Mechanistic insights into SARS-CoV-2 epidemic via revealing the features of SARS-CoV-2 coding proteins and host responses upon its infection

Shijie Qin[1#], Xinyi Xia[2,3,4#], Xuejia Shi[1], Xinglai Ji[5*], Fei Ma[1*], Liming Chen[1*]

[1]Department of Biochemistry, School of Life Sciences, Nanjing Normal University, Nanjing, China 210023;

[2]COVID-19 Research Center, Institute of Laboratory Medicine, Jinling Hospital, Nanjing University School of Medicine, Southern Medical University, Nanjing, Jiangsu 210002, China

[3]Department of Laboratory Medicine & Blood Transfusion, Wuhan Huoshenshan Hospital, Wuhan, Hubei 430100, China

[4]Joint Expert Group for COVID-19, Wuhan Huoshenshan Hospital, Wuhan, Hubei 430100, China

[5]Smart Health Big Data Analysis and Location Services Engineering Lab of Jiangsu Province, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

[#]These authors contributed equally.

*To whom correspondence may be addressed. Email: chenliming1981@njnu.edu.cn; rich@njupt.edu.cn; 08258@njnu.edu.cn.

Abstract

There are seven known coronaviruses that infect humans: four mild coronaviruses, including HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1, only cause mild respiratory diseases, and three severe coronaviruses, including SARS-CoV, MERS-CoV and SARS-CoV-2, can cause severe respiratory diseases even death of infected patients. Both infection and death caused by SARS-CoV-2 are still rapidly increasing worldwide. In this study, we demonstrate that viral coding proteins of SARS-CoV-2 have distinct features and are most, medium and least conserved with SARS-CoV, MERS-CoV, and the rest four mild coronaviruses (HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1), respectively. Moreover, expression of host responsive genes (HRG), HRG-enriched biological processes, and HRG-enriched KEGG pathways upon infection of SARS-CoV-2 show slightly overlapping with SARS-CoV and MERS-CoV but distinctive to the four mild coronaviruses. Interestingly, enrichment of overactivation of neutrophil by HRGs is only and commonly found in infections of severe SARS-CoV-2, SARS-CoV, and MERS-CoV but not in the other four mild coronaviruses, and the related gene networks show different patterns. Clinical data supports that overactivation of neutrophil for severe patients can be one major factor for the similar clinical symptoms observed in SARS-CoV-2 infection compared to infections of the other two severe coronavirus (SARS-CoV, and MERS-CoV). Taken together, our study provides a mechanistic insight into SARS-CoV-2 epidemic via revealing the conserved and distinct features of SARS-CoV-2, raising the critical role of dysregulation of neutrophil for SARS-CoV-2 infection.

Keyword: Coronavirus; COVID-19; SARS-CoV-2; host response; neutrophil

Contact: example@example.org

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

1   Introduction

Coronaviruses are mantle-coated positive-stranded single-stranded RNA viruses that are broad-spectrum found in humans and other mammals (Wang, et al., 2020). There are seven known coronaviruses that infect humans (Wang, et al., 2020). In this study, we classified these seven coronaviruses into two groups: mild coronavirus group (MCG) four coronaviruses (HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1) and severe coronavirus group (SCG) consists of remaining three coronaviruses (SARS-CoV, MERS-CoV and SARS-CoV-2). Patients with infection of MCG and SCG viruses will develop mild and severe respiratory diseases, respectively. Infection of SCG viruses have significantly higher mortality due to multiple organ failure in the infected patients (Corman, et al., 2018; Pillaiyar, et al., 2020). WHO data show that the mortality rates for SARS-CoV and MERS-CoV infections are 10% and 30%, respectively (Singh, 2016; Wang, et al., 2020). So far, the new coronavirus (SARS-CoV-2) has caused more than 2,500,000 infections and more than 180,000 deaths. The case numbers of coronavirus disease 2019 (COVID-19) caused by SARS-CoV-2 are still rapidly increasing worldwide.

It is well documented that the immune system plays an important role in the body's response to viral infections. However, excessive immune responses can cause pathological immune responses and cause tissue damage (Garcia-Sastre and Biron, 2006). The interaction of viral genes with the host immune system, especially the innate immune system, is a determinant of coronavirus toxicity and patient prognosis (Frieman, et al., 2008; Shi, et al., 2014). Compared with MCG, SCG shows much stronger ability in promoting cytokine storms, resulting in severe respiratory diseases, multiple organ failure and even death in patients (Garcia-Sastre and Biron, 2006; Huang, et al., 2020). In addition, there is still a lack of research on the response of human coronavirus (whether MCG or SCG) to the host of system biology. In particular, it is necessary for revealing the pathology and pandemic of COVID-19 to systematically construct a human signal transduction network and identify master regulators (MRs) in response to SARS-CoV-2 infection.

