ABSTRACT

Introduction: The coronavirus disease 2019 (COVID-19) pandemic has caused extensive disruption of public health worldwide. There were reports of COVID-19 patients having multiple complications. This study investigated COVID-19 from a genetic perspective.

Methods: We conducted RNA sequencing (RNA-Seq) analysis of respiratory tract samples from 24 patients with COVID-19. Eight patients receiving mechanical ventilation or extracorporeal membrane oxygenation were regarded as severe cases; the remaining 16 patients were regarded as non-severe cases. After quality control, statistical analyses were performed by logistic regression and the Kolmogorov–Smirnov test to identify genes associated with disease severity.

Results: Six genes were associated with COVID-19 severity in both statistical tests, namely RPL15, BACE1-AS, CEPT1, EIF4G1, TMEM91, and TBCK. Among these genes, RPL15 and EIF4G1 played roles in the regulation of mRNA translation. Gene ontology analysis showed that the differentially expressed genes were mainly involved in nervous system diseases.

Conclusion: RNA sequencing analysis showed that severe acute respiratory syndrome coronavirus 2 infection is associated with the overexpression of genes involved in nervous system disorders.

New knowledge added by this study
• Differentially expressed genes between patients with severe and non-severe cases of coronavirus disease 2019 (COVID-19) were reported.
• Overexpression of genes involved in cell proliferation, viral binding and replication, and neurological and lung diseases was observed, suggesting a pathophysiological mechanism by which severe acute respiratory syndrome coronavirus 2 induces lung inflammation and neurological complications.

Implications for clinical practice or policy
• Future studies that involve gene expression profiling with larger sample sizes, in vitro infection experiments, and animal models can help to elucidate the mechanisms and corresponding therapeutic approaches for neurological complications of COVID-19.

Introduction
Coronavirus disease 2019 (COVID-19) has spread to >500 million people and caused 6.2 million deaths worldwide as of 22 April 2022. Approximately 20% of patients with COVID-19 develop severe symptoms and 5% of patients require intensive care. A wide range of complications were reported with COVID-19 infection, including nervous system diseases, circulatory system diseases, urinary system diseases, and digestive system diseases.
通過RNA定序對香港住院患者進行新冠肺炎重症與非重症病例的基因關聯性研究

李奇、陳子桂、張業先、陳韻怡、莊家俊、徐仲鍈、凌若崴、雷頌恩、陳基湘、王海天

引言：新冠肺炎大流行對全球公共衛生造成了廣泛破壞。新冠肺炎患者出現多種併發症，本研究由基因角度探究新冠肺炎。

方法：我們對24名新冠肺炎患者進行了呼吸道樣本RNA定序分析，其中8位接受機械通氣或體外膜氧合的患者被視為重症，餘下16名患者則被視為非重症病例。我們在控制樣本品質後，通過邏輯迴歸和柯爾莫哥洛夫—斯米爾諾夫檢定來找出與疾病嚴重程度相關的基因。

結果：在上述兩個檢定中均與新冠肺炎的嚴重程度顯著相關，包括RPL15、BACE1-AS、CEPT1、EIF4G1、TMEM91和TBCK，其中RPL15和EIF4G1調控mRNA的翻譯。基因本體分析顯示差異表現的基因主要涉及神經系統疾病。

結論：RNA定序分析顯示嚴重急性呼吸系統綜合症冠狀病毒2型與神經系統疾病涉及的基因過度表現相關。

Various genetic associations with COVID-19 outcomes have been explored. A whole-genome sequencing study of germline mutations revealed a cluster of six genes (SLC6A20, CCR9, FYCO1, CXC6, XCR1, and LZTFL1) that increased susceptibility to severe COVID-19 with respiratory failure. In a Chinese population, a whole-genome sequencing study of 332 patients with COVID-19 identified loci in the genes TMEM189 and LBE2V1 with potential genome-wide implications through the IL-1 signalling pathway. In an intensive care unit cohort of 15 patients with severe COVID-19, analysis of RNA sequencing (RNA-Seq) data from blood samples showed that the immune-modulating genes PD-L1 and PD-L2 were differentially expressed among patients with fatal outcomes.

Thus far, studies of gene expression at initial sites of infection in patients with severe and non-severe COVID-19 remain limited. To investigate COVID-19 from a genetic perspective, we conducted RNA-Seq analysis of respiratory tract samples from patients with COVID-19; we sought to identify genes associated with disease severity.

Methods

Patients

Twenty-four patients were recruited from Prince of Wales Hospital in Hong Kong between 7 February and 10 April 2020. All patients had severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, confirmed by two independent real-time reverse transcription–polymerase chain reaction assays targeting the N gene. Symptoms on admission were recorded, and medical histories were collected from clinical health records. Among the recruited patients, eight who received mechanical ventilation or extracorporeal membrane oxygenation were regarded as severe cases; the remaining 16 patients showed asymptomatic or mild (no pneumonia) to moderate (pneumonia but not requiring oxygen supplementation) disease and were regarded as non-severe cases. RNA sequencing was performed on upper and lower respiratory swab samples collected within 3 days after hospitalisation.

