

# Host transcriptome alterations in vitro and in vivo following severe acute respiratory syndrome coronavirus 2 infection

**Yannian Luo**

Lanzhou University Second Hospital

**Juan Xu**

Lanzhou University Second Hospital

**Mingzhen Zhou**

Lanzhou University Second Hospital

**Xiaomei Lei**

Gansu Provincial Hospital

**Wen Cao**

Lanzhou University Second Hospital

**Xiaojun Wang** (✉ [wangxj19@lzu.edu.cn](mailto:wangxj19@lzu.edu.cn))

Gansu Provincial Hospital

**Ruili Qiao** (✉ [qiaorl@lzu.edu.cn](mailto:qiaorl@lzu.edu.cn))

Lanzhou University First Hospital

---

## Research Article

**Keywords:** biomarkers, SARS-CoV-2, COVID-19, Gene Expression Omnibus, integrated bioinformatics, microarray analysis

**Posted Date:** April 15th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-37567/v2>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Host transcriptome alterations in vitro and in vivo following severe acute respiratory  
syndrome coronavirus 2 infection**

Yannian Luo<sup>1</sup>, Juan Xu<sup>1</sup>, Mingzhen Zhou<sup>1</sup>, Xiaomei Lei<sup>2</sup>, Wen Cao<sup>1</sup>, Xiaojun Wang<sup>2</sup>, Ruili Qiao<sup>3\*</sup>

<sup>1</sup> Lanzhou University Second Hospital. No. 82, Cuiyingmen, Chengguan District, Lanzhou,  
Gansu, China.

<sup>2</sup> Gansu Provincial Hospital. No. 204, Donggang West Road, Chengguan District, Lanzhou,  
Gansu, China.

<sup>3</sup> Lanzhou University First Hospital. No. 1, Donggang West Road, Chengguan District, Lanzhou,  
Gansu, China.

\* Corresponding author:

Ruili Qiao

No. 1, Donggang West Road, Chengguan District, Lanzhou, Gansu, China.

E-mail address: qiaorl@lzu.edu.cn

**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.

**Keywords.** biomarkers; SARS-CoV-2; COVID-19; Gene Expression Omnibus; integrated bioinformatics; microarray analysis

**Introduction**

The coronavirus disease 2019 (COVID-19) epidemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is one of the major public health emergencies worldwide. As of June 16, 2020, more than 8,300,000 cases were confirmed, with 446,000 deaths and about 5.3% estimated mortality rate. The rapid increase in the number of patients, especially critically ill patients, has brought a big challenge to the public health. Generally, the infection with SARS-CoV-2 spreads 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 our focus<sup>1</sup>. Although, nonspecific gastrointestinal symptoms such as diarrhea is observed in several patients with infection<sup>2</sup>, majority of the patients have flu-like initial symptoms. Among the patient population, smokers, and male above 65 years of age are generally at a greater risk of disease progression that involves critical or lethal conditions such as severe pneumonia, acute respiratory distress syndrome (ARDS), multiple organ failure, and even death<sup>3</sup>. Based on the data published from the Chinese Center for Disease Control and Prevention, more than 80% of patients have mild conditions, and less than 20% have severe conditions. Alarmingly, the mortality rate was nearly 50% in patients with severe conditions<sup>4</sup>. According to the current evidence, the pathogenesis of the SARS-CoV-2 infection is associated with dysregulated inflammatory response, tissue injury, and multiple organ dysfunction; since the details of infection remain obscure, it is necessary to explore at the gene level to know how the disease exactly progresses from mild to severe conditions. This is not only warrant to help understand disease progression, but also provides new perspective for screening effective antiviral drugs, finding new therapeutic targets, and evaluating the risk of systemic inflammatory response syndrome. Hence, we designed this study of integrated bioinformatics

analysis based on the Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>) database<sup>5,6</sup>, using gene expression data of lung tissues from healthy individuals and patients with infection, with a view to discover valuable mRNA biomarkers, potential pathogenic mechanisms, and finally to guide clinical diagnosis and treatment.

## **Methods**

### **Microarray data**

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<sup>7</sup> (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE147507>) data sets from GEO database.

In case of in vivo level exploration, we extracted the gene expression data of adults from GSE148815 and GSE150316 data sets labeling as healthy samples and SARS-CoV-2 infected samples, respectively. Then, we integrated two sets of raw-counts data to compose the patient-microarray data used in the study. And in case of in vitro exploration, 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 chips 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/>): 1) standardizing microarray data utilizing DESeq2 package<sup>8</sup>, 2) log<sub>2</sub> transformation, 3) handling missing or unknown values, 4) checking and adjusting batch effects, 5) correction of value of log fold change (logFC) to reduce the error caused by a gene with a constant expression, such as zero. The detailed methods of data

processing are described in Text S1. For the microarray data were downloaded from a public database, patient consent or ethics committee approval was not necessary.

### **Differential analysis**

Differential analysis was performed on the three microarray data using the DESeq2 package to identify the differentially expressed genes (DEGs) between the samples with SARS-CoV-2 infection and without infection. As a result, logFC and statistical p values were obtained. When  $\log_{2}FC > 0$ , that the gene is highly expressed in the SARS-CoV-2 infected group, and when  $\log_{2}FC < 0$  the gene is expressed in a low level. The result was visualized as a heatmap with hierarchical clustering using the pheatmap package (<https://CRAN.R-project.org/package=pheatmap>).

### **Gene set enrichment analysis (GSEA) and identification of core genes**

GSEA, based on functional class scoring methods, was performed with the results of DEGs using phenotype labels “SARS-CoV-2 infected” versus “non-infected” by the clusterProfiler package<sup>9</sup> to elucidate relevant biological significance. First, we converted the gene symbols to Entrez IDs, and then analyzed by comparing it with the data of hallmark gene set (Version 7.1)<sup>10</sup> downloaded from the Molecular Signatures Database (MSigDB, <https://www.gsea-msigdb.org/gsea/msigdb/>)<sup>11</sup>. The results with p value  $< 0.05$  are plotted as a bubble chart. After that, we obtained co-enriched hallmark gene sets and their co-core genes from the GSEA results of the three microarray datasets.

