System Biology

COVID-2019 associated overexpressed Prevotella proteins mediated host-pathogen interactions and their role in coronavirus outbreak

Abdul Arif Khan[1,*] and Zakir Khan[2]

[1]Division of Microbiology, Indian Council of Medical Research-National AIDS Research Institute, Pune, Maharashtra 411026, India, [2]Department of Pathology and Laboratory Medicine; Department of Biomedical Sciences, Cedars-Sinai Medical Centre, Los Angeles, CA 90048, USA.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: The outbreak of COVID-2019 initiated at Wuhan, China has become a global threat by rapid transmission and severe fatalities. Recent studies have uncovered whole genome sequence of SARS-CoV-2 (causing COVID-2019). In addition, lung metagenomic studies on infected patients revealed overrepresented Prevotella spp. producing certain proteins in abundance. We performed host-pathogen protein-protein interaction analysis between SARS-CoV-2 and overrepresented Prevotella proteins with human proteome. We also performed functional overrepresentation analysis of interacting proteins to understand their role in COVID-2019 severity.

Results: It was found that over-expressed Prevotella proteins can promote viral infection. As per the results, Prevotella proteins, but not viral proteins are involved in multiple interactions with NF-kB, which is involved in increasing clinical severity of COVID-2019. Prevotella may have role in COVID-2019 outbreak and should be given importance for understanding disease mechanisms and improving treatment outcomes.

Contact: abdularifkhan@gmail.com

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Recently at the end of 2019, cases of pneumonia with unknown cause appeared in Wuhan city, the capital of Central China's Hubei province. The atypical clinical features were suggestive of viral pneumonia and causing worldwide panic due to its severity, rapid transmission with the ability to affect both lungs, and the chances of missed diagnosis (Zhou and Liu, 2020). This severe epidemic initiated on 12th December 2019 and the novel coronavirus (nCoV-2019 later renamed to SARS-CoV-2 and the disease named as COVID-2019) was found to be responsible for these infections. COVID-2019 spread across the world within few months and become public health emergency due to increasing number of cases and mortality in the absence of effective treatment strategy (Cheng and Shan, 2020). Normally, coronavirus caused respiratory and enteric infections, but it gained worldwide attention after the discovery of its involvement in outbreak of severe acute respiratory syndrome (SARS) followed by middle east respiratory syndrome (MERS) (Habibzadeh and Stoneman, 2020). Coronoaviruses are positive stranded enveloped RNA viruses belonging to the family Coronaviridae. Generally, coronoviruses emerged as zoonotic infections, where a viral strain evolved to cause human infections with severe fatality. The previous outbreak of SARS, MERS and recent COVID-2019 is also suggested to be of zoonotic origin. The current COVID-2019 was found

to be closely associated with bat coronavirus with more than 96% whole genome sequence identity (Zhou, et al., 2020), but its spike (S) protein shows 99% sequence similarity with Malaysian pangolin coronavirus. Therefore, it is assumed that it is a combination of bat and pangolin coronavirus (Hassanin, 2020; Zhou, et al., 2020).

The reduction of immune response by coronavirus or some secondary factors lead to increased mortality as evidenced with recent COVID-2019 outbreak, in addition to previous findings (Chen, et al., 2020; Fehr, et al., 2016). The modulation of microflora under the influence of pathogens is a major factor deciding fate of infections. During a recent metagenomic study carried out on COVID-2019 patients (NCBI Bioproject Accession ID: PRJNA603194), it was found that some bacteria are overrepresented in the lungs of infected patients (Chakraborty, 2020). The over expressed proteins in the most abundant Prevotella species were also found. We determined host-pathogen protein-protein interactions (HP-PPI) between SARS-CoV-2 and human through computational methods. We also predicted HP-PPI maps of over expressed Prevotella proteins in COVID-2019 patients using bioinformatics tools. We elucidated a mechanistic insight of these host-pathogen interactions in the severity of COVID-2019.

2 Methods

2.1 Database

SARS-CoV-2 tax ID 2697049 was downloaded from NCBI Taxonomy browser. Only Ref seq source database proteins were considered in order to avoid redundant sequences in analysis. Total 28 sequences were found in NCBI Ref Seq for COVID-2019 at the time of analysis. Human tax ID 9606 was used for host-pathogen PPI prediction. Only reviewed human proteins were taken into account in order to increase validity of predictions. The sequences of Prevotella proteins over expressed in COVID-2019 patients were downloaded from NCBI after getting their IDs from lung metagenomic study (Chakraborty, 2020).

