Alterations in the host transcriptome in vitro and in vivo following severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection

Xiaomei Lei  
Gansu Provincial Hospital

Zhijun Feng  
Lanzhou University Second Hospital

Xiaojun Wang (wangxj19@lzu.edu.cn)  
Gansu Provincial Hospital

Xiaodong He  
Lanzhou University Second Hospital

Research Article

Keywords: biomarkers, SARS-CoV-2, COVID-19, GEO, integrated bioinformatics, microarray analysis

DOI: https://doi.org/10.21203/rs.3.rs-37567/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background. Exploring alterations in the host transcriptome following SARS-CoV-2 infection is not only highly warranted to help us understand molecular mechanisms of the disease, but also provide new prospective for screening effective antiviral drugs, finding new therapeutic targets, and evaluating the risk of systemic inflammatory response syndrome (SIRS) early.

Methods. We downloaded three gene expression matrix files from the Gene Expression Omnibus (GEO) database, and extracted the gene expression data of the SARS-CoV-2 infection and non-infection in human samples and different cell line samples, and then performed gene set enrichment analysis (GSEA), respectively. Thereafter, we integrated the results of GSEA and obtained co-enriched gene sets and co-core genes in three various microarray data. Finally, we also constructed a protein-protein interaction (PPI) network and molecular modules for co-core genes and performed Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis for the genes from modules to clarify their possible biological processes and underlying signaling pathway.

Results. A total of 11 co-enriched gene sets were identified from the three various microarray data. Among them, 10 gene sets were activated, and involved in immune response and inflammatory reaction. 1 gene set was suppressed, and participated in cell cycle. The analysis of molecular modules showed that 2 modules might play a vital role in the pathogenic process of SARS-CoV-2 infection. The KEGG enrichment analysis showed that genes from module one enriched in signaling pathways related to inflammation, but genes from module two enriched in signaling of cell cycle and DNA replication. Particularly, necroptosis signaling, a newly identified type of programmed cell death that differed from apoptosis, was also determined in our findings. Additionally, for patients with SARS-CoV-2 infection, genes from module one showed a relatively high-level expression while genes from module two showed low-level.

Conclusions. We identified two molecular modules were used to assess severity and predict the prognosis of the patients with SARS-CoV-2 infection. In addition, these results provide a unique opportunity to explore more molecular pathways as new potential targets on therapy in COVID 19.

Introduction

The epidemic of Coronavirus Disease 2019 (COVID–19) caused by SARS-CoV–2 is one of the public health emergencies all over the world and severely affects human health. As of June 16, 2020, more than 8,300,000 human cases were confirmed, with 446,000 deaths and about 5.3% estimated mortality rate. The rapidly increasing of patients, especially the critical or lethal patients, brought a big challenge to the public health. Generally, the infection with SARS-CoV–2 is spread from person to person via respiratory droplets, however, other special transmission vectors, such as aerosol, fecal–oral route, and intermediate fomites from both symptomatic and asymptomatic patients during the incubation period, are also be our focuses 1. In terms of the characteristics of symptoms, it has been demonstrated that majority of
patients with SARS-CoV–2 infection have flu-like symptoms as their initial symptom. But non-specific gastrointestinal symptoms such as dyspnea and diarrhea are also observed in several patients 2. In terms of the characteristics of patient population, male, aged over 65 and smoking patients are generally at a greater risk of disease progression, which result in the critical or lethal condition and present with severe pneumonia, acute respiratory distress syndrome (ARDS), multiple organ failure, and even death 3. Based on published data from the Chinese Center for Disease Control and Prevention, more than 80% patients were classified as mild, and 20% cases were classified as severe or critical illness. Alarmingly, the mortality was up to nearly 50% in patients with critical illness 4.

According to current evidences, the pathogenesis of lethal SARS-CoV–2 infection is associated with dysregulated inflammatory responses, tissue injury, and multiple organ dysfunction (MODS), but the details remain obscure. Therefore, at a gene level, exploring how disease exactly changes from mild to severe is not only highly warranted to help us understand disease progress, but also provide new prospective for screening effective antiviral drugs, finding new therapeutic targets, and evaluating the risk of systemic inflammatory response syndrome (SIRS) early. For this purpose, we designed this study of integrated bioinformatics analysis based on GEO (https://www.ncbi.nlm.nih.gov/geo/) database 5,6 using gene expression profile data of lung tissues from healthy individuals and patients with SARS-CoV–2 infection, with a view to discovering valuable mRNA biomarkers potential pathogenic mechanisms, and finally in order to guiding clinical diagnosis and treatment.

