Received: 11 December 2020 | Revised: 4 January 2021 | Accepted: 22 January 2021

DOI: 10.1002/jmv.26824

SHORT COMMUNICATION

JOURNAL OF MEDICAL VIROLOGY WILEY

A parsimonious approach for recognizing SARS-CoV-2 and host interactions

Bhaskar Ganguly

Department of Clinical Research, Research and Development Division, Ayurvet Limited, Baddi, Himachal Pradesh, India

Correspondence
Bhaskar Ganguly, Department of Clinical Research, Research and Development Division, Ayurvet Limited, Baddi, HImachal Pradesh – 173205, India.
Email: vetbhaskar@gmail.com

Abstract

Effective countermeasures against the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) demand a better understanding of the pathogen-host interactions. However, such information about the targets, responses, and effects in the host due to the virus is limited, especially so in the case of newly emerged pathogens. The peptide domains that form the interfaces of host and pathogen interacting proteins being evolutionarily conserved, it may be hypothesized that such interactions can be inferred from the similarities in the nucleotide sequences between the host and the pathogen. This communication reports the results of a study based on a parsimonious approach for the identification of the host-virus interactions, where sequence complementarity between the human and SARS-Cov-2 genomes was used to predict several interactions between the host and SARS-CoV-2 at different levels of biological organization. In particular, the findings are suggestive of a direct effect of SARS-CoV-2 on cardiac health. The existing literature on host responses to SARS-CoV-2 and other viruses attest to many of these predicted interactions, supporting the utility of the proposed approach for the identification of host interactions with other novel pathogens.

KEYWORDS
COVID-19, host, interaction, parsimony, pathology, SARS-CoV-2

1 | INTRODUCTION

Following the first report from the city of Wuhan in the Hubei province of China in December 2019, a novel coronavirus-induced disease, coronavirus disease 2019 (COVID-19), has spread rapidly, triggering a global pandemic.[1] COVID-19 is caused by a hitherto unknown beta-coronavirus which has been named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) due to its high sequence similarity with SARS-CoV.[2] Besides the pneumonia-like manifestations, such as cough, fever, and acute respiratory failure, evidence of the attack to multiple organs, such as digestive, cardiovascular, urinary, and reproductive systems have been reported.[3–6] Still, information on most aspects of this virus, including its interactions with the host, is limited.

Knowledge of the complete repertoire of host cell molecules that a pathogen can interact with can be extremely helpful in understanding the pathobiology of the disease but is distinctly lacking for newly emerged pathogens. The peptide domains forming the interfaces of host-pathogen interactions are evolutionarily conserved and parsimonious comparison of host and pathogen genomes can uncover hitherto unknown interaction networks.[7] Beyond the protein-protein interactions, complementary RNA–RNA interactions between the host and the pathogen that may result in altered expression of certain host genes can also be identified,[8] but not distinguished from the former, by parsimonious associations. This communication reports the results of a study based on a parsimonious, sequence complementarity-based approach for the identification of the human and SARS-CoV-2 interactions.

2 | METHODS

All complete genome sequences of SARS-CoV-2 available as of March 25, 2020 at the Virus Pathogen Database and Analysis Resource (VIPR)[9] were added to the workbench and aligned using MUSCLE[10] with Unclust function at default settings. After completion of the multiple sequence alignment (MSA), the consensus sequence was generated using VIPR workbench analysis tools and gaps were removed. BioEdit[11] was used for visualization of the MSA and calculation of its entropy.

The consensus genome sequence of SARS-CoV-2 was used to query the human RefSeq Gene, genomic plus transcript (G + T), and PDB nucleotide databases, respectively, by *blastn* at default settings.[12] Irrespective of their expect values, all nonredundant hits from the *blastn* outputs were selected for further analysis; the hits from the human RefSeq Gene and transcript datasets were used to query STRING database[13] for network reconstruction and enrichment.

3 | RESULTS AND DISCUSSION

It was hypothesized that if a viral gene *abc*, coding for a protein ABC, was similar in sequence to a host gene *xyz*, coding for protein XYZ, then the viral protein ABC could possibly interact with some of the proteins in the host that the protein XYZ interacted with and thereby, compromise the functions and pathways served by XYZ. Otherwise, the sequence similarity between *abc* and *xyz* could also possibly result in a gene silencing event in the host. Therefore, a consensus sequence of SARS-CoV-2 was derived and its local similarity to the human genome was investigated.

As of 25 March 2020, 110 complete genomes of SARS-CoV-2 were available at VIPR. The entropy of the MSA (Supplementary File 1) of these 110 genomes is shown in Figure 1A. Overall low entropy for the MSA implies that the findings based on the consensus sequence can be generalized for all other sequences. The *blastn* search against the human RefSeq Gene database yielded 69 hits, and that against the G + T database yielded 93 hits, respectively. In all, 73 nonredundant hits were identified from the human RefSeq Gene and transcript databases (Supplementary File 2).

