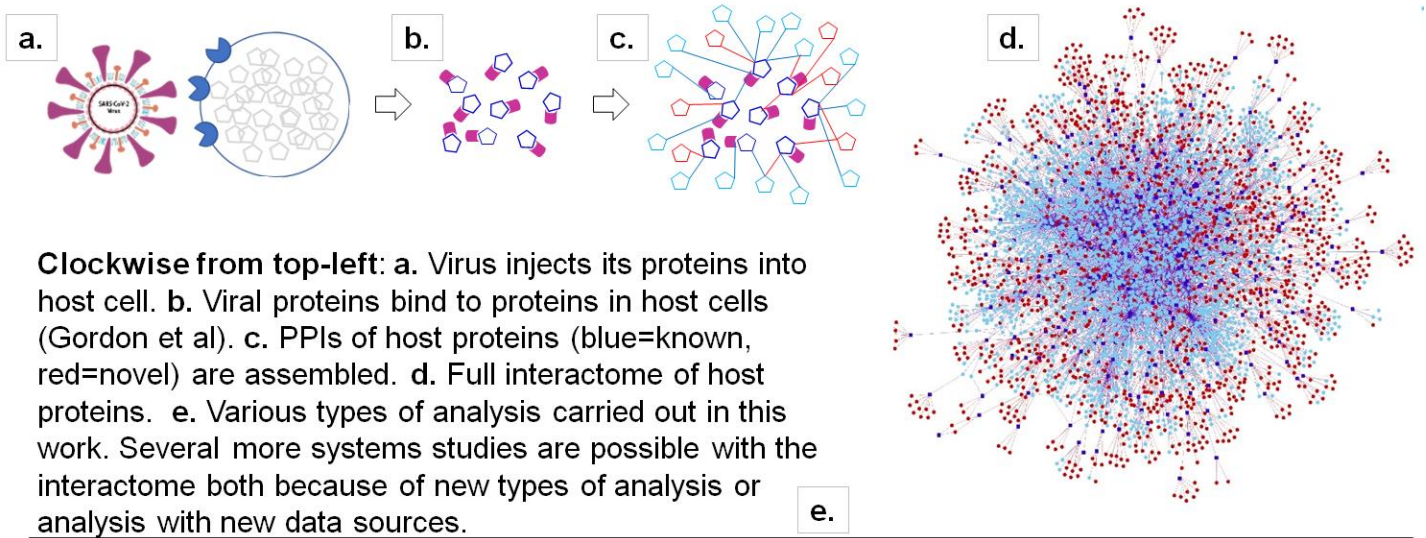
World over, people are looking for solutions to tackle the pandemic coronavirus disease (COVID-19) caused by the virus SARS-CoV-2/nCoV-19. Notable contributions in biomedical field have been characterizing viral genomes, host transcriptomes and proteomes, repurposable drugs and vaccines. In one such study, 332 human proteins targeted by nCoV19 were identified. We expanded this set of host proteins by constructing their protein interactome, including in it not only the known protein-protein interactions (PPIs) but also novel, hitherto unknown PPIs predicted with our High-precision Protein-Protein Interaction Prediction (HiPPIP) model that was shown to be highly accurate. In fact, one of the earliest discoveries made possible by HiPPIP is related to activation of immunity upon viral infection. We found that several interactors of the host proteins are differentially expressed upon viral infection, are related to highly relevant pathways, and that the novel interaction of NUP98 with CHMP5 may activate an antiviral mechanism leading to disruption of viral budding. We are making the interactions available as downloadable files to facilitate future systems biology studies and also on a web-server at <http://hagrid.dbmi.pitt.edu/corona> that allows not only keyword search but also queries such as “PPIs where one protein is associated with ‘virus’ and the interactors with ‘pulmonary’”.

Introduction

COVID-19 (Coronavirus Disease 2019) is an infectious virus outbreak which emerged as an epidemic in one city in December 2019 and within 3 months swept across 213 countries and territories into a pandemic global health crisis with more than 3 million confirmed infected cases and 217,000 deaths as of April 30, 2020 ([WHO](#)). The novel coronavirus (SARS-CoV-2/nCoV-19) has been identified as the causative agent of this coronavirus disease named COVID-19 (19 refers to the year 2019 when it started).¹ nCoV19 belongs to the large family of coronaviruses (coronaviridae) known to be pathogenic in mammals and birds. They are enveloped viruses with a positive-sense single-stranded RNA genome of 26.2-31.7 kilobases.² In humans, they cause respiratory infections ranging from the common cold to possibly-fatal acute respiratory distress syndrome (ARDS) and acute lung injury (ALI), which are noted in COVID-19 as well as in its predecessors, namely, SARS (severe acute respiratory syndrome, 2002-2003) and MERS (middle east respiratory syndrome, 2012).³ nCoV19 is airborne, and causes no symptoms in several infected people who may become silent carriers of the disease to more vulnerable population; COVID-19 is spreading at an exponential rate globally, prompting efforts to develop technologies, therapeutics and vaccines to prevent it from reaching its projected peak. Scientists across the globe are studying the genetic features of nCoV19 and investigating the mechanisms of its host invasion and host-response to viral infection, in hopes of discovering treatment strategies to combat the outbreak.

The viral infection sets off a cascade of interactions among multiple genes and proteins in the host cell. This complex network has the potential to restrict viral replication in host cells, or conversely to be taken over by the virus for its replication. Several research groups have studied the effects of nCoV19 on the host from a systems-level perspective.⁴⁻⁶ 332 human proteins that bind to nCoV19 proteins were identified through affinity purification – mass spectrometry (AP-MS) by Gordon et al.⁶ Melo et al. identified 120 genes differentially expressed in the A549 human cell line on nCoV19 infection.⁵ Bojkova et al. monitored nCoV19 infection in Caco2 cell line and generated temporal infection profiles of 2,687 genes in the host transcriptome and 6,258 gene in the proteome.⁷ Data generated by these studies can be employed to

conduct a systematic, unbiased and data-driven investigation into COVID-19 from the perspective of the host, by constructing the relevant protein interactome (i.e. protein-protein interaction network).



Dissemination of Interactome on webserver: <http://severus/dbmi.pitt.edu/corona>

Figure 1. Concept diagram of the analysis presented in the paper.

Protein-protein interactions (PPIs) drive the cellular machinery and facilitate biological processes including signal transduction, formation of cellular structures and enzymatic complexes. When viral proteins bind to some proteins in the host cell, this effect spreads through regulatory and biophysical interactions along the interactome affecting other proteins in the PPI network, posing deeper implications for the infection, host immunity, and the effect of therapeutics.⁸ Despite being critical to unravelling novel disease mechanisms and drugs, ~85% of estimated PPIs are currently unknown and several disease-associated genes have no known PPIs. More than ~600,000 PPIs are said to exist in the human

interactome⁹ and only ~90,000 PPIs are known from PPI repositories such as HPRD¹⁰ and BioGRID¹¹. Detecting the PPIs using experimental techniques such as co-immunoprecipitation (Co-IP)^{12,13} are prohibitively laborious and time-consuming at large scale. Tens of thousands of PPIs are being added into the interactome through systematic high throughput studies with yeast two hybrid system¹⁴ and AP-MS¹⁵. A large part of the interactome (>70-80%) remains unknown yet. Hence, computational algorithms have been developed to predict PPIs in human as well as model organisms.¹⁶⁻²⁴ We developed a computational model called **HiPPIP** (high-precision protein-protein interaction prediction) that was deemed highly accurate by computational evaluations, and experimental validations of some pairs (sixteen PPIs were tested by experiments and all sixteen were shown to be true PPIs (**Supplementary Table S1**)).^{25,26} HiPPIP computes features of protein pairs such as cellular localization, molecular function, biological process membership, genomic location of the gene, and gene expression in microarray experiments, and classifies the pairwise features as *interacting* or *non-interacting* based on a random forest model.²⁵ The threshold of HiPPIP to classify a protein-pair as “a PPI” was set high in such a way that it yields very high-precision even if low recall. Some of the PPIs predicted by HiPPIP proved to be of high translational impact. For example, we predicted that the human OASL protein (*IFN-inducible oligoadenylate synthetases-like*) interacts with RIG-I (*retinoic acid-inducible gene D*); it was validated to be a true PPI. Further investigations conclusively showed that this interaction is responsible for activating cellular innate immunity to virus infections.²⁶ OASL enhances antiviral signalling mediated by the viral RNA sensor RIG-I by binding through its C-terminal ubiquitin-like domain.²⁶ Other high-impact results from interactome analysis include shared PPIs explaining inverse epidemiological relationship between schizophrenia and rheumatoid arthritis²⁷ and cilia-transduced cell signaling in congenital heart disease,^{28,29} and more.³⁰

In this work, we present the host protein interactome (HoP Interactome) of nCoV19, namely PPIs of the 332 human proteins that the virus proteins bind to (Gordon et al.⁶). Key contributions of this work are a set of novel PPIs that are computationally predicted with high-precision, combined with known PPIs from public databases, and the results of analyzing the network of interactions with functional annotations and with nCoV19-relevant transcriptomic and proteomic data. Importantly, we are making this interactome, including about two thousand hitherto unknown PPIs with rich annotations on a webserver with search capabilities to the scientific community and in graph formats amenable for computational analyses. A concept diagram of analysis carried out here is shown in Figure 1.