**Methods**

We downloaded GSE148815 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc = GSE148815), GSE150316 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi), and GSE147507 datasets from GEO database. In vivo level, we extracted the gene expression data of the adults in GSE148815 as normal healthy samples, and the data of patients with SARS-CoV–2 infection in GSE150316 as infected samples. Then, we integrated two sets of Raw-Counts data to compose the patient-microarray data used in the study. In vitro level, we extracted the gene expression data of normal human bronchial epithelial (NHBE) cell line, and A549 (lung adenocarcinoma) cell line to form two cell line- microarray data (NHBE-data and A549-data). In order to reduce the effect of errors caused by gene chip on the pattern of results and our conclusions, data preprocessing was performed according to the following procedures using R software (R 3.6.1, https://cran.r-project.org/): standardized microarray data utilizing DESeq2 package 8. log2 transformed. handled missing or unknown values. checked and adjusted batch effects. corrected value of log fold change (logFC) to reduce the error caused by gene with a constant expression like zero. The detailed methods of data processing were described in the supplementary Text S1. For the microarray data was downloaded from a public database, the patient consent or ethics committee approval was not necessary.

**Differential analysis**
Differential analysis was performed in the three microarray data using DESeq2 package to identify the differentially expressed genes (DEGs) between the samples with SARS-CoV–2 infection and non-infection. And a logFC and a statistical p value were obtained. When a gene's logFC>0, we defined it as high-expression in the SARS-CoV–2 infected group; when logFC<0 as low-expression. The result was visualized as a heat map with hierarchical clustering using pheatmap package (https://CRAN.R-project.org/package = pheatmap).

**GSEA and core genes identification**

GSEA was performed with the results of DEGs using phenotype labels “SARS-CoV–2 infected” versus “non-infected” by the clusterProfiler package 9 to elucidate relevant biological significance. Firstly, we converted the gene symbols to Entrez IDs, and then comparatively analyzed with the data of hallmarks gene set (Version 7.1) 10 downloaded from the Molecular Signatures Database (MSigDB, https://www.gsea-msigdb.org/gsea/msigdb/) 11. And the results with a p value<0.05 as a bubble chart. After that, we obtained co-enriched hallmarks gene sets and their co-core genes from the GSEA results of the three microarray datasets.

**PPI network and molecular modules construction**

The PPI network of co-core genes was constructed in the STRING database (https://string-db.org/) through the following setting: meaning of network edges was set as “confidence”, active interaction sources were all, minimum required interaction score was selected as “highest confidence (0.4500)” and finally display simplification was to hide disconnected nodes in the network 12,13. The PPI-data was downloaded and then identified key molecular modules by using MCODE plug-in Cytoscape software (version 3.7.1, https://cytoscape.org/), respectively. In MCODE, filters were based on the parameters as “Network Scoring ticked Include Loops and Degree Cutoff = 2,” “Cluster Finding ticked Haircut, Node Score Cutoff = 0.2, K-Core = 2, and Max. Depth = 100” 14.

**KEGG pathway enrichment analysis**

KEGG pathway enrichment analysis was conducted with genes in each molecular module to identify the biological functions of modules using clusterProfiler 9 package. Adjusted p value<0.05 was considered to be statistically significant, and the results was visualized by GOplot package. In addition, we used patient-data to evaluate the expression levels of genes from the modules in patients with SARS-CoV–2 infection compared to normal healthy individuals, and the results were shown as box plot.

**Results**
The data processing workflow was illustrated in Figure 1. The results corresponding to data processing were explained in Text S1, which reflected that the gene expression used in our study was uniformly distributed.

**Differential analysis**

The heatmaps (Figure 2A-C) reflected the variance in gene expression relative to the mean across all samples with high-level expressed genes in red and low-level expressed genes in green. For this outcome, we found that there was a significant difference in gene expression level between samples with SARS-CoV–2 infection and non-infection, which also provided a foundation for us to further explore the genetic changes caused by SARS-CoV–2 in the host.

**GSEA and core genes identification**

The GSEA results between samples with SARS-CoV–2 infection and non-infection were shown in Figure 3A-C. For patient-data as in vivo level (Figure 3A), 29 of the hallmark gene sets were enriched, covering 14 activated and 15 suppressed items. In vitro level, 15 of the hallmark gene sets (activated: 14, and suppressed: 1) were enriched in NHBE-data (Figure 3B), and 19 ((activated: 16, and suppressed: 3)) in A549-data (Figure 3C). We merged the three GSEA results and obtained 11 hallmark gene sets co-enriched, including 10 activated and 1 suppressed (Figure 3D), and the details were listed in Table 1. Then, we extracted all core genes in the co-enriched gene sets from the three datasets (499 core enriched genes in patient-data, 293 in NHBE-data, and 438 in A549-data), and obtained 56 co-core genes (Figure 3D) that were described in Table S1.