Here, we reveal the features of SARS-CoV-2 in viral genome-coded proteins, expression of host responsive genes (HRG), signal transduction network, master regulator and clinical data to give a mechanical insight into the SARS-CoV-2 epidemic.

2   Methods and Materials

2.1 Data sources

The genome sequences, gene sequences and protein sequences of seven coronaviruses were retrieved from NCBI (https://www.ncbi.nlm.nih.gov/). Human gene expression data with virus infection were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) (Table S1). Specifically, the transcriptome data after SARS-CoV-2 infected lung epithelial cells and lung alveolar (A549) cells was numbered GSE147507. The transcriptome data after SARS-CoV infected lung-derived cells was numbered GSE56192. The sera (PBMC) transcriptome data of patients with SARS-CoV infection and healthy persons is GSE1739. The transcriptome data after MERS-CoV infection of lung-derived cells is GSE56189. The gene expression chip number of HCoV-229E after infecting liver-derived cells is GSE89166. The gene expression chip data of HCoV-OC43 after infecting neuron-derived cells is GSE13879. The transcriptome data of 430 nasopharyngeal swabs positive for SARS-CoV-2 and 54 negative

controls came from GSE152075. The blood neutrophils and clinical data of 2979 COVID-19 patients are available in Table S7, which were derived from COVID-19 Research Center, Institute of Laboratory Medicine, Jinling Hospital, Nanjing University School of Medicine, Southern Medical University, Nanjing, Jiangsu 210002, China and Department of Laboratory Medicine & Blood Transfusion, Wuhan Huoshenshan Hospital, Wuhan, Hubei 430100, China.

2.2 Data analysis

The transcriptome differentially expressed genes after virus infection were calculated using the edgeR R package (Robinson, et al., 2010). The differentially expressed genes of the chip expression were calculated by the GEO2R included in the GEO database. Filtering criteria for differentially expressed genes are a fold change of 1.5 and a p-value of less than 0.01. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) of differentially expressed genes were enriched by the clusterProfiler package (Yu, et al., 2012). Protein interaction network analysis is derived from the STRING database (https://string-db.org/). The protein-coding gene similarities among the seven viruses were compared and calculated by BLAST software. Based on the ARACne mutual information algorithm, 484 human nasopharyngeal swab gene expression profiles were used to assemble an interactive set of hosts in response to SARS-CoV-2 infection (Margolin, et al.). ARACne was ran with 100 bootstraps, p value threshold of $10-8$ and no DPI tolerance and consolidate with bonferroni correction. Master Regulators were detected using the algorithm called VIPER (Virtual Inference of Protein-activity by Enriched Regulon analysis), implemented in R and Bioconductor (Alvarez, et al.). Statistical significance was estimated by permuting the samples uniformly at random 1,000 times and 100 bootstrap interactions was performed to reduce the effect of outlier samples on the gene expression signature for discriminating the enriched MRs. The list of all human transcription factors is downloaded from AnimalTFDB 3.0 (Hu, et al.).

2.3 Detecting patients with clinical SARS-CoV-2 infection and testing patients' blood cells

All COVID-19 patients are implemented in accordance with the New Coronavirus Pneumonia Diagnosis and Treatment Plan (6th edition) issued by the National Health Commission of China (http://www.nhc.gov.cn/yzygj/s7653p/202002/8334a8326dd94d329df35 1d7da8aefc2.shtml). All patients were diagnosed in the laboratory using the real-time PCR method of SARS-CoV-2 throat swab. Immediately after obtaining each throat swab, it was placed in the delivery tube. All tested samples were handled under airborne prevention. The SARS-CoV-2 nucleic acid was detected by reverse transcription and real-time PCR assays using commercial detection kits (Changsha Sansure Biotech). using two independent primers, respectively with open reading frame 1ab (ORF1ab) and nuclear Capsid protein (N) fragments are matched. RNase-P is used as an internal standard gene to monitor the sample collection and extraction process by detecting whether it is normal or not, to avoid false negative results. Reverse transcription and real-time PCR were performed according to the manufacturer's recommendations. Each transcript provides a cycle threshold (Ct value), which is the number of cycles required for the fluorescent signal. A Ct value less than 40 is positive, and a Ct value greater than 40 is negative. All patients' blood cell ratios and counts were tested with a Mindray automatic blood cell analyzer (BC-5390CRP). The instrument can detect lymphocytes, monocytes, and eosinophils from basophils and neutrophils by laser scattering and flow

cytometry in the DIFF laser channel. During operation, only 2ml EDTA-K2 anticoagulated venous blood needs to be put into the sample holder for testing. This study was reviewed and approved by the Medical Ethics Committee of Wuhan Huoshenshan Hospital and is in compliance with the regulations issued by the China National Health Commission and the Declaration of Helsinki. Each participant's informed consent was obtained.