Network reconstruction (Figure 1B) and enrichment (Figure 1C,D; Table 1) of the 73 hits, performed with STRING, identified the prominent interactions of the pathogen with the host at the bioprocess, molecular function, and pathway levels. Thirty of the 73 hits were found not to interact with each other. The proteins (ADA2, ADD1, HDAC9, JUP, MRC1, PPP4R2, and SIN3A), their paralogs (GNB1, KCNA4, KRT14, KRT17, MYO18A, NLRP12, PSMA6, SLC25A43, TLN2, and TTBK2), or regulatory subunits thereof (CACNA1A), coded by 18 of these 73 genes have been shown to be expressed differentially during SARS-CoV-2 infection.[14] Intriguingly, sequence complementarity of any segment of the consensus viral genome was not seen with*ace2* or *tmprss2*. A *blastn* query of the RefSeq genes of the family *Coronaviridae* with human *ace2* reference sequence at default settings showed sequence complementarity within *orf6* of SARS-related bat Coronaviruses but neither with any sequence of SARS-CoV-2 nor with any spike protein of the viral family (data not shown). However, at the molecular function level, an interaction between the host and the virus, involving PDZ domain binding (GO:0030165), was inferred; previously, the PDZ-binding motif of the SARS envelope protein has been established as a determinant of viral pathogenesis.[15]

Sequence similarity with dentin sialophosphoprotein (*dspp*) gene as well as the involvement of enamel mineralization (GO:0070166) could be also inferred, hinting at the possibility of developmental defects in dentition and tooth decay in COVID-19 patients. The canine distemper virus, a member of *Paramyxoviridae*, is also known to interfere with enamel mineralization in its host, resulting in poor dentition in animals that suffer from the disease while their adult teeth are forming.[16]

The involvement of two pathways viz. arrhythmogenic right ventricular cardiomyopathy (ARVC) (hsa05412) and glutamatergic synapse (hsa04724) was also identified (Table 1). The enrichment of ARVC at the pathway level and of regulation of heart rate (GO:0002027), regulation of cardiac muscle contraction (GO:0055117), regulation of heart rate by cardiac conduction (GO:0086091), regulation of atrial cardiac muscle cell membrane repolarization (GO:0060372), and cardiac muscle cell action potential (GO:0086001) at the bioprocess level is highly suggestive of a direct cardiomyopathic effect of SARS-CoV-2; myocardial injury associated with in-hospital mortality has already been reported in confirmed and suspected COVID-19 patients.[17] Xiong et al.[14] have also reported the enrichment of ARVC-related transcriptional responses in COVID19 patients. It is important to note that cardiac comorbidities have been a major risk factor for COVID19-related deaths[18] and that the penetrance of ARVC-related mutations is very high in the populations of the Mediterranean basin, especially Italy,[19] where the highest death rates due to this disease have been experienced.

The involvement of glutamatergic synapse (hsa04724) at the pathway level was predicted with much lower confidence than the ARVC pathway. However, some other viruses, including the neurotropic and neuroinvasive human coronaviruses, the HCoV strain OC43, for example, are already known to elicit glutamatergic excitotoxicity. Based on this finding, *N*-methyl-D-aspartate receptor antagonists such as memantine may be tested in the management of COVID-19 patients should they show specific signs of excitotoxicity.[20] Interestingly, our analysis was also able to identify the sequence similarity between viral NSPs 14-15 and host*ntng1*, previously reported by Lehrer and Rheinstein.[21] Given the interactions of host*ntng1* with *cntn5* and *cdh13* (Figure 1B), the competitions arising from this sequence similarity between the virus and *ntng1* may account for the sensory disturbances, such as hyposmia/anosmia and dysgeusia observed in COVID-19 patients. Clustering analysis of SARS-CoV-2, SARS, and MERS virus genes with human genes based on the codon usage and molecular features also associated the human genes with diseases of the nervous and cardiovascular systems.[22]