From these findings, we found that in samples with SARS-CoV–2 infection, inflammation-related hallmark gene sets like IL6-Jak-Stat3-signaling, IL2-Stat5-signaling, complement, coagulation, angiogenesis, Kras-signaling-up, and TGF-β-signaling were activated while cell cycle-related hallmarks gene set like E2F-targets was suppressed. Additionally, we also found inflammation-related genes like IL6, ICAM1, IL1B, and TNF were involved in the co-enriched gene sets. These findings indicated that, to some extent, inflammatory genes contribute greatly to the biological process of disease progression and the interaction of molecular function in samples with SARS-CoV–2 infection.

**PPI network and molecular models**

The PPI network for 56 co-core genes was built through STRING database and the result contained 54 nodes and 234 edges (Figure 4A). Through MCODE, we identified two key molecular modules, and their gene make-ups were shown in Figure 4B-C (Figure 4B as module one, Figure 4C as module two). To our surprise, nearly a third of the co-core genes were members of module one suggesting that module one plays a crucial role in the PPI network, which deserved special attention.
KEGG pathway enrichment analysis

50 of various KEGG pathways were enriched by the genes from the module one (Table S2), including multiple specific infection-related signaling pathways caused by malaria, legionellosis, measles, tuberculosis, influenza A, and hepatitis B. In addition, it also included multiple inflammation-related signaling pathways like TNF signaling pathway, IL–17 signaling pathway, PI3K-Akt signaling pathway, Jak-Stat signaling pathway (Figure 5A). For genes from module two, four pathways (Figure 5B) related to cell cycle and DNA replication were identified. Finally, through the results of gene expression level (Figure 5C), we found that, for patients with SARS-CoV–2 infection, genes from module one showed a relatively high-level expression while genes from module two showed low-level. Remarkably, our data source was from patients who died due to SARS-CoV–2 infection. These results signified that high expression of genes in module one and low expression of genes in module two might predict a worse prognosis in patients with in SARS-CoV–2 infection.

Discussion

As obligate intracellular parasites, viruses must rely on host cell for replication, and its infection also cause changes in the morphology and function of host cells. SARS-CoV–2 is a monopartite, single-stranded, and positive-sense RNA virus with a genome size of 29,903 nucleotides, making it the second-largest known RNA genome 15. Severe COVID–19 was characterized by respiratory failure caused by hyper-inflammation, which not only bring great difficulties to clinical treatment, but also increase mortality for disease 16. It is now well established that a cytokine storm syndrome (CSS) involved in molecules like interleukin 6 (IL6), interleukin 8 (IL8), E-cadherin, MCP–1, and VEGF, probably may be the reason for hyper-inflammation in severe cases 2,17.

Investigating the differences in genetic level during the disease occurrence and development not only help us understand disease more clearly, but also guide for developing countermeasures. In the present study, candidate datasets were consisted of two parts, in vivo level (patient-data) and in vitro level (NHBE-data and A549 data), and they were annotated by the same platform of GPL18573 [Illumina NextSeq 500 (Homo sapiens)], which eliminated the adverse effect of the bias of gene coverage and algorithm usage from different sequencing platforms on the results and conclusion. After completing the differential analysis and GSEA, we found that in samples with SARS-CoV–2 infection, inflammation-related hallmark gene sets like IL6-Jak-Stat3-signaling, IL2-Stat5-signaling, complement, coagulation, angiogenesis, Kras-signaling-up, and TGF-β-signaling were activated while cell cycle-related hallmark gene set like E2F-targets was suppressed both in vivo and vitro level. Subsequently, inflammation-related genes were found in the core genes of those co-enriched gene sets, such as IL6, ICAM1, IL1B, CSF1, TLR2, VEGFA, TNF, and so on. We constructed PPI network for the co-core genes, and analyzed molecular modules through MCODE, and identified two key molecular modules. Module one consisted of 17 genes: BCL2A1, BCL2L1, SOD2, PTGS2, TNF, PLAUR, MYC, LIF, VEGFA, LOX, IL6, CSF1, C3, TLR2, ICAM1, IL1B, and TNFAIP3, which were mainly involved in inflammatory response pathway and complement system cascade, such as such as TNF signaling, IL–17 signaling, PI3K-Akt signaling, Jak-Stat signaling, NOD-like receptor signaling,
Toll-like-receptor signaling, and complement signaling. Module two consisted of 9 genes: EZH2, PTTG1, DLGAP5, RFC3, DUT, RPA2, CCNB2, PCNA, and MCM4, were mainly involved in cell cycle and DNA replication signaling. Finally, by assessing the level of gene expression in the patient-data, we found genes from module one showed a relatively high-level expression while genes from modules two showed low-level. All of these findings may provide novel evidences selecting effective mRNA biomarkers to evaluate disease progression and predict prognosis for patients with SARS-CoV–2 infection. Also, these findings improve our understanding of additional targets for anti-SARS-CoV–2 agents.

