RNA Sequencing Analysis of Peripheral Blood Mononuclear Cells Reveals a Novel Pathway Associated with COVID-19 Deterioration

Mami Ikemachi  
Kobe City Nishi-Kobe Medical Center

Rina Ogura  
Kobe City Nishi-Kobe Medical Center

Kazuya Miyagawa (✉ kazuya_miyagawa@kcho.jp)  
Kobe City Nishi-Kobe Medical Center

Article

Keywords:

Posted Date: July 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1849733/v1

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

The severity of COVID-19 depends on the host immune cell response. Most patients with COVID-19 had mild disease, but up to 30% of patients developed a severe disease. Risk factors associated with severe diseases such as age, smoking, and comorbidities are known, while molecular signatures that predict the disease progression are still emerging. Here, we investigated gene expression profiles in peripheral blood mononuclear cells (PBMCs) from patients with COVID-19 using RNA-sequencing analysis (RNA-seq) based on the hypothesis that deteriorating patients have a unique molecular signature in PBMCs at an early stage of the disease. Twelve patients were included in RNA-seq (deteriorated vs. nondeteriorated, each n = 6). RNA-seq demonstrated 156 differentially expressed genes, and pathway analysis showed that "Nucleosome assembly" was the most important pathway, with 24 HIST1 cluster genes upregulated in the deteriorated group. In silico prediction of upstream transcription factors exhibited protein inhibitors of activated STAT 4 (PIAS4) as the most significant. TP53, STAT1, and IKBKG associated with COVID-19 pathophysiology are predicted to interact with PIAS4. Our results show that elevated HIST1 cluster gene expression predicts COVID-19 deterioration, which may induce histone-mediated lung injury and coagulopathy. The molecular mechanism could involve PIAS4-mediated transcriptional regulation and inflammatory signal transduction.

Introduction

The coronavirus disease 2019 (COVID-19) pandemic is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which spread rapidly and globally, killing 5.8 million people in two years. Up to 80% of SARS-CoV-2 infections are asymptomatic or mild, but 15–30% of the patients, especially the elders, develop severe diseases that include pneumonia, respiratory failure, and death\(^1,2\). Additionally, thrombosis, cerebral infarction, and multiple organ failure are frequent complications\(^2-4\).

The severity of COVID-19 depends on the host immune cell response and dysregulation\(^5\). Risk factors associated with severe diseases such as age, cancer, chronic obstructive pulmonary disease, chronic kidney disease, diabetes mellitus, hypertension, hyperlipemia, obesity, and smoking are well known\(^6,7\). Also, laboratory test findings such as lymphopenia and thrombocytopenia and elevated D-dimer, C-reactive protein (CRP), and lactate dehydrogenase (LDH) may be associated with severe respiratory failure and death\(^6\). Nevertheless, molecular signatures related to the dysregulation of the immune response and the disease deterioration are not fully uncovered.

The first line of host defense against viral infection relies on the innate immune system\(^5,8\) in which cells such as monocytes, dendritic cells, natural killer cells, and neutrophils are involved. In response to pathogen-associated molecular patterns derived from viruses, innate immune cells activate inflammatory signaling and cytokine production via pattern recognition receptors (PRRs) on the cell surface\(^8\). PRRs typified by Toll-like receptors (TLRs) also recognize damage-associated molecular patterns (DAMPs), endogenous molecules, including nucleic acids, HMGB1, heat shock proteins, S100 proteins, and
histones, and amplify inflammatory signaling\textsuperscript{9}. These events in immune cells are coordinated to promote influenza-induced lung injury\textsuperscript{10} and are suggested to play a critical role in COVID-19\textsuperscript{11}; however, little is known about specific molecules and mechanisms related to disease progression and deterioration.

Based on this background, we reasoned that a comprehensive search for genes dysregulated in peripheral blood mononuclear cells (PBMCs) could improve our understanding of the mechanisms underlying immune response dysfunction. Since PBMCs include the gatekeeper immune cells of viruses spread during systemic viral infections and several subpopulations that may cooperate in the immune response\textsuperscript{12}, this might facilitate the development of biomarkers and targeted therapies in COVID-19. We, therefore, applied RNA-sequencing analysis (RNA-seq) to investigate gene expression profiles of PBMCs derived from patients with COVID-19. PBMCs were collected from early-stage nonsevere patients with COVID-19 and analyzed by grouping patients into two groups: deteriorated and nondeteriorated, based on the subsequent clinical courses, to explore the profile predictive of COVID-19 deterioration. RNA-seq revealed 156 differentially expressed genes (DEGs) between groups. Pathway analysis of the DEGs exhibits “Nucleosome assembly” as the top pathway, including \textit{HIST1} cluster genes that are mostly upregulated in the deteriorated group. \textit{In silico} analysis demonstrated that protein inhibitors of activated STAT 4 (PIAS4) could be a possible upstream regulator in gene expression and inflammatory response by interacting with TP53, STAT1, and IKBKG.