At a time when biomedical science researchers have closed their labs to practice social distancing,³¹ this digital resource presents new computationally derived information at their fingertips, presenting an avenue for continued scientific enquiry. A biomedical scientist typically has undeterred focus, say on a specific protein or a pathway even when broadly interested in the study of a specific disease. When presented with numerous computational results about the protein or pathway, the biomedical scientist typically needs to formulate a functional hypothesis about what the prediction entails if/after validated to be true. That is, the experiment is built around a hypothesis formulated by the scientist, for which one of the computational results serves as the seed. We believe that it would be very timely and of immense value to present these computational results to the biomedical community, when the scientists are currently having to work in a digital mode away from their wet labs. Many of these two thousand novel PPIs may serve as seeds for new directions to science, similar to how OASL-DDX58 interaction has advanced the study of the role of OASL proteins in viral immunity.²⁶

Results

We collected 332 host proteins that were identified to interact with 27 nCoV19 viral proteins by Gordon et al.⁶ To assemble the interactome of these host proteins, we compiled known PPIs from HPRD¹⁰ (Human Protein Reference Database) and BioGRID¹¹ (Biological General Repository for Interaction Datasets) and predicted novel PPIs by applying the HiPPIP algorithm described earlier.³² Note that the interactome is human protein interactome, and not a host-virus interactome; the relevance to COVID19 is that the core proteins for which the interactome is assembled are those that viral proteins bind to. HiPPIP predicted ~2,600 PPIs of which ~600 PPIs are previously cataloged in HPRD and BioGRID, leaving ~2,000 PPIs to be considered as novel PPIs of the host proteins. There are an additional 3,500 PPIs that are known and not predicted by HiPPIP (this is as it should be: HiPPIP prediction threshold has already been fixed³² to achieve *high precision* by compromising *recall*, which is required for adoption into biology; in other words, it is set to

predict only a few PPIs out of the hundreds of thousands of unknown PPIs, but those will be highly accurate). As reported in Supplementary File 1, prior validations of 16 PPIs predicted by HiPPIP in our other studies validated all 16 to be true; the experiments were carried out by diverse research labs. Overall, the host protein (HoP) interactome consists of 4,408 proteins and 6,076 interactions (**Supplementary Data File 1**). A partial network of host proteins and their novel interactors is shown in **Figure 2A** (see **Supplementary Figure S1** for the full network of novel interactors).

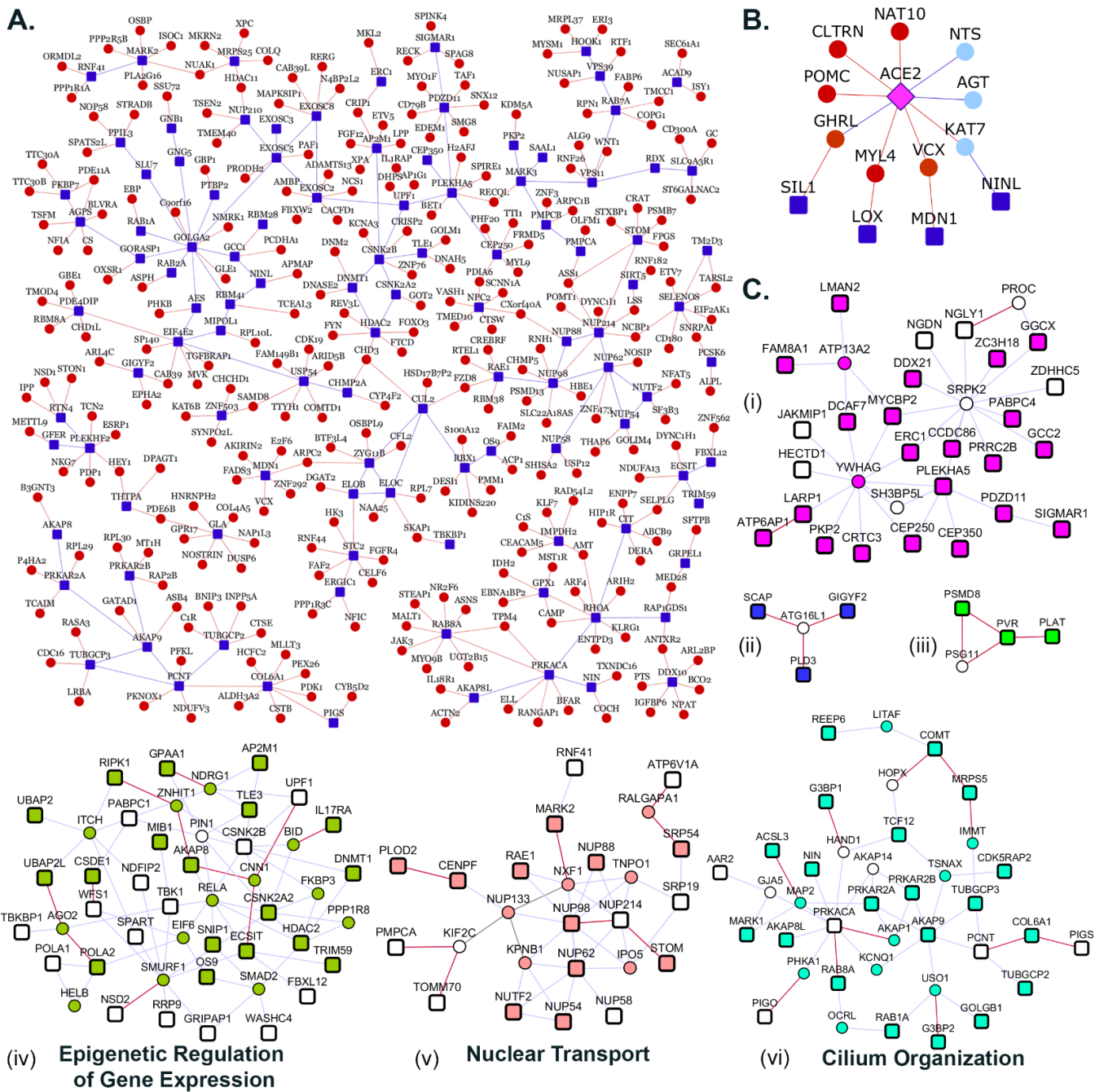


Figure 2. Network views of protein-protein interactions: **A. Partial view of the HoP interactome:** Genes are shown as nodes and PPIs as edges. As the full network is very large, only a partial view showing a large connected component of novel interactors and their neighbors, all of which have transcriptomic/proteomic evidence related to nCoV19 (Supplementary Data File S2), are shown. Legend: Dark blue square-shaped nodes: host proteins; red nodes/edges: novel interactors/interactions; blue nodes/edges: known interactors/interactions. **B. ACE2 Interactome:** PPIs of ACE2 protein, extended to show four host proteins that are two-edges away from it. Color legend is as in A. **C. Modules identified from network topology:** Six out of seventeen total modules each with 3 or more nodes are shown, whereas remaining modules are shown in Supplementary Figure S2. Each module is depicted in a different color. Within each module, colored nodes depict genes with at least one evidence of transcriptomic/proteomic relevance to nCoV19.

We verified whether any of the 2,000 novel PPIs came up in recent interactome maps HuRI (HI-Union)¹⁴ and BioPlex¹⁵. While there was no overlap with HuRI union dataset, there were 8 PPIs in the BioPlex map (ADAM9-ADAM32, P3H3-OS9, PVR-NECTIN2, SRRM2-SNIP1, PABPC4-LUC7L2, PRKACA-AKAP1, NDUFA13-ECSIT, and NPTX1-NPTX2). The small overlap is not surprising because even high-throughput biotechnological methods discover different parts of the interactome with only small overlaps with each other⁶, this demonstrating complementary strengths.¹⁴

Applications of this network are two-fold: (1) biologists, who typically focus their research on specific proteins or a pathway may look up the novel interactions relevant to that protein or pathway (e.g.²⁶) (2) computational systems biologists may investigate it in conjunction with transcriptomic/proteomic data (e.g.).^{30,33-35} To facilitate '(1)', we are making these results available over an interactive webserver, and to facilitate '(2)', we are releasing the data as downloadable files in various formats.

We employed 'Netbox'³⁶ to identify modules based on network topology. It expands the core proteins by adding nodes from the interactome whose number of links to core proteins are statistically significant compared to its degree in the human interactome. From this network, it identifies highly interconnected modules. It was able to connect 323 proteins (220 host proteins and 103 linker proteins) into 21 modules, of which 14 modules had 4 or more nodes each (**Supplementary Figure S2**). For comparison, when novel PPIs are not included, it connects 199 proteins (138 host proteins and 61 linker proteins) into 18 modules of which 10 had 4 or more proteins each. Scaled modularity score (z-score compared to random networks) was 17.0 with novel PPIs, and it was 14.5 without novel PPIs (z-score compared to corresponding random networks). Bioinformatic analysis of the computed modules showed that five modules formed with novel interactors had statistically significant enrichment of Gene Ontology biological process terms: *epigenetic regulation of gene expression* (p-value=3.3E-04, odds ratio=10.4), *nuclear transport* (p-value=2.4E-12, odds ratio=21.6), *cilium organization* (p-value=1.28E-03, odds ratio= 7.8) *ribonucleoprotein complex biogenesis* (p-value=0, odds ratio=22.4), and *vesicle-mediated transport between endosomal compartments* (p-value=9.4E-06, odds ratio=123.4) (**Figure 2C i-vi**). When novel PPIs are excluded, some of these associations were missed and the modules were smaller, but three additional functional modules were found: *cell cycle G2/M phase transition* (p-value=0.0019, odds ratio=21.7, 20 proteins), *DNA replication* (p-value=0.0049, odds ratio=55.25, 3 proteins) and *cell-cell signaling by Wnt* (p-value=0.0049, odds ratio=9.3, 24 proteins) (**Supplementary Table S2**).