2.2 Prediction of host-pathogen protein-protein interaction (HP-PPI) using homology modeling

The homology modeling was used to predict HP-PPI between SARS-CoV-2 and human. In addition, the HP-PPIs were also determined between over expressed Prevotella proteins in COVID-2019 patients and host. The HP-PPI detection was based on Biologic Interaction and Network Analysis database (BIANA) using interlog prediction (Garcia-Garcia, et al., 2010; Garcia-Garcia, et al., 2012). The curated HP-PPI were predicted against host-pathogen interaction database v 3.0 (Ammari, et al., 2016). The homology modeling of interacting pairs, known as interlog was performed at 30% sequence identity and E value 1e-5. Our approach for PPI findings based on certain homology conditions, for example, A and B is predicted to interact with each other if their homolog A' and B' are already known to interact in interaction databases. These interacting partners are known as interlogs. The redundant PPI entries were filtered out in order to find unique PPI. In case of several PPI with interacting human proteins merged in to a single uniprot ID, the one PPI was used for further analysis. This all filtering was done to remove computational problems with finding same interaction repeatedly. This approach provides high quality data for unique HP-PPI.

2.3 Functional overrepresentation analysis

To understand the role of SARS-CoV-2 and over expressed Prevotella proteins in COVID-2019 outbreak, functional overrepresentation analyses of interacting human proteins was performed. The GOnet was used to perform functional overrepresentation analysis based on gene ontology (Pomaznoy, et al., 2018). Role of human interactors in SARS-CoV-2-human and over expressed Prevotella protein-human PPI network was predicted with ontology version 2019-07-01 using q value threshold * ($\leq$ 0.05). The relevant gene ontologies-biological processes were selected for understanding their involvement in COVID-2019 severity. The HP-PPI maps were constructed using cytoscape v 3.7.2 (Smoot, et al., 2011)

3 Results

The details of strains overrepresented in COVID-2019 patients are available on NCBI Bioproject Accession ID: PRJNA603194 and therefore not presented here in this manuscript. Different species of Prevotella represented highest 23% share of lung metagenome after 61% unidentified spots (dark matter). Moreover, the overexpressed proteins in Prevotella strains associated with COVID-2019 patients identified in above study is presented in Table 1 in order to give overview about key proteins (Chakraborty, 2020).

Table 1. The overexpressed proteins of Prevotella strains found in COVID-2019 patients. The list is arranged as per their relative expression reads. The EF-Tu was highly expressed followed by ClpB as per the study (PRJNA603194).

Name of protein	NCBI ID	Abbreviation in network
elongationfactor Tu	WP_009012371.1	EF-Tu
ATP-dependent Clp protease ATP-binding subunit ClpB	SNR97511.1	ClpB
elongation factor G	WP_009012398.1	EF-G
molecular chaperone DnaK	SNR93756.1	DnaK
transcription termination/antitermination factor NusG	WP_009012372.1	NusG
hypothetical protein SAMN06265364_1044	SNR67224.1	
phosphoenolpyruvate carboxykinase (ATP)	SNS04478.1	PEPCK
glyceraldehyde 3-phosphate dehydrogenase	SNR80995.1	G3PDH
cysteine synthase A	SNR91480.1	csA
chaperonin GroEL	SNR94701.1	GroEL
pyruvate-ferredoxin/ flavodoxin oxidoreductase	SNR97358.1	PFOR
energy transducer TonB	WP_009012245.1	TonB
DNA-directed RNA polymerase subunit alpha	WP_004361631.1	RNAP alpha
DNA-directed RNA polymerase subunit beta'	SNR93143.1	RNAP beta
Pyruvate phosphate dikinase, PEP/pyruvate binding domain	SNR97473.1	PPDK
Outer membrane protein OmpA	SNR93835.1	OmpA
Peroxiredoxin	WP_089365394.1	
ATP-dependent Clp protease ATP-binding subunit ClpC	SNS00213.1	ClpC
preprotein translocase subunit SecA	SNR60000.1	SecA

hypothetical protein SAMN06265364	SNR84878.1	
translation initiation factor IF-3	WP_089366830.1	IF-3
Biopolymer transport protein ExbD/TolR	SNR68043.1	ExbD/TolR 1
Biopolymer transport protein ExbD/TolR	SNR68055.1	ExbD/TolR 1
elongation factor Ts	WP_009010992.1	EF-Ts