Cascade reactions of inflammation and immunity are the two key aspects in the pathogenic process of virus infecting the host, which often requires multiple cells and cellular components to participate, and lead to changes in the level of gene expression. It is no exception for SARS-CoV-2. GSEA, based upon functional class scoring methods, is capable of solving biological problem by focusing on gene sets rather than individual genes 11. In our study, we did use this method for analyzing genetic changes related to inflammation and immune response after SARS-CoV–2 infection, which preserved gene-gene correlations, and provided an improved understanding of the biological functional enrichment in the groups of high-level and low-level expressed genes 18. For hallmarks gene sets, we found two gene sets involved in two significant molecules in the interleukin super family, gene set of IL6-Jak-Stat3-signaling (genes were up-regulated by IL6 via Stat3 during acute phase response) and IL2-Stat5-signaling (genes up-regulated by Stat5 in response to IL2 stimulation), were activated in samples of SARS-CoV–2 infection. Several studies have been reported that IL6, as one of the genes involved in cytokine release syndrome (CRS), was used to assess severity of COVID 19 like respiratory failure, ARDS, and adverse clinical outcomes 19–24. Similar to these conclusions, we identified IL6 not only was at a high-level expression in data of cell lines and cases died due to SARS-CoV–2 infection, but also interacted the most with other genes in the PPI network, which confirmed again its value of as an active cytokine with a wide range of biological functions and as an effective biomarker for the prognosis of COVID 19. IL2 is a secreted cytokine that was important for the proliferation of T and B lymphocytes25,26. Here, no core role of IL2 was observed even though we found activated IL2- Stat5-signaling in the infected samples, but IL2-related genes with the example of CSF1 and LIF, were among the co-core genes with a high-level expression, which revealed that the action of IL2 might be as a 'homeostatic' cytokine, and mainly involved in the immune response but not in CRS. Moreover, the gene sets of TNFA-signaling (genes regulated by NF-κB in response to TNF) and Kras-signaling-up (genes up-regulated by Kras activation) were also activated in patients with COVID 19, which further uncover the roles of genes and cytokines such as NF-κB, TNF, and Kras in the pathogenesis and progression of SARS-CoV–2 infection. These factors together with IL6 might lead to the possibility of severe and/or fatal status for COVID 19. Notably, gene set of epithelial-mesenchymal-transition (EMT) (genes defining EMT) was represented in activated state in SARS-CoV–2 infected samples, which implied the potential of SARS-CoV–2 to lead to pulmonary interstitial fibrosis (PIF) as the disease progress. PIF was also an important feature of COVID 19 cases with poor prognosis, as previously reported 27,28. Apart from this, gene set of complement (genes encoding components of the complement system, which is part of the innate immune system) and gene set of coagulation (genes up-regulated during formation of blood vessels) were also found as activated...
status in our result, which indicated that the disease evolution of SARS-CoV–2 infection was a multiple pathway, complicated process, comprising of various dynamic changes in the genome.

E2F transcription factors play critical roles in the control of transcription, cell cycle and apoptosis29–31. For samples with SARS-CoV–2 infection, we found that the gene set of E2F targets was suppressed obviously, which indicated that cell cycle disorders might occur in COVID 19.

From a clinical perspective, this might be related to hypoxia caused by SARS-CoV–2 infection in human bodies 32, because hypoxia could directly affect cell differentiation and energy metabolism 33. From gene level, we identified that genes like PCNA, CCNB2, PTTG1, as were at a low-level expression in the samples of SARS-CoV–2 infection. According to literature report, PCNA and PTGG1 involved in regulating the biological function of p21 that was a CDK inhibitor to trigger cell-cycle arrest in the G1 and G2 phases 34,35. CCNB2 involved in the formation of CCNB2/CDK1 complex that controlled separase activity through inhibition of phosphorylation and regulated the biological process of G2 phase of cell cycle 36. Incorporating previous studies, we proposed that in the process of SARS-CoV–2 infecting the host, genetic changes caused cell cycle disorders, which might manifest as hyper-active S-phase and hypo-active G2 phase facilitating viral DNA replication.

