Identification and Ontology Term Enrichment Analysis of Genes Associated with COVID-19 and Acute Kidney Disease

Guirui Huang[1,2], Suyuan Peng[1,3], Luxia Zhang[1,4,5], Yongqun He[6]

1 National Institute of Health Data Science, Peking University, Beijing 100191, China
2 Department of Respiratory, The Third Affiliated Hospital, Beijing University of Chinese Medicine, Beijing 100029, China
3 School of Public Health, Peking University, Beijing 100191, China
4 Department of Medicine, Renal Division, Peking University First Hospital, Peking University Institute of Nephrology, Beijing 100034, China
5 Advanced Institute of Information Technology, Peking University, Hangzhou 311215, China
6 University of Michigan Medical School, Ann Arbor, MI 48109, USA

Abstract

Acute kidney injury (AKI) is the main comorbidity of COVID-19, and the pathogenesis remains unclear. This study first performed a gene set enrichment analysis of 6 AKI-related Gene Expression Omnibus (GEO) studies and identified 3,876 AKI-associated genes. By incorporating COVID-19 related interactions from BioGRID, we further found 1,027 genes associated with both COVID-19 and AKI. Our Gene ontology (GO) enrichment analysis of these genes showed that viral and inflammation-related biological processes played important roles on COVID-19 related AKI. Furthermore, the COVID-19 pathways ranked second in the top 5 KEGG-enriched pathways, in which 66 enriched genes were all up-regulated in the kidney tissue of the above 6 GEO studies. Ontology modeling is currently undergoing to systematically and logically represent the AKI pathogenesis process in COVID-19 patients.

Keywords
COVID-19, Acute Kidney Injury, Gene Ontology

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first described in December 2019 and is responsible for the coronavirus disease 2019 (COVID-19) pandemic. COVID-19 was initially characterized as a febrile respiratory disease, and increasing evidence has shown that it can also result in several extrapulmonary manifestations [1].

Acute kidney injury (AKI) is recognized as a common complication of COVID-19, and it is also an independent risk factor for all-cause mortality of COVID-19 inpatients [2-4]. The initial reports from China suggested relatively low rates (0.5%-15%) of kidney involvement and high rates of hematuria and proteinuria in COVID-19 patients [5-8]. The subsequent reports from the USA and Europe indicate much higher rates of AKI, particularly in the intensive care setting, with up to 45% of patients in the intensive care unit (ICU) requiring kidney replacement therapy (KRT).

The long-term follow-up data from hospitalized patients with COVID-19 study showed that 13% of patients with normal renal function during hospitalization showed renal insufficiency within six months

after discharge (estimated glomerular filtration rate, eGFR < 90 mL/min /1.73m2) [9]. In another cohort study of US patients, COVID-19-associated AKI was associated with a greater rate of eGFR decrease after discharge compared with AKI in patients without COVID-19 [10]. Since the evidence of renal tissue morphologic correlates are few and limited to patient reports or autopsy series, the pathogenesis of COVID-19-associated AKI remains unclear.

This study aimed to identify and analyze significantly up-regulated genes associated with COVID-19 associated AKI through bioinformatics data mining, followed by ontology-based functional term enrichment analyses and ontological modeling.

2. Materials and Methods

2.1 GEO Analysis of AKI-associated Genes

We obtained the gene expression datasets of AKI through the Gene Expression Omnibus (GEO) database. Datasets were included according to the following eligibility criteria: (1) Containing at least ten total samples; (2) Raw data or gene expression profiling by array was available in GEO; (3) Gene expression profiling was extracted from human renal tissue. We determined the differentially expressed genes（DEGs) between acute kidney injury tissues and normal control kidney tissues. The significance of differential expression for each microarray was performed by the "limma" package in R. The |log2fold change (FC)| > 0.5 and P-value < 0.05 were regarded as the cut-off criteria to determine DEGs. The DEGs were combined, and duplicate genes were deleted to obtain all the AKI-associated genes.

2.2 Detection of AKI and COVID-19 Associated Genes

We obtained COVID-19 related genes from Biological General Repository for Interaction Datasets (BioGRID)[11]. The COVID-19 Coronavirus Curation Project in BioGRID provides comprehensive datasets of curated interactions for the viral proteins encoded by SARS-CoV-2 and related coronaviruses, SARS-CoV and MERS-CoV. This database provides the 32 viral proteins and the different host interactors with which each of 32 proteins interacts. The intersection of AKI-associated genes obtained from GEO analysis above and COVID-19 associated genes were identified as the potential key gene set of COVID-19-associated AKI.