2.4 Data Statistics and Visualization

All data analysis was done on R software (version 3.5.2). Cytoscape software (version 3.7.0) is used to visualize the network and analyze the topological properties of the network (Shannon, et al., 2003).

3   Results and discussion

3.1 Features of viral coding proteins of SARS-CoV-2

We obtained and analyzed the viral coding proteins of seven coronaviruses (MCG: HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1; SCG: SARS-CoV, MERS-CoV and SARS-CoV-2) (Table 1). Overall, viral coding proteins of SARS-CoV-2 show greater conservation in viruses of SCG compared to those of MCG (Table 1). The nucleocapsid phosphoprotein was strikingly conserved in SCG and different between SCG and MCG, indicating that the nucleocapsid phosphoprotein will contribute to the different pathogenic responses between SCG infection and MCG infection. Among SCG coronaviruses, we found the highest similarity in protein sequences between SARS-CoV-2 and SARS-CoV. The ORF8 protein of SARS-CoV-2 shows low similarities to all other coronavirus-encoded proteins (<25%), suggesting ORF8 protein is likely to be unique to SARS-CoV-2. Surface glycoproteins are well conserved among MCG and SCG, implicating its important role for the common susceptibility in the population.

Table 1. Similarity of the protein sequences obtained by comparing SARS-CoV-2 with other six infected human coronavirus proteins sequence and genomic sequence.

SARS-CoV-2	SARS-CoV	MERS-CoV	HCoV-229E	HCoV-NL63	HCoV-OC43	HCoV-HKU1
ORF6	93.44%	22.95%	27.87%	18.03%	29.51%	18.03%
Membrane protein	98.20%	62.61%	66.05%	63.16%	70.84%	63.16%
Envelope protein	97.37%	58.67%	48.05%	29.33%	54.86%	29.33%
ORF3a	90.55%	11.21%	6.18%	8.00%	5.45%	8.00%
ORF10	92.11%	44.74%	28.95%	42.11%	42.10%	42.11%
ORF8	20.94%	15.70%	11.57%	13.22%	19.01%	13.22%
Surface glycoprotein	87.35%	39.51%	28.66%	24.04%	39.26%	24.04%
ORF7a	95.90%	14.05%	9.92%	14.05%	15.70%	14.05%
Nucleocapsid phosphoprotein	97.16%	68.49%	37.14%	27.98%	53.26%	27.98%

3.2 Expression features of host responsive genes (HRG) for SARS-CoV-2

Then, we analyzed and compared the host responsive genes' expression upon infection of SARS-CoV-2 and other coronaviruses. For SCG, we found 566, 766 and 3,671 differentially expressed genes upon SARS-CoV-2, SARS-CoV and MERS-CoV infection, respectively, and these genes are regarded to be host responsive genes (HRG) for SCG (Table S2-4). For MCG, we found 549 and 3,210 differentially expressed genes upon infection of HCoV-229E and HCoV-OC43, respectively, and these genes are regarded to be HRGs for SCG (Table S5-6). HRGs for SARS-CoV-2 contains 6 genes (CXCL2,C1S,NFKBIA,INHBA,BIRC3,PDE5A) common to those in SARS-CoV and MERS-CoV, compared to 4 genes to those found in SCG viruses (Fig 2). No HRG was found between SCG and MCG viruses. These results suggest that SARS-CoV-2 infection has distinct HSG expression patterns.
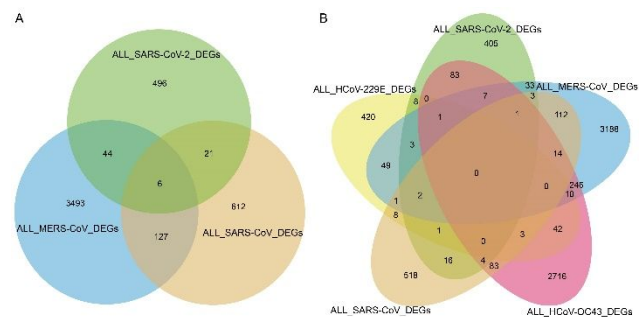

Fig 1. Comparison of differentially expressed genes after five coronaviruses infect human. A: Intersection of SARS-CoV-2, SARS-CoV and MERS-CoV; B: Intersection of SARS-CoV-2, SARS-CoV, MERS-CoV, H229E-CoV and HCoV-OC43.