ACE2 Interactome

SARS-CoV-2 engages the host receptor ACE2 (*angiotensin-converting enzyme 2*) for cell entry.³⁷ Viral entry happens prior to the interaction of the viral proteins with host cellular proteins; it was the latter that was studied by Gordon et al.⁶ Therefore, it was not part of the 332 core genes considered in constructing the interactome. Owing to its crucial role in nCov19 infection, we assembled its known and novel PPIs separately and found that it was connected to four host proteins (SIL1, LOX, MDN1 and NINL) through an intermediate interactor, i.e. separated by two edges, where one or both intermediary PPIs are novel predicted ones (see red edges in **Figure 2B**).

These connections reveal interesting insights: ACE2 is a key player of the renin-angiotensin hormone system that regulates blood pressure and electrolyte balance.³⁸ In line with this, we found that its interactors AGT (*angiotensin*), GHRL, CLTRN and POMC are associated with the Reactome Pathway *peptide hormone metabolism* (p-value=2.9E-05). ACE2 and its interactors were also enriched in the Gene Ontology Biological Process *circulatory system process* (ACE2, AGT, NTS, POMC, GHRL and the host protein MYL4; p-value=0.001). Three host proteins are associated with numerous vascular and cardiac phenotypes: LOX with *abnormality of blood volume homeostasis*, *aortic root aneurysm*, *ascending aortic dissection*, *carotid artery dilatation*, *coronary artery atherosclerosis*, *cystic medial necrosis of the aorta*, *descending thoracic aorta aneurysm*, *dilatation of the cerebral artery*, *left ventricular failure*, *peripheral arterial stenosis*, MYL4 with *paroxysmal atrial fibrillation* and *bradycardia*, and SIL1 with *abnormal aldolase level*.

The co-morbidity of hypertension, diabetes and cardiovascular among the group of COVID19 patients with high fatality rate warrants a closer look at ACE2 and other host proteins linked to cardiac and vascular phenotypes.

Wiki-CORONA: A web server of novel host PPIs

The HoP interactome is available on a website called *Wiki-CORONA* (<http://hagrid.dbmi.pitt.edu/corona/>). It has advanced-search capabilities, and presents comprehensive annotations, namely Gene Ontology, diseases, drugs and pathways, of the two proteins of each PPI side-by-side. Here, a user can query for results such as “PPIs where one protein is anti-viral and the other is involved in immunity”, and then see the results with the functional details of the two proteins side-by-side. The PPIs and their annotations also get indexed in major search engines like Google and Bing. Querying by biomedical associations is a unique feature which we developed in *Wiki-Pi* that presents known interactions of all human proteins.³⁹

Transcriptome Analysis

Significantly large number of proteins in the interactome were differentially expressed in epithelial cells infected with SARS coronavirus (GSE17400, Calu-3 cell, 48 hours post-infection; p-value=4.76E-12). Several proteins also showed differential expression in the transcriptome level after infection by Urbani strain of SARS coronavirus (GSE37827, Calu-3 cells, 72 hours post-infection) and in peripheral blood mononuclear cells of SARS patients (GSE1739⁴⁰). These latter two datasets of differential expression did not show statistically significant overlaps; yet, the transcriptomic evidence highlights key protein-encoding genes associated with viral infection that interact with the core proteins considered in this study. As several of the interactors here are revealed through computational prediction, the information that they are differentially expressed in SARS/SARS-Cov-2 infections presents opportunity to prioritize consideration of novel PPIs for further study.

Melo et al. had identified 120 differentially expressed genes (DEGs) associated with nCoV19 infection in the A549 cell line.⁵ Of these, only 2 were common with the 332 host proteins identified through AP-MS study⁶ (‘host proteins’). However, our study revealed several interesting links between the two sets: (a) 31 DEGs are *direct interactors* of 38 host proteins, with some DEGs interacting with multiple host proteins; (b) Thirteen novel PPIs exist between the two sets: AAR2-SAMHD1, TUBGCP2-C1R, IMPDH2-C1S, GOLGA7-TCIM, RAB8A-STEAP1, GDF15-EHF, REEP5-PDK4, FAM162A-PARP14, STOML2-CDH1, FGA-RAB14, FBXL12-C19orf66, ECSIT-C19orf66 and EIF4H-PTPN12. (c) 108 DEGs and 285 host proteins are highly interconnected through 808 common interactors (statistically significant overlap with odds ratio=1.5, p-value=7.12E-54). (d) Pathway enrichment analysis of overlapping interactome (consisting of shared interactors and the DEGs and host proteins that they interact with) revealed several immune-related pathways with FDR-corrected p-value<0.05.

2,630 proteins in the interactome that are supported by the above mentioned transcriptomic and proteomic evidence are listed in **Supplementary Data File S2**. In fact, the selected novel interactors shown in **Figure 2A** all have transcriptomic/proteomic evidence.

We studied tissue specific expression of the proteins in the interactome using GTEx data.⁴¹ Genes with an expression level greater than 1 TPM (transcripts per million) and relative expression at least 5-fold higher in a particular tissue (tissue-enriched) or a group of 2-7 tissues (group-enriched) were considered. As expected, many genes showed specific expression in lung which is the target tissue of the virus, and in spleen which regulates the immune response of the host (**Figure 3**). New PPIs were found of host proteins with 37 lung-specific proteins and 49 spleen-specific proteins. Host proteins also interacted with several brain and heart tissue specific proteins, which is of importance as cerebrovascular diseases and coronary heart diseases are co-morbidities among COVID-19 non-survivors.⁴² (**Figure 3**).

Gene Ontology Term Enrichment

PML bodies and the midbody may function as subcellular targets of nCoV19

Gene Ontology enrichment analysis of the interactome identified several subcellular locations that may be targeted by nCoV19. Cellular locations included points of virus entry such as the *cell-substrate junction*, *nuclear periphery* and specific sites from where viral proteins may potentiate viral replication, gene expression and modulate the immune response of the host such as the *midbody*, *nuclear chromatin* and *PML body* (each term with p-value<0.0001). PML

(promyelocytic leukaemia bodies are nuclear sub-compartments that repress viral replication through entrapment or epigenetic silencing of the viral genomes.⁴³ Components of PML bodies activate interferon-stimulated genes and cytokines, and may also be upregulated on induction of interferons.⁴³ Therefore, it is conceivable that viruses target PPIs

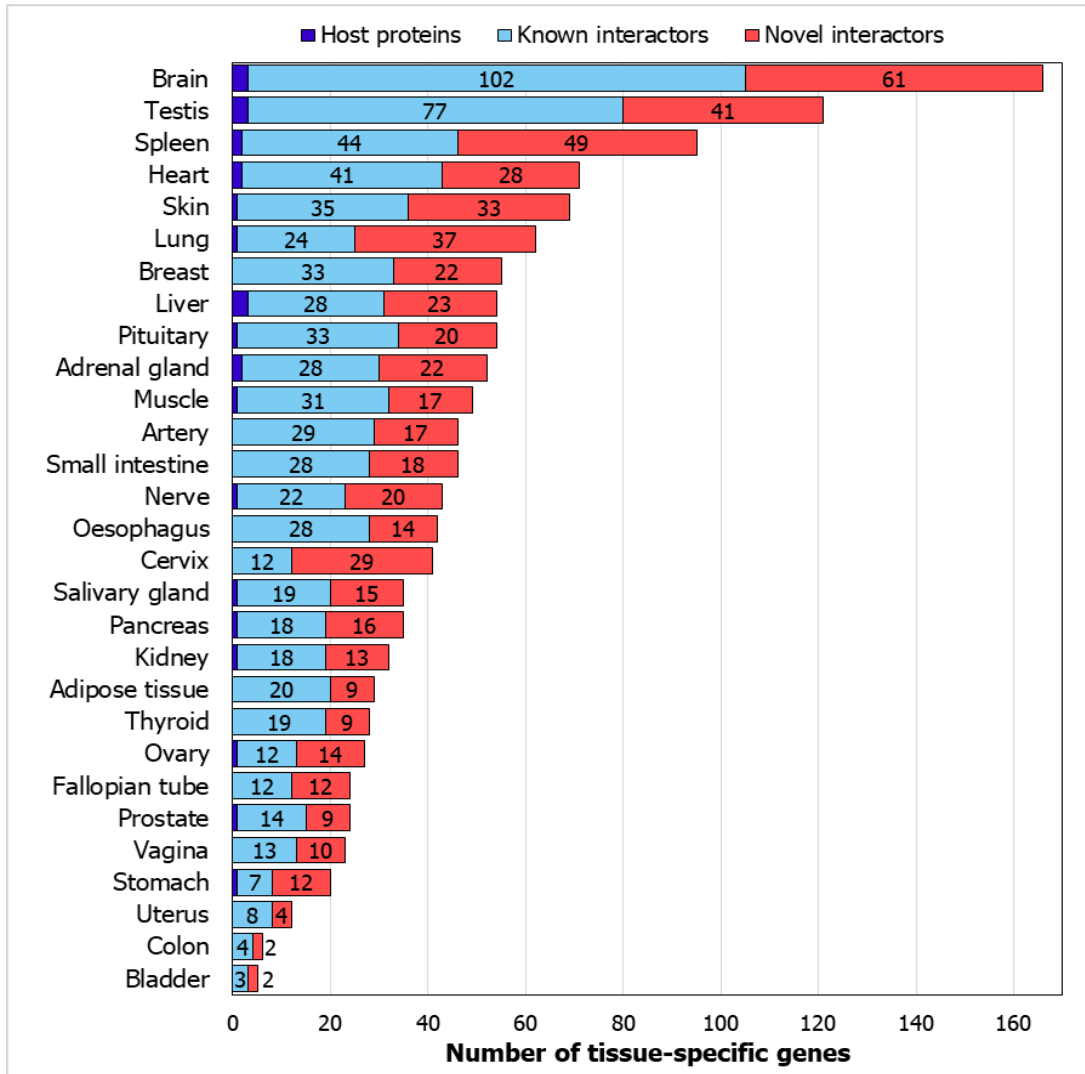


Figure 3. Tissue-specific genes in CoV-HP interactome:

Number of genes from the interactome which show tissue specificity are shown. The genes show at least 5-fold

in these structures to circumvent anti-viral defences of the host cell. Sixty-one proteins in the HoP interactome are PML components. These include the host protein AKAP8L, which has been known to promote retroviral gene expression, and 55 known interactors and 5 novel interactors (RNF111, SP140, ELF4, NFE2, CIART) of other host proteins. Our model predicted an interaction of EIF4E2 with SP140, an interferon-inducible PML component; nCoV19 may perturb this PPI. The midbody is a microtubule-rich stricture that connects the daughter cells and marks the site of abscission during cytokinesis. Viruses have been known to recruit certain protein complexes that also localize to the midbody during cytokinesis, to the host cell membrane to promote its scission and thereby the release of viruses.⁴⁴ This co-opting of proteins may explain the enrichment of midbody proteins in the HoP interactome. 83 proteins in the HoP interactome, including 11 host proteins (RHOA, CENPF, CIT, RAB8A, NUP62, SCCPDH, SPART, RDX, ARF6, CNTRL and RALA), 63 known interactors and 9 novel interactors (KIF4A, BIRC5, INCENP, ALKHB4, DNMT2, DDX11, ARL2BP, ABRAXAS2 and WIS) localize to the midbody.

Cell cycle phase transitions in the host may be modulated by nCoV19

Enriched biological processes in the interactome included (*G1/S and G2/M*) mitotic cell cycle phase transitions, regulation of vesicle-mediated transport, covalent chromatin modification and nuclear transport (p-value<0.0001). The response of the host cell to nCoV19 infection has been shown to be significantly delayed and devoid of several anti-viral mechanisms.⁵ During early stages of the infection, it is possible that the virus induces a G1/S phase transition to surreptitiously synergize the replication of the viral genome with that of the host genome.⁴⁵ In the later stages, it may block the G2/M phase transition to maximise the levels of viral genome.⁴⁵ We found novel (predicted) interactions of host proteins with 34 proteins involved in cell cycle phase transition: ANAPC4, ANAPC7, ARPP19, CCNB3, CDC14B, CDC16, CDC7, CEP164, CETN2, CLSPN, CRLF3, DCTN1, DNMT2, DYNC1H1, E2F6, ENSA, FBXL7, GF11, GML, HYAL1, INHBA, JADE1, NEUROG1, NPAT, ORC2, PPM1D, RAD17, SPDYA, TAOK2, TICRR, TRIAP1, XPC, ZFP36L1, ZNF655.

Pathway Associations

Resveratrol-modulated sub-network of genes involved in the tristetraproline pathway

Using WebGestalt,⁴⁶ we compiled the list of the Reactome pathways (**Figure 4**), which showed a statistically significant enrichment of several pathways related to viral entry and infection such as *infectious disease*, *HIV life cycle*, *vesicle-mediated transport* and *membrane trafficking*. Several immunity-related pathways which mediate the host response such as *MyD88 dependent TLR4 signalling* and *ISG15 anti-viral mechanism* were also identified.

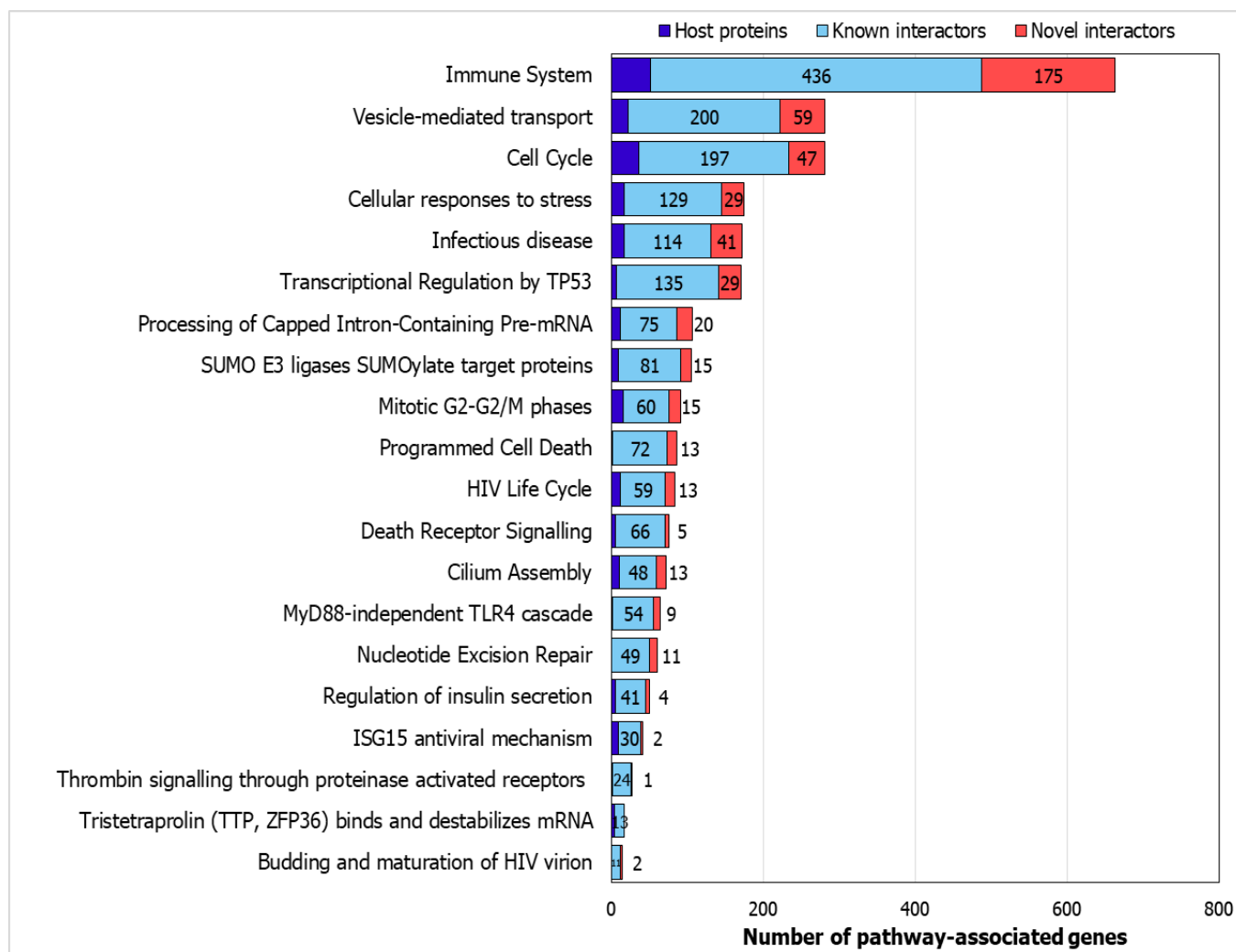


Figure 4. Pathways associated with the CoV-HP interactome; Number of genes from the interactome associated with selected pathways are shown.

The transcriptional profile of the host cell after nCoV19 infection had revealed a remarkably limited anti-viral response compared to that elicited by seasonal influenza-A and respiratory syncytial viruses.⁵ This prompted us to inspect a post-transcriptional regulatory pathway that was enriched in the HoP interactome, namely, *tristetraproline (ZFP36) binds and destabilizes mRNA* (p-value<0.0001). ZFP36 is an RNA-binding protein that targets AU-rich sites in the mRNA transcripts coding for immune proteins and destabilizes them by promoting the deadenylation of their poly(A) tails.^{47,48} YWHAB increases cytoplasmic localization of ZFP36, possibly preventing destabilization of these genes and attenuation of immune response.⁴⁹ We extracted the direct PPIs of the 17 genes belonging to this pathway from the HoP interactome and isolated this sub-network for further inspection (**Figure 5**). Our predictions show that the host protein DCAF7, which is known to function as a scaffold protein and a facilitator of PPIs, interacts with YWHAB and ACE1, belonging to the class of receptors targeted by nCoV19 (**Figure 5**). This raises the possibility that the virus protein Nsp9 (which interacts with DCAF7) may somehow perturb YWHAB-induced cytoplasmic localization of ZFP36 through its action on DCAF7. Nsp9 may activate or promote the sequestration of YWHAB with DCAF7, thereby reducing its capacity to form a complex with ZFP36. YWHAB-mediated destabilization of immune genes may then lead to a weakened immune response, creating a conducive environment for nCoV19 infection. We also identified 3 drugs targeting the proteins in this sub-network using Drug Bank:⁵⁰ resveratrol targeting KHSRP and APP, known interactors of the host protein EXOSC2, which is involved in the tristetraproline (TTP) pathway, staurosporine targeting TTP-associated MAPKAPK2 which has been predicted to interact with PABPC1 and dacarbazine targeting the host protein POLA2 (**Figure 5**). Gene expression profiles induced by these drugs in various cell lines were found to have a negative correlation with SARS-associated gene expression profiles, namely, that of lung fibroblast MRC5 cells infected with SARS-CoV and in peripheral blood mononuclear cells of SARS patients (analysis using NextBio; <https://www.nextbio.com>).^{51,52} Resveratrol has been proposed as a therapeutic option for nCoV19 based on its antagonistic properties against MERS-CoV.⁵³