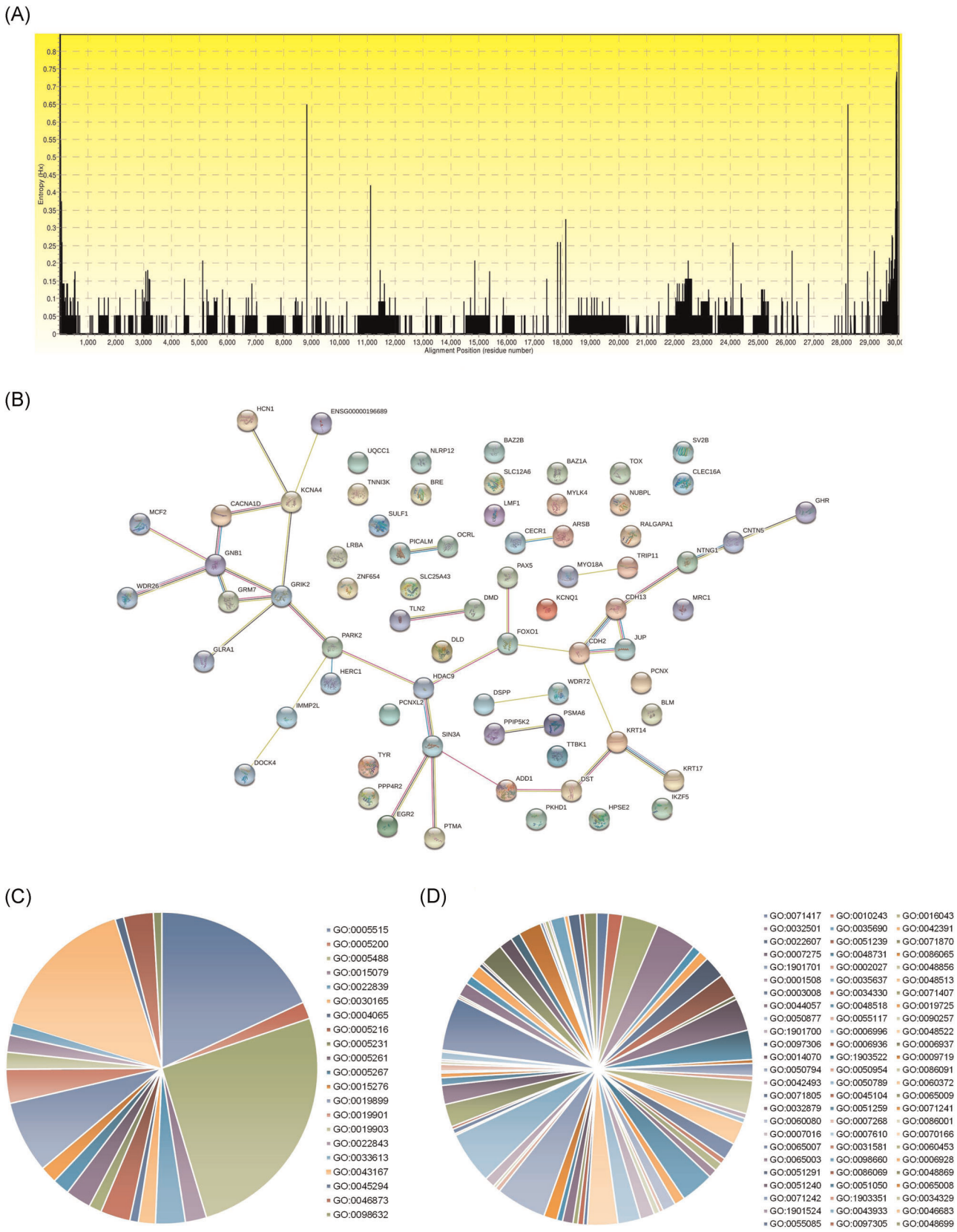

FIGURE 1 (A) Entropy of multiple sequence alignment (MSA) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genomes. One hundred and ten complete genomes of SARS-CoV-2 available at VIPR were selected and aligned with MUSCLE. The MSA was visualized in BioEdit and an entropy (Hx) plot was generated. (B) Interaction network of the host molecules showing parsimonious association with the consensus SARS-CoV-2 genome. STRING was used for network reconstruction and gene enrichment of the 73 nonredundant *blastn* hits. Thirty of the 73 hits did not interact with each other. (C) Gene ontologies (GO) of molecular functions in the host associated with SARS-CoV-2 infection. (D) GO of biological processes in the host associated with SARS-CoV-2 infection. The numbers of genes assigned to a particular molecular function by STRING, that is, observed gene counts have been depicted; for details of GO identifiers, please refer to Table 1. VIPR, Virus Pathogen Database and Analysis Resource

TABLE 1 Gene ontologies and molecular pathways enriched in the host associated with SARS-CoV-2 infection