3.3 Biological process annotation and enrichment of HRGs for SARS-CoV-2 compared to other coronaviruses

When we performed functional enrichment analysis on HRGs of SARS-CoV-2 and other coronaviruses, we found that the enrichment of HRGs of SARS-CoV-2 biological processes are distinctive to those of other coronaviruses. Interestingly, enrichment of neutrophil dysregulation and activation was only found to be enriched in HRGs of SARS-CoV-2 and other two SCG viruses (SARS-CoV and MERS-CoV) but not in those of MCG viruses (Fig 2). Moreover, the 6 genes shared by 3 severe viruses are mainly involved in multiple immune-related processes such as NF−kappaB, Toll like, and inflammatory response (Fig S1). Among them, CXCL2 was found to be directly involved in neutrophil imbalance (Fig 2). This finding strongly suggests that the immune disorders with main neutrophil overactivation is may be one of the main causes for similar severe illness symptoms upon infection of SARS-CoV-2 to those upon infection of other two SCG viruses (SARS-CoV and MERS-CoV).

3.4 KEGG pathway annotation and enrichment of HRGs for SARS-CoV-2 compared to other coronaviruses

We further investigated the signaling pathways involved in HRGs of SARS-CoV-2 and compared to those of other coronaviruses. The results show that HRGs of SARS-CoV-2 are enriched in distinctive pathways compared to both other SCG viruses and MCG viruses (Fig 3). HRGs of SARS-CoV-2 show specially enriched in IL-17 signaling, Rheumatoid arthritis, Cytokine−cytokine receptor interaction and NF−kappa B pathways and so on. Most of the above signal pathways involve the disturbance of immune signals and the release and recruitment of

inflammatory factors. It is well known that their disorders are the cause of various inflammatory diseases such as rheumatoid arthritis, systemic lupus erythematosus, acute respiratory pneumonia and the like. Rheumatoid arthritis and systemic lupus erythematosus are well-documented to be caused by excessive activation of the immune system via releasing amounts of cytokines to attack normal tissues (Arriens, et al.; Smolen, et al.). Several specific HRGs such as CXCL1, TLR2,etc shared by SARS-CoV-2 and SARS-CoV, and IL6 by SARS-CoV-2 and MERS-CoV involved in rheumatoid arthritis pathway (Fig S2). Furthermore, C1S response to SARS-CoV-2 and HIST4H4, HISTH2BD, HIST2H2BE shared by SARS-CoV and MERS-CoV, are involved in severe hyperimmune activation (Fig S3). However, rare commonalities can be found between SARS-CoV-2 and the other MCG viruses. These results suggest that HRG-enriched pathways in SARS-CoV-2 infection have slightly common and no common features to other SCG viruses and MCG
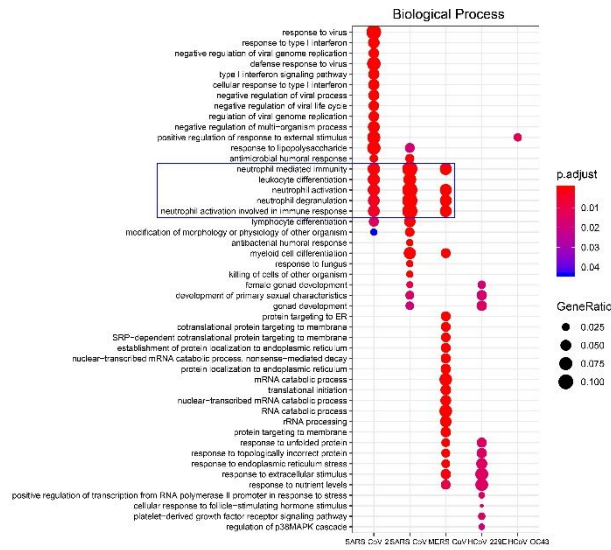

viruses repetitively,implicating the mechanism behind the similarities and differences of clinical symptoms upon infection of SARS-CoV-2, the SCG viruses and the MCG viruses.

Fig 2. Gene ontology functional annotation of differentially expressed genes that specifically respond to coronavirus SARS-CoV-2/SARS-CoV/MERS-CoV.

Fig 3. Signal pathway enrichment analysis of differentially expressed genes that specifically respond to coronavirus SARS-CoV-2/SARS-CoV/MERS-CoV.