2.3 GO and KEGG Enrichment Analysis

Database for Annotation, Visualization, and Integrated Discovery (DAVID), a commonly used functional annotation tool, was used for Gene Ontology (GO) functional enrichment analysis and KEGG pathway analysis. We uploaded the COVID-19-associated AKI genes obtained from the above GEO and BioGRID analysis to investigate the potential functions. P-value <0.05 and false discovery rate (FDR) <0.05 were used as the cut-off criteria. Our domain experts further evaluated the analysis results, and COVID-19 and AKI-related terms were then selected for further mechanism study.

3. Results

3.1 3,876 AKI-associated Genes Identified

According to the previously established inclusion criteria for AKI-associated genes, six GEO studies, including GSE1563, GSE26578, GSE30718, GSE37838, GSE61739, and GSE141864, were included. In total, 90 AKI patients and 164 non-AKI controls participated in these studies. The "limma" package screened out the DEGs in R software according to the cut-off criteria. Our study identified 1,308 DEGs, 20 DEGs, 624 DEGs, 17 DEGs, 18 DEGs, and 2,302 DEGs from these six studies. By removing duplicate genes from these DEGs, we obtained 3,876 AKI-associated genes in total.

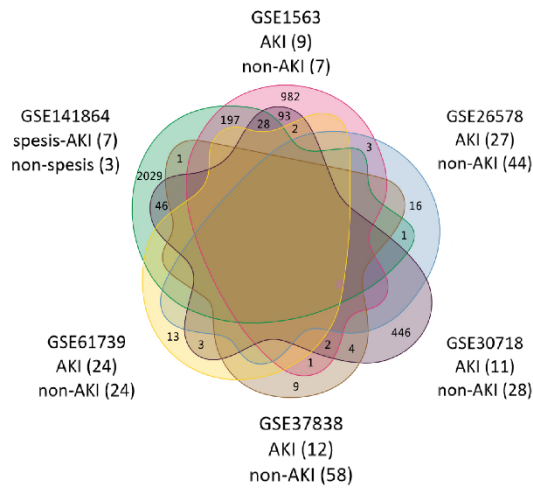Figure 1 provides a Venn Diagram analysis of these above-described DEGs and their association with AKI.

Figure 1: Venn Diagram analysis of AKI-associated genes from the DEGs in six GEO studies. AKI(9) means the AKI group contains nine human subjects, and non-AKI(7) means seven human subjects in the control group.

3.2   1,027 COVID-19 and AKI-associated Genes Identified

As of July 10, 2021, we obtained the complete SARS-CoV-2 and coronaviruses-related interactions on BioGRID, including 22,372 interactions between the 32 viral proteins. Figure 2 illustrates 5,075 COVID-19 associated genes (green area), 3,876 AKI-associated genes (red area). Finally, a total of 1,027 genes were found to be shared between COVID-19 associated genes as defined by BioGRID and AKI-associated genes as identified in our GEO data analysis.
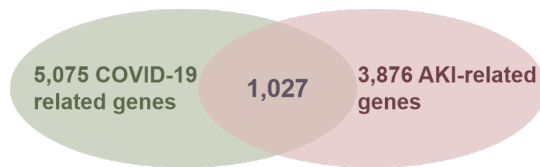

Figure 2: Intersection between COVID-19 associated genes from BioGRID (green area) and AKI-associated genes (red area) by analyzing expression profiling from the GEO database.

3.3   GO and KEGG Enrichment Analysis

We utilized the 1,027 COVID-19 associated AKI genes with performing the GO and KEGG analysis. The outcomes of GO analysis revealed that viral and inflammation-related biological processes such as viral gene expression and transcription and NF-kappaB signaling were significantly enriched (Table 1).