**Results**

**Baseline characteristics of patients with COVID-19 at PBMCs collection**

PBMCs were collected from early-stage nonsevere patients with COVID-19 (Median 4 days, range 3–6 days after symptom onset). Based on the subsequent clinical courses, patients were grouped into two groups, deteriorated and nondeteriorated, and PBMCs from deteriorated (n = 6) and nondeteriorated (n = 6) patients were analyzed by RNA-seq (Figure 1a). There were no significant differences in clinical characteristics such as body temperature, pulse rate, and oxygen saturation between the groups at sample collection. Still, laboratory testing showed higher CRP and LDH in the deteriorated group compared with the nondeteriorated group (Table 1).

**RNA-seq of PBMCs and validation with reverse-transcribed quantitative polymerase chain reaction (RT-qPCR)**

RNA-seq was performed on RNA isolated from PBMCs under deteriorated versus nondeteriorated groups (Fig. 1a) and revealed 156 DEGs between groups analyzed by DESeq2. Among 156 DEGs, 102 were upregulated, and 54 were downregulated in the deteriorated group compared with the nondeteriorated group (Fig. 1b, Fig. S1, Table S1). We also validated gene expression of \textit{HIST1H2BO} and \textit{HIST1H2AE}, the top two statistically significant genes, by RT-qPCR analysis and confirmed that those genes were differentially expressed with statistical significance (Fig. 1c).

**Functional Enrichment analysis based on the DEGs**
Functional enrichment analysis based on the 156 DEGs were performed to explore pathways related to COVID-19 deterioration. Significant enrichment was observed in pathways related to nucleosome or chromatin assembly, including terms such as “Nucleosome assembly,” “Nucleosome organization,” and “DNA packaging” (Fig. 2a). Pathways related to immune response such as “Innate immune response in mucosa” and “Organ or tissue specific immune response” were also shown but with less significance. Gene set enrichment analysis (GSEA) also demonstrated “Nucleosome assembly” as the top pathway, and the pathway included 24 HIST1 cluster genes upregulated in the deteriorated group (Fig. 2b). No term regarding cellular proliferation or apoptosis is shown. Thus, the results indicate that PBMCs from deteriorating COVID-19 heightened HIST1 cluster gene transcription more than genes related to proinflammatory response at the early stage of disease, which may result in histone-induced lung injury.

In silico predictions of upstream transcription factors associated with COVID-19 deterioration

We next explored upstream transcription factors that induced the differential gene expression in PBMCs associated with COVID-19 deterioration using GSEA. We found four transcription factors of significance, false discovery rate below 0.05 (Fig. 3a). Among those factors, PIAS4 exhibited the highest enrichment score, including multiple HIST1 cluster genes as core enrichment genes (Fig. 3b). PIAS4 is an E3-type small ubiquitin-related modifier (SUMO) ligase and can modify the function of signal transducers or transcription factors essential for the immune response to viral infection\textsuperscript{13,14}. In line with this, we focused on PIAS4, explored its interaction partners using the STRING database, and found that PIAS4 interacts strongly and physically with signal transducers and transcription factors associated with COVID-19 pathogenesis, including TP53, STAT1, and IKBKG\textsuperscript{15,16} (Fig. 4). Thus, our in silico analysis indicates that PIAS4 is an essential factor in the pathogenesis of COVID-19 and disease progression, possibly through modifying the inflammatory response and HIST1 cluster gene expression.

Discussion

In this study, we identified a PBMCs gene set differentially expressed between deteriorated and nondeteriorated COVID-19 at the early stage of the disease. The gene set is highly enriched in HIST1 cluster genes, mostly upregulated in severe diseases before disease deterioration. Based on the gene set, we demonstrate that PIAS4 could be an upstream regulator of HIST1 cluster gene transcription and a modulator of an inflammatory response by interacting with inflammatory signaling molecules. Our results show that dysregulated PIAS4 function and HIST1 cluster gene expression in PBMCs could contribute to COVID-19 exacerbation.