### **Protein-protein interaction (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”, all active interaction sources were used, minimum required interaction score was selected as “highest

confidence (0.4500)", and finally display simplification was to hide disconnected nodes in the network<sup>12,13</sup>. The PPI data were downloaded and key molecular modules were identified using MCODE plug-in Cytoscape software (version 3.7.1, <https://cytoscape.org/>). 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"<sup>14</sup>.

### **Kyoto Encyclopedia of Genes and Genomes (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<sup>9</sup> package. An adjusted value of  $p < 0.05$  was considered to be statistically significant, and the results were visualized by GOplot package. In addition, we used data of patients to evaluate the expression levels of genes from the modules in patients with SARS-CoV-2 infection compared to healthy individuals, and the results are shown as a box plot.

## **Results**

### **Microarray data**

The data processing workflow is illustrated in Figure 1. The results corresponding to data processing are explained in Text S1, which indicated 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 represented in red and low-level expressed genes represented in green. For this outcome, we found that there was a significant difference in gene

expression levels between samples with SARS-CoV-2 infection and without infection, which also provided a foundation for us to further explore the genetic changes.

### **GSEA and identification of core genes**

The GSEA results between samples with SARS-CoV-2 infection and without infection are shown in Figure 3A-C. In case of in vivo level gene alteration analysis from the data of patients (Figure 3A), 29 hallmark gene sets were enriched, covering 14 activated and 15 suppressed gene sets, whereas at the 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 are 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 data of patients, 293 in NHBE data, and 438 in A549 data), and obtained 56 co-core genes (Figure 3D), which are described in Table S1. From these findings, we found that in samples with SARS-CoV-2 infection, inflammation-related hallmark gene sets such as IL6-Jak-Stat3-signaling, IL2-Stat5-signaling, complement, coagulation, KRAS-signaling-up, and TNF  $\alpha$  - signaling were activated while cell cycle related hallmark gene set such as E2F-targets was suppressed. Additionally, we also found that inflammation-related genes such as 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 using the 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 1, Figure 4C as module 2). To our surprise, nearly one third of the co-core genes were members of module 1, suggesting that the module plays a crucial role in the PPI network, which requires special attention.

### **KEGG pathway enrichment analysis**

50 of the KEGG pathways were enriched by the genes from module 1 (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 such as TNF signaling pathway, IL-17 signaling pathway, PI3K-Akt signaling pathway, and Jak-Stat signaling pathway (Figure 5A). For genes from module 2, four pathways (Figure 5B) related to cell cycle and DNA replication were identified. Finally, with the results of gene expression level (Figure 5C), we found that, for patients with SARS-CoV-2 infection, genes from module 1 showed a relatively high level of expression in contrast to genes from module 2. It should be noted specifically that our data source was from patients who died due to SARS-CoV-2 infection. These results signified that high and low expression of genes in module 1 and 2, respectively, might signal severe negative condition in patients with SARS-CoV-2 infection.

### **Discussion**

As obligate intracellular parasites, viruses must rely on host cells for replication, and hence their infection causes 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<sup>15</sup>. Severe COVID-19 is characterized by respiratory failure caused by hyper-inflammation, which causes great difficulties in clinical treatment leading to increased mortality<sup>16</sup>. It is now well established that a cytokine storm syndrome, involving molecules such as interleukin 6 (IL6), interleukin 8 (IL8), E-cadherin, MCP-1, and VEGF, may be the reason for hyper-inflammation in severe cases<sup>2,17</sup>. Investigating differences at a genetic level during disease occurrence and prognosis helps us understand the disease more clearly thereby guiding the development of countermeasures. In the present study, candidate datasets consisted of two parts, in vivo level (data of patients) and in vitro level (NHBE data and A549 data), and they were annotated using the same platform as 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 such as IL6-Jak-Stat3-signaling, IL2-Stat5-signaling, complement, coagulation, KRAS-signaling-up, and TNF  $\alpha$  -signaling were activated while the cell cycle related hallmark gene set such as E2F-targets was suppressed in both in vitro and in vivo levels. 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 a PPI network for the co-core genes, and analyzed molecular modules through MCODE, and identified two key molecular modules. Module 1 consisted of 17 genes: BCL2A1, BCL2L1, SOD2, PTGS2, TNF, PLAUR, MYC, LIF, VEGFA, LOX, IL6, CSF1, C3, TLR2, ICAM1, IL1B, and TNFAIP3, which are mainly involved in the inflammatory response pathway

and complement system cascade, such as TNF signaling, IL17 signaling, PI3K-Akt signaling, Jak-Stat signaling, NOD-like receptor signaling, Toll-like-receptor signaling, and complement signaling. Module 2 consisted of nine genes: EZH2, PTTG1, DLGAP5, RFC3, DUT, RPA2, CCNB2, PCNA, and MCM4, which were mainly involved in cell cycle and DNA replication signaling. Finally, by assessing the level of gene expression in the data of patient, we found that genes from module 1 showed a relatively high level expression in contrast to genes from modules 2. These findings may provide novel evidence for the selection of effective mRNA biomarkers to evaluate disease progression and predict prognosis for patients with SARS-CoV-2 infection. In addition, 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 pathogenesis of virus infecting the host, which often requires the participation of multiple cells and cellular components, and lead to changes in the level of gene expression. This is no exception for SARS-CoV-2. GSEA is capable of solving biological problems by focusing on gene sets rather than individual genes<sup>11</sup>. In our study, we use this method to analyze 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<sup>18</sup>. For hallmark gene sets, we found gene sets involved in significant molecules in the interleukin superfamily, 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 reported that IL6, as one of the genes involved in

cytokine release syndrome (CRS), was used to assess the severity of COVID 19 like respiratory failure, ARDS, and adverse clinical outcomes<sup>19-24</sup>. We also identified from cell line data and cases died due to SARS-CoV-2 infection that IL6 gene was not only highly expressed but also interacted the most with other genes in the PPI network, which confirmed its importance 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 is important for the proliferation of T and B lymphocytes<sup>25,26</sup>. Here, even though IL2 was not observed in the core gene, but we found activated IL2-Stat5-signaling in the infected samples, and IL2-related genes such as CSF1 and LIF were among the co-core genes with a high level of expression, which revealed that the action of IL2 might be as a “homeostatic” cytokine, and is 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 conditions for patients with COVID 19. Notably, the gene set of epithelial mesenchymal transition (EMT) (genes defining EMT) was represented in the 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 progresses. PIF was also an important feature of COVID 19 cases with poor prognosis, as previously reported<sup>27,28</sup>. Meanwhile, we should be alert to the possibility of sequela caused by EMT-activating. 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 to be activated in our study, which indicated that the disease evolution of SARS-CoV-2 infection was a multiple pathway, complicated process, comprising various dynamic changes in the genome.

