LETTER TO THE EDITOR

Inferring SARS-CoV-2 functional genomics from viral transcriptome with identification of potential antiviral drugs and therapeutic targets

Xu Pan†, Xin Li1,2†, Shangwei Ning* and Hui Zhi*†

Abstract

Coronavirus disease 2019 (COVID-19) is an emerging infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and has posed a serious threat to global health. Here, we systematically characterized the transcription levels of the SARS-CoV-2 genes and identified the responsive human genes associated with virus infection. We inferred the possible biological functions of each viral gene and depicted the functional landscape based on guilt-by-association and functional enrichment analyses. Subsequently, the transcription factor regulatory network, protein–protein interaction network, and non-coding RNA regulatory network were constructed to discover more potential antiviral targets. In addition, several potential drugs for COVID-19 treatment and prevention were recognized, including known cell proliferation-related, immune-related, and antiviral drugs, in which proteasome inhibitors (bortezomib, carfilzomib, and ixazomib citrate) may play an important role in the treatment of COVID-19. These results provided novel insights into the understanding of SARS-CoV-2 functional genomics and host-targeting antiviral strategies for SARS-CoV-2 infection.

Dear Editor,

The rapid spread of coronavirus disease 2019 (COVID-19) caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has resulted in worldwide concern. High-throughput sequencing studies have identified SARS-CoV-2 as a positive-sense RNA virus with a genome size of 29,903 nucleotides. Similar to other coronaviruses, a set of subgenomic mRNAs are produced during the SARS-CoV-2 replication, including the structural protein mRNAs (S, M, E, and N) and other small accessory protein mRNAs (ORF1ab, ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10). These subgenomic mRNAs have common eukaryotic mRNA-like structures, such as the 5′ leader sequence and 3′ poly-A tail structure [1]. However, to date, the function of viral genes in COVID-19 infection remains unclear.

To systematically investigate the viral transcriptome, we first obtained RNA-seq datasets of SARS-CoV-2-infected cells from GSE147507. The dataset included three independent biological replicates of primary human lung epithelium NHBE cells, human lung adenocarcinoma Calu3 and A549 cells, and A549 cells transduced with a vector expressing human ACE2. All cells were mock-treated or infected with SARS-CoV-2 (USA-WA1/2020) at an MOI (multiplicity of infection) of 2 for 24 h, except for the transduced A549 cells, which were treated at an MOI of 0.2. We next quantified human and viral gene expression abundance, respectively. We aligned the raw

© The Author(s) 2021. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.
reads to the GENCODE GRCh38 human genome assembly using Bowtie2 and normalized the counts of human genes to TPM (Transcripts Per Million) using RSEM. Next, the unaligned reads were assigned to the SARS-CoV-2 reference genome (NCBI Reference Sequence: NC_045512.2). Viral genes were quantified using the viGEN pipeline (Additional file 1). Compared with other viral genes, the nucleocapsid (N) gene expression level was the highest and ORF7b was the lowest in all SARS-CoV-2-infected cells (Fig. 1A). Moreover, we identified differentially expressed human genes in five SARS-CoV-2-infected and mock-infected control experiments using the paired Student’s t test. Genes with $P$ value < 0.05 in more than three experiments were recognized as differentially expressed genes ($n = 185$) (Fig. 1B; Additional file 2: Table S1).

To better understand the host transcriptional response to SARS-CoV-2, we interrogated the viral gene-related human genes based on the linear regression model, which was constructed between the transcription levels of differentially expressed human gene $i$ ($H$) and viral gene $j$ ($V$) and corrected for the effect of cell types and MOI as co-variables, i.e., $H_i \sim \alpha V_j + \beta + \varepsilon$ (Additional file 1). Human genes with $P < 0.05$ were identified as the responsive genes (SR genes) associated with SARS-CoV-2 infection at the transcriptional level ($n = 180$; Fig. 1C; Additional file 2: Table S1). To validate the association between SR genes and SARS-CoV-2, we tested the correlation between the SR gene transcription levels and SARS-CoV-2 concentration detected by RT-PCR in a COVID-19 cohort of 93 patients (GSE156063) (Fig. 1C). The results showed that the transcription levels of 123 SR genes in total varied with the concentration of SARS-CoV-2 ($P < 0.05$, Spearman’s correlation test). Furthermore, we calculated the coupled correlation coefficient among COVID-19 patients and healthy controls based on the transcription levels of SR genes and performed a hierarchical clustering analysis, which