Term ID	Term description	False discovery rate	Matching proteins in network
			Molecular function
GO:0005515	protein binding	0.02	ADD1, BLM, BRE, CACNA1D, CDH13, CDH2, CECR1, DMD, DOCK4, DST, EGR2, ENSG00000196689, FOXO1, GHR, GLRA1, GNB1, GRIK2, GRM7, HCN1, HDAC9, HERC1, HPSE2, IKZF5, JUP, KCNQ1, KRT17, MCF2, MYO18A, NTNG1, OCRL, PARK2, PICALM, PPP4R2, PSMA6, PTMA, RALGAPA1, SIN3A, SLC12A6, TLN2, TNNI3K, TYR
GO:0005200	structural constituent of cytoskeleton	0.0481	DMD, KRT14, KRT17, TLN2
GO:0005488	binding	0.0481	ADD1, ARSB, BAZ1A, BAZ2B, BLM, BRE, CACNA1D, CDH13, CDH2, CECR1, DLD, DMD, DOCK4, DSPP, DST, EGR2, ENSG00000196689, FOXO1, GHR, GLRA1, GNB1, GRIK2, GRM7, HCN1, HDAC9, HERC1, HPSE2, IKZF5, JUP, KCNA4, KCNQ1, KRT14, KRT17, MCF2, MRC1, MYLK4, MYO18A, NLRP12, NTNG1, NUBPL, OCRL, PARK2, PAX5, PICALM, PPIP5K2, PPP4R2, PSMA6, PTMA, RALGAPA1, SIN3A, SLC12A6, SULF1, TLN2, TNNI3K, TOX, TTBK1, TYR, ZNF654
GO:0015079	potassium ion transmembrane transporter activity	0.0481	GRIK2, HCN1, KCNA4, KCNQ1, SLC12A6
GO:0022839	ion gated channel activity	0.0481	CACNA1D, ENSG00000196689, GLRA1, GRIK2, HCN1, KCNA4, KCNQ1
GO:0030165	PDZ domain binding	0.0481	DOCK4, GRIK2, GRM7, PARK2
GO:0004065	arylsulfatase activity	0.0482	ARSB, SULF1
GO:0005216	ion channel activity	0.0482	CACNA1D, ENSG00000196689, GLRA1, GRIK2, HCN1, KCNA4, KCNQ1
GO:0005231	excitatory extracellular ligand-gated ion channel activity	0.0482	ENSG00000196689, GLRA1, GRIK2
GO:0005261	cation channel activity	0.0482	CACNA1D, ENSG00000196689, GRIK2, HCN1, KCNA4, KCNQ1
GO:0005267	potassium channel activity	0.0482	GRIK2, HCN1, KCNA4, KCNQ1
GO:0015276	ligand-gated ion channel activity	0.0482	ENSG00000196689, GLRA1, GRIK2, HCN1
GO:0019899	enzyme binding	0.0482	CDH2, DMD, DOCK4, EGR2, FOXO1, GHR, GNB1, GRIK2, HDAC9, HERC1, JUP, KCNQ1, MCF2, OCRL, PARK2, PICALM, SLC12A6
GO:0019901	protein kinase binding	0.0482	CDH2, DOCK4, GHR, HDAC9, JUP, KCNQ1, PARK2, SLC12A6
GO:0019903	protein phosphatase binding	0.0482	CDH2, FOXO1, JUP, KCNQ1
GO:0022843	voltage-gated cation channel activity	0.0482	CACNA1D, HCN1, KCNA4, KCNQ1
GO:0033613	activating transcription factor binding	0.0482	EGR2, PTMA, SIN3A
GO:0043167	ion binding	0.0482	ARSB, BAZ1A, BAZ2B, BLM, CACNA1D, CDH13, CDH2, CECR1, DLD, DMD, DSPP, DST, EGR2, ENSG00000196689, GLRA1, GRM7, HCN1, HDAC9, HPSE2, IKZF5, KCNA4, KCNQ1, MYLK4, MYO18A, NLRP12, NUBPL, PARK2, PICALM, PPIP5K2, PSMA6, SULF1, TNNI3K, TTBK1, TYR, ZNF654
GO:0045294	alpha-catenin binding	0.0482	CDH2, JUP
GO:0046873	metal ion transmembrane transporter activity	0.0482	CACNA1D, ENSG00000196689, GRIK2, HCN1, KCNA4, KCNQ1, SLC12A6
GO:0098632	cell-cell adhesion mediator activity	0.0482	JUP, NTNG1
			Biological process
GO:0071417	cellular response to organonitrogen compound	0.00056	BLM, ENSG00000196689, FOXO1, GHR, GLRA1, GNB1, HCN1, HDAC9, JUP, KCNQ1, PARK2, SIN3A
GO:0010243	response to organonitrogen compound	0.00077	BLM, CDH13, EGR2, ENSG00000196689, FOXO1, GHR, GLRA1, GNB1, HCN1, HDAC9, JUP, KCNQ1, PARK2, SIN3A, TYR

(Continues)

TABLE 1 (Continued)