3.5 Analysis of protein networks of HRGs of SARS-CoV-2 compared to other coronaviruses
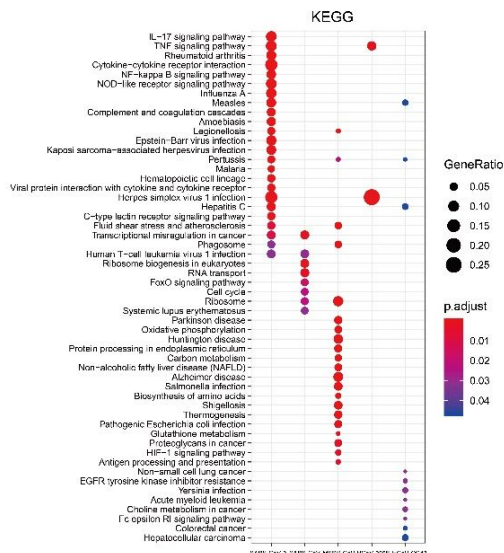
To further investigate the protein-protein interactions among specific HRGs may causing severe symptom, we performed a protein network analysis on the proteins coded by HRGs in SARS-CoV-2 infections and compared to those in the other coronaviruses' infection. The protein network formed by HRGs for SARS-CoV-2 are distinctive to other coronaviruses, including SCG viruses. We looked into the specific HRGs related to neutrophil disorder, which are only commonly founded between SARS-CoV-2 and other two SCG viruses but not between SARS-CoV-2 and other MCG viruses. SARS-CoV-2 and other two SCG viruses show common in neutrophil disorder involving following common biological processes: a wide range of common protein-protein interactions response to neutrophil activation, neutrophil activation involved in immune response, neutrophil degranulation and neutrophil mediated immunity (Fig 4). We found that the above public biological processes mainly involve OLR1, TLR2, S100A12, CXCL1, VNN1, TCN1, MMP9, LCN2 proteins of SARS-2-CoV/SARS-CoV, and CSTB, IL6 proteins of SARS-CoV-2/MERS, and S100A11, GNS, GDI2, QSOX1, CCT2, TIMP2, SLC2A3, CKAP4 proteins of MERS/SARS, suggesting the different molecular mechanism behind the common biological processes (Fig 4). It's worth to be noticed that compared with two other severe viruses, SARS-CoV-2 has a unique neutrophil migration and neutrophil chemotactic process which included 27 special proteins such as CXCL2, CXCL5, CXCL6, IL1A, IL1B, S100A9, CCL5 and SAA1 and so on (Fig 4). Taken together, these findings provide an insight into the molecular mechanism behind current SARS-CoV-2 epidemic.

Fig 4. Diagram of protein interaction network for HRGs that specifically respond to neutrophil disorder. SARS-CoV-2/SARS-CoV/MERS-CoV. Blue diamonds represent biological processes and circles represent proteins. The size of the circle represents the
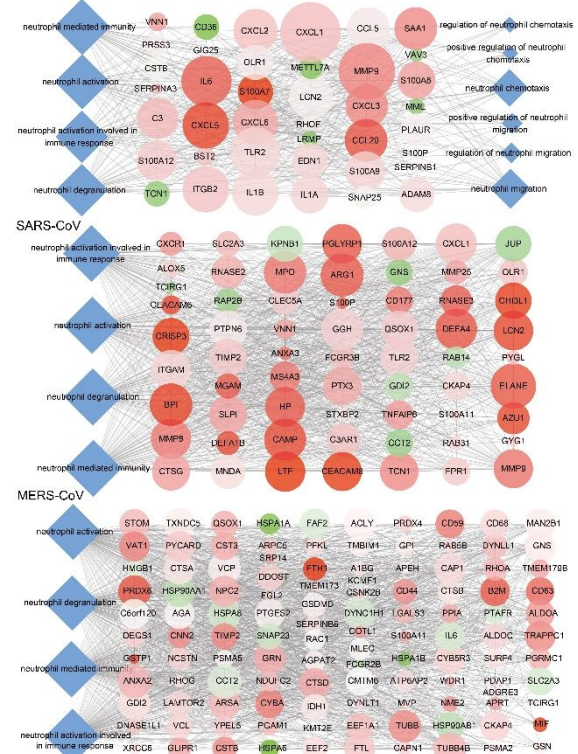
degree of protein centrality, that is, the degree of connection with other proteins. Red represents up-regulated genes and green represents the down-regulated.