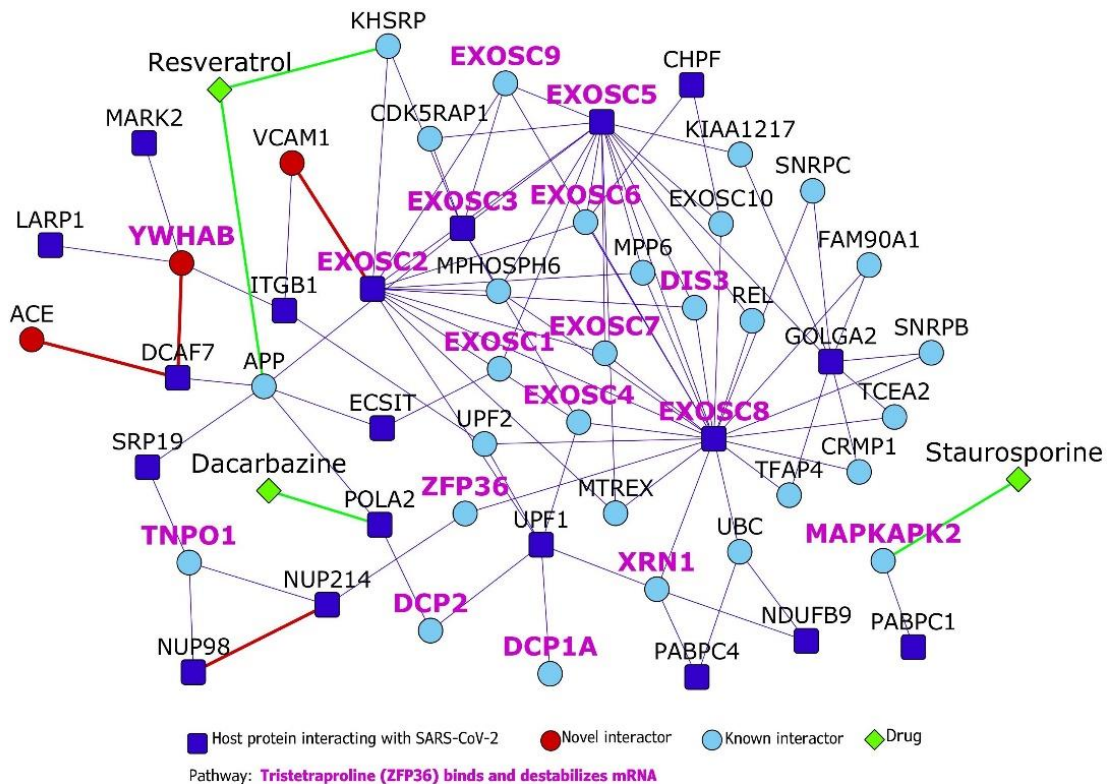


Figure 5. Tristetraproline pathway: Dark blue nodes are host proteins, light-blue nodes are known interactors, and red nodes are novel interactors. Diamond-shaped green nodes depict drugs. Nodes with purple labels are proteins involved in the tristetraproline pathway.

Genetic Disorder Enrichment Analysis

Network proximity of genes associated with diabetes and hypertension to the host proteins

We studied the association of interactome genes with any genetic disorders/traits in the OMIM database. 155 genes in the interactome, including 9 host protein-encoding genes, and 121 known interactors and 25 novel interactors of host proteins, were found to be associated with 35 disorders (overlap of each disease had p -value <0.05). This included 13 types of cancers, 7 metabolic disorders, 4 neurological disorders, 3 developmental disorders, 2 eye-related disorders, 2 vascular diseases, 1 infectious disease, 1 inflammatory disorder, 1 respiratory disorder and 1 skin disease (**Figure 6 and Table 1**). Some of these diseases enriched in the interactome are co-morbidities among non-survivors and critically ill COVID patients (e.g. diabetes, hypertension, cerebrovascular events and cancer).^{42,54} Thirteen genes in the interactome were associated with *non-insulin dependent diabetes mellitus* (odds ratio=10.8, p -value=4.38E-10), 6 genes with *essential hypertension* (odds ratio=12, p -value=2.34E-05), 3 genes with *ischemic stroke* (odds ratio=14.4, p -value=0.0017) and 10 genes with *lung cancer* (odds ratio=14.1, p -value=2.36E-09). Network proximity of the proteins associated with these co-morbid conditions to the nCoV19 host proteins may explain why patients with these conditions are increasingly affected by the viral infection. Further investigations are necessary to dissect these co-morbidities. Treatment strategies that prevent the deterioration of the underlying genetic conditions must be devised to combat COVID-19 in susceptible individuals. Additionally, neurological disorders such as *Alzheimer's disease* (odds ratio=15.3, p -value=5.13E-07) and *schizophrenia* (odds ratio=12, p -value=4.19E-06) were also found to be enriched in the interactome, warranting further investigations into these potential co-morbidities.

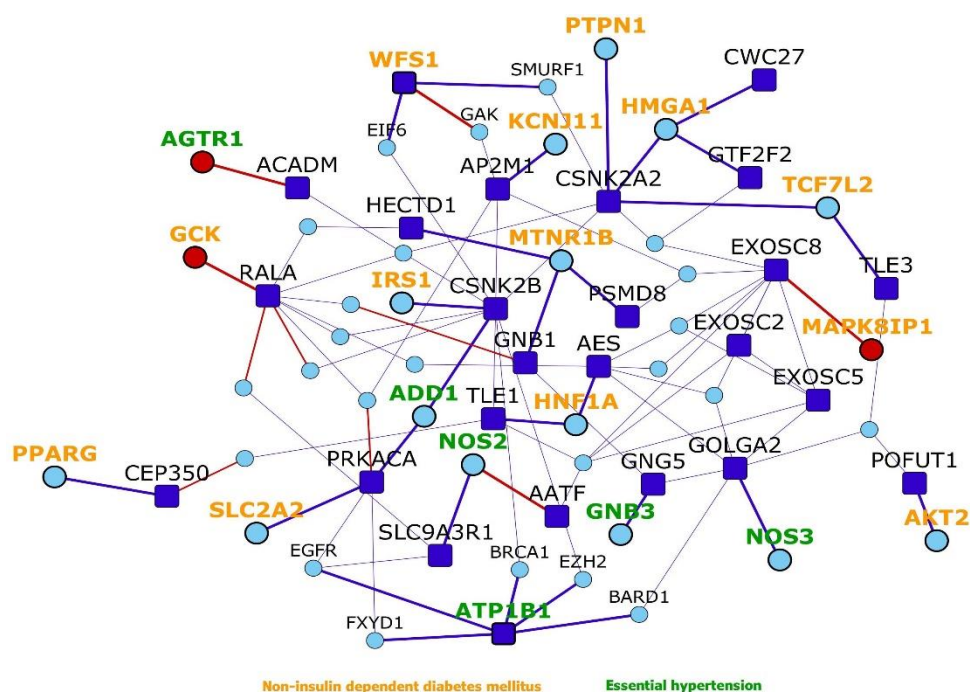


Figure 6. Network proximity of diabetes and hypertension to host proteins: Dark blue nodes are host proteins, light blue nodes are known interactors and red nodes are novel interactors. Nodes with orange labels are diabetes-associated genes, whereas nodes with dark green labels are hypertension-associated genes.

Interconnections to Ciliary Proteins

SARS coronavirus which emerged in 2002 has been known to induce necrosis in ciliated airway epithelium of humans in a species-specific manner.⁵⁵ nCoV19's host receptor ACE2 is highly expressed in ciliated respiratory cells.⁵⁶ Cilia may serve as virus entry points and potential modulators of viral pathogenesis. This conjecture prompted us to investigate the ciliary association of the host proteins and their interactors in the HoP interactome. For this, we studied its overlap with an

interactome of 165 ciliary proteins that we constructed in a similar manner. The ciliary protein interactome contained 1,665 proteins. 617 of these proteins, and specifically 30 core ciliary proteins, are also found in nCoV19's host protein interactome, and the overlap was found to be statistically significant (p-value=2.24E-10, odds ratio=1.22). Thirteen novel predicted interactions connected host proteins to ciliary proteins: NUP98-CHMP5, GG3BP1-DNAH1, SEPSECS-DNAH1, NEK9-IFT43, TLE1-DNAH5, ATP6AP1-CETN2, C1orf50-ZMYND12, RAB10-IFT172, TOR1AIP1-GPR161, DNAJC19-CETN3, NLRX1-IFT46, FKBP7-TTC30B, POLA2-TMEM216 and NDUFB9-DRC7.

Pathway analysis of the 617 common proteins (i.e., common to HoP and cilia interactomes) revealed two interesting pathways: *budding and maturation of HIV virion* (p-value=1.29E-06; odds ratio=8.8) and *anti-viral mechanism by IFN-stimulated genes* (p-value=0.013; odds ratio=2.98). We predicted that ciliary protein CHMP5 involved in the former pathway interacts with host protein NUP98 which is involved in the latter pathway. This prompted us to ask whether the predicted interaction connected the functional modules of viral budding to interferon (IFN) signaling.

Novel interaction of NUP98 with CHMP5 may activate an IFN-stimulated pathway that interferes with viral budding

We extracted the PPIs of the 20 proteins belonging to viral budding and IFN pathways and isolated this sub-network, containing 171 proteins and 176 PPIs, for further analysis. Firstly, we identified 343 functional interactions (i.e. activation, inhibition etc.) among 98 proteins in the network. Strikingly, distinct functional modules were identified for both the pathways; CHMP5 seemed to serve as a connector from the viral budding pathway to the IFN pathway through NUP98 (**Figure 7**). The gene UBC was shared between the clusters.