Histone proteins are intranuclear proteins that form nucleosomes. They play key roles in DNA replication and gene transcription. Once translocated to the extracellular space, they act as DAMPs that mediate inflammatory response and organ injury via PRRs typified by TLRs\textsuperscript{9,17,18}. Previous articles showed that in patients with ARDS, circulating histones and histone proteins in the bronchoalveolar lavage fluids were increased and contributed to the pathogenesis of ARDS\textsuperscript{19}. Histone proteins induce endothelial or
epithelial cell injury in the lungs and kidneys, which results in pulmonary edema and multiple organ failure with neutrophil infiltration in capillaries\textsuperscript{17,20,21}.

Furthermore, histones induce coagulopathy and thrombosis via active thrombin production by reducing thrombomodulin-dependent protein C activation\textsuperscript{22} and recruiting fibrinogen to induce platelet aggregation\textsuperscript{23}. In COVID-19, histone-mediated lung injury and thrombosis related to neutrophil extracellular traps (NETs) are also expected to be involved in the pathogenesis\textsuperscript{24,25}. Our observations reinforce the importance of histones in the COVID-19 deterioration and the pathogenesis.

Extracellular histones originate from dead cells, and NETs of activated neutrophil origin as chromatin\textsuperscript{18} or are actively released in soluble or vesicle-associated form by immune cells such as lymphocyte and monocyte/macrophage without cell death\textsuperscript{26}. Furthermore, histone modulates proinflammatory responses on the cell surface in lymphocytes and monocyte/macrophages by translocating them from the nucleus, which is also unrelated to cell death\textsuperscript{27}. Here, we show that in PBMCs collected from deteriorated patients, at the early stage of the disease before deterioration, \textit{HIST1} cluster gene expression is upregulated without a cell death signature. Therefore, it highlights the importance of histone expression and secretion or translocation to the membrane in PBMCs, other than NETs-mediated histone released, in the pathogenesis and detection of COVID-19 exacerbation at the early stages of the disease.

Our work further demonstrates that PIAS4 can be an upstream mediator associated with COVID-19 deterioration. PIAS4 is a transcriptional corepressor that physically interacts with transcription factors such as TP53, STAT1, and NF\textsubscript{κ}B\textsuperscript{14–16}. Our \textit{in silico} analysis predicts PIAS4 as an upstream transcription factor strongly enriched by \textit{HIST1} cluster genes that are upregulated, indicating that gene expression may be derepressed. Its function can explain this as a SUMO E3 ligase or redundant interaction network that modulates functions and subcellular localization of many proteins and regulates numerous signaling pathways, including inflammatory pathways such as NF-kB signaling in a context-dependent manner\textsuperscript{13,14}. Intriguingly, PIAS4 is also physically interacts with viral proteins or DNA sites upon the infection, which regulates viral replication\textsuperscript{28,29}. These functions may inhibit PIAS4 as an intranuclear transcriptional corepressor, resulting in the upregulation of \textit{HIST1} cluster gene expression. Thus, our conclusion is still hypothetical but can propose candidate molecules, a new mechanistic insight into the pathogenesis, and a marker for the disease progression of COVID-19.

This study is not without limitations. PBMCs includes several subpopulations such as monocytes, lymphocytes and dendric cells, but the specific cell types that exhibit \textit{HIST1} cluster gene upregulation are not determined. Additionally, the distribution of upregulated histone protein, e.g., cellular surface or circulation, remains to be uncovered. Future studies will validate the function of histone proteins and the expression and function of upstream regulators.

In conclusion, we have reported that \textit{HIST1} cluster gene expression is upregulated in PBMCs from patients with deteriorating COVID-19 at the early stage of the disease. Our results propose a molecular signature and mechanisms that involve PIAS4-mediated transcriptional regulation and inflammatory
response. Histone-mediated lung injury could be linked to the pathogenesis and the deterioration of COVID-19.

**Methods**

**Study participants**

Patients with mild-to-moderate COVID-19 admitted to our COVID-19 subunit at Nishi-Kobe Medical Center within six days after the disease onset were eligible for inclusion between August 2020 and February 2021. SARS-CoV-2 infection was confirmed by RT-qPCR analysis. Also, blood for RNA-seq analysis was collected at admission. Disease severity was evaluated using the WHO definition (meet any of the followings: oxygen saturation < 90%, respiratory rate > 30 breaths/min, and signs of severe respiratory distress or not). Based on the subsequent clinical courses, patients were grouped into two groups, deteriorated (developed severe diseases) and nondeteriorated (had only mild-to-moderate diseases). The baseline characteristics of patients with COVID-19, which include age, sex, the day of disease onset and sample collection, body temperature, pulse rate, SpO$_2$, white blood cells, neutrophils, lymphocytes, monocytes, CRP, LDH, and D-dimer were extracted from medical records. This study was conducted following the World Medical Association's Declaration of Helsinki and approved by the ethical review board of Nishi-Kobe Medical Center (#2021–22). Informed consent was obtained from all study participants.