3.6 Verifying the excessive activation of neutrophils after SARS-CoV-2 infection

In order to further determine whether the neutrophils really increased significantly after severe coronavirus infection, we collected blood neutrophil gauge test data of 2976 patients who have been diagnosed with SARS-CoV-2 at Wuhan Huoshenshan Hospital, Wuhan, Hubei 430100, China (Table S7). The results are shown in Fig 5, compared with patients with mild symptoms, severe and critical patients had significantly higher the percentage and the absolute of blood neutrophils (Fig 5A~5B). The same increase was observed in patients in the death group vs patients in the surviving group (Fig 5C~5D). Moreover, critical patients are also significantly higher than severe patients (Fig 5A~5B) and with the abnormal increase of neutrophils, the proportion of critical and dead patients also increases (Fig 5E) indicating that the deterioration of the disease was related to the increase in the abundance and proportion of neutrophils.

In addition, we also analyzed the dynamic changes of neutrophils throughout the hospitalization period by using the patient's multiple test results. As shown in Figs 5F~5G, the percentage of neutrophils and the absolute value of neutrophils in patients with critical illness and death are always higher than those of non-critically ill patients and surviving patients, which indicates that the continued excessive activation of neutrophils is a key accomplice to SARS-CoV-2 leading to severe illness and death.
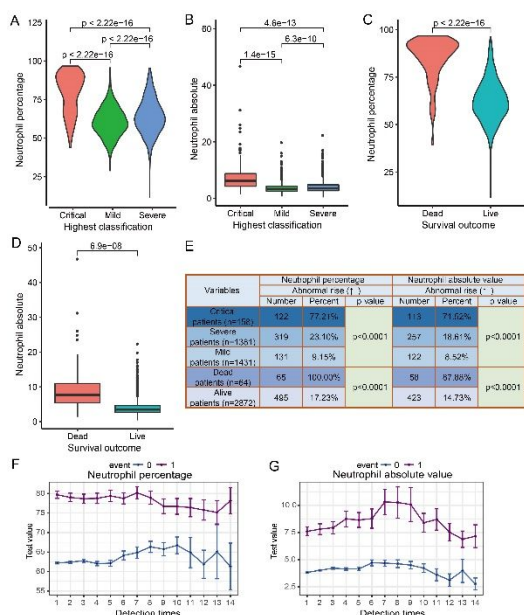

Fig 5. Verify the correlation between neutrophils and clinical severe diseases caused by SARS-CoV-2. A: The violin diagram shows neutrophil percentage in different classification. B: The box chart shows neutrophil absolute value in different classification. C: The violin diagram shows neutrophil percentage in different survival outcome. D: The box chart shows neutrophil absolute value in different survival outcome. E: Statistics on the number and proportion of neutrophil abnormalities in different classification and survival outcome. The first test value of admission is used. The reference ranges of normal neutrophil percentage and absolute value are 40 ~ 75 (40% ~ 75%) and 1.8 ~ 6.3 (*109 / L), respectively. F: The dynamic change curve of neutrophil percentage during hospitalization. G: The dynamic curve of neutrophil absolute value during hospitalization. Each point represents the average of different detection times increasing with length of hospitalization, and the standard error was displayed. The two-tailed t-test was used to calculate the significance of continuous variables, and the chi-square test was used to analyze the significance of frequency. Event 1 represents the patients who developed critical illness or died or admitted to the ICU while event 0 represents patients with better symptoms and prognosis.