Other co-core genes, such as ICAM1, a leukocyte adhesion molecule, mainly encoded a cell surface glycoprotein typically expressed on endothelial cells and immune cells. As reported, ICAM1 plays an important role in enhancing CD16+ monocyte adhesion to the endothelium 37, and regulating IL6/Akt/Stat3/NF-κB-dependent pathway 38. For roles in viral infection, ICAM1 had been determined to involve in DC-mediated HIV–1 transmission to CD4(+) T cells 39 and regulate interferon-gamma and IL17 in hepatitis B virus infection 40. Our results showed that ICAM1 participated in the signaling pathways of NF-κB, IL17, and TNF, which was in line with similar previous studies. TNFAIP3, induced by the TNF, plays a role in inhibiting the activation of NF-κB as well as TNF-mediated apoptosis. Literature data reported TNFAIP3 was closely associated with the replication of viruses like influenza A 41, and hepatitis B virus 42. In our study, we identified TNFAIP3, TNF, and IL1B were involved in the signaling pathway of necroptosis.

Similar to apoptosis, necroptosis was executed via distinctive signaling mechanism comprising a cascade of specified proteins, resulting in regulated necrotic cell death 43. Physiologically, necroptosis induced an innate immune response as well as premature assembly of viral particles in cells infected with virus that abrogates host apoptotic machinery, which was advantageous for the host. On the other hand, necroptosis was also deleterious because it can cause various diseases such as sepsis, neurodegenerative diseases and ischemic reperfusion injury 43. Qin et al. observed that necroptosis of the pulmonary epithelium was associated with severe H7N9 infection leading to ARDS, and the final conclusion indicated that necroptosis inhibition might be a novel therapy for H7N9 influenza virus 44. However, the benefits of necroptosis to the host may sometimes be outweighed by the potentially deleterious hyperinflammatory consequences of activating this death modality in pulmonary and other tissues45,46. All this evidence remined us that necroptosis signaling was a double-edged sword in the
defense to microbial infection, or even might be the likely culprit of the critical and/or lethal SARS-CoV–2 infected cases. This also provided a novel clue for necroptosis signaling in the treatment of COVID 19 and served it as a potential therapeutic target. BCL2A1, one of the pro-inflammatory cytokines, encoded a member of the BCL–2 protein family. In our study, we found BCL2A1 participated in the signaling of NF-κB and apoptosis, which were similar to previous reports in the literature 47,48. Nevertheless, more details of how BCL2A1 affected the pathogenesis of SARS-CoV–2 infection still needed to be explored.

For signaling pathway, besides mentioned above, signaling pathway of AGE-RAGE, MAPK, cytokine-cytokine receptor interaction, EGFR tyrosine kinase inhibitor resistance, viral protein interaction, were enriched in our results. From these findings, we believed that in the course of SARS-CoV–2 infection, the dysregulation was gene-specific rather than pathway-specific. Hence, we tried to construct molecular modules not only for accurate prediction but also for the evaluation of the effects of genes on COVID 19 patients’ prognosis. Our data indicated that in severe and/or fatal SARS-CoV–2 infection cases, immune responses tended to be gentle while inflammatory cascades were hyper-activated, which might be attributable to aberrant gene expression. However, in terms of how genes found in our results participated in the pathogenesis of SARS-CoV–2, it is still an unresolved problem, and further needed to verify in vivo data.

Conclusions

In conclusion, with integrated bioinformatics analysis for microarray datasets, we identified two molecular modules were used to assess severity and predict the prognosis of the patients with SARS-CoV–2 infection. In addition, these results provide a unique opportunity to explore more molecular pathways as new potential targets on therapy in COVID 19.

Declarations

Funding

The authors received no funding for this work.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

W.X.J. and H.X.D. designed the overall study with contributions from L.X.M and F.Z.J.. L.X.M and F.Z.J. collected and analyzed data, and co-wrote the paper. L.X. and S.P.P. collected and analyzed data, prepared figures and/or tables. W.X.J. and H.X.D. conceived and designed the study, authored or reviewed drafts of the paper. All authors read and approved the final manuscript.

Acknowledgements
The authors thank the contributors of the GEO (https://www.ncbi.nlm.nih.gov/geo/) database for sharing the microarray data on open access.

Data Availability

Data is available at NCBI GEO (https://www.ncbi.nlm.nih.gov/geo/), accession numbers: GSE147507, GSE150316, and GSE148815.

References

1. Zhou, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273, doi:10.1038/s41586-020-2012-7 (2020).

2. Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, *Lancet (London, England)* **395**, 497-506, doi:10.1016/s0140-6736(20)30183-5 (2020).

3. Li, J. et al. The epidemic of 2019-novel-coronavirus (2019-nCoV) pneumonia and insights for emerging infectious diseases in the future. *Microbes and infection* **22**, 80-85, doi:10.1016/j.micinf.2020.02.002 (2020).

4. Wu, & McGoogan, J. M. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. *Jama*, doi:10.1001/jama.2020.2648 (2020).

5. Barrett, et al. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res* **41**, D991-995, doi:10.1093/nar/gks1193 (2013).