Table 1. The GO enrichment of the 1,027 COVID-19 and AKI-associated genes

ID	Description	Gene Count	*p*-value
GO:0019080	viral gene expression	62	3.52E-33
GO:0019083	viral transcription	57	2.22E-30
GO:0070498	interleukin-1-mediated signaling pathway	27	5.51E-13
GO:0060071	Wnt signaling pathway, planar cell polarity pathway	27	4.08E-12
GO:0002223	stimulatory C-type lectin receptor signaling pathway	25	4.30E-10
GO:0071347	cellular response to IL-1	32	5.62E-10
GO:0046718	viral entry into the host cell	25	1.97E-09
GO:0033209	tumor necrosis factor-mediated signaling pathway	29	7.15E-09
GO:0038061	NIK/NF-kappaB signaling	27	7.24E-07

GO:0002429	immune response-activating cell surface receptor signaling pathway	48	5.31E-06
GO:0002224	toll-like receptor signaling pathway	21	1.80E-05
GO:0060337	type I interferon signaling pathway	16	2.47E-05
GO:0034341	response to interferon-gamma	25	3.24E-05
GO:0070741	response to interleukin-6	11	3.64E-05
GO:0070102	interleukin-6-mediated signaling pathway	7	5.29E-04

KEGG pathway enrichment analysis was performed with the functional annotation tool DAVID. Our study found a total of 1,027 DEGs of COVID-19 associated AKI in kidney tissue, including 839 up-regulated and 188 down-regulated genes. Hierarchical clustering was performed for the 1,027 DEGs to identify the sets of co-regulated or functionally related genes. The resulting dendrogram was transformed to be suitable for a visualization with ggplot2 package. A circular layout was chosen, because it is not only effective but also visually appealing. The inner ring next to the dendrogram represents the logFC of the genes, which are actually the leaves of the clustering tree. The logFC values are colour- coded, which the up-regulated genes were set to red color, and the blue color represent the down-regulated genes. The outer ring represents the top 5 GO terms assigned to the genes. The terms are colour- coded as well. For example, the yellow part of the outer ring represents the COVID-19 pathways. The results showed that COVID-19 was an essential term of the top 5 enriched pathways, in which the 66 enriched genes were up-regulated in the kidney tissue from the above GSE studies, including NRP1, TNFR, gp130, Jak1, NFKβ, STAT3, MAPK, TAK1 and so on. The general profile is illustrated in Figure 3.
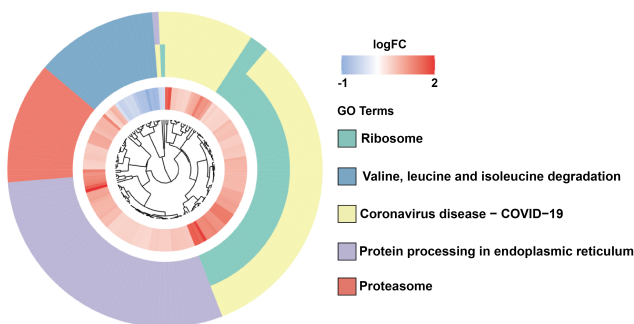

Figure 3: Top 5 KEGG enriched domains from the 1,027 COVID-19 and AKI-associated genes. The yellow part of the outside ring in Figure 3 represents the Coronavirus disease - COVID-19 pathways; The red part of the inside ring represents the up-regulated genes of the COVID-19 associated AKI, and the blue part of the inside ring represents the down-regulated genes of the COVID-19 associated AKI (839 up-regulated and 188 down-regulated).

4. Discussion

Acute respiratory distress syndrome and respiratory failure are the main manifestations of COVID-19, and kidney involvement is also common. Existing evidence supports several potential pathophysiological pathways through which AKI can develop in the context of SARS-CoV-2 infection. Histopathological findings have highlighted similarities and differences between AKI in patients with COVID-19 and those with AKI in non-COVID-related sepsis[12]. However, the potential pathophysiological mechanisms of COVID-19-associated AKI remain unclear. This study aimed to investigate the mechanism by analyzing the potential gene sets followed by bioinformatics mining of GEO gene expression studies and the COVID-19 associated interaction knowledge available in BioGRID. Our GO and KEGG enrichment analyses have generated promising results as described in this short paper.

Our study found a total of 1,027 DEGs of COVID-19 associated AKI in kidney tissue, including 839 up-regulated and 188 down-regulated genes. The top rank GO enrichment focused more on the up-regulation of viral-related biological processes (e.g., viral gene expression and transcription), which

were directly relevant to the whole life cycle of SARS-CoV-2 and the inflammation-related pathway. The inflammation-related pathway could induce acute kidney injury by activating the excessive immune response in cytokine storm and epithelial-mesenchymal transition.