---

**Fig. 1** Overview of SARS-CoV-2 genes and human SR genes in SARS-CoV-2 infected human cells. **A** Viral gene expression levels in SARS-CoV-2-infected human cells. Error bars show mean ± SE of results of replicates. **B** Differentially expressed human genes responding to virus infection. The x-axis shows all the human genes. The y-axis shows the -log$_{10}$($P$ value) based on the paired t-test between SARS-CoV-2 infected and mock-infected cells. **C** Circosplot showing the SR genes of each viral gene. Circos plot’s tracks, from outer to inner, are the heatmap of log, Fold Change between SARS-CoV-2- and mock- infected cells in GSE147507, and the $P$ values of Spearman’s correlation test between SR genes and RT-PCR results of SARS-CoV-2 in GSE156063. **D** Hierarchical clustering of the pairwise correlation of SR genes (rows and columns) in GSE156063. **E** Heatmap showing the clustering results based on SR genes among healthy and severe COVID-19 patients in GSE171110. **F** Heatmap showing the expression levels of SR genes among healthy, mild, and severe COVID-19 patients in GSE164805.
showed that COVID-19 patients tended to be clustered together (Fig. 1D, Additional file 1). The same result was also observed for distinguishing healthy individuals and patients with severe disease in a different COVID-19 cohort (GSE171110), suggesting that the dysregulation of SR genes might play an important role in COVID-19. Remarkably, in a dataset on COVID-19 disease progression (GSE164805), we found that transcription levels of SR genes also presented a dynamic variable trend with the progression of COVID-19 (from healthy to mild to severe) (Fig. 1E). These results highlighted the potential function of SR genes in response to SARS-CoV-2 infection.

Subsequently, we inferred viral functional genomics by performing functional enrichment analysis of the SR genes for each viral gene based on “guilt-by-association” (GBA) (Additional file 1). GBA states that genes which are associated are more likely to share similar functions [2]. Therefore, we could establish connections between each viral gene and host biological process based on the enriched functional terms of the viral SR gene (P < 0.05, hypergeometric test). Furthermore, we constructed the functional landscape of the host response to SARS-CoV-2 infection and similar functional terms were clustered (Fig. 2A). The SARS-CoV-2 genes extensively affected various cellular programs, such as immune response, metabolic processes, and signaling pathways (Table 1). Notably, one of the functional annotations for the viral spike (S) gene enriched in the reproductive system-related terms (GO:0048608 reproductive structure development, GO:0003006 developmental process involved in reproduction, and GO:0008585 female gonad development), indicated that S gene-related SR genes might be engaged in reproductive system-related biological processes. The S protein can help SARS-CoV-2 entry into host cells by interacting with cell surface entry factors, ACE2 and NRP1, to mediate the cell membrane fusion process (Fig. 2B) [3]. We found that both ACE2 and NRP1 were generally expressed at higher levels in the human reproductive system than in other organs at the transcriptome and proteome levels, implying a high risk of SARS-CoV-2 infection in the reproductive system (Fig. 2C). In addition, some tissue-specific expression of S gene-related SR genes in the reproductive system, e.g., Testis Tissue Sperm-Binding Protein Li 44a (WDR77), Testis-Specific Gene A2 Protein (RSPH1), Follicle-Stimulating Hormone-Releasing Protein (INHBA), suggested that the viral S protein might affect reproductive system-related processes (Fig. 2D). Recently, some case and autopsy reports have also demonstrated an involvement of the reproductive system in patients with COVID-19 [4, 5]. Although it is unclear whether the reproductive system of patients functioned normally prior to SARS-CoV-2 infection, the evidence prompts that we should be aware of any possible impact of SARS-CoV-2 on the reproductive system. As SARS-CoV-2 has evolved into a long-term problem, some S protein-based vaccines are widely used to prevent COVID-19. There are currently four main COVID-19 vaccines available: whole virus, protein subunit, viral vector, and nucleic acid (mRNA). Except for the whole virus vaccine, which is produced by inactive pathogens, the other three vaccines are designed and manufactured on the basis of the viral S protein. They all work by exposing the body to molecules from the target pathogens to trigger host immune response, however, they did not contain the live components (for whole virus, nucleic acid, and protein subunit vaccines) or produce the complete SARS-CoV-2 structure (for viral vector vaccine). Therefore, these vaccinations should be safe to develop immunity against SARS-CoV-2 infection and are unlikely to affect the reproductive system. Nevertheless, as these results are based on reasonable speculation and retrospective analysis, systematic validation in a large sample cohort is still required.