6. Edgar, , Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207-210, doi:10.1093/nar/30.1.207 (2002).

7. Blanco-Melo, D. et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*, doi:10.1016/j.cell.2020.04.026 (2020).

8. Love, M. I., Huber, & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

9. Yu, , Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).

10. Liberzon, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).

11. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

12. Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental *Nucleic Acids Res* **47**, D607-d613, doi:10.1093/nar/gky1131 (2019).
13. Feng, et al. Could CTSK and COL4A2 be specific biomarkers of poor prognosis for patients with gastric cancer in Asia?—a microarray analysis based on regional population. *J Gastrointest Oncol* **11**, 386-401 (2020).

14. Otasek, D., Morris, J. H., Boucas, J., Pico, A. R. & Demchak, B. Cytoscape Automation: empowering workflow-based network *Genome Biol* **20**, 185, doi:10.1186/s13059-019-1758-4 (2019).

15. Helmy, A. *et al.* The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control. *Journal of clinical medicine* **9**, doi:10.3390/jcm9041225 (2020).

16. Dhama, K. *et al.* An update on SARS-CoV-2/COVID-19 with particular reference to its clinical pathology, pathogenesis, immunopathology and mitigation strategies. *Travel medicine and infectious disease*, 101755, doi:10.1016/j.tmaid.2020.101755 (2020).

17. Moore, B. & June, C. H. Cytokine release syndrome in severe COVID-19. *Science (New York, N.Y.)* **368**, 473-474, doi:10.1126/science.abb8925 (2020).

18. Wu, Y. *et al.* Internal driving factors leading to extrahepatic manifestation of the hepatitis C virus infection. *International journal of molecular medicine* **40**, 1792-1802, doi:10.3892/ijmm.2017.3175 (2017).

19. Chen, *et al.* Clinical and immunological features of severe and moderate coronavirus disease 2019. *The Journal of clinical investigation* **130**, 2620-2629, doi:10.1172/jci137244 (2020).

20. Ruan, Q., Yang, , Wang, W., Jiang, L. & Song, J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive care medicine* **46**, 846-848, doi:10.1007/s00134-020-05991-x (2020).

21. Gubernatorova, O., Gorshkova, E. A., Polinova, A. I. & Drutskaya, M. S. IL-6: Relevance for immunopathology of SARS-CoV-2. *Cytokine & growth factor reviews*, doi:10.1016/j.cyto.2020.05.009 (2020).

22. Liu, B., Li, M., Zhou, Z., Guan, X. & Xiang, Can we use interleukin-6 (IL-6) blockade for coronavirus disease 2019 (COVID-19)-induced cytokine release syndrome (CRS)? *Journal of autoimmunity* **111**, 102452, doi:10.1016/j.jaut.2020.102452 (2020).

23. Zhang, C., Wu, , Li, J. W., Zhao, H. & Wang, G. Q. Cytokine release syndrome in severe COVID-19: interleukin-6 receptor antagonist tocilizumab may be the key to reduce mortality. *International journal of antimicrobial agents* **55**, 105954, doi:10.1016/j.ijantimicag.2020.105954 (2020).

24. Choy, H. *et al.* Translating IL-6 biology into effective treatments. *Nature reviews. Rheumatology* **16**, 335-345, doi:10.1038/s41584-020-0419-z (2020).

25. Zhang, , Yin, Y., Zhang, S., Luo, H. & Zhang, H. HIV-1 Infection-Induced Suppression of the Let-7i/IL-2 Axis Contributes to CD4(+) T Cell Death. *Scientific reports* **6**, 25341, doi:10.1038/srep25341 (2016).

26. Henneman, *et al.* Genetic susceptibility for cow’s milk allergy in Dutch children: the start of the allergic march? *Clinical and translational allergy* **6**, 7, doi:10.1186/s13601-016-0096-9 (2015).

27. George, M., Wells, A. U. & Jenkins, R. G. Pulmonary fibrosis and COVID-19: the potential role for antifibrotic therapy. *The Lancet. Respiratory medicine*, doi:10.1016/s2213-2600(20)30225-3 (2020).
28. Spagnolo, et al. Pulmonary fibrosis secondary to COVID-19: a call to arms? *The Lancet. Respiratory medicine*, doi:10.1016/s2213-2600(20)30222-8 (2020).

29. DeGregori, J. & Johnson, D. G. Distinct and Overlapping Roles for E2F Family Members in Transcription, Proliferation and Apoptosis. *Current molecular medicine* **6**, 739-748, doi:10.2174/1566524010606070739 (2006).