During our initial HP-PPI detection, total 24258 HP-PPI were found in over-expressed Prevotella proteins in COVID-2019 patients. These 24258 PPI included 24168 PPI detected against Biological Interactions and Network Analysis (BIANA) database, and 90 curated PPI detected against Host-Pathogen Interaction Database (HPI-DB). However, most of these interactions were redundant and filtered during detection of unique HP-PPI between over expressed Prevotella proteins and human leaving only 6775 unique HP-PPI. The SARS-CoV-2 was predicted to perform total 125 HP-PPI with human host. Among these, 90 redundant HP-PPI were present leaving only 35 unique PPI. The detail about HP-PPI and interacting partners is presented in Table 2. The exhaustive list of HP-PPI between SARS-CoV-2-human and over expressed Prevotella proteins-human is presented as Table S1 and S2 respectively. Among these HP-PPI between SARS-CoV-2 and human, many of them are already proved in literature giving strong support for our prediction method.

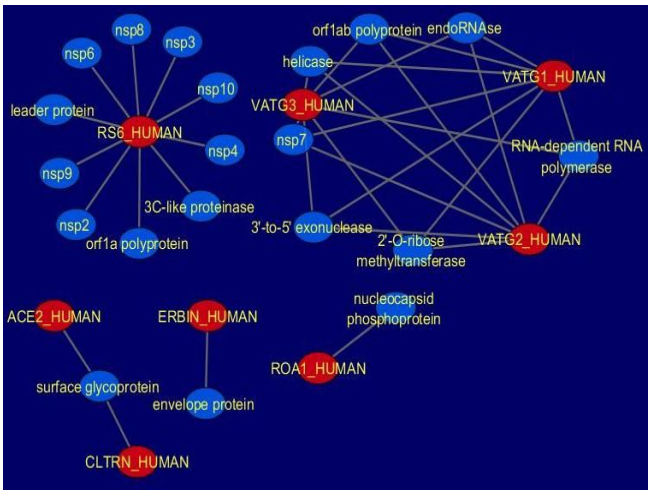
Table 2 The details of HP-PPI and interacting partners between SARS-CoV-2 and over expressed Prevotella proteins with human.

HP-PPI network	Total number of Predicted PPI	Total number of unique PPI	Number of participating pathogen proteins	Number of participating host protiens
COVID2019-HUMAN	126	35	20	8
Overexpressed Prevotella proteins-Human	24258	6775	18	3431

3.1 HP-PPI network

HP-PPI network between SARS-CoV-2 and human is presented in Fig. 1, while Fig. 2 represents HP-PPI network between over expressed Prevotella proteins with human proteome. Multiple interactions of over expressed proteins were found with NF-kB indicating its possible role in COVID-2019 severity.

Fig. 1: HP-PPI network between SARS-CoV-2 and human proteome. Viral proteins are indicated with light blue color while human protein with red color.



3.2 Identification of SARS-CoV-2 and over expressed Prevotella proteins role in COVID-2019

Functional overrepresentation analysis of human proteins found in COVID-2019 HP-PPI network revealed that the viral proteins are majorly interacting with human proteins involved in viral attachment to host cell, regulation of pH and phagosome related processes (Table S3). In contrast, the over expressed Prevotella proteins were found to affect several human proteins involved in viral growth, life cycle and gene expression. These proteins involved in different processes are shown in Fig. 2, however the exhaustive list is presented as Table S4.While all human proteins found in over expressed Prevotella proteins HP-PPI network were scanned for their presence in proteins involved in SARS (as per querying SARS in uniprot), only one protein (ECHA_HUMAN) out of 3431 proteins was found.

4 Discussion

However, the COVID-2019 patients were not found to have increased procalcitonin level (Huang, et al., 2020), as found with bacterial infections, but the data obtained from lung metagenomic study of COVID-2019 patients revealed that Prevotella is overrepresented in these patients. In addition, these Prevotella spp. are showing over expression of certain proteins. This paradoxical behavior can be justified by the facts that procalcitonin is produced in response to invasion by bacterial, fungal and parasitic pathogens (Lee, 2013), but it has poor diagnostic value during detection of certain bacterial pneumonia (El-Solh, et al., 2011). Therefore, it is necessary to understand alternative mechanism for direct or indirect involvement of Prevotella in COVID-2019 disease severity.