RNA sequencing data

Total RNA was extracted from respiratory swab samples using the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany), pre-treated with DNase I and depleted of human rRNA and globin genes using the QIAseq FastSelect ribosomal RNA and globin mRNA Removal Kit (Qiagen, Hilden, Germany). Illumina libraries for RNA-Seq next-generation sequencing were prepared using the KAPA HyperPrep Kit (Roche, Pleasanton [CA], US) in accordance with the manufacturer’s instructions, then sequenced on an Illumina NextSeq 500 system (Illumina, San Diego [CA], US) using 150 bp paired-end reads. The raw data consisted of 58,735 Ensembl-annotated genes. Quality control was performed to remove patients with low numbers of RNA-Seq reads (three samples) and genes with zero reads in >20% of samples (55,751 genes). Thus, 3,164 genes remained available for differential expression analysis. Raw read count data were summarised as fragments per million reads mapped and then log2-transformed.

Statistical analysis

Logistic regression was utilised to identify genes associated with severity outcomes. Subsequent evaluations by the Kolmogorov–Smirnov (KS) test were performed to test for differences in gene expression between groups. The Bonferroni corrected significance threshold was 1.58 × 10−5.

Functional analysis of genes

According to their disease relevance in the GeneCards and MalaCards databases, genes were categorised into the following 10 disease groups: nervous system, integumentary system, circulatory system, urinary system, digestive system, respiratory system, musculoskeletal system, endocrine system, reproductive system, and infectious diseases.

Results

Demographics and baseline characteristics

The patients’ demographics and baseline characteristics are summarised in Table 1. The mean age of patients in the severe group was 62.13 years (95% confidence interval [CI]=53.34-70.91), which was significantly higher than that in the non-severe...
group (29.73 years; 95% CI=22.11-37.36). Compared with the non-severe group, the severe group had higher prevalences of complications including cardiovascular, liver, endocrine, and metabolic disorders, as well as higher rates of respiratory, fever, and diarrhoea symptoms. The COVID-19 World Health Organization score\textsuperscript{25} was significantly higher in the severe group than in the non-severe group. Lopinavir, antibiotics, conventional oxygen therapy, and mechanical ventilation were more commonly used for treatment in the severe group than in the non-severe group (Table 1).

Differentially expressed genes according to RNA sequencing

Six genes, namely RPL15, BACE1-AS, CEPT1, EIF4G1, TMEM91, and TBCK, were differentially expressed between the severe and non-severe groups (all P values <0.05 in both logistic regression and the KS test) [Table 2]. Fold-change and odds ratio results indicated that these genes were consistently highly expressed in the severe group. The complete list of genes with P values <0.05 in KS test is provided in online supplementary Table 1.

Gene ontology and enrichment analysis

The functions of the identified genes were summarised through database and literature searches. Two genes, 60S ribosomal protein L15 (RPL15) and eukaryotic translation initiation factor 4 gamma 1 (EIF4G1), play roles in host translation\textsuperscript{26-28} Furthermore, the top genes were involved in neurological disorders. RPL15 is involved in the life cycle of human immunodeficiency virus\textsuperscript{29,30} and baculovirus infection reportedly disrupts the expression of this gene.\textsuperscript{31,32} EIF4G1 plays a role in viral binding and affects the pathogenicity and virulence of H5N1 influenza A virus, foot-and-mouth disease virus, and vaccinia virus\textsuperscript{33-35}; it also contains multiple mutations among patients with familial Parkinson's disease.\textsuperscript{36} TBCK encodes a conserved protein kinase that regulates cell size and proliferation.\textsuperscript{37} CEPT1 encodes choline/ethanolamine phosphotransferase, which is used in the synthesis of choline- or ethanolamine-containing phospholipids. The function of TMEM91, a transmembrane protein, is unclear; however, the results of genome-wide association studies suggest that loci containing this gene are involved in lung diseases.

The non-coding gene BACE1-AS regulates the stability of the BACE1 protein and directly increases the abundance of amyloid beta-peptide (Aβ\textsubscript{1-42}) in Alzheimer's disease.\textsuperscript{38} The implications of this gene in severe COVID-19 are unclear. For the top 15 overexpressed genes (P values in KS test <0.05), disease relevance data were retrieved from GeneCards\textsuperscript{22}; 14 of the 15 genes (93.3%) have been linked to neurological diseases, followed by eye (80.0%) and psychiatric (73.3%) diseases. Thus, all of the top genes were involved in nervous system disorders (Fig).