E2F transcription factors play critical roles in the control of transcription, cell cycle and apoptosis<sup>29-31</sup>. For samples with SARS-CoV-2 infection, we found that the gene set of E2F targets was suppressed, 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 humans<sup>32</sup>, because hypoxia could directly affect cell differentiation and energy metabolism<sup>33</sup>.

From the gene level, we identified that genes such as PCNA, CCNB2, and PTTG1, were at a low level in the samples of SARS-CoV-2 infection. According to the literature, PCNA and PTTG1 are involved in regulating the biological function of p21, which is a CDK inhibitor that triggers cell-cycle arrest in the G1 and G2 phases<sup>34,35</sup>. CCNB2 is involved in the formation of the CCNB2/CDK1 complex that controls separate activity through inhibition of phosphorylation and regulates the biological process of the G2 phase of the cell cycle<sup>36</sup>. 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 encode 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<sup>37</sup>, and regulating IL6/Akt/Stat3/NF- $\kappa$ B-dependent pathway<sup>38</sup>. For roles in viral infection, ICAM1 had been determined to involve in DC-mediated HIV-1 transmission to CD4(+) T cells<sup>39</sup> and regulate interferon-gamma and IL17 in hepatitis B virus

infection<sup>40</sup>. Our results showed that ICAM1 participated in the signaling pathways of NF- $\kappa$ B, IL17, and TNF, which was in line with similar previous studies. TNFAIP3, induced by TNF, plays a role in inhibiting the activation of NF- $\kappa$ B and TNF-mediated apoptosis. Literature data reported that TNFAIP3 was closely associated with the replication of viruses such as influenza A<sup>41</sup>, and hepatitis B virus<sup>42</sup>. In our study, we identified that TNFAIP3, TNF, and IL1B were involved in the signaling pathway of necroptosis. Similar to apoptosis, necroptosis was executed via a distinctive signaling mechanism comprising a cascade of specified proteins, resulting in regulated necrotic cell death<sup>43</sup>. Physiologically, necroptosis induces an innate immune response as well as premature assembly of viral particles in cells infected with virus that abrogates host apoptotic machinery, which is advantageous for the host. On the other hand, necroptosis is also deleterious because it can cause various diseases such as sepsis, neurodegenerative diseases and ischemic reperfusion injury<sup>43</sup>. 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<sup>44</sup>. However, the benefits of necroptosis to the host may sometimes be outweighed by the potentially deleterious hyper-inflammatory consequences of activating this death modality in pulmonary and other tissues<sup>45,46</sup>. All these evidences suggest that necroptosis signaling is a double-edged sword providing defense against microbial infection, or might even be the likely culprit that increases mortality of the SARS-CoV-2 infected cases. This also provided a novel clue for necroptosis signaling in the treatment of COVID 19 and serves as a potential therapeutic target. BCL2A1, a pro-inflammatory cytokine, encode a member of the BCL-2 protein family. In our study, we found that BCL2A1 participated in the signaling of NF- $\kappa$ B and apoptosis, which was similar to previous reports in the

literature<sup>47,48</sup>. Nevertheless, more details of how BCL2A1 affects the pathogenesis of SARS-CoV-2 infection still needs 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 believe 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 needs to be verified through in vivo data.

## **Conclusions**

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

## **Funding**

The authors received no funding for this work.

## **Competing Interests**

The authors declare that they have no competing interests.

## Author Contributions

Contributions: (I) Conception and design: R Qiao, Y Luo; (II) Administrative support: R Qiao; (III) Provision of study materials or patients: Y Luo, J Xu; (IV) Collection and assembly of data: M Zhou, M Lei; (V) Data analysis and interpretation: Y Luo, W Cao, X Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

## 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. And an earlier version of this work has been released as a preprint at <https://www.researchsquare.com/article/rs-37567/v1>.