It has been found that SARS-CoV-2 uses angiotensin-converting enzyme 2 (ACE2) receptor to get entry into human cells (Wan, et al., 2020). ACE2 receptor is expressed on the surface of airway epithelium and parenchyma of lung, which serves as a receptor for SARS-Coronavirus (CoV). Our study efficiently predicted this interaction of ACE2 with surface glycoprotein of coronavirus. In addition, the SARS-CoV nucleocapsid protein is known to interact with heterogeneous nuclear ribonucleoprotein A1 (ROA1_HUMAN) with high binding affinity (Luo, et al., 2005). We have also predicted this interaction in our HP-PPI network. Moreover, V-ATPase G1 (VATG1_HUMAN) subunit of human is already known to interact with SARS-CoV and it is considered as an important interaction for coronavirus pathogenesis (Lin, et al., 2005). Our study predicted multiple protein interactions of SARS-CoV with VATG1, VATG2, and VATG3 of human. All these findings validate our prediction method and support other predicted interactions not yet known experimentally. In addition, it is already known that use of more than 30% sequence similarity can detect interlogs efficiently and this approach has been used in several other studies (Espadaler, et al., 2005). Considering detection of several experimentally known interactions with the use of our methods, it gives strong support for other predicted interaction not yet experimentally known.

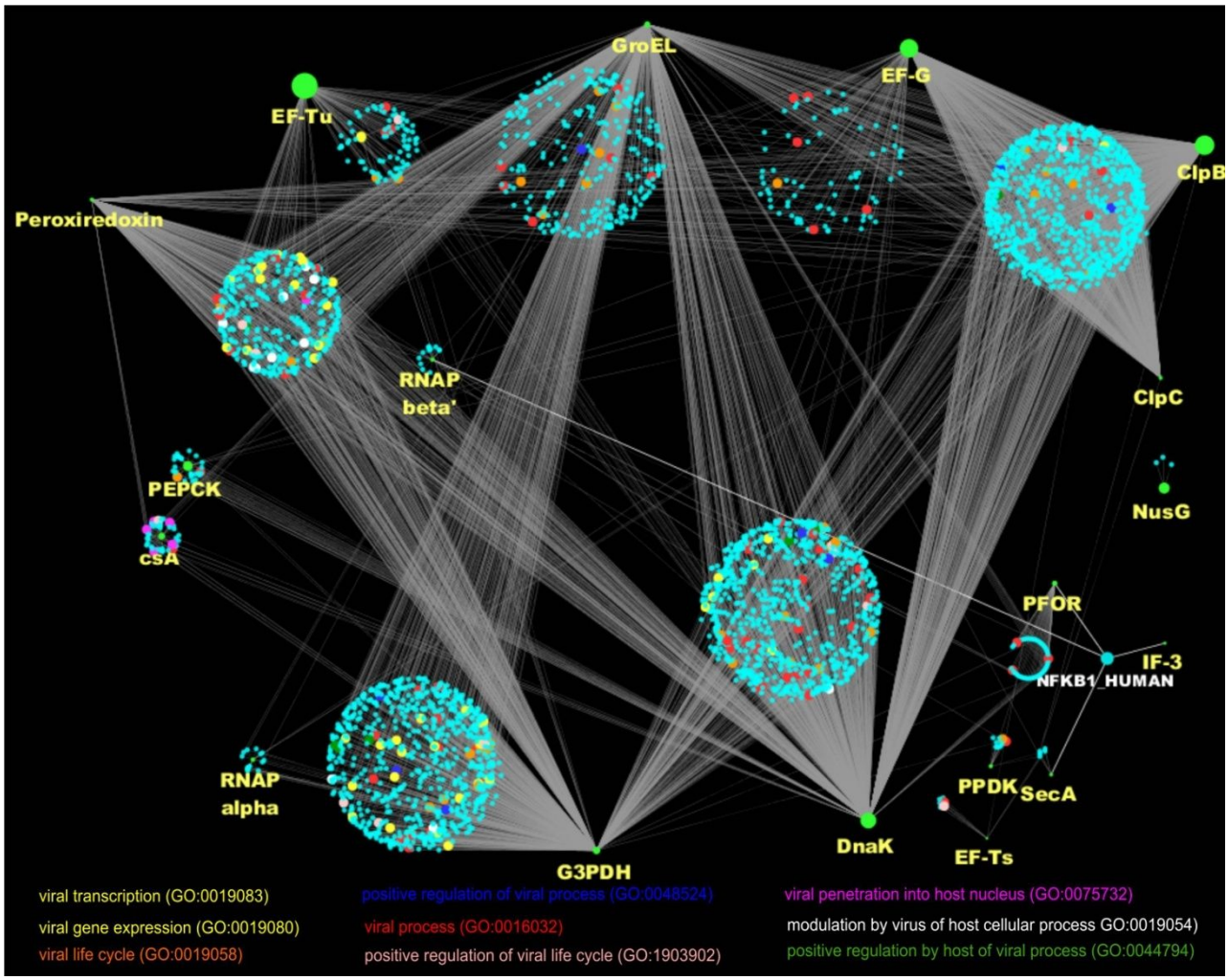


Fig. 2: HP-PPI between over expressed Prevotella proteins in COVID-2019 patients. Human proteins are normally indicated in deep sky blue colors while Prevotella proteins are represented by fluorescent green color (Size of the nodes represent their relative expression raw reads). The human proteins involved in supporting viral growth are indicated by their respective color. However, the same proteins were repeated in multiple categories, therefore this figure is indicative of Prevotella involvement in COVID-2019. The NF-kB was found to perform multiple interactions and therefore showed in large size node. Please refer Table S2 for detail information about interactors and their function.

NF-kB plays a major role in CoV-induced SARS. It was found that inhibition of NF-kB mediated inflammation can lead to increased survival during SARS infection (DeDiego, et al., 2014). Our prediction also indicates multiple interactions of Prevotella over expressed proteins with NF-kB (Fig. 2), and in contrast, SARS-CoV2 is not showing any direct interaction with NF-kB. In addition, Prevotella over expressed proteins are supporting viral growth at multiple levels. These findings predict that Prevotella can play several significant roles in the severity of current outbreak ranging from supporting SARS-CoV-2 growth to directly interacting with certain mediators of disease progression.