Moreover, the KEGG enrichment revealed that 66 COVID-19 associated AKI genes were up-regulated in the currently known COVID-19 pathway (hsa05171; P-value = 1.01E-22). Many of these 66 genes appear to be critical to AKI pathogenesis. For example, NRP1, one of 66 up-regulated genes, is a receptor for SARS-CoV-2. Its up-regulation increased human beings' susceptibility. The occurrence of tissue injury in terms of fibrosis, inflammation, oxidation, and the downstream upregulation of NF-κB, likely cause the release of inflammatory cytokines, including TNFa, IL-6, IL-1β, IL-12, MMP-3, MMP-1, and IL-8, which form the key components of the cytokine storm. In addition, the up-regulation of TNFR, gp130, Jak1, NFKβ, STAT3, MAPK, and TAK1 also contribute to the cytokine storm through various molecular mechanisms. On the one hand, the cytokine storm over-activated inflammatory cells to release active mediators such as reactive oxygen species (ROS) and nitric oxide (NO). Furthermore, the overreaction of effector cells, such as T lymphocytes and NK cells, would induce acute tissue injury jointly, including AKI.

Altogether, the infection of SARS-CoV-2 may induce kidney tissue injury directly or indirectly by regulating the 1,027 potential genes identified in our data mining. Our integrated analysis concluded novel gene signatures and contributed to understanding comprehensive molecular changes in COVID-19 associated AKI. The future work would identify the DEGs on the different pathophysiological mechanisms of COVID-19-associated AKI.

Given the complexity of the AKI pathogenesis in COVID-19 patients, it is critical to systematically represent the complex condition-dependent molecular and cellular interactions using an integrative ontological approach. Currently, we are applying the Coronavirus Infectious Disease Ontology (CIDO) in our modeling of various molecular interactions given specific conditions[13]. More information about ontology-based knowledge representation will be introduced at the ICBO-2021 conference.

5. Acknowledgments

6. References

[1] A. Gupta et al., Extrapulmonary manifestations of COVID-19, Nat Med, vol. 26, no. 7, pp. 1017-1032, Jul 2020, doi: 10.1038/s41591-020-0968-3.

[2] J. H. Ng et al., Outcomes Among Patients Hospitalized With COVID-19 and Acute Kidney Injury, (in English), Am J Kidney Dis, vol. 77, no. 2, pp. 204-215 e1, Feb 2021, doi: 10.1053/j.ajkd.2020.09.002.

[3] L. Chan et al., AKI in Hospitalized Patients with COVID-19, J Am Soc Nephrol, vol. 32, no. 1, pp. 151-160, Jan 2021, doi: 10.1681/ASN.2020050615.

[4] C. Ronco, T. Reis, and F. Husain-Syed, Management of acute kidney injury in patients with COVID-19, Lancet Respir Med, vol. 8, no. 7, pp. 738-742, Jul 2020, doi: 10.1016/S2213-2600(20)30229-0.

[5] N. Chen et al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, Lancet, vol. 395, no. 10223, pp. 507-513, Feb 15 2020, doi: 10.1016/S0140-6736(20)30211-7.

[6] W. J. Guan et al., Clinical Characteristics of Coronavirus Disease 2019 in China, N Engl J Med, vol. 382, no. 18, pp. 1708-1720, Apr 30 2020, doi: 10.1056/NEJMoa2002032.

[7] D. Wang et al., Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China, Jama, vol. 323, no. 11, pp. 1061-1069, Mar 17 2020, doi: 10.1001/jama.2020.1585.

[8] F. Zhou et al., Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study, Lancet, vol. 395, no. 10229, pp. 1054-1062, Mar 28 2020, doi: 10.1016/S0140-6736(20)30566-3.

[9] C. Huang et al., 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study, The Lancet, 2021.

[10] J. Nugent et al., Assessment of Acute Kidney Injury and Longitudinal Kidney Function After Hospital Discharge Among Patients With and Without COVID-19, JAMA network open, vol. 4, no. 3, pp. e211095-e211095, 2021, doi: 10.1001/jamanetworkopen.2021.1095.

[11] R. Oughtred et al., The BioGRID interaction database: 2019 update, Nucleic acids research, vol. 47, no. D1, pp. D529-D541, 2019, doi: 10.1093/nar/gky1079.

[12] S. Peng et al., Early versus late acute kidney injury among patients with COVID-19-a multicenter study from Wuhan, China, Nephrol Dial Transplant, vol. 35, no. 12, pp. 2095-2102, Dec 4 2020, doi: 10.1093/ndt/gfaa288.

[13] Y. He et al., CIDO, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis, Sci Data, vol. 7, no. 1, p. 181, Jun 12 2020, doi: 10.1038/s41597-020-0523-6.