30. McClellan, K. A. *et al.* The p107/E2F pathway regulates fibroblast growth factor 2 responsiveness in neural precursor cells. *Molecular and cellular biology* **29**, 4701-4713, doi:10.1128/mcb.01767-08 (2009).

31. Fischer, M. & Müller, A. Cell cycle transcription control: DREAM/MuvB and RB-E2F complexes. *Critical reviews in biochemistry and molecular biology* **52**, 638-662, doi:10.1080/10409238.2017.1360836 (2017).

32. Rello, J., Storti, E., Belliato, M. & Serrano, R. Clinical phenotypes of SARS-CoV-2: implications for clinicians and researchers. *The European respiratory journal* **55**, doi:10.1183/13993003.01028-2020 (2020).

33. Hubbi, M. E. & Semenza, G. L. Regulation of cell proliferation by hypoxia-inducible factors. *American journal of physiology. Cell physiology* **309**, C775-782, doi:10.1152/ajpcell.00279.2015 (2015).

34. Mansilla, F., de la Vega, M. B., Calzetta, N. L., Siri, S. O. & Gottifredi, V. CDK-Independent and PCNA-Dependent Functions of p21 in DNA Replication. *Genes* **11**, doi:10.3390/genes11060593 (2020).

35. Tong, & Eigler, T. Transcriptional targets for pituitary tumor-transforming gene-1. *Journal of molecular endocrinology* **43**, 179-185, doi:10.1677/jme-08-0176 (2009).

36. Li, J., Ouyang, C., Zhang, C. H., Qian, W. P. & Sun, Q. Y. The cyclin B2/CDK1 complex inhibits separase activity in mouse oocyte meiosis I. *Development (Cambridge, England)* **146**, doi:10.1242/dev.182519 (2019).

37. Regal-McDonald, K., Xu, B., Barnes, J. & Patel, R. P. High-mannose intercellular adhesion molecule-1 enhances CD16(+) monocyte adhesion to the endothelium. *American journal of physiology. Heart and circulatory physiology* **317**, H1028-h1038, doi:10.1152/ajpheart.00306.2019 (2019).

38. Liu, C. *et al.* PM2.5-induced oxidative stress increases intercellular adhesion molecule-1 expression in lung epithelial cells through the IL-6/AKT/STAT3/NF-kappaB-dependent pathway. *Particle and fibre toxicology* **15**, 4, doi:10.1186/s12989-018-0240-x (2018).

39. Wang, H., Kwas, C. & Wu, L. Intercellular adhesion molecule 1 (ICAM-1), but not ICAM-2 and -3, is important for dendritic cell-mediated human immunodeficiency virus type 1 transmission. *Journal of virology* **83**, 4195-4204, doi:10.1128/jvi.00006-09 (2009).

40. Chen, H., Zhang, D., Wang, , Wang, X. & Yang, C. Significance of correlation between interferon-gamma and soluble intercellular adhesion molecule-1 and interleukin-17 in hepatitis B virus-related cirrhosis. *Clinics and research in hepatology and gastroenterology* **37**, 608-613, doi:10.1016/j.clinre.2013.05.010 (2013).

41. Maelfait, J. *et al.* A20 (Tnfaip3) deficiency in myeloid cells protects against influenza A virus infection. *PLoS pathogens* **8**, e1002570, doi:10.1371/journal.ppat.1002570 (2012).
42. Li, N. et al. Association of the tandem polymorphisms (rs148314165, rs200820567) in TNFAIP3 with chronic hepatitis B virus infection in Chinese Han population. *Virology journal* **14**, 148, doi:10.1186/s12985-017-0814-5 (2017).

43. Cho, S. The role of necroptosis in the treatment of diseases. *BMB reports* **51**, 219-224, doi:10.5483/bmbrep.2018.51.5.074 (2018).

44. Qin, C. et al. Close Relationship between cIAP2 and Human ARDS Induced by Severe H7N9 *BioMed research international* **2019**, 2121357, doi:10.1155/2019/2121357 (2019).

45. Balachandran, S. & Rall, G. Benefits and Perils of Necroptosis in Influenza Virus Infection. *Journal of virology* **94**, doi:10.1128/jvi.01101-19 (2020).

46. Xia, X., Lei, L., Wang, , Hu, J. & Zhang, G. Necroptosis and its role in infectious diseases. *Apoptosis : an international journal on programmed cell death* **25**, 169-178, doi:10.1007/s10495-019-01589-x (2020).

47. Vogler, BCL2A1: the underdog in the BCL2 family. *Cell death and differentiation* **19**, 67- 74, doi:10.1038/cdd.2011.158 (2012).

48. Mandal, M. et al. The BCL2A1 gene as a pre-T cell receptor-induced regulator of thymocyte survival. *The Journal of experimental medicine* **201**, 603-614, doi:10.1084/jem.20041924 (2005).