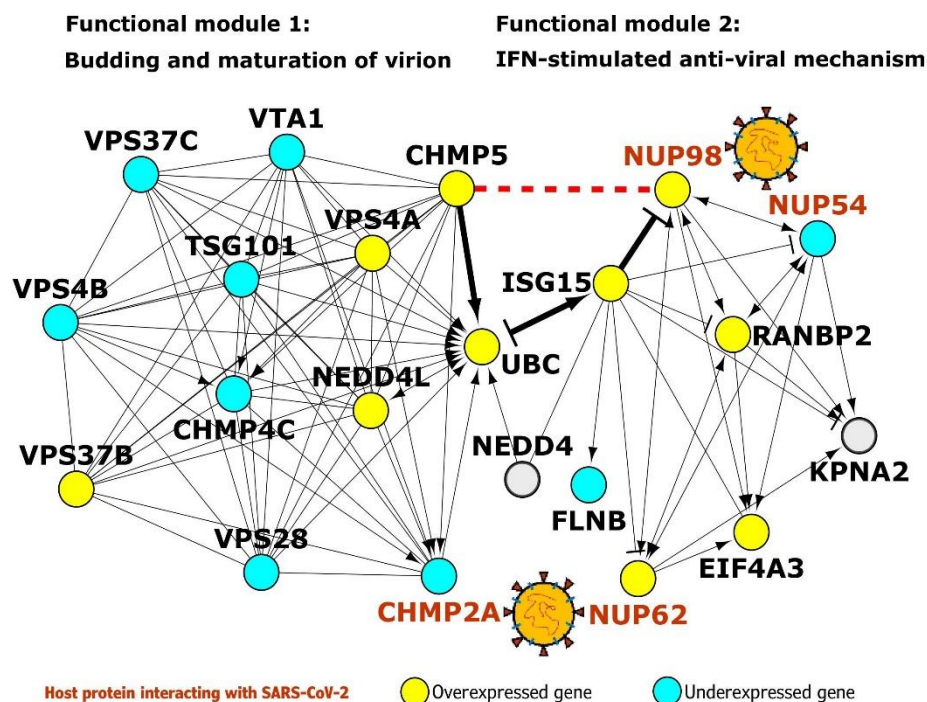


Figure 7. Functional modules of viral budding and interferon-mediated anti-viral pathway: The novel predicted interaction of the host protein NUP98 with CHMP5 (red dashed line) may set off an ISG15-mediated pathway that culminates in interference of viral budding. Yellow nodes and light blue colored nodes indicate overexpression and underexpression in SARS-CoV-affected Calu-3 lung cells. → indicates ‘activation’, -| indicates ‘inhibition’ and – indicates ‘part of the same complex/physical association’.

We then checked whether the genes in these modules were differentially expressed in Calu-3 lung cells infected with SARS CoV Urbani (for 72 hours) versus mock infected cells. This was done to identify the functional interactions that remain active during viral infection. It was assumed that differential expression of the genes would directly impact the proteins encoded by them and their interactions. 20 genes including NUP98 and CHMP5 were found to be differentially expressed (**Figure 7**). Viruses hijack the ESCRT/VPS4 (*endosomal sorting complex required for transport*) machinery of

the host cell to release viral particles through membrane scission.⁵⁷ This machinery is normally recruited during endocytic and membrane repair processes in the host cell. The process of membrane scission is catalyzed by various ESCRT-III proteins including CHMP5.⁵⁷ VPS4 is an ATPase that is found in the cytoplasm in its inactive form. Activation of the VPS4 and its ATPase activity is essential for membrane budding and the release of viral particles.⁵⁷ VPS4 is activated on membranes in the presence of its co-activator VTA (also known as LIP5). VTA is delivered to the membranes by ESCRT-III proteins such as CHMP5.⁵⁷ Hence, the interaction of VPS4 and VTA is facilitated by CHMP5. However, when interferons are induced in the host cell following viral infection, ISGs (*interferon stimulated genes*) such as ISG15, a dimer homologue of ubiquitin, may be activated.⁵⁷ ISG15 may then conjugate to CHMP5 and promote its accumulation in the membrane, effectively blocking the interaction of VTA with VPS4 and preventing viral budding.⁵⁷ The novel interaction of CHMP5 with NUP98 may serve as the critical juncture at which the IFN-stimulated anti-viral mechanism interferes with viral budding. NUP98, a protein induced on viral expression, has been shown to promote anti-viral gene expression in drosophila.⁵⁸ Both CHMP5 and NUP98 are overexpressed following SARS CoV Urbani infection. This interaction may serve as a signal for the initiation of ISG15-mediated interference of viral budding. ISG15 may further regulate this mechanism through feedback inhibition of NUP98. Hence, potentiation of this anti-viral mechanism through administration of recombinant interferon alfa-2b and interferon alfacon-1 may be a feasible therapeutic option for nCoV19. Both these interferons induce gene expression profiles negatively correlated with SARS-associated profiles. The machinery of ESCRT-III and VPS4 is co-opted into two subcellular structures that are intricately linked to cilia function, namely, the centrosomes and the midbody.⁴⁴ It is important to study these structures as potential modulators of viral infections.

Discussion

In this study, to gain insight into the biological processes and pathways that may be involved in host upon nCoV19 infection, we assembled the interactome of the host proteins targeted by the virus. The host protein (HoP) interactome has ~4,000 previously known PPIs in addition to ~2,000 PPIs we computationally predicted. The interactome and its annotations are made available on the website that is freely accessible, *Wiki-Corona*. The HoP interactome was found to share a large and statistically significant overlap with a SARS-specific gene expression profile. Proteins with tissue-specific gene expression in lungs, spleen, brain and heart were also found in the interactome. Topologically connected modules in the network showed functional association to cilium organization, nuclear transport, ribonucleoprotein complex biogenesis, endosomal transport and epigenetic regulation of gene expression. The interactome is enriched for subcellular locations and host cellular processes that may be targeted by nCoV19. Host proteins were found to directly interact with proteins associated with two co-morbidities, hypertension and diabetes, that are commonly found among COVID-19 non-survivors. The nCoV19 host proteins and ciliary proteins shared several common interactors. The role of cilia as viral entry points and modulators of viral infections should be investigated further on this premise. With further analysis of the shared interactome, we hypothesize that the novel interaction of NUP98 with CHMP5, a ciliary (and centrosome and midbody-localizing) protein, may activate an IFN-stimulated pathway with the potential to interfere with viral budding.

The HoP interactome with nCoV19 associated host proteins and their interacting partners will help biologists, bioinformaticians and clinicians to piece together an integrated view on how host genes in various high throughput studies are functionally linked. To facilitate analysis by both computational and biomedical scientists, all the results are being released in multiple data formats in open access and via an interactive webserver (see **Data Availability**). While we carried out several analyses and presented the results here, the network of HoP interactome will facilitate several future systems biology studies derived from overlaying the interactome with data generated for research on coronaviruses in general and COVID-19 in particular. For example, consider comparing the 332 SARS-CoV-2 host proteins⁶ considered in this work, with the 65 host proteins of SARS coronavirus⁵⁹ ('SARS-2' and 'SARS-1' respectively, for clarity). There are only 4 proteins common to them (BZW2, MARK2, MARK3 and SMOC1). However, the interactome reveals that 41 SARS-2 host proteins interact with 29 SARS-1 host proteins and that eight of these PPIs are novel PPIs (N4BP2L2-EXOSC8, NMB-MRPS5, MKRN2-MRPS25, HOXC6-BRD2, XPA-AP2M1, VKORC1- DCTPP1, RSRP1-CEP350 and TPSAB1-ADAMTS1). Gene Ontology Biological Processes such as *autophagic mechanism* (odds ratio=4.5, p-

value=2.21E-05) *regulation of mitochondrion organization* (odds ratio=7.5, p-value=5.49E-05) and *protein localization to mitochondrion* (odds ratio=7.8, p-value=3.74E-04) may be commonly targeted by both these viruses as they were enriched among the proteins that are targeted by SARS-CoV-1 that also interact with host proteins of SARS-CoV-2. Mitochondria may be directly targeted by viral proteins, influenced by the cellular changes arising from viral infection or may even play a crucial role in viral pathogenesis due to their function as immune signalling hubs.⁶⁰ These organelles are constantly eliminated and recycled through a process called mitophagy. Viruses may modulate mitochondrial function and mitophagy to exacerbate infection.⁶⁰ Thus, in addition to the results presented here, several future studies may be carried out with the interactome to generate biologically insightful results that may be translated to biomedically actionable results.

Methods

Compilation of host proteins and prediction of novel interactions

Lists of 332 host proteins identified to interact with 27 nCoV19 proteins were compiled from data files in Gordon et al.⁶ Novel PPIs of these proteins were predicted using the HiPPIP model that we developed.³² Each host protein (say N1) was paired with each of the other human protein say, (M1, M2,...Mn), and each pair was evaluated with the HiPPIP model.³² The predicted interactions of each of the host proteins were extracted (namely, the pairs whose score is >0.5, a threshold which through computational evaluations and experimental validations was revealed to indicate interacting partners with high confidence). This resulted in 1941 newly discovered PPIs of the host proteins.

The significance of the overlap of this interactome with two datasets, namely, with the ciliary protein interactome and the interactome of 120 genes differentially expressed in nCoV19-infected A549 cell line,⁵ was computed based on hypergeometric distribution.