**PBMCs isolation and total RNA extraction**

Density gradient centrifugation was used to prepare PBMCs. In detail, Lymphoprep (STEMCELL TECHNOLOGIES, Canada) was added to blood samples and poured into SepMate(STEMCELL TECHNOLOGIES), followed by dilution using an equal volume of saline. After centrifugation at 1,200 g for 10 min, the layer containing PBMCs was transferred into a new tube and washed with saline. Then, the pellet was collected by centrifugation at 300 g for 8 min. Total RNA was extracted using Maxwell RSC instrument and Maxwell RSC simplyRNA Blood Kit (Promega, WI, USA) according to the manufacturer's instructions and stored at −80 °C for future analysis.

**RNA-seq analysis**

Total RNA from six deteriorated and six nondeteriorated patients with COVID-19 were analyzed. Libraries were prepared with the TruSeq Stranded mRNA Library Prep Kit (Illumina, San Diego, CA, USA) and subjected to sequencing with an Illumina Novaseq 6000 instrument (Illumina). Reads were trimmed using FASTQ groomer$^{30}$ and aligned to the GENCODE human reference genome GRCh37 (hg19, GENCODE website) using the HISAT2 function$^{31}$ on the Galaxy platform$^{32}$. Gene transcripts were quantified using DESeq2$^{33}$, and the volcano plot was generated by ggplot2 on Galaxy. Heatmap was generated by ClustVis$^{34}$. The enrichment of DEGs compared with the deteriorated and nondeteriorated group was
analyzed using ShynyGO\textsuperscript{35} and GSEA\textsuperscript{36}. \textit{In silico} protein–protein interaction analysis was conducted using STRING\textsuperscript{37}. RNA-seq data have been deposited on GEO and can be accessed at the following link.


**RT-qPCR analysis**

RT-qPCR was performed using iTaq Universal SYBR Green One-Step Kit (BIO-RAD Laboratories, CA, USA) according to the manufacturer’s instructions. Each measurement was duplicated using a CFX96 Real-Time System (BIO-RAD). The assay conditions were: 50°C for 10 min for reverse transcription, and 95°C for 1 min, followed by 40 cycles of 95°C for 10 s and 60°C for 30 s. Primer pairs were designed using PrimerBank (http://pga.mgh.harvard.edu). \textit{b-actin} (forward, 5′-CATGCCATCTCTGCTGGA-3′, reverse, 5′-CCGTGGCCATCTCTTGCTCG-3′), \textit{HIST1H2BO} (forward, 5′-GACCCGGCTAAATCTGCTCC-3′, reverse, 5′-GGCCTTGGTTACGGCTTTC-3′), \textit{HIST1H2AE} (forward, 5′-CTACTCGAACGAGTCGGG-3′, reverse, 5′-GATGGTCACCGACCTAGAGG-3′) were used.

**Statistical analysis**

All data represent mean ± standard deviation (SD). The normality of the data was evaluated by Kolmogorov–Smirnov normality test. Statistical significance was determined by a two-sided unpaired t-test or Mann–Whitney test. A \( p \) value of < 0.05 was considered significant.

**Declarations**

**Acknowledgements**

The authors would like to thank Enago (www.enago.jp) for the English language review.

**Author contributions**

M. I. and K. M. designed and performed the experiments, analyzed the data, wrote the article, and prepared the figures. R. O. analyzed the data. K. M. oversaw the article preparation and editing. All authors reviewed the manuscript.