The synergic role of Prevotella in inducing severity of *Streptococcus pneumonia* mediated pneumonia is already known. During this study, researcher proposed that *Prevotella intermedia* (a periodontopathic bacteria) can lead to its pathogenic effects due to saliva aspiration in respiratory tract and proposed that presence of this bacteria in lower respiratory tract can be a risk factor for severe pneumococcal pneumonia (Nagaoka, et al., 2014).

As per the results of our study, 18 over expressed Prevotella proteins were predicted to interact with 3431 human proteins, while 20 SARS-CoV-2 proteins were predicted to interact with only 8 human proteins with multiple interactions (Table 2). We must consider that HP-PPI detection is based on identification of known homologous PPI in several interaction databases. For example BIANA uses detection of known interactions from DIP, MINT, HPRD, IntAct, PHI_base, Mpact, BIND, PIG, VirusMINT and BioGRID interaction databases and is known to provide 72-98% sensitivity and up to 59% specificity under different homology conditions (Garcia-Garcia, et al., 2012). In addition, HPIDB includes 69,787 unique curated PPI derived from 66 host and 668 pathogens (Ammari, et al., 2016). As the over-expressed Prevotella proteins are obtained from several Prevotella species, while the SARS-CoV-2 proteins are non redundant and specific, the known homologous interactions must be comparatively less for viral proteins. This may be one of the reasons for relatively large HP-PPI prediction in Prevotella-human network. In addition, the information about viral structure is still growing and significant updates are arising everyday through global scientific efforts. More studies about viral structure can add more valuable inputs about its host-pathogen interactions. Nevertheless, the

current study holds its value as it focuses on contribution of overexpressed Prevotella proteins on COVID-2019. The inclusion of BIANA and HPIDB leads to detection of homologous PPI in several interaction databases in comparison to simple homology detection. The use of two complementary tools provided us several common HP-PPI and we filtered those common interactions in order to find unique HP-PPI for further analysis (Table 2). These all the steps can be assumed to provide additional reliability on our prediction data.

However, the exact role of Prevotella in COVID-2019 mediated pneumonia severity needs to be explored, but our findings indicate that proper attention should be given for management of such bacteria in order to increase treatment outcomes in COVID-2019 outbreak. The management of Prevotella in addition to COVID-2019 can give promising results for management of this outbreak. Although, the results of this study provides a large overview about involvement of microflora modulation in COVID-2019 but computational methods have their own limitations and the resulted data must be correlated with experimental findings, but under current panic situation these findings can provide valuable inputs for devising strategies to increase treatment outcomes of COVID-2019.