Network functional module analysis

Network modules among host proteins and their interactors were identified using Netbox.³⁶ Netbox reports modularity and a scaled modularity score, as compared with the modularity observed in 1000 random permutations of the subnetwork. Scaled modularity refers to the standard deviation difference between the observed subnetwork and the mean modularity of the random networks.⁶¹

Gene expression analysis

The severe acute respiratory syndrome (SARS) dataset used was obtained from GSE17400 (Calu-3 epithelial cells infected with SARS coronavirus for 48 hours versus mock infected cells). Statistical significance of the overlap between genes in the HoP interactome and those differentially expressed in SARS was computed based on hypergeometric distribution. In this method, P-value is computed from the probability of k successes in n draws (without replacement) from a finite population of size N containing exactly k objects with an interesting feature.

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

N= Total number of genes whose expression was assayed

K= Number of genes differentially expressed in SARS

n= Number of genes in the HoP interactome

k= $K \cap n$

Differential gene expression in Calu-3 lung cells infected with SARS CoV Urbani (GSE37827) and in peripheral blood mononuclear cells of SARS patients (GSE1739⁴⁰) was checked using the Base Space Correlation Engine, a software suite for analysis of publicly available gene expression datasets.⁶² Genes with fold change >2 or <1/2 were considered as significantly overexpressed and underexpressed respectively at p-value<0.05.

Tissue-specificity analysis

Tissue-specificity of the genes in the HoP interactome were checked using TissueEnrich.⁶³ The analysis was based on genes from the GTEx database.⁴¹ This included ‘tissue-enriched genes’ with at least 5-fold higher mRNA levels in a particular tissue compared to all the other tissues, ‘group-enriched genes’ with at least 5-fold higher mRNA levels in a group of 2-7 tissues and ‘tissue-enhanced genes’ with at least 5-fold higher mRNA levels in a particular tissue compared to average levels in all tissues.

Functional enrichment analysis

Gene Ontology (GO), Pathway and genetic disorder enrichments were computed using WebGestalt.⁴⁶ P-values indicating significance of enrichment of the Reactome pathways, GO Biological Process and Cellular Component and OMIM diseases were corrected using the Benjamini-Hochberg test for multiple test adjustment. Annotations with FDR-corrected p-value<0.05 were considered significant. ReactomeFIViz, a Cytoscape plugin, was used to extract known functional interactions among genes in the HoP interactome.⁶⁴ It was also used to analyse the connections between the functional modules of viral budding and interferon signalling pathways.

Data Availability¹

The interactome, consists of 332 host proteins, their known interactions and computationally predicted interactions. It is being released via a webserver (<http://hagrid.dbmi.pitt.edu/corona>) as well as in the following electronic data formats.

The list of Proteins and Protein-Protein Interactions is available in **Supplementary Data File S1**. The nodes and edges are labeled to indicate their type (host protein, or known or novel interactor, and known or novel PPI). The membership of the proteins in additional data sources referred to in this work can be found in **Supplementary Data File S2**.

Acknowledgements

MKG acknowledges the contribution of past graduate and undergraduate students (specifically Adam Handen, Naoki Orii and Sai Supreetha Varanasi) who have worked on website and database development and system administration, whose neat work has made it possible to post these results on this website. This work is derived from results from MKG’s project funded by NIMH ([R01MH094564](https://www.ncbi.nlm.nih.gov/pubmed/31111111)).

References

- 1 Rothan, H. A. & Byrareddy, S. N. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun*, 102433, doi:10.1016/j.jaut.2020.102433 (2020).
- 2 Lai, M. M. & Cavanagh, D. The molecular biology of coronaviruses. *Adv Virus Res* **48**, 1-100 (1997).
- 3 Gralinski, L. E. & Baric, R. S. Molecular pathology of emerging coronavirus infections. *J Pathol* **235**, 185-195, doi:10.1002/path.4454 (2015).
- 4 Zhou, Y. *et al.* Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discovery* **6**, 14, doi:10.1038/s41421-020-0153-3 (2020).
- 5 Blanco-Melo, D. *et al.* SARS-CoV-2 launches a unique transcriptional signature from in vitro, ex vivo, and in vivo systems. *bioRxiv*, 2020.2003.2024.004655, doi:10.1101/2020.03.24.004655 (2020).
- 6 Gordon, D. E. *et al.* A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing. *bioRxiv*, 2020.2003.2022.002386, doi:10.1101/2020.03.22.002386 (2020).

¹ Cytoscape and NDEX data will be made available at the time of publication.

- 7 Denisa Bojkova, K. K., Benjamin Koch et al. . SARS-CoV-2 infected host cell proteomics reveal potential therapy targets. *PREPRINT (Version 1) available at Research Square* [[+https://doi.org/10.21203/rs.3.rs-17218/v1](https://doi.org/10.21203/rs.3.rs-17218/v1)] (2020).
- 8 Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature reviews genetics* **12**, 56 (2011).
- 9 Keskin, O., Tuncbag, N. & Gursoy, A. Predicting protein–protein interactions from the molecular to the proteome level. *Chemical reviews* **116**, 4884-4909 (2016).
- 10 Keshava Prasad, T. *et al.* Human protein reference database—2009 update. *Nucleic acids research* **37**, D767-D772 (2008).
- 11 Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic acids research* **34**, D535-D539 (2006).
- 12 Blasche, S. & Koegl, M. Analysis of Protein–Protein Interactions Using LUMIER Assays. *Virus-Host Interactions: Methods and Protocols*, 17-27 (2013).
- 13 Trepte, P. *et al.* DULIP: a dual luminescence-based co-immunoprecipitation assay for interactome mapping in mammalian cells. *Journal of molecular biology* **427**, 3375-3388 (2015).
- 14 Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature*, 1-7 (2020).
- 15 Huttlin, E. L. *et al.* Dual Proteome-scale Networks Reveal Cell-specific Remodeling of the Human Interactome. *bioRxiv* (2020).
- 16 Kotlyar, M. *et al.* In silico prediction of physical protein interactions and characterization of interactome orphans. *Nature methods* **12**, 79-84 (2015).
- 17 Hopf, T. A. *et al.* Sequence co-evolution gives 3D contacts and structures of protein complexes. *Elife* **3**, e03430 (2014).
- 18 Emamjomeh, A. *et al.* Protein-protein interaction prediction by combined analysis of genomic and conservation information. *Genes & genetic systems* **89**, 259-272 (2014).
- 19 Garzón, J. I. *et al.* A computational interactome and functional annotation for the human proteome. *Elife* **5**, e18715 (2016).
- 20 You, Z.-H., Lei, Y.-K., Zhu, L., Xia, J. & Wang, B. Prediction of protein-protein interactions from amino acid sequences with ensemble extreme learning machines and principal component analysis. *BMC bioinformatics* **14**, S10 (2013).
- 21 Jia, J., Liu, Z., Xiao, X., Liu, B. & Chou, K.-C. iPPI-Esml: an ensemble classifier for identifying the interactions of proteins by incorporating their physicochemical properties and wavelet transforms into PseAAC. *Journal of theoretical biology* **377**, 47-56 (2015).
- 22 Li, Y. & Ilie, L. SPRINT: Ultrafast protein-protein interaction prediction of the entire human interactome. *arXiv preprint arXiv:1705.06848* (2017).
- 23 Deng, M., Zhang, K., Mehta, S., Chen, T. & Sun, F. Prediction of protein function using protein–protein interaction data. *Journal of Computational Biology* **10**, 947-960 (2003).
- 24 Raja, K., Subramani, S. & Natarajan, J. PPIInterFinder—a mining tool for extracting causal relations on human proteins from literature. *Database* **2013** (2013).
- 25 Ganapathiraju, M. K. *et al.* Schizophrenia interactome with 504 novel protein–protein interactions. *npj Schizophrenia* **2**, 16012 (2016).
- 26 Zhu, J. *et al.* Antiviral activity of human OASL protein is mediated by enhancing signaling of the RIG-I RNA sensor. *Immunity* **40**, 936-948 (2014).
- 27 Malavia, T. *et al.* Generating testable hypotheses for schizophrenia and rheumatoid arthritis pathogenesis by integrating epidemiological, genomic and protein interaction data *npj Schizophrenia in Press* (2017).
- 28 Li, Y. *et al.* Global genetic analysis in mice unveils central role for cilia in congenital heart disease. *Nature* **521**, 520-524, doi:10.1038/nature14269 (2015).
- 29 Liu, X. *et al.* The complex genetics of hypoplastic left heart syndrome. *Nat Genet* **49**, 1152-1159, doi:10.1038/ng.3870 (2017).
- 30 Karunakaran, K. B., Chaparala, S. & Ganapathiraju, M. K. Potentially repurposable drugs for schizophrenia identified from its interactome. *Scientific Reports* **9**, 12682, doi:10.1038/s41598-019-48307-w (2019).
- 31 Black, S. COVID-19 shuts down research labs as concern among scientists. (2020).