**Table**
| Description of hallmark gene set | Patient-data | NHBE-data | A549-data |
|----------------------------------|-------------|-----------|-----------|
| | Size | NES | pvalue | Size | NES | pvalue | Size | NES | pvalue |
| ALL OGR AFT_ REJECTION | 189 | 1.67 | 0.00 1 | 150 | 2.02 | 0.00 1 | 157 | 2.38 | 0.00 1 |
| COAGULATION | 132 | 1.77 | 0.00 3 | 116 | 1.79 | 0.00 1 | 121 | 1.97 | 0.00 1 |
| COMPLEMENT | 197 | 1.70 | 0.00 1 | 174 | 2.09 | 0.00 1 | 178 | 2.18 | 0.00 1 |
| E2F_TARGETS | 195 | -2.03 | 0.00 6 | 195 | -1.44 | 0.01 2 | 194 | -2.26 | 0.00 6 |
| EMT | 197 | 2.20 | 0.00 1 | 188 | 1.70 | 0.00 2 | 187 | 2.23 | 0.00 1 |
| IL2_STAT5_SIGNALING | 194 | 1.46 | 0.00 7 | 174 | 1.74 | 0.00 1 | 178 | 2.02 | 0.00 1 |
| IL6_JAK_STAT3_SIGNALING | 86 | 1.81 | 0.00 1 | 67 | 2.12 | 0.00 1 | 79 | 2.44 | 0.00 1 |
| INFLAMMATORY RESPONSE | 197 | 1.94 | 0.00 1 | 165 | 2.26 | 0.00 1 | 176 | 2.66 | 0.00 1 |
| INTERFERON_GAMMA | 198 | 1.36 | 0.02 2 | 179 | 2.35 | 0.00 1 | 186 | 2.90 | 0.00 1 |
Table 1 Summary of co-GSEA results in the three microarray data. NHBE, normal human bronchial epithelial cell line. A549, lung adenocarcinoma cell line A549. Size, the number of the core enrichment gene. NES, normalized enrichment score. ALLOGRAFT_REJECTION, genes up-regulated during transplant rejection. COAGULATION, genes encoding components of blood coagulation system; also up-regulated in platelets. COMPLEMENT, genes encoding components of the complement system, which is part of the innate immune system. E2F_TARGETS, genes encoding cell cycle related targets of E2F transcription factors. EMT, genes defining epithelial-mesenchymal transition (EMT), as in wound healing, fibrosis and metastasis. IL2_STAT5_SIGNALING, genes up-regulated by STAT5 in response to IL2 stimulation. IL6_JAK_STAT3_SIGNALING, genes up-regulated by IL6 via STAT3, during acute phase response. INFLAMMATORY_RESPONSE, genes defining inflammatory response. INTERFERON_GAMMA_RESPONSE, genes up-regulated in response to IFNG. KRAS_SIGNALING_UP, Genes up-regulated by KRAS activation. TNFA_SIGNALING_VIA_NFKB, genes regulated by NF-kB in response to TNF.

Figures
Figure 1

Network map of enriched KEGG pathways, and expression levels for genes from module one and module two. (A) network of enriched KEGG pathways for genes from module one. (B) network of enriched KEGG pathways for genes from module two. (C) expression levels for genes from module one and module two in patient-data.
Figure 2

The PPI network and the molecular modules for co-core genes. (A) PPI network of co-core genes constructed in STRING database. (B) molecular module one for co-core genes. (C) molecular module two for co-core genes. Circles represent genes, lines represent interactions between gene-encoded proteins and line width represents evidence of interactions between proteins.
**Figure 3**

The GSEA results (A-C) and distributions of co-enriched gene sets and co-core genes in the three microarray data. (A) is the result from patient-data, (B) is from NHBE-data, (C) is from A549-data. Upper half of (D) is Venn plot for distribution of co-enriched gene sets, and lower half of (D) is Venn plot for distribution of co-core genes. Patients, patient-data. NHBE, normal human bronchial epithelial cell line. A549, lung adenocarcinoma cell line A549.

**Figure 4**

The cluster heat map of the three microarray data. (A) is for patient-data, (B) is for NHBE-data, (C) is for A549-data. Red represents gene expression with a high-level, green represents gene expression with a low-level, and black represents gene expression with an intermediate-level.
Figure 5

The data processing workflow. N, number. NHBE, normal human bronchial epithelial cell line. A549, lung adenocarcinoma cell line A549. vs, versus. GSEA, gene set enrichment analysis. PPI, protein-protein interaction. KEGG, kyoto encyclopedia of genes and genomes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS2.docx
• TableS1.docx
• TextS1Detaileddataprocessinginformation.pdf