## Data Availability

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

## References

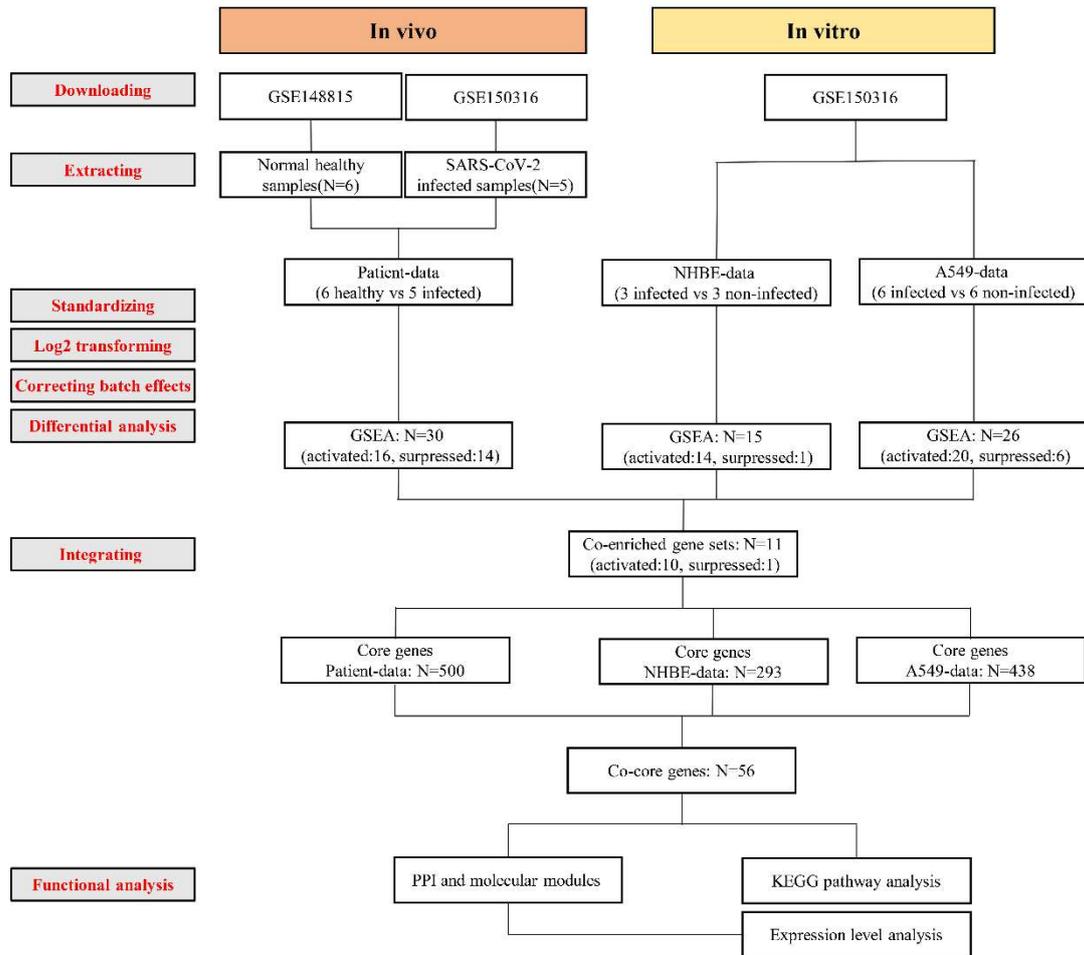
- 1 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273, doi:10.1038/s41586-020-2012-7 (2020).
- 2 Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet (London, England)* **395**, 497-506, doi:10.1016/s0140-6736(20)30183-5 (2020).
- 3 Li, J. Y. *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, Z. & 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, T. *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, R., 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, W. & 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, G., 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, A. *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 U S A* **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 datasets. *Nucleic Acids Res* **47**, D607-d613, doi:10.1093/nar/gky1131 (2019).
  - 13 Feng, Z. *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 analysis. *Genome Biol* **20**, 185, doi:10.1186/s13059-019-1758-4 (2019).
  - 15 Helmy, Y. 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, J. 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, Z. 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, G. *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, K., 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, E. 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.cytogfr.2020.05.009 (2020).
  - 22 Liu, B., Li, M., Zhou, Z., Guan, X. & Xiang, Y. 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, Z., 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).

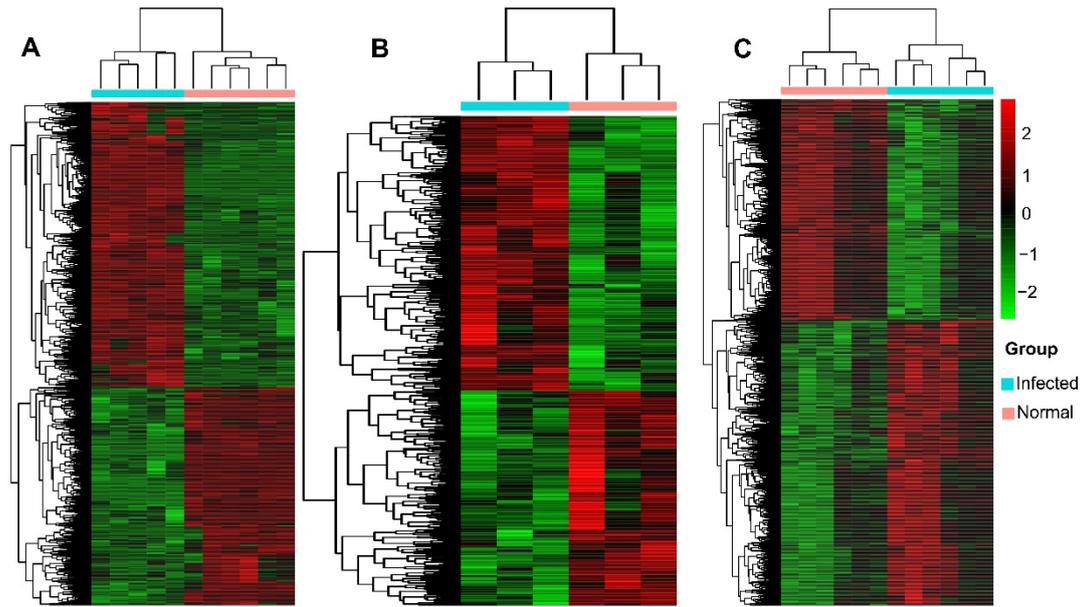
- 24 Choy, E. H. *et al.* Translating IL-6 biology into effective treatments. *Nature reviews. Rheumatology* **16**, 335-345, doi:10.1038/s41584-020-0419-z (2020).
- 25 Zhang, Y., Yin, Y., Zhang, S., Luo, H. & Zhang, H. HIV-1 Infection-Induced Suppression of the Let-7i/IL-2 Axis Contributes to CD4(+) T Cell Death. *Scientific reports* **6**, 25341, doi:10.1038/srep25341 (2016).
- 26 Henneman, P. *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, P. 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, P. *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, G. 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, S. 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, Y. & 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, Y. 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. W. & 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. W. *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, J. 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, S., 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, Y. 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 Infection. *BioMed research international* **2019**, 2121357, doi:10.1155/2019/2121357 (2019).
- 45 Balachandran, S. & Rall, G. F. 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, S., 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, M. 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).