			Biological process
GO:0016043	cellular component organization	0.0011	ADD1, ARSB, BAZ1A, BAZ2B, BLM, BRE, CDH13, CDH2, DMD, DSPP, DST, EGR2, ENSG00000196689, GLRA1, GNB1, HCN1, HDAC9, HERC1, HPSE2, IKZF5, IMMP2L, JUP, KCNA4, KRT14, KRT17, MYO18A, NTNG1, NUBPL, OCRL, PARK2, PICALM, PKHD1, PTMA, SIN3A, SULF1, TLN2, TRIP11, UQCC1
GO:0032501	multicellular organismal process	0.0012	ADD1, ARSB, CACNA1D, CDH13, CDH2, CECR1, CNTN5, DLD, DMD, DOCK4, DSPP, EGR2, ENSG00000196689, FOXO1, GLRA1, GNB1, GRIK2, GRM7, HCN1, HDAC9, HERC1, IMMP2L, JUP, KCNQ1, KRT14, KRT17, LMF1, MCF2, NTNG1, OCRL, PARK2, PAX5, PICALM, PKHD1, SIN3A, SLC12A6, SULF1, TNNI3K, TOX, TRIP11, TTBK1, TYR, WDR72
GO:0035690	cellular response to drug	0.0012	BLM, ENSG00000196689, FOXO1, GLRA1, GNB1, JUP, KCNQ1, PARK2, SIN3A
GO:0042391	regulation of membrane potential	0.0012	CACNA1D, DLD, DMD, ENSG00000196689, GLRA1, GRIK2, HCN1, JUP, KCNQ1, PARK2
GO:0022607	cellular component assembly	0.0016	ADD1, BLM, CDH13, CDH2, DMD, DST, ENSG00000196689, GLRA1, GNB1, HCN1, IKZF5, IMMP2L, JUP, KCNA4, KRT14, NUBPL, OCRL, PARK2, PICALM, PKHD1, TLN2, TRIP11, UQCC1
GO:0051239	regulation of multicellular organismal process	0.0025	ADD1, ARSB, CACNA1D, CDH2, DMD, DOCK4, DSPP, EGR2, ENSG00000196689, FOXO1, GHR, GLRA1, HDAC9, JUP, KCNQ1, KRT17, MCF2, NLRP12, PARK2, PICALM, SIN3A, SULF1, TNNI3K, TOX, TTBK1
GO:0071870	cellular response to catecholamine stimulus	0.0025	GNB1, KCNQ1, PARK2, SIN3A
GO:0007275	multicellular organism development	0.0029	ADD1, ARSB, CDH13, CDH2, CECR1, DLD, DMD, DSPP, EGR2, FOXO1, GNB1, HCN1, HDAC9, HERC1, IMMP2L, JUP, KCNQ1, KRT14, KRT17, MCF2, NTNG1, OCRL, PARK2, PAX5, PICALM, PKHD1, SIN3A, SLC12A6, SULF1, TOX, TRIP11, TTBK1, TYR, WDR72
GO:0048731	system development	0.0032	ADD1, ARSB, CDH13, CDH2, DMD, DSPP, EGR2, FOXO1, GNB1, HCN1, HDAC9, HERC1, IMMP2L, JUP, KCNQ1, KRT14, KRT17, MCF2, NTNG1, PARK2, PAX5, PICALM, PKHD1, SIN3A, SLC12A6, SULF1, TOX, TRIP11, TTBK1, TYR, WDR72
GO:0086065	cell communication involved in cardiac conduction	0.0032	CACNA1D, JUP, KCNQ1, TNNI3K
GO:1901701	cellular response to oxygen-containing compound	0.0032	BLM, ENSG00000196689, FOXO1, GHR, GLRA1, GNB1, HCN1, HDAC9, JUP, KCNQ1, MRC1, PARK2, SIN3A
GO:0002027	regulation of heart rate	0.0038	CACNA1D, DMD, JUP, KCNQ1, TNNI3K
GO:0048856	anatomical structure development	0.0038	ADD1, ARSB, CDH13, CDH2, CECR1, DLD, DMD, DSPP, EGR2, FOXO1, GNB1, HCN1, HDAC9, HERC1, IMMP2L, JUP, KCNQ1, KRT14, KRT17, MCF2, NTNG1, NUBPL, OCRL, PARK2, PAX5, PICALM, PKHD1, SIN3A, SLC12A6, SULF1, TOX, TRIP11, TTBK1, TYR, WDR72
GO:0001508	action potential	0.0052	CACNA1D, DMD, GLRA1, GRIK2, KCNQ1
GO:0035637	multicellular organismal signaling	0.0064	CACNA1D, GRIK2, JUP, KCNQ1, TNNI3K
GO:0048513	animal organ development	0.0077	ADD1, CDH2, DMD, DSPP, EGR2, FOXO1, GNB1, HCN1, HDAC9, HERC1, IMMP2L, JUP, KCNQ1, KRT14, KRT17, PAX5, PICALM, PKHD1, SIN3A, SULF1, TOX, TRIP11, TYR, WDR72
GO:0003008	system process	0.0081	CACNA1D, CNTN5, DMD, DOCK4, EGR2, ENSG00000196689, GLRA1, GNB1, GRIK2, GRM7, HERC1, IMMP2L, KCNQ1, PARK2, PICALM, SULF1, TTBK1, TYR
GO:0034330	cell junction organization	0.0089	CDH13, CDH2, DST, JUP, KRT14, TLN2

TABLE 1 (Continued)