Discussion

Nervous system disorders such as encephalopathy, impaired consciousness, seizure, ataxia,
neuropathies, neurodegenerative diseases, and anosmia have been extensively documented in patients with COVID-19.39-41 The two major potential pathogenesis pathways are direct viral invasion and immune-mediated injury. Direct viral entry to the central nervous system can travel through hematogenous or olfactory routes, or by transneuronal spread from the lungs.42 Post-mortem analysis of brain tissue from patients with COVID-19 encephalitis reportedly contained SARS-CoV-2 viral particles.43,44 Furthermore, a series of autopsy studies showed that localised inflammation of the brainstem nuclei, as well as the cytokine storm associated with SARS-CoV-2 infection, could disrupt the blood–brain barrier and cause necrosis in the brains of patients with severe COVID-19.45,46 In patients with COVID-19, anosmia may be caused by an inflammation-mediated decrease in odorant receptor expression.47 Several studies have utilised RNA-Seq to characterise the transcriptomic profiles of patients with COVID-19.46,47 Significant downregulation of genes related to the hypoxia-inducible factor system was observed during periods of infection and oxygen deprivation.48 Additionally, transcriptomic profiles of peripheral blood mononuclear cells revealed that patients with COVID-19 shared several dysregulated genes with individuals who had bipolar illness, post-traumatic stress disorder, or schizophrenia.51

TABLE 2. Differentially expressed genes between patients with severe and non-severe cases of coronavirus disease 2019 according to RNA sequencing analysis

<table>
<thead>
<tr>
<th>Gene</th>
<th>Log2(FC)*</th>
<th>Odds ratio</th>
<th>P value in logistic regression</th>
<th>P value in Kolmogorov–Smirnov test</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPL15</td>
<td>0.33</td>
<td>1.48</td>
<td>0.022</td>
<td>0.005</td>
</tr>
<tr>
<td>BACE1-AS</td>
<td>0.27</td>
<td>1.38</td>
<td>0.034</td>
<td>0.013</td>
</tr>
<tr>
<td>CEP1</td>
<td>0.11</td>
<td>1.42</td>
<td>0.022</td>
<td>0.030</td>
</tr>
<tr>
<td>EIF4G1</td>
<td>0.57</td>
<td>1.70</td>
<td>0.015</td>
<td>0.044</td>
</tr>
<tr>
<td>TMEM91</td>
<td>0.36</td>
<td>1.42</td>
<td>0.027</td>
<td>0.044</td>
</tr>
<tr>
<td>TBCK</td>
<td>0.52</td>
<td>1.34</td>
<td>0.036</td>
<td>0.044</td>
</tr>
</tbody>
</table>

Abbreviation: Log2(FC) = log2(fold-change)
* Value >0 indicates higher mean expression in the severe group

FIG. Frequency of diseases related to the top identified genes (n=15)*
Abbreviation: MSSD = musculoskeletal system disease
* Data are shown as No. (%)
present findings suggest that SARS-CoV-2 infection is associated with differential expression of genes involved in nervous system disorders. Future studies that involve gene expression profiling with larger sample sizes, in vitro infection experiments, and animal models can help to elucidate the mechanisms and corresponding therapeutic approaches for neurological complications of COVID-19.

Limitations
A major limitation of this study was its small sample size. Patient age distributions considerably differed between groups. However, age-stratified analysis showed effects consistent with the directions reported in Table 2, although the statistical significance was hindered by the small sample size (online supplementary Table 2 and online supplementary Fig). Further sequencing of samples collected from respiratory tract sites may provide stronger evidence of protein expression abnormalities at the initial site of SARS-CoV-2 infection.

Conclusion
In this study, we conducted RNA-Seq analysis to identify differentially expressed genes between patients with severe and non-severe cases of COVID-19. We observed overexpression of genes involved in cell proliferation, viral binding and replication, and neurological and lung diseases, suggesting a pathophysiological mechanism by which SARS-CoV-2 induces lung inflammation and neurological complications.

Author contributions
Concept or design: BCY Zee, PKS Chan, MH Wang. Acquisition of data: Z Chen, PKS Chan. Analysis or interpretation of data: Q Li, Z Chen, Y Zhang, RWY Chan, MKC Chong, PKS Chan, G Lui, I. Ling. Drafting of the manuscript: Q Li, MH Wang. Critical revision of the manuscript for important intellectual content: Q Li, Y Zhang, MH Wang.

All authors had full access to the data, contributed to the study, approved the final version for publication, and take responsibility for its accuracy and integrity.

Conflicts of interest
BCY Zee is a shareholder of Health View Bioanalytic Limited. As a statistical adviser of the journal, MKC Chong was not involved in the peer review process. MH Wang is a shareholder of Beth Bioinformatics Co, Ltd. Other authors have disclosed no conflicts of interest.

Funding/support
This research was partially supported by the Health and Medical Research Fund of the former Food and Health Bureau, Hong Kong SAR Government (Ref Nos.: COVID190103, COVID190112 and INF-CUHK-1), The Chinese University of Hong Kong (CUHK) Project Impact Enhancement Fund (Ref No.: CUPIEF/Ph2/COVID/06) and CUHK Direct Grant (Ref No.: 2020.025). The funders had no role in study design, data collection/analysis/interpretation, or manuscript preparation.

Ethics approval
The study protocol of this research was approved by the Joint Chinese University of Hong Kong–New Territories East Cluster Clinical Research Ethics Committee (Ref No.: 2020.076). All patients provided written informed consent for participation in this research.

Supplementary material
The supplementary material was provided by the authors and some information