**Data availability**

The datasets generated during the current study are available in the GEO repository, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE207625

Token: wbpkesittwbzgn

**Additional information**

The authors declare no competing interests.
References


### Tables

Table 1. Baseline characteristics of patients with COVID-19 at PBMCs collection

<table>
<thead>
<tr>
<th>Clinical manifestations</th>
<th>Nondeteriorated (n = 6)</th>
<th>Deteriorated (n = 6)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age (range) -yr</td>
<td>61 (50–75)</td>
<td>61.5 (44–76)</td>
<td>0.7624</td>
</tr>
<tr>
<td>Male no. ( يصل )</td>
<td>3 (50)</td>
<td>3 (50)</td>
<td></td>
</tr>
<tr>
<td>Median day of sample collection after disease onset (range)</td>
<td>4 (3–6)</td>
<td>4.5 (3–6)</td>
<td></td>
</tr>
<tr>
<td>BT (°C), mean ± SD</td>
<td>38.1 ± 0.9</td>
<td>38.7 ± 0.4</td>
<td>0.0877</td>
</tr>
<tr>
<td>PR, mean ± SD</td>
<td>98 ± 6.0</td>
<td>87 ± 11.1</td>
<td>0.1215</td>
</tr>
<tr>
<td>SpO₂ (%), mean ± SD</td>
<td>95 ± 1.4</td>
<td>94.5 ± 1.8</td>
<td>0.2287</td>
</tr>
<tr>
<td>WBC (/ μL), mean ± SD</td>
<td>4.50 ± 1.16</td>
<td>3.90 ± 1.59</td>
<td>0.6586</td>
</tr>
<tr>
<td>Neutrophil (/ μL), mean ± SD</td>
<td>3.10 ± 0.60</td>
<td>2.76 ± 1.30</td>
<td>0.6758</td>
</tr>
<tr>
<td>Lymphocyte (/ μL), mean ± SD</td>
<td>1.37 ± 0.75</td>
<td>0.86 ± 0.30</td>
<td>0.1903</td>
</tr>
<tr>
<td>Monocyte (/ μL), mean ± SD</td>
<td>0.39 ± 0.16</td>
<td>0.17 ± 0.16</td>
<td>0.1897</td>
</tr>
<tr>
<td>CRP (mg / dL), mean ± SD</td>
<td>0.9 ± 0.6</td>
<td>5.0 ± 2.6</td>
<td>0.0044*</td>
</tr>
<tr>
<td>LDH (U / L), mean ± SD</td>
<td>236.0 ± 40.1</td>
<td>355.5 ± 98.5</td>
<td>0.0196*</td>
</tr>
<tr>
<td>D-dimer (μg / mL), mean ± SD</td>
<td>0.00 ± 0.29</td>
<td>0.87 ± 1.49</td>
<td>0.0898</td>
</tr>
</tbody>
</table>

PBMCs, peripheral blood mononuclear cells; BT, body temperature; PR, pulse rate; WBC, white blood cell; CRP, C-reactive protein; LDH, lactate dehydrogenase, *p < 0.05 by t-test*

### Figures
RNA-seq analysis of PBMCs from patients with COVID-19 revealed differences in gene expression related to disease deterioration.

**a,** Experimental design. PBMCs were collected from early-stage nonsevere patients with COVID-19. Based on the subsequent clinical courses, patients were grouped into two groups, deteriorated and nondeteriorated. RNA-seq analyzed PBMCs from each group (each n = 6). **b,** Volcano plots illustrated fold differences in individual gene expression in deteriorated versus nondeteriorated COVID-19 and associated p-values (negative log10). **c,** Validation of mRNA expression relative to β-actin by reverse transcription-quantitative polymerase chain reaction (RT-qPCR), of two genes with the top statistical significance by DESeq2 analysis. Data are mean ± SD, n = 6; *p < 0.05 by t-test.
Figure 2

Pathway analysis of the differentially expressed genes (DEGs) by RNA-seq a, Pathway analysis of 156 DEGs. Circles indicate the number of genes, and colors depict the fold enrichment. The threshold of significance was set at $p < 0.05$. b, The top pathway of gene set enrichment analysis (GSEA) from RNA-seq and the heatmap of the analyzed data set. ND indicates nondeteriorated, and D indicates the deteriorated group. ES, enrichment score; NES, normalized enrichment score; FDR, false discovery rate.
Figure 3

The prediction of upstream transcription factors by GSEA.

a. Statistically significant upstream transcription factors analyzed by GSEA using the gene set of KEGG_BASAL_TRANSCRIPTION_FACTORS from RNA-seq. b. The heatmap of PISA4 target genes analyzed. ND indicates nondeteriorated, and D indicates the deteriorated group. ES, enrichment score; NES, normalized enrichment score; FDR, false discovery rate
Figure 4

*In silico* prediction of the interactive partners of PIAS4

STRING-based prediction of protein–protein interaction of PIAS4 and the score. The edges indicate that the proteins are part of a physical complex.

**Supplementary Files**

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

- SupplementaryFigureS1070122.docx
- SupplementaryTableS1.xlsx