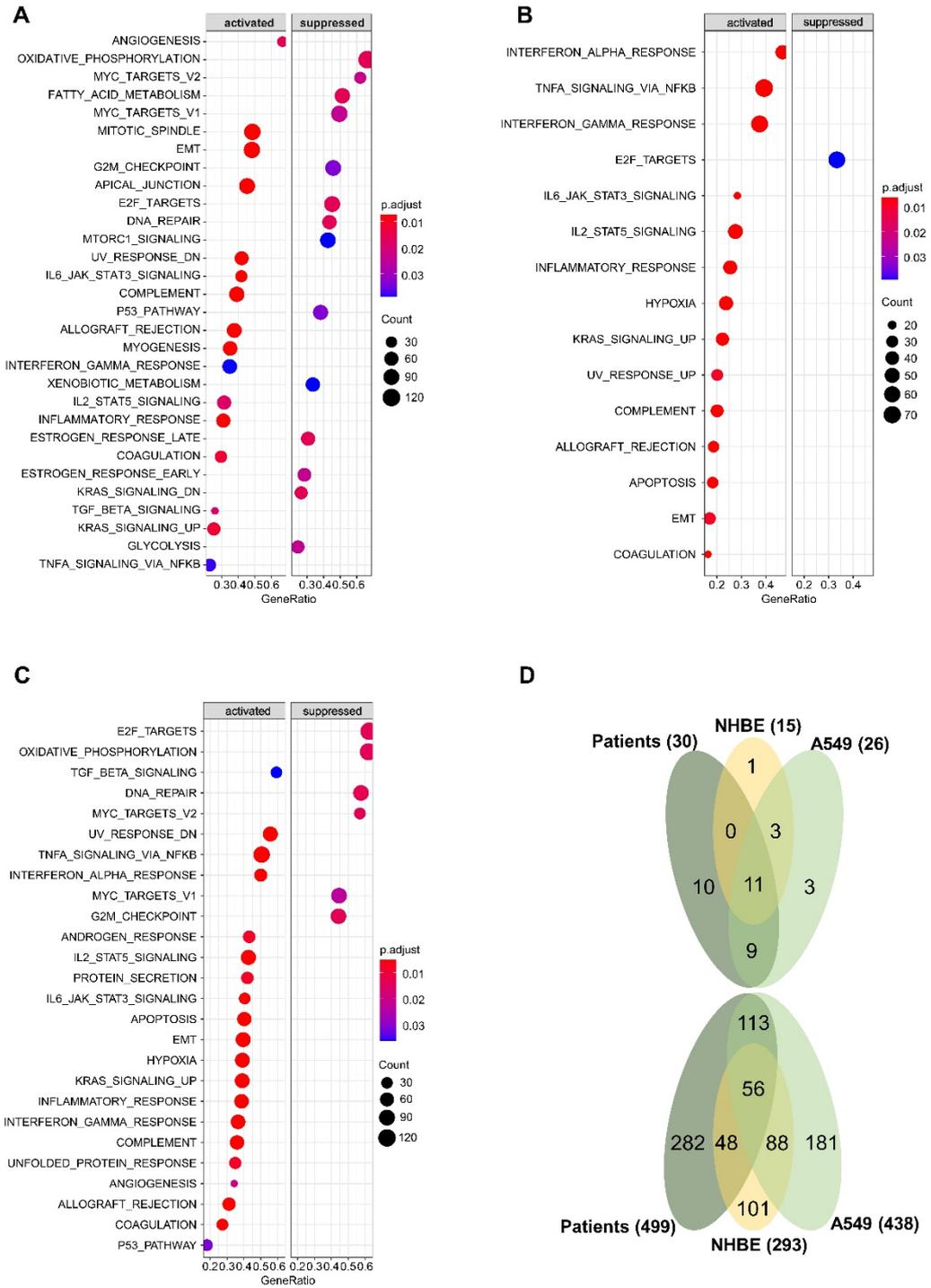
Figures:



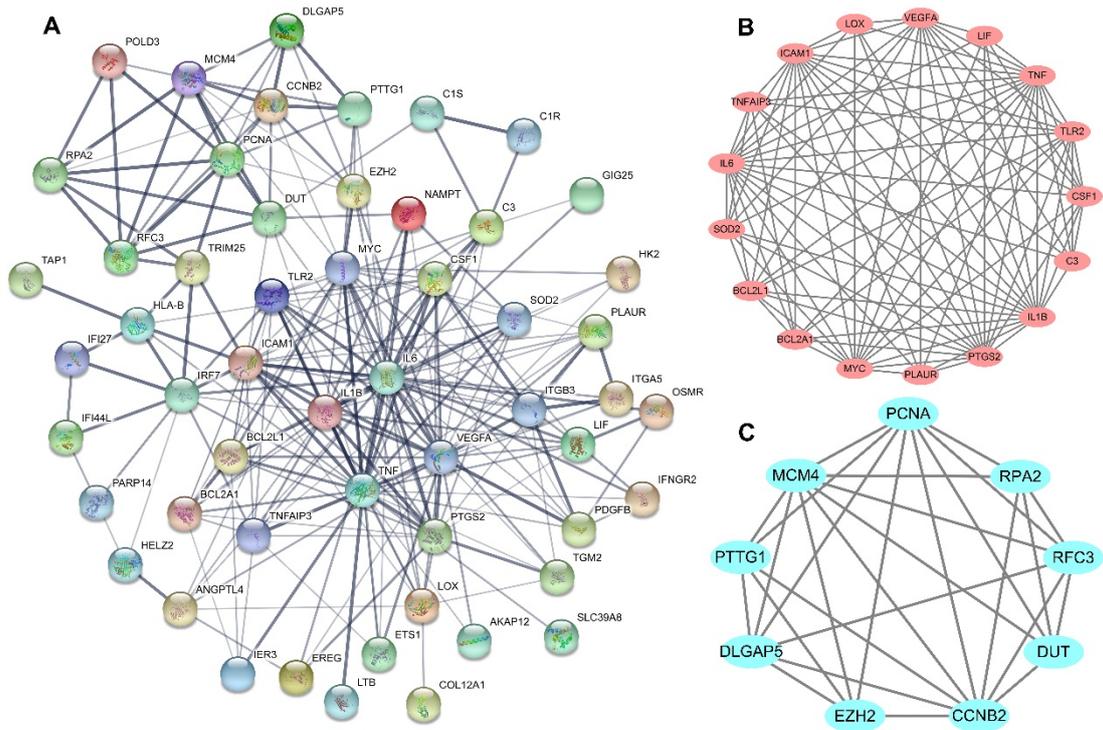
**Figure 1** The data processing workflow. N, number. NHBE, normal human bronchial epithelial cell line. A549, lung adenocarcinoma cell line A549. vs, versus. GSEA, gene set enrichment analysis. PPI, protein-protein interaction. KEGG, kyoto encyclopedia of genes and genomes.