			Biological process
GO:0071407	cellular response to organic cyclic compound	0.009	BLM, ENSG00000196689, FOXO1, GNB1, HCN1, JUP, KCNQ1, PARK2, SIN3A
GO:0044057	regulation of system process	0.0099	CACNA1D, DMD, DOCK4, ENSG00000196689, FOXO1, GLRA1, JUP, KCNQ1, TNNI3K
GO:0048518	positive regulation of biological process	0.0118	ADD1, ARSB, BLM, BRE, CACNA1D, CDH13, CDH2, CLEC16A, DMD, DOCK4, EGR2, ENSG00000196689, FOXO1, GHR, GLRA1, GRIK2, HDAC9, HPSE2, JUP, KCNQ1, KRT17, MCF2, MYO18A, NLRP12, PARK2, PAX5, PICALM, PKHD1, PSMA6, SIN3A, SULF1, TOX, TRIP11, TTBK1, UQCC1
GO:0019725	cellular homeostasis	0.013	ADD1, DLD, DMD, ENSG00000196689, FOXO1, GNB1, GRIK2, PARK2, PICALM, PKHD1, SIN3A
GO:0050877	nervous system process	0.0136	CACNA1D, CNTN5, EGR2, ENSG00000196689, GLRA1, GNB1, GRIK2, GRM7, HERC1, KCNQ1, PARK2, PICALM, TTBK1, TYR
GO:0055117	regulation of cardiac muscle contraction	0.0136	DMD, JUP, KCNQ1, TNNI3K
GO:0090257	regulation of muscle system process	0.0136	DMD, DOCK4, FOXO1, JUP, KCNQ1, TNNI3K
GO:1901700	response to oxygen-containing compound	0.0136	BLM, EGR2, ENSG00000196689, FOXO1, GHR, GLRA1, GNB1, HCN1, HDAC9, JUP, KCNQ1, MRC1, PARK2, SIN3A, TYR
GO:0006996	organelle organization	0.0147	ADD1, ARSB, BAZ1A, BAZ2B, BLM, BRE, DMD, DST, HDAC9, IMMP2L, JUP, KRT14, KRT17, MYO18A, NUBPL, OCRL, PARK2, PICALM, PKHD1, PTMA, SIN3A, TLN2, TRIP11, UQCC1
GO:0048522	positive regulation of cellular process	0.0165	ADD1, ARSB, BLM, BRE, CDH13, CDH2, CLEC16A, DMD, DOCK4, EGR2, ENSG00000196689, FOXO1, GHR, GRIK2, HDAC9, HPSE2, JUP, KCNQ1, KRT17, MCF2, MYO18A, NLRP12, PARK2, PAX5, PICALM, PKHD1, SIN3A, SULF1, TOX, TRIP11, TTBK1, UQCC1
GO:0097306	cellular response to alcohol	0.0175	BLM, GLRA1, GNB1, JUP
GO:0006936	muscle contraction	0.0181	CACNA1D, DMD, ENSG00000196689, GLRA1, KCNQ1, SULF1
GO:0006937	regulation of muscle contraction	0.0181	DMD, DOCK4, JUP, KCNQ1, TNNI3K
GO:0014070	response to organic cyclic compound	0.0192	BLM, ENSG00000196689, FOXO1, GHR, GNB1, HCN1, JUP, KCNQ1, PARK2, SIN3A, TYR
GO:1903522	regulation of blood circulation	0.0192	CACNA1D, DMD, DOCK4, JUP, KCNQ1, TNNI3K
GO:0009719	response to endogenous stimulus	0.0198	BLM, CDH13, EGR2, ENSG00000196689, FOXO1, GHR, GLRA1, GNB1, HCN1, HDAC9, JUP, KCNQ1, PARK2, SIN3A
GO:0050794	regulation of cellular process	0.0198	ADD1, ARSB, BAZ1A, BAZ2B, BLM, BRE, CACNA1D, CDH13, CDH2, CECR1, CLEC16A, DLD, DMD, DOCK4, DSPP, DST, EGR2, ENSG00000196689, FOXO1, GHR, GLRA1, GNB1, GRIK2, GRM7, HDAC9, HERC1, HPSE2, IKZF5, JUP, KCNQ1, KRT17, LMF1, MCF2, MRC1, MYLK4, MYO18A, NLRP12, OCRL, PARK2, PAX5, PICALM, PKHD1, PPP4R2, PSMA6, RALGAPA1, SIN3A, SULF1, TNNI3K, TOX, TRIP11, TTBK1, UQCC1, ZNF654
GO:0050954	sensory perception of mechanical stimulus	0.0198	CACNA1D, CNTN5, ENSG00000196689, GRM7, KCNQ1
GO:0086091	regulation of heart rate by cardiac conduction	0.0207	CACNA1D, JUP, KCNQ1
GO:0042493	response to drug	0.0212	BLM, ENSG00000196689, FOXO1, GHR, GLRA1, GNB1, HDAC9, JUP, KCNQ1, PARK2, SIN3A
GO:0050789	regulation of biological process	0.0212	ADD1, ARSB, BAZ1A, BAZ2B, BLM, BRE, CACNA1D, CDH13, CDH2, CECR1, CLEC16A, DLD, DMD, DOCK4, DSPP, DST, EGR2, ENSG00000196689, FOXO1, GHR, GLRA1, GNB1, GRIK2, GRM7, HCN1, HDAC9, HERC1, HPSE2, IKZF5, JUP, KCNA4,

(Continues)

TABLE 1 (Continued)