It can be concluded from our study that Prevotella can play an important role in COVID-2019 outbreak and should be assessed fairly for its involvement in COVID-2019 pathogenesis. Due to synergistic role of Prevotella in increasing severity of pneumonia, a proper attention must be given to these bacteria for controlling COVID-2019 outbreak.

Author contributions

AAK searched literature, designed and performed the experiment, collected and analyzed data, wrote the manuscript, edited and approved the final version; ZK designed the experiment, analyzed data, wrote the manuscript, edited and approved the final version

Funding

None

Availability of data and materials

All the data related to manuscript is available with journal as supplementary materials.

Conflict of Interest: none declared.

References

Ammari, M.G., et al. (2016) HPIDB 2.0: a curated database for host-pathogen interactions, Database : the journal of biological databases and curation, 2016.baw103

Chakraborty, S. (2020) The 2019 Wuhan outbreak could be caused by the bacteriaPrevotella, which is aided by the coronavirus, possibly to adhere to epithelial cells –Prevotella is present in huge amounts in patients from both China and Hong Kong.DOI: 10.31219/osf.io/usztn

Chen, N., et al. (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, Lancet. 395, P507-513

Cheng, Z.J. and Shan, J. (2020) 2019 Novel coronavirus: where we are and what we know, Infection, 48, 155-163.

DeDiego, M.L., et al. (2014) Inhibition of NF-kappaB-mediated inflammation in severe acute respiratory syndrome coronavirus-infected mice increases survival, J Virol, 88, 913-924.

El-Solh, A.A., et al. (2011) Diagnostic use of serum procalcitonin levels in pulmonary aspiration syndromes, Crit Care Med, 39, 1251-1256.

Espadaler, J., et al. (2005) Prediction of protein-protein interactions using distant conservation of sequence patterns and structure relationships, Bioinformatics, 21, 3360-3368.

Fehr, A.R., et al. (2016) The Conserved Coronavirus Macrodomain Promotes Virulence and Suppresses the Innate Immune Response during Severe Acute Respiratory Syndrome Coronavirus Infection, mBio, 7.

Garcia-Garcia, J., et al. (2010) Biana: a software framework for compiling biological interactions and analyzing networks, BMC Bioinfo., 11, 56.

Garcia-Garcia, J., et al. (2012) BIPS: BIANA Interolog Prediction Server. A tool for protein-protein interaction inference, Nucleic acids research, 40, W147-151.

Habibzadeh, P. and Stoneman, E.K. (2020) The Novel Coronavirus: A Bird's Eye View, Int J Occup Environ Med, 11, 65-71.

Hassanin, A. (2020) Coronavirus origins: genome analysis suggests two viruses may have combined. The Conversation.

Huang, C., et al. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, Lancet, 395, 497-506.

Lee, H. (2013) Procalcitonin as a biomarker of infectious diseases, Korean J Intern Med, 28, 285-291.

Lin, C.W., et al. (2005) Binding interaction of SARS coronavirus 3CL(pro) protease with vacuolar-H+ ATPase G1 subunit, FEBS Lett., 579, 6089-6094.

Luo, H., et al. (2005) The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1, FEBS Lett., 579, 2623-2628.

Nagaoka, K., et al. (2014) Prevotella intermedia induces severe bacteremic pneumococcal pneumonia in mice with upregulated platelet-activating factor receptor expression, Infect Immun, 82, 587-593.

Pomaznoy, M., Ha, B. and Peters, B. (2018) GOnet: a tool for interactive Gene Ontology analysis, BMC Bioinfo., 19, 470.

Smoot, M.E., et al. (2011) Cytoscape 2.8: new features for data integration and network visualization, Bioinformatics, 27, 431-432.

Wan, Y., et al. (2020) Receptor recognition by novel coronavirus from Wuhan: 3 An analysis based on decade-long structural studies of SARS, J. Virol., DOI: 10.1128/JVI.00127-20

Zhou, L. and Liu, H.G. (2020) Early detection and disease assessment of patients with novel coronavirus pneumonia, Zhonghua Jie He He Hu Xi Za Zhi, 43, E003

Zhou, P., et al. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature, 579, 270-273.