To better understand the response mechanisms within SARS-CoV-2 host cells, we performed an integrative network analysis (Additional file 1). First, we constructed the viral gene-SR gene- transcription factors (TFs) regulatory network, which consisted of five viral genes, 27 TFs, and 50 SR genes (Fig. 3A; Additional file 2: Table S2). These TFs might be involved in the regulation of SR gene expression in SARS-CoV-2 infected cells, where TFs NFKB1, RELA, and STAT3 have been demonstrated could respond to COVID-19 by previous studies [3]. Next, we established the protein–protein interaction (PPI) network, which included 236 interaction relationships between 117 SR genes (Fig. 3B). We found the Complement component 3 (C3) gene was a hub node in the PPI network. C3 was identified as the SR gene of viral ORF3a, ORF7a, S, E, and N genes, suggesting that its transcription levels might vary with these viral gene expression levels. In addition, the C3 gene could interact with 14 other SR genes (Fig. 3B). Recent studies have suggested that complement C3 activation is an initial effector mechanism that contributes to thromboinflammation and organ damage in COVID-19. Indeed, patients with severe acute respiratory distress syndrome (ARDS) caused by COVID-19 pneumonia have also been safely and successfully treated with AMY-101, a compstatin-based complement C3 inhibitor [4]. These results highlighted the potential value of identifying SR genes in the development of antiviral treatment strategies. Besides, we constructed a non-coding RNA-SR gene co-expression network in COVID-19 patient peripheral blood (GSE166552), which consisted of 460 co-expression relationships between 125 ncRNAs (99 IncRNAs and 26
circRNAs) and 55 SR genes (Fig. 3C). Some hub circRNAs and IncRNAs were identified, which may be important in screening novel blood biomarkers of COVID-19 (Additional file 2: Table S3).

In addition, we constructed the drug-SR gene interaction network to examine potential antiviral drugs (Fig. 3D; Additional file 2: Table S4). The networks contained 48 SR genes and 204 FDA-approved drugs, some of which have shown clinical efficacy for the treatment of COVID-19. For example, dexamethasone and methylprednisolone have been widely used to treat COVID-19 with promising benefits [5]. Indomethacin, a PTGS2 inhibitor (SR gene PTGS2 had been reported as a pro-viral host factor of SARS-CoV-2 [6]), has been used in several clinical trials for the treatment of COVID-19 and has shown promising clinical efficacy [7]. Bevacizumab, a monoclonal anti-vascular endothelial growth factor antibody, which exhibits immunomodulatory effects and has been used in combination with antineoplastic agents for the treatment...
of multiple cancers, has also been shown to have high clinical benefits for patients with severe COVID-19 in a single-arm trial (NCT04275414) [8]. We subsequently recognized some candidate drugs for SARS-CoV-2 infection based on the network cluster analysis, which resulted in two drug-gene clusters (Table 2; Additional file 1). Cluster A contained three proteasome inhibitors (bortezomib, carfilzomib, and ixazomib citrate), in which Carfilzomib was identified as a suitable candidate for the treatment of COVID-19 by computational methods. The main biological processes of each SARS-CoV-2 gene are listed in Table 1. The processes associated with the viral genes and their regulatory networks are illustrated in Fig. 3.
drug repurposing studies [9]. Cluster B consisted of two potential drugs zinc chloride and methyldopa. Zinc, acting an immune booster, may prevent SARS-CoV-2 infection. Previous studies have indicated that zinc plays an important role in immune system development, and that increased intracellular zinc concentrations efficiently impair replication in a number of RNA viruses [10]. Several retrospective analyses of COVID-19 patients have also indicated that zinc deficiency is associated with a prolonged hospital stay and increased mortality [11, 12]. Thus, together, these findings suggested that zinc might be useful in preventing SARS-CoV-2 infection. In summary, SR gene-associated drugs provided novel insights into the development of antiviral treatment and prevention strategies in the post-COVID-19 era.

**Abbreviations**
COVID-19: Coronavirus disease 2019; SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; SR gene: Human genes responding to SARS-CoV-2 infection.

**Supplementary Information**
The online version contains supplementary material available at https://doi.org/10.1186/s13578-021-00684-4.

| Clusters | SARS-CoV-2 gene | SR gene | Drug | Interaction types |
|----------|----------------|---------|------|-------------------|
| Cluster A | S, E, M, N, ORF7a | C3 | Zinc chloride | Inhibitor, ligand |
|          |                | C3  | Methyldopa | Unknown |
|          |                | C1R | Zinc chloride | Modulator |
|          |                | C1R | Methyldopa | Unknown |
|          |                | C1S | Zinc chloride | Modulator |
|          |                | C1S | Methyldopa | Unknown |
|          |                | CFB | Methyldopa | Unknown |
| Cluster B | S, E, M | PSMB3 | Bortezomib | Inhibitor |
|          | S, E, N, ORF1ab, ORF3a | PSMB3 | Carfilzomib | Inhibitor |
|          |            | PSMB3 | Ixazomib citrate | Inhibitor |
|          | M, ORF1ab, ORF7a, ORF8 | PSMC5 | Carfilzomib | Inhibitor |
|          |                | PSMC5 | Bortezomib | Inhibitor |
|          |                | PSMC5 | Ixazomib citrate | Inhibitor |