3.7 Identify key transcriptional regulators of host response to SARS-CoV-2 infection

In order to further explore the molecular mechanisms behind the host response to SARS-CoV-2 infection, we intended to identify potential key transcriptional regulators in the host response to SARS-CoV-2 infection. ARACNe is an mutual information theory-based method that can infer the mechanistic interaction between transcription factors (TF) and target genes based on a large amount of gene expression data (Basso, et al., 2005; Margolin, et al., 2006). The interacting genomes from ARACNe has proved to be very effective for further identification the master regulators in the host responses to virus infection (Basso, et al., 2010; Carro, et al., 2010). To further explore the key transcriptional regulators in the host in response to SARS-CoV-2 infection, we analyzed the gene expression profile of 484 human nasopharyngeal swabs using ARACNe and VIPER. ARACNe reveals a complex interactome, where 1168 TFs regulates 13262 target genes through 147600 transcriptional interactions involved in the host responses to SARS-CoV-2 infection. Then, we further analyzed the interactome using VIPER via prioritizing regulatory proteins by evaluating the degree of enrichment of direct targets in gene expression signatures of SARS-CoV-2 positive patients and negative controls. We defined the top 30 proteins enriched for SARS-CoV-2 infection from VIPER analysis as master regulators (MR) in the host responses to SARS-CoV-2 infection. In these 30 MR proteins, the main activity-enhancing MRs are TEF, THAP7, ELF3, ESRRA, YBX3, MTA1, SLC2A4RG, RFXANK, E4F1, FEZF1, CENPX, IRX3, THAP3, ID1, ZF524, JUN and RXRB, and the main activity-reducing MRs are ZNF506, ZNF562, ZNF766, ZNF121, SP1, ZNF107, ETV6, STAT1, ZNFX1, TFEC, ZNF587, REL and NR2C2 (Fig 6A). GO functional enrichment analysis on the 30 MRs shows that the 30 MRs are mainly involved in "rhythmic process," "positive regulation by host of viral transcription", "DNA-templated transcription, initiation", "modulation by host of viral transcription" (Fig S4), suggesting the identified MRs play an important role in the host's response to SARS-CoV-2. Furthermore, KEGG analysis reveals that the 30 MRs are enriched in "Herpes simplex virus 1 infection", "Human papillomavirus infection", "Kaposi sarcoma-associated herpesvirus infection", "Epstein-Barr virus infection", "Viral carcinogenesis", "Human T-cell Leukemia virus 1 infection", "Influenza A", "Human immunodeficiency virus 1 infection" and other virus-related pathogenic processes (Fig 6B and Table S8).

In addition, these MRs are also enriched in many immune and inflammation-related pathways, such as "Th17 cell differentiation", "Inflammatory bowel disease (IBD)", "Toll-like receptor signaling pathway", "TNF signaling pathway", "Rheumatoid arthritis" ((Fig 6B and Table S8). In particular, although the first-ranked TEF has not been reported to be related to COVID-19, it has been reported to be essential for the transcriptional activation of human papillomavirus HPV-16 (Ishiji, et al.). In vivo, TEF (TEF-1) binding is necessary for HPV-16 P97 promoter activity (Ishiji, et al.). In addition, the activation of the original TEF is necessary for the activation of the pseudorabies virus glycoprotein X gene promoter (Ou, et al.). There are also reports that TEF is involved in the simian virus transcriptional replication, when the binding site of TEF and viral enhancer is destroyed, the transcription of viral genes will be weakened (Berger, et al.). These studies suggest that TEF is likely to play an important role in the replication and proliferation of SARS-CoV-2 in humans. Interestingly, STAT1 protein activity shows to be decreased upon SARS-CoV-2 infection. STAT1 was reported to be a key anti-viral transcription factor (Kostanian Ia Fau - Vonarshenko, et al.; Raftery and Stevenson), indicating inhibition of STAT1 activity will contribute to severe clinical symptoms caused by SARS-CoV-2 infection.
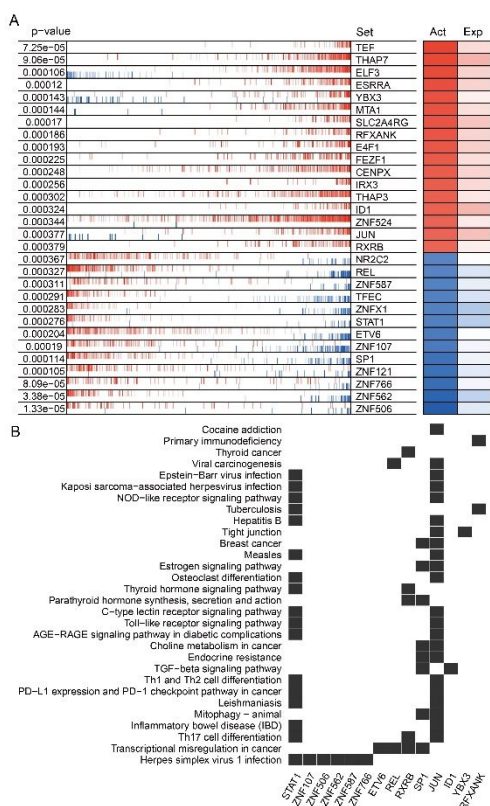
Fig 6. A: VIPER plot showing the projection of the negative (repressed, shown in blue color) and positive (activated, shown in red color) targets for top 30 MRs, as inferred by ARACNe and correlation analysis when reverse engineering the regulatory network (vertical lines resembling a bar-code), on the GES (x-axis), where the genes in the GES were rank-sorted from the one most down-regulated to the one most upregulated in the 'SARS-CoV-2 positive' vs 'SARS-CoV-2 negative' conditions. The two-columns heatmap displayed on the right side of the Fig shows the inferred differential activity (first column) and differential expression (second column), with the rank of the displayed genes in the GES (shown all the way to the right). B: The KEGG signaling pathway analysis of the top 30 MRs. The first 30 significantly enriched KEGG items are displayed and the detailed contents refer to Table S8.