**Figure 2** 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 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 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.





Description of hallmark gene set	Patient-data			NHBE-data			A549-data		
	Size	NES	pvalue	Size	NES	pvalue	Size	NES	pvalue
ALLOGRAFT_REJECTION	189	1.67	0.001	150	2.02	0.001	157	2.38	0.001
COAGULATION	132	1.77	0.003	116	1.79	0.001	121	1.97	0.001
COMPLEMENT	197	1.70	0.001	174	2.09	0.001	178	2.18	0.001
E2F_TARGETS	195	-2.03	0.006	195	-1.44	0.012	194	-2.26	0.006
EMT	197	2.20	0.001	188	1.70	0.002	187	2.23	0.001
IL2_STAT5_SIGNALING	194	1.46	0.007	174	1.74	0.001	178	2.02	0.001
IL6_JAK_STAT3_SIGNALING	86	1.81	0.001	67	2.12	0.001	79	2.44	0.001
INFLAMMATORY_RESPONSE	197	1.94	0.001	165	2.26	0.001	176	2.66	0.001
INTERFERON_GAMMA_RESPONSE	198	1.36	0.022	179	2.35	0.001	186	2.90	0.001
KRAS_SIGNALING_UP	190	1.55	0.002	166	2.19	0.001	182	2.32	0.001
TNFA_SIGNALING_VIA_NFKB	198	1.36	0.020	191	2.43	0.001	196	2.93	0.001

**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

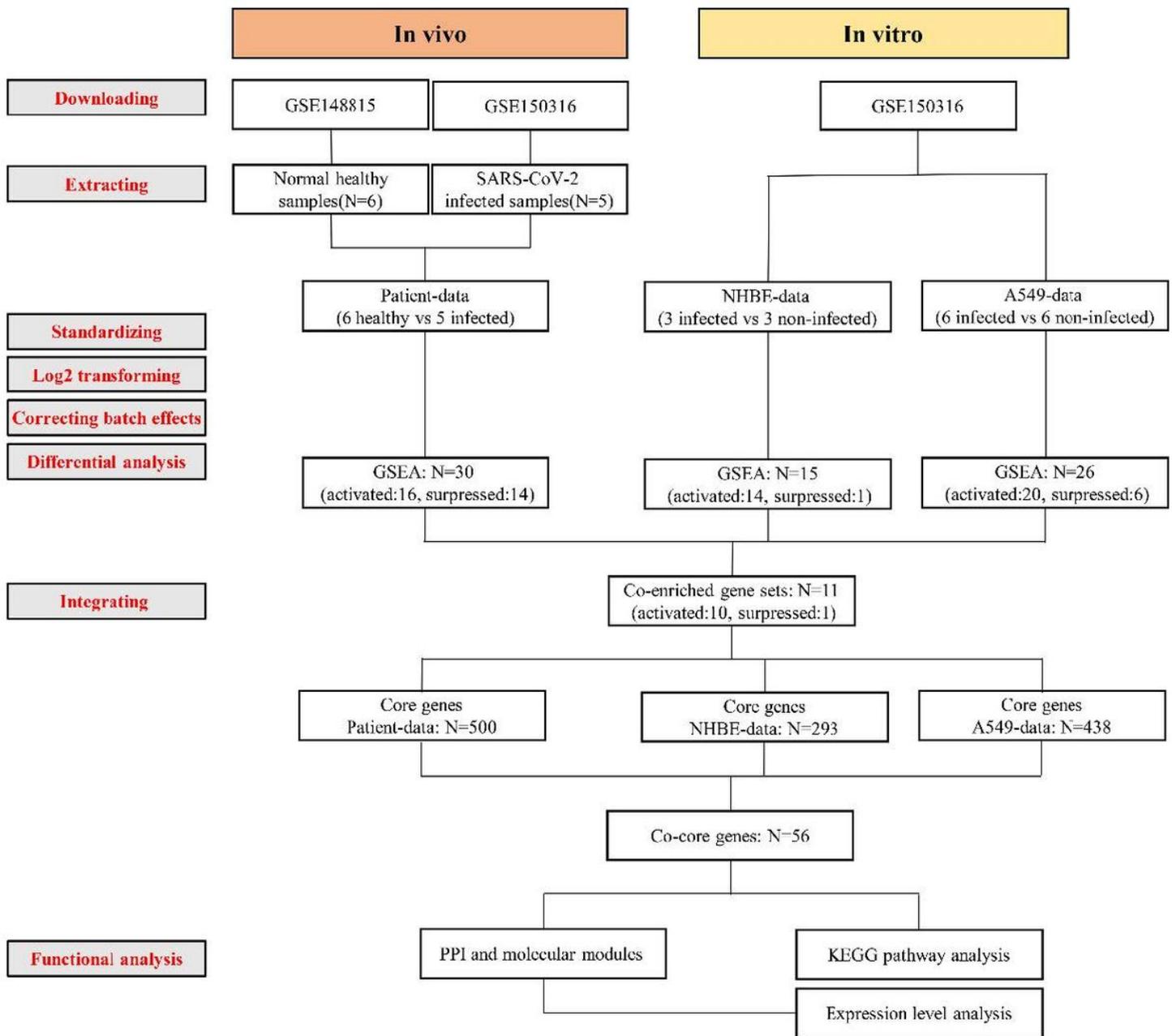
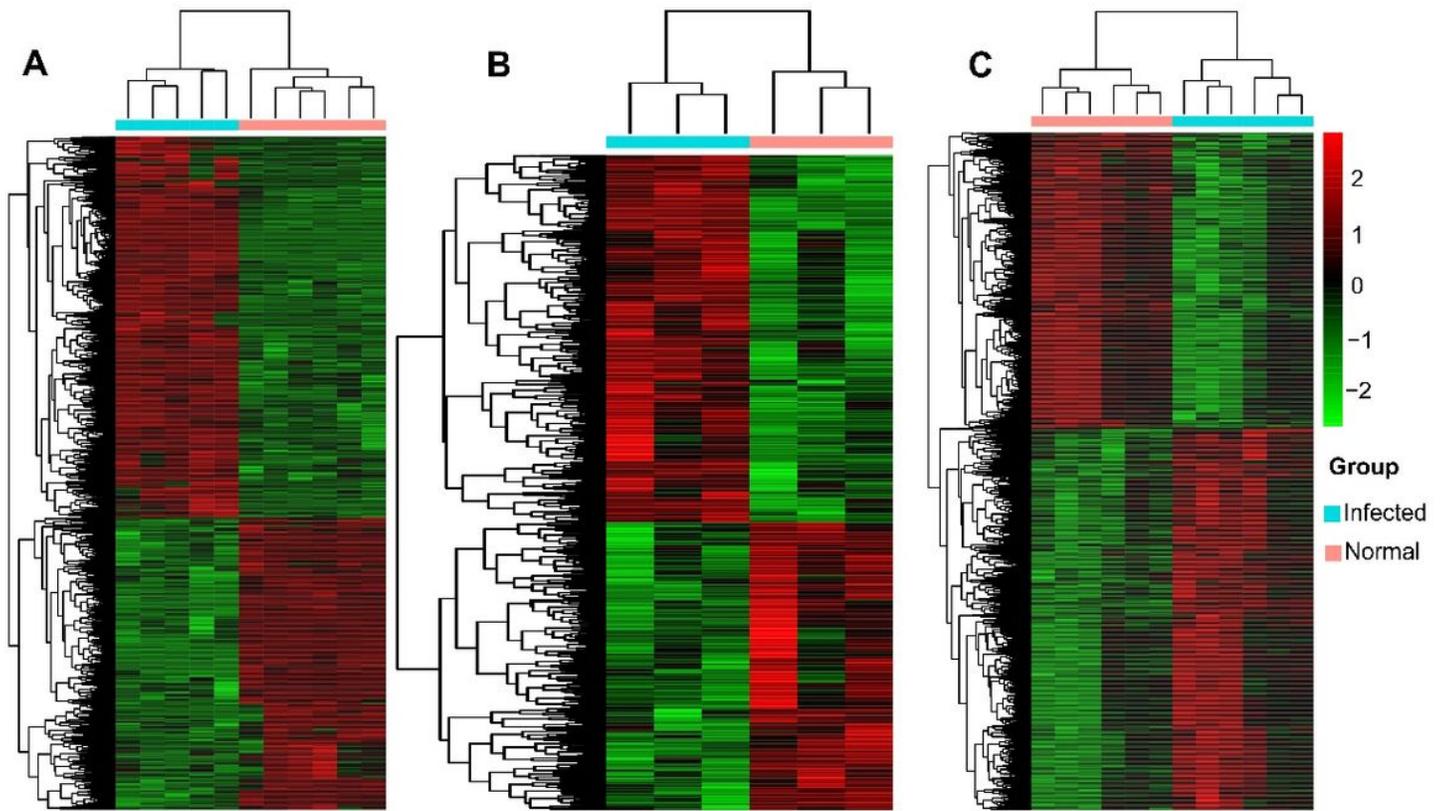