			Biological process
			KCNQ1, KRT17, LMF1, MCF2, MRC1, MYLK4, MYO18A, NLRP12, OCRL, PARK2, PAX5, PICALM, PKHD1, PPP4R2, PSMA6, RALGAPA1, SIN3A, SULF1, TNNI3K, TOX, TRIP11, TTBK1, UQCC1, ZNF654
GO:0060372	regulation of atrial cardiac muscle cell membrane repolarization	0.0212	CACNA1D, KCNQ1
GO:0071805	potassium ion transmembrane transport	0.0212	GRIK2, HCN1, KCNA4, KCNQ1, SLC12A6
GO:0045104	intermediate filament cytoskeleton organization	0.0249	DST, KRT14, KRT17
GO:0065009	regulation of molecular function	0.0249	ADD1, BLM, CACNA1D, CECR1, DMD, DOCK4, ENSG00000196689, GHR, GRM7, HERC1, JUP, KCNQ1, LMF1, MCF2, NLRP12, OCRL, PARK2, PICALM, PKHD1, PPP4R2, PSMA6, RALGAPA1, SIN3A, TTBK1
GO:0032879	regulation of localization	0.0254	ARSB, CACNA1D, CDH13, CDH2, DMD, DOCK4, ENSG00000196689, GRM7, HCN1, HDAC9, JUP, KCNA4, KCNQ1, MYO18A, NLRP12, PARK2, PICALM, PKHD1, SIN3A, SULF1
GO:0051259	protein complex oligomerization	0.0254	BLM, ENSG00000196689, GLRA1, GNB1, HCN1, IKZF5, JUP, KCNA4
GO:0071241	cellular response to inorganic substance	0.0265	ADD1, BLM, FOXO1, GLRA1, PARK2
GO:0060080	inhibitory postsynaptic potential	0.0276	GLRA1, GRIK2
GO:0007268	chemical synaptic transmission	0.0282	ENSG00000196689, GLRA1, GRIK2, GRM7, PARK2, SLC12A6, SV2B
GO:0086001	cardiac muscle cell action potential	0.0283	CACNA1D, DMD, KCNQ1
GO:0007016	cytoskeletal anchoring at plasma membrane	0.0296	JUP, TLN2
GO:0007610	behavior	0.0301	EGR2, ENSG00000196689, GLRA1, GRIK2, PARK2, PAX5, PICALM, TTBK1
GO:0070166	enamel mineralization	0.0334	FOXO1, WDR72
GO:0065007	biological regulation	0.0371	ADD1, ARSB, BAZ1A, BAZ2B, BLM, BRE, CACNA1D, CDH13, CDH2, CECR1, CLEC16A, DLD, DMD, DOCK4, DSPP, DST, EGR2, ENSG00000196689, FOXO1, GHR, GLRA1, GNB1, GRIK2, GRM7, HCN1, HDAC9, HERC1, HPSE2, IKZF5, JUP, KCNA4, KCNQ1, KRT17, LMF1, MCF2, MRC1, MYLK4, MYO18A, NLRP12, OCRL, PARK2, PAX5, PICALM, PKHD1, PPP4R2, PSMA6, RALGAPA1, SIN3A, SULF1, TLN2, TNNI3K, TOX, TRIP11, TTBK1, UQCC1, ZNF654
GO:0031581	hemidesmosome assembly	0.0374	DST, KRT14
GO:0060453	regulation of gastric acid secretion	0.0374	ENSG00000196689, KCNQ1
GO:0065003	protein-containing complex assembly	0.0374	BLM, DMD, ENSG00000196689, GLRA1, GNB1, HCN1, IKZF5, IMMP2L, JUP, KCNA4, NUBPL, PARK2, PICALM, UQCC1
GO:0098660	inorganic ion transmembrane transport	0.0374	CACNA1D, ENSG00000196689, GLRA1, GRIK2, HCN1, KCNA4, KCNQ1, PICALM, SLC12A6
GO:0006928	movement of cell or subcellular component	0.0392	ARSB, CACNA1D, CDH13, CDH2, DMD, DOCK4, DST, EGR2, JUP, KCNQ1, MYO18A, NLRP12, TRIP11
GO:0051291	protein heterooligomerization	0.0403	GLRA1, GNB1, IKZF5, JUP
GO:0086069	bundle of His cell to Purkinje myocyte communication	0.0403	JUP, TNNI3K

TABLE 1 (Continued)

			Biological process
GO:0048869	cellular developmental process	0.0413	ADD1, ARSB, CDH2, DLD, DMD, EGR2, FOXO1, HCN1, HDAC9, HERC1, JUP, KRT14, KRT17, MCF2, NTNG1, NUBPL, PARK2, PAX5, PICALM, SIN3A, SULF1, TOX, TRIP11, TTBK1
GO:0051240	positive regulation of multicellular organismal process	0.0424	ARSB, DMD, ENSG00000196689, GHR, HDAC9, JUP, KCNQ1, KRT17, NLRP12, PARK2, PICALM, SULF1, TOX, TTBK1
GO:0051050	positive regulation of transport	0.0443	CACNA1D, CDH2, DMD, ENSG00000196689, JUP, KCNQ1, MYO18A, NLRP12, PARK2, PICALM
GO:0065008	regulation of biological quality	0.0443	ADD1, BLM, CACNA1D, CDH2, DLD, DMD, DOCK4, EGR2, ENSG00000196689, FOXO1, GHR, GLRA1, GNB1, GRIK2, HCN1, JUP, KCNQ1, PARK2, PICALM, PKHD1, SIN3A, TLN2, TNNI3K, TTBK1
GO:0071242	cellular response to ammonium ion	0.0443	GNB1, PARK2, SIN3A
GO:1903351	cellular response to dopamine	0.0443	PARK2, SIN3A
GO:0034329	cell junction assembly	0.0457	DST, JUP, KRT14, TLN2
GO:1901524	regulation of mitophagy	0.0462	CLEC16A, PARK2
GO:0043933	protein-containing complex subunit organization	0.0473	BLM, DMD, ENSG00000196689, GLRA1, GNB1, HCN1, IKZF5, IMMP2L, JUP, KCNA4, NUBPL, PARK2, PICALM, PTMA, UQCC1
GO:0046683	response to organophosphorus	0.0473	ENSG00000196689, HCN1, KCNQ1, TYR
GO:0055085	transmembrane transport	0.0473	ADD1, CACNA1D, ENSG00000196689, GLRA1, GRIK2, HCN1, KCNA4, KCNQ1, PICALM, SLC12A6, SLC25A43, SV2B
GO:0097305	response to alcohol	0.0473	BLM, GHR, GLRA1, GNB1, JUP
GO:0048699	generation of neurons	0.0495	ARSB, CDH2, DMD, EGR2, HCN1, HDAC9, HERC1, MCF2, NTNG1, PARK2, PICALM, TRIP11, TTBK1
			Molecular pathways
hsa05412	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	0.0183	CACNA1D, CDH2, DMD, JUP
hsa04724	Glutamatergic synapse	0.0458	CACNA1D, GNB1, GRIK2, GRM

Abbreviation: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Finally, the *blastn* search against the PDB nucleotide database yielded 6 redundant hits (Supplementary File 2), all showing complementarity over a 16 base-long stretch, in the immediate proximity of an RGD domain-coding sequence and within the receptor binding domain-coding sequence,[23] of the viral spike glycoprotein gene and the human 18S rRNA. The occurrence of an 18S rRNA-complementary sequence juxtaposed with an integrin-binding domain-coding sequence within the receptor-binding domain of SARS-CoV-2 may have strategic implications in the successful arrest of host translational machinery and virulence that is worthy of further investigations.