4 Conclusions

Our findings strongly suggest that SARS-CoV-2 codes distinctly viral coding proteins to promote distinctive host response gene expression pattern, leading to current SARS-CoV-2 epidemic. Amount efforts are still required to reveal the mechanism behind SARS-CoV-2 epidemic. Nevertheless, our study provides a pioneer mechanistic insight into SARS-CoV-2 epidemic via revealing the features of SARS-CoV-2 coding proteins and host responses upon its infection, with highlighted the critical role of dysregulation of neutrophils in SARS-CoV-2 epidemic. Finally, using ARACNe and VIPER, we revealed the signal transduction network and key transcription factors in the host response to SARS-CoV-2 infections. Our current study provides important mechanistic implications for the pathogenesis of SARS-CoV-2, and these implications would be valuable for development of new strategies against SARS-CoV-2.

Acknowledgements

Funding

References

Alvarez, M.J., et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. (1546-1718 (Electronic)).

Arriens, C., et al. Systemic lupus erythematosus biomarkers: the challenging quest. (1462-0332 (Electronic)).

Basso, K., et al. Reverse engineering of regulatory networks in human B cells. Nature genetics 2005;37(4):382-390.

Basso, K., et al. Integrated biochemical and computational approach identifies BCL6 direct target genes controlling multiple pathways in normal germinal center B cells. Blood 2010;115(5):975-984.

Berger, L.C., et al. Interaction between T antigen and TEA domain of the factor TEF-1 derepresses simian virus 40 late promoter in vitro: identification of T-antigen domains important for transcription control. (0022-538X (Print)).

Carro, M.S., et al. The transcriptional network for mesenchymal transformation of brain tumours. Nature 2010;463(7279):318-325.

Corman, V.M., et al. Hosts and Sources of Endemic Human Coronaviruses. Advances in virus research 2018;100:163-188.

Frieman, M., Heise, M. and Baric, R. SARS coronavirus and innate immunity. Virus research 2008;133(1):101-112.

Garcia-Sastre, A. and Biron, C.A. Type 1 interferons and the virus-host relationship: a lesson in detente. Science 2006;312(5775):879-882.

Hu, H., et al. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. (1362-4962 (Electronic)).

Huang, C., et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395(10223):497-506.

Ishiji, T., et al. Transcriptional enhancer factor (TEF)-1 and its cell-specific co-activator activate human papillomavirus-16 E6 and E7 oncogene transcription in keratinocytes and cervical carcinoma cells. (0261-4189 (Print)).

Kostanian Ia Fau - Vonarshenko, A.V., Vonarshenko Av Fau - Lipkin, V.M. and Lipkin, V.M. [STAT1: a many-sided transcription factor]. (0132-3423 (Print)).

Margolin, A.A., et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC bioinformatics 2006;7 Suppl 1:S7.

Margolin, A.A., et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. (1471-2105 (Electronic)).

Ou, C.J., Wong Ml Fau - Chang, T.-J. and Chang, T.J. A TEF-1-element is required for activation of the promoter of pseudorabies virus glycoprotein X gene by IE180. (0920-8569 (Print)).

Pillaiyar, T., Meenakshisundaram, S. and Manickam, M. Recent discovery and development of inhibitors targeting coronaviruses. Drug discovery today 2020.

Raftery, N. and Stevenson, N.A.-O. Advances in anti-viral immune defence: revealing the importance of the IFN JAK/STAT pathway. (1420-9071 (Electronic)).

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26(1):139-140.

Shannon, P., et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research 2003;13(11):2498-2504.

Shi, C.S., et al. SARS-coronavirus open reading frame-9b suppresses innate immunity by targeting mitochondria and the MAVS/TRAF3/TRAF6 signalosome. Journal of immunology 2014;193(6):3080-3089.

Singh, S.K. Middle East Respiratory Syndrome Virus Pathogenesis. Seminars in respiratory and critical care medicine 2016;37(4):572-577.

Smolen, J.S., et al. Rheumatoid arthritis. (2056-676X (Electronic)).

Wang, H.J., et al. Review and Prospect of Pathological Features of Corona Virus Disease. Fa yi xue za zhi 2020;36(1):16-20.

Yu, G., et al. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics : a journal of integrative biology 2012;16(5):284-287.