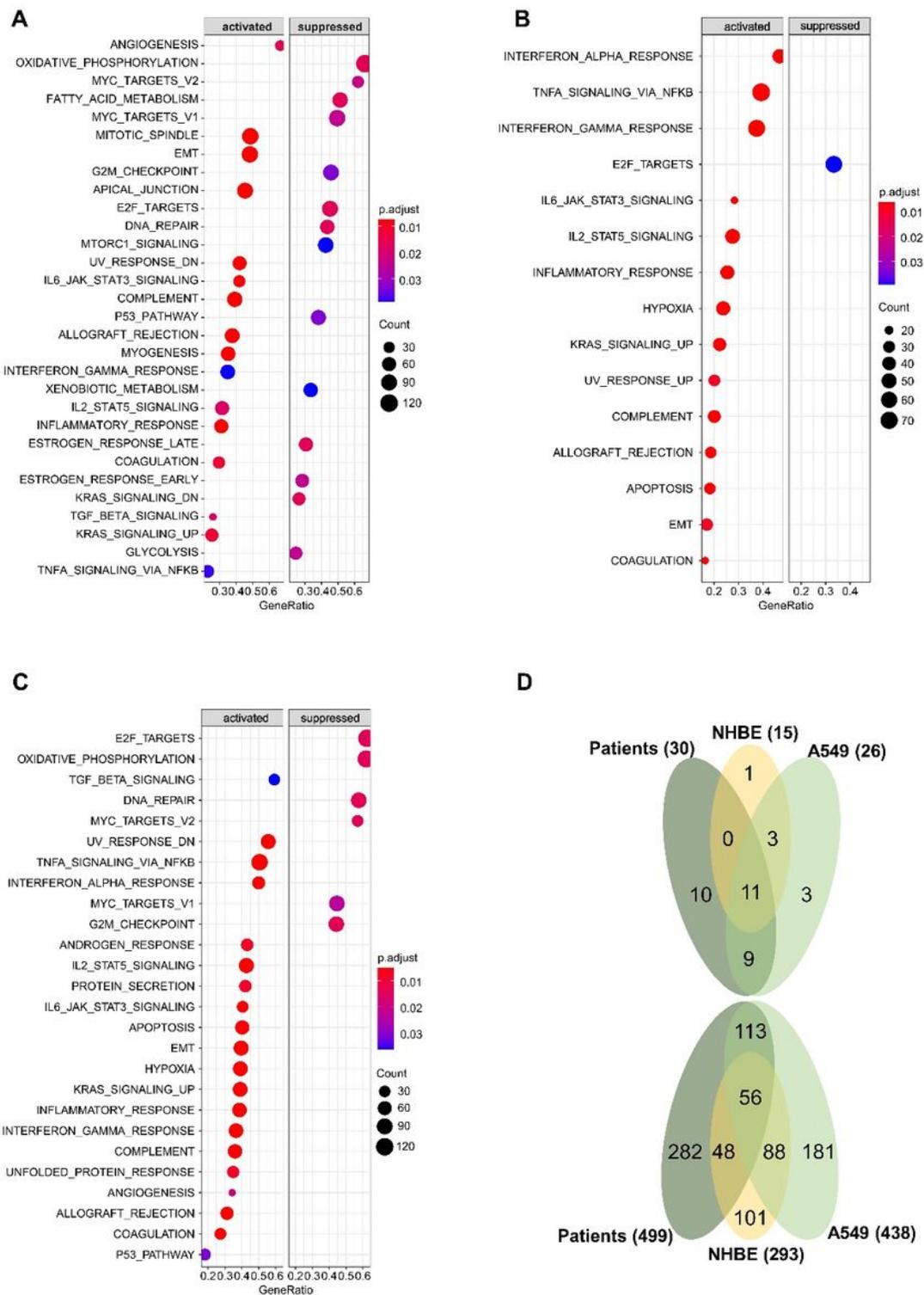
Figure 1

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.



**Figure 2**

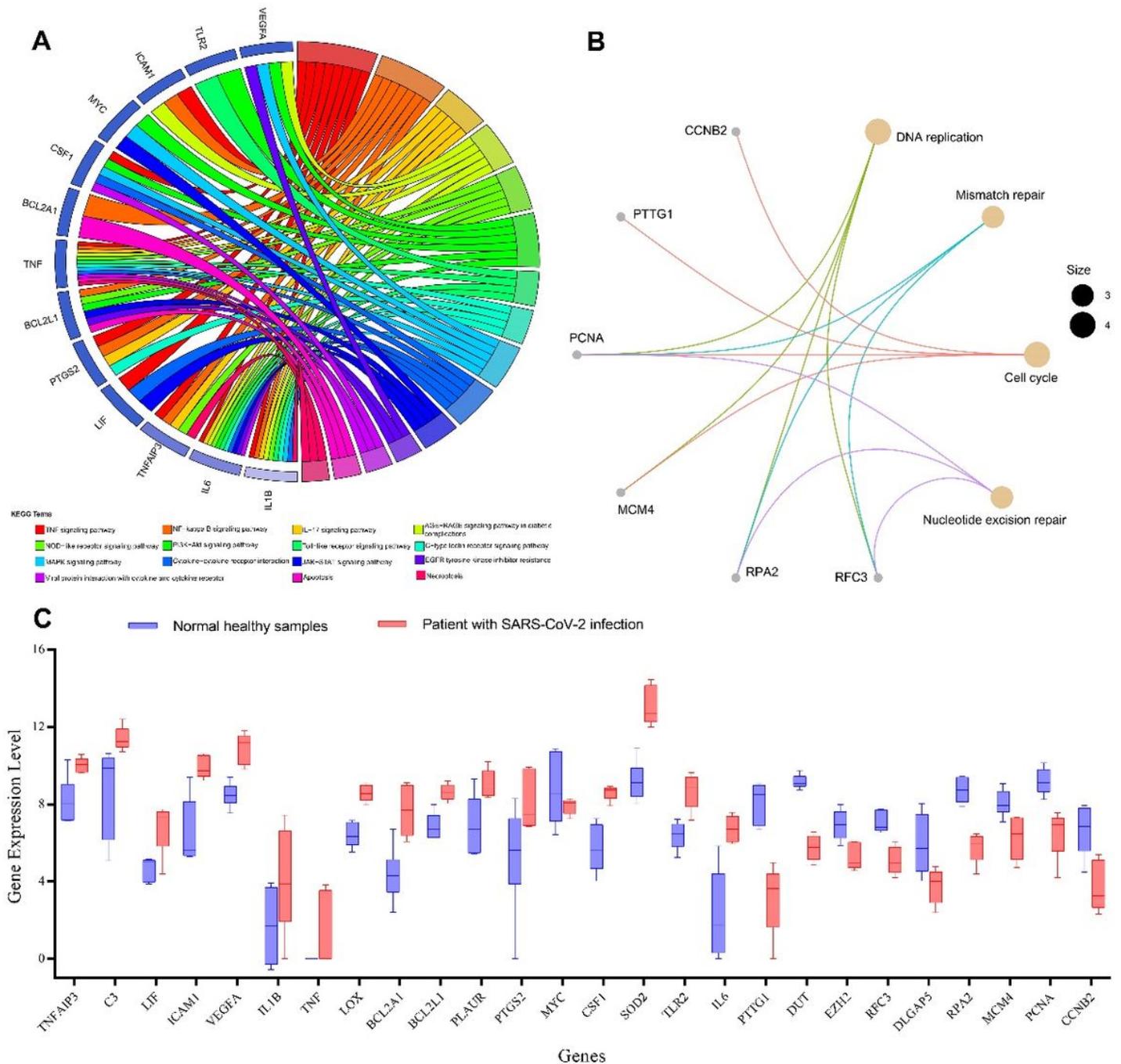
The cluster heat map of the three microarray data. (A) is for patient-data, (B) is for NHBE-data, (C) is for A549-data. Red represents gene expression with a high-level, green represents gene expression with a low-level, and black represents gene expression with an intermediate-level.



**Figure 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 5**

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 (TNFAIP3, C3, LIF, ICAM1, VEGFA, IL1B, TNF, LOX, BCL2A1, BCL2L1, PLAUR, PTGS2, MYC, CSF1, SOD2, TLR2, and IL6) and module two (PTTG1, DUT, EZH2, RFC3, DLGAP5, RPA2, MCM4, PCNA, and CCNB2) in patient-data.

## Supplementary Files

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

- [TextS1Detaileddataprocessinginformation.pdf](#)
- [TableS1.docx](#)
- [TableS2.docx](#)