**Acknowledgements**
The authors gratefully thank the GEO for providing data for this work.

**Authors’ contributions**
XP and XL designed the study, performed analysis, wrote and revised the manuscript. HZ and NS share the senior authorship of this study. The authors read and approved the final manuscript.

**Funding**
This work was supported by the National Natural Science Foundation of China (32170674) and Natural Science Foundation of Heilongjiang Province, China (LH2020C05).

**Availability of data and materials**
All data analysed during this study are included in the article.

**Declarations**

**Ethics approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Competing interests**
The authors declared no potential competing interests in terms of the research, authorship, and/or publication of this article.

**Author details**
1. College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China.
2. Department of Dermatology, Xiangya Hospital, Central South University, Changsha 410008, China.

Received: 20 April 2021   Accepted: 24 August 2021
Published online: 08 September 2021

**References**
1. Davidson AD, Williamson MK, Lewis S, Shoemark D, Carroll MW, Heesom K, Zambon M, Ellis J, Lewis PA, Hiscox JA, Matthews DA. Characterisation of the transcriptome and proteome of SARS-CoV-2 using direct
RNA sequencing and tandem mass spectrometry reveals evidence for a cell passage induced in-frame deletion in the spike glycoprotein that removes the furin-like cleavage site. bioRxiv. 2020;69:6219.

2. Oliver S. Guilt-by-association goes global. Nature. 2000;403:601–3.

3. Fagone P, Ciarleo R, Lombardo SD, Iacobello C, Palermo CI, Shoenfeld Y, Bendtzen K, Bramanti P, Nicolletti F. Transcriptional landscape of SARS-CoV-2 infection dismantles pathogenic pathways activated by the virus, proposes unique sex-specific differences and predicts tailored therapeutic strategies. Autonomn Rev. 2020;19:102571–102571.

4. Mastaglio S, Ruggeri A, Risitano AM, Angelillo P, Yancopoulou D, Mastellos DC, Huber-Lang M, Piemontese S, Asianelli A, Garlanda C, et al. The first case of COVID-19 treated with the complement C3 inhibitor AMY-101. Clinical Immunol. 2020;215:108450–108450.

5. Ahmed MH, Hassan A. Dexamethasone for the treatment of coronavirus disease (COVID-19): a review. SN Compr Clin Med. 2020;2:2637–46.

6. Loganathan T, Ramachandran S, Shankaran P, Nagarajan D, Mohan SS. Host transcriptome-guided drug repurposing for COVID-19 treatment: a meta-analysis based approach. PeerJ. 2020;8:e9357.

7. Oh KK, Adnan M, Cho DH. Network pharmacology approach to decipher signalling pathways associated with target proteins of NSAIDs against COVID-19. Sci Rep. 2021;11:9606.

8. Pang J, Xu F, Aondio G, Li Y, Fumagalli A, Lu M, Valmadre G, Wei J, Bian Y, Canesi M, et al. Efficacy and tolerability of bevacizumab in patients with severe Covid-19. Nat Commun. 2021;12:814.

9. Wang J. Fast identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study. J Chem Inf Model. 2020;60:3277–86.

10. te Velthuis AJ, van den Worm SH, Sims AC, Baric RS, Snijder EJ, van Hemert MJ. Zn(2+) inhibits coronavirus and arterivirus RNA polymerase activity in vitro and zinc ionophores block the replication of these viruses in cell culture. PLoS Pathog. 2010;6:e1001176.

11. Jothimani D, Kailasam E, Danielraj S, Nallathambi B, Ramachandran H, Sekar P, Manoharan S, Ramani V, Narasimhan G, Kaliamoorthy I, Rela M. COVID-19: poor outcomes in patients with zinc deficiency. Int J Infect Dis. 2020;100:343–9.

12. Ali N, Farhia KA, Islam F, Mohanto NC, Ahmad I, Hosen MJ, Ahmed S. Assessment of the role of zinc in the prevention of COVID-19 infections and mortality: a retrospective study in the Asian and European population. J Med Virol. 2021;93:4326–33.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.