In conclusion, a computational, sequence complementarity-based parsimonious approach was used to identify different types of host-virus interactions. Existing literature on SARS-CoV-2 and related viruses is in support of many of these predicted interactions whereas further studies are warranted for attesting some of the other predicted interactions. It is believed that such studies will help in the development of new antiviral and disease management strategies.

CONFLICT OF INTERESTS

The author declares that there are no conflict of interests.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Bhaskar Ganguly http://orcid.org/0000-0002-4224-1887

REFERENCES

1. Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med*. 2020: 29. https://doi.org/10.1056/NEJMoa2001316

2. Gralinski LE, Menachery VD. Return of the coronavirus: 2019-nCoV. *Viruses*. 2020;12(2):pii:E135. https://doi.org/10.3390/v12020135
3. Chai X, Hu L, Zhang Y, et al. Specific ACE2 expression in cholangiocytes may cause liver damage after 2019-nCoV infection. *bioRxiv* (2020. https://doi.org/10.1101/2020.02.03.931766
4. Zhang H, Kang Z, Gong H, et al. The digestive system is a potential route of 2019-nCov infection: a bioinformatics analysis based on single-cell transcriptomes. *bioRxiv*. 2020;30:927806. https://doi.org/10.1101/2020.01
5. Zou X, Chen K, Zou J, Han P, Hao J, Han Z. The single cell RNA seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to Wuhan 2019 nCoV infection. *Front Med*. 2020;14:185-192. https://doi.org/10.1007/s11684-020-0754-0
6. Ma L, Xie W, Li D, et al. Effect of SARS-CoV-2 infection upon male gonadal function: a single center-based study. *medRxiv* 2020. https://doi.org/10.1101/2020.03.21.20037267
7. Patro R, Kingsford C. Predicting protein interactions via parsimonious network history inference. *Bioinformatics*. 2013;29:i237-i246.
8. Mumtaz PT, Bhat SA, Ahmad SM, et al. LncRNAs and immunity: watchdogs for host pathogen interactions. *Biol Proced Online*. 2017;19(1):1-2.
9. Pickett B, Greer D, Zhang Y, et al. Virus pathogen database and analysis resource (ViPR): a comprehensive bioinformatics database and analysis resource for the coronavirus research community. *Viruses*. 2012;4(11):3209-3226.
10. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792-1797.
11. Hall T, Biosciences I, Carlsbad C. BioEdit: an important software for molecular biology. *GERF Bull Biosci*. 2011;2(1):60-61.
12. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 2008;36(suppl_2):W5-W9.
13. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43(D1):D447-D452.
14. Xiong Y, Liu Y, Cao L, et al. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *SRN*. 2020:3549993.
15. Jimenez-Guardeño JM, Nieto-Torres JL, DeDiego ML, et al. The PDZ-binding motif of severe acute respiratory syndrome coronavirus envelope protein is a determinant of viral pathogenesis. *PLOS Pathog*. 2014;10(8):e1004320.
16. Dubielzig RR. The effect of canine distemper virus on the ameloblastic layer of the developing tooth. *Vet Pathol*. 1979;16(2):268-270.
17. Zhang F, Yang D, Li J, et al. Myocardial injury is associated with in-hospital mortality of confirmed or suspected COVID-19 in Wuhan China: a single center retrospective cohort study. medRxiv. 2020.
18. Onder G, Rezza G, Brusaferro S Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. JAMA. 2020.
19. Elmaghawry M, Alhashemi M, Zorzi A, Yacoub MH. A global perspective of arrhythmogenic right ventricular cardiomyopathy. *Global Cardiology Science and Practice*. 2013;2012(2):26.
20. Brison E, Jacomy H, Desforges M, Talbot PJ. Novel treatment with neuroprotective and antiviral properties against a neuroinvasive human respiratory virus. *J Virol*. 2014;88(3):1548-1563.
21. Lehrer S, Rheinstein PH. Human gene sequences in SARS-CoV-2 and other viruses. *In Vivo*. 2020;34(3 suppl):1633-1636.
22. Maldonado L, Kamenetzky L Molecular features similarities between SARS-CoV-2, SARS, MERS and key human genes could favour the viral infections and trigger collateral effects. bioRxiv. 2020.
23. Sigrist CJ, Bridge A, Le Mercier P. A potential role for integrins in host cell entry by SARS-CoV-2. *Antiviral Res*. 2020;177:104759.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Ganguly B. A parsimonious approach for recognizing SARS-CoV-2 and host interactions. *J Med Virol*. 2021;93:4576-4584.
https://doi.org/10.1002/jmv.26824