- 32 Ganapathiraju, M. K., Karunakaran, K. B. & Correa-Menendez, J. Predicted protein interactions of IFITMs may shed light on mechanisms of Zika virus-induced microcephaly and host invasion. *F1000Res* **5**, 1919, doi:10.12688/f1000research.9364.2 (2016).
- 33 Karunakaran, K. B., Yanamala, N., Boyce, G. & Ganapathiraju, M. Mesothelioma Interactome with 367 Novel Protein-Protein Interactions. *bioRxiv*, 459065 (2018).
- 34 Liu, J., Li, M., Luo, X.-J. & Su, B. Systems-level analysis of risk genes reveals the modular nature of schizophrenia. *Schizophrenia research* **201**, 261-269 (2018).
- 35 Chellappa, S. A. *et al.* Meta-analysis of genomic variants and gene expression data in schizophrenia suggests the potential need for adjunctive therapeutic interventions for neuropsychiatric disorders. *Journal of genetics* **98**, 60 (2019).
- 36 Cerami, E., Demir, E., Schultz, N., Taylor, B. S. & Sander, C. Automated network analysis identifies core pathways in glioblastoma. *PLoS one* **5** (2010).
- 37 Hoffmann, M. *et al.* SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* (2020).
- 38 Burrell, L. M., Johnston, C. I., Tikellis, C. & Cooper, M. E. ACE2, a new regulator of the renin-angiotensin system. *Trends in Endocrinology & Metabolism* **15**, 166-169 (2004).
- 39 Orii, N. & Ganapathiraju, M. K. Wiki-pi: a web-server of annotated human protein-protein interactions to aid in discovery of protein function. *PLoS One* **7**, e49029, doi:10.1371/journal.pone.0049029 (2012).
- 40 Reghunathan, R. *et al.* Expression profile of immune response genes in patients with severe acute respiratory syndrome. *BMC immunology* **6**, 2 (2005).
- 41 Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nature Genetics* **45**, 580-585, doi:10.1038/ng.2653 (2013).
- 42 Fang, L., Karakioulakis, G. & Roth, M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet Respir Med*, doi:10.1016/s2213-2600(20)30116-8 (2020).
- 43 Scherer, M. & Stamminger, T. Emerging Role of PML Nuclear Bodies in Innate Immune Signaling. *Journal of Virology* **90**, 5850, doi:10.1128/JVI.01979-15 (2016).
- 44 Morita, E. *et al.* Human ESCRT-III and VPS4 proteins are required for centrosome and spindle maintenance. *Proceedings of the National Academy of Sciences* **107**, 12889, doi:10.1073/pnas.1005938107 (2010).
- 45 Fan, Y., Sanyal, S. & Bruzzone, R. Breaking Bad: How Viruses Subvert the Cell Cycle. *Front Cell Infect Microbiol* **8**, 396-396, doi:10.3389/fcimb.2018.00396 (2018).
- 46 Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic acids research* (2019).
- 47 Moore, M. J. *et al.* ZFP36 RNA-binding proteins restrain T cell activation and anti-viral immunity. *Elife* **7**, doi:10.7554/eLife.33057 (2018).
- 48 Blackshear, P. J. Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover. *Biochem Soc Trans* **30**, 945-952, doi:10.1042/bst0300945 (2002).
- 49 Brook, M. *et al.* Posttranslational regulation of tristetraprolin subcellular localization and protein stability by p38 mitogen-activated protein kinase and extracellular signal-regulated kinase pathways. *Mol Cell Biol* **26**, 2408-2418, doi:10.1128/mcb.26.6.2408-2418.2006 (2006).
- 50 Wishart, D. S. *et al.* DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids research* **36**, D901-D906, doi:10.1093/nar/gkm958 (2008).
- 51 Kupersmidt, I. *et al.* Ontology-based meta-analysis of global collections of high-throughput public data. *PLoS one* **5**, doi:10.1371/journal.pone.0013066 (2010).
- 52 Chattopadhyay, A. & Ganapathiraju, M. K. Demonstration Study: A Protocol to Combine Online Tools and Databases for Identifying Potentially Repurposable Drugs. *Data* **2**, 15 (2017).
- 53 Lin, S.-C. *et al.* Effective inhibition of MERS-CoV infection by resveratrol. *BMC infectious diseases* **17**, 144 (2017).
- 54 Sidaway, P. COVID-19 and cancer: what we know so far. *Nature Reviews Clinical Oncology*, 1-1 (2020).
- 55 Sims, A. C. *et al.* Severe acute respiratory syndrome coronavirus infection of human ciliated airway epithelia: role of ciliated cells in viral spread in the conducting airways of the lungs. *Journal of virology* **79**, 15511-15524, doi:10.1128/JVI.79.24.15511-15524.2005 (2005).
- 56 Sungnak, W., Huang, N., Bécavin, C., Berg, M. & Network, H. C. A. SARS-CoV-2 Entry Genes Are Most Highly Expressed in Nasal Goblet and Ciliated Cells within Human Airways. (2020).

- 57 Pincetic, A., Kuang, Z., Seo, E. J. & Leis, J. The Interferon-Induced Gene ISG15 Blocks Retrovirus Release from Cells Late in the Budding Process. *Journal of Virology* **84**, 4725, doi:10.1128/JVI.02478-09 (2010).
- 58 Panda, D. *et al.* Nup98 promotes antiviral gene expression to restrict RNA viral infection in *Drosophila*. *Proc Natl Acad Sci U S A* **111**, E3890-3899, doi:10.1073/pnas.1410087111 (2014).
- 59 Pfefferle, S. *et al.* The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. *PLoS pathogens* **7** (2011).
- 60 Khan, M., Syed, G. H., Kim, S.-J. & Siddiqui, A. Mitochondrial dynamics and viral infections: a close nexus. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1853**, 2822-2833 (2015).
- 61 Wang, Z. & Zhang, J. In search of the biological significance of modular structures in protein networks. *PLoS computational biology* **3** (2007).
- 62 Kupersmidt, I. *et al.* Ontology-based meta-analysis of global collections of high-throughput public data. *PLoS one* **5**, e13066 (2010).
- 63 Jain, A. & Tuteja, G. TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics* **35**, 1966-1967, doi:10.1093/bioinformatics/bty890 (2019).
- 64 Wu, G., Dawson, E., Duong, A., Haw, R. & Stein, L. ReactomeFIViz: a Cytoscape app for pathway and network-based data analysis. *F1000Research* **3** (2014).

Interactome of SARS-CoV-2 / nCoV19 modulated host proteins with computationally predicted PPIs

Kalyani B. Karunakaran¹, N. Balakrishnan¹ and Madhavi K. Ganapathiraju^{2,3*}

¹Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore, 560 012, India

²Department of Biomedical Informatics, School of Medicine, and ³Intelligent Systems Program, School of Computing and Information, University of Pittsburgh, Pittsburgh, USA

Correspondence should be sent to madhavi@pitt.edu

SUPPLEMENTARY MATERIAL

Supplementary Table S1. List of experimentally validated protein-protein interactions predicted by HiPPIP. The method of experimental validation and links to source publications are also shown.

Protein 1	Protein 2	Method of validation	Publications
DDX58	OASL	Co-immunoprecipitation	https://doi.org/10.1016/j.immuni.2014.05.007
HMGB1	FLT1	Co-immunoprecipitation	https://www.nature.com/articles/npjischz201612
HMGB1	KL	Co-immunoprecipitation	https://www.nature.com/articles/npjischz201612
STT3A	RPS25	Co-immunoprecipitation	https://www.nature.com/articles/npjischz201612
STT3A	SYCP3	Co-immunoprecipitation	https://www.nature.com/articles/npjischz201612
STT3A	MCAM	Co-immunoprecipitation	https://www.nature.com/articles/npjischz201612
PDCD1	<hidden>	Co-immunoprecipitation	Unpublished validation by collaborator
ALB	KDR	In-vitro pull down and mass spectrometry	https://www.biorxiv.org/content/10.1101/459065v1 (under review at Scientific Reports)
ALB	PDGFRA	In-vitro pull down and mass spectrometry	https://www.biorxiv.org/content/10.1101/459065v1 (under review at Scientific Reports)
BAP1	PARP3	In-vitro pull down and mass spectrometry	https://www.biorxiv.org/content/10.1101/459065v1 (under review at Scientific Reports)
CLPS	CUTA	In-vitro pull down and mass spectrometry	https://www.biorxiv.org/content/10.1101/459065v1 (under review at Scientific Reports)
HMGB1	CUTA	In-vitro pull down and mass spectrometry	https://www.biorxiv.org/content/10.1101/459065v1 (under review at Scientific Reports)
STX3	LPXN	Co-localization	https://www.nature.com/articles/npjischz201612
STX4	MAPK3	Co-localization	https://www.nature.com/articles/npjischz201612
IFT88	KL	Co-localization	---
WDR5	IGFBP3	Co-localization	---

Supplementary Table S2. List of modules detected exclusively with known protein-protein interactions that are enriched for functions. Gene Ontology Biological Processes enriched in these modules are also mentioned in Column 2. Genes with * code for known interactors whereas all the other genes code for host proteins.

Genes in modules	Enriched Gene Ontology Biological Process
UBAP2L, SMURF1*, NDFIP2, PRKACA, C14orf166*, AKAP8L, UBAP2, FAM98A, STOM, RIPK1, WFS1, ASIC1*, ITPKB*, ITCH*, RHOA, RAP1GDS1, CIT, NIN, GRIPAP1, RRP9	Cell cycle G2/M phase transition (p-value=0.0019, odds ratio=21.7, 20 proteins)
PABPC1, NDRG1*, CSDE1, CSNK2A2, AQP4*, GTF2F2, STC2, GPAA1, CHPF, SIX3*, GSPT2*, TCTEX1D2*, G3BP2, HDAC2, UPF1, ACSL3, AP2M1, FKBP3*, DNMT1, PPP1R8*, TLE3, CSNK2B, G3BP1, PIN1*	Cell-cell signaling by Wnt (p-value=0.0049, odds ratio=9.3, 24 proteins)
POLA2, HELB*, POLA1	DNA replication (p-value=0.0049, odds ratio=55.25, 3 proteins)

Supplementary Figure S1. Network view of novel protein-protein interactions: Network view of the host protein interactome is shown as a graph, where genes are shown as nodes and PPIs as edges connecting the nodes. Host proteins are shown as dark blue square-shaped nodes and novel interactors as red colored circular nodes. Red edges are the novel interactions, whereas blue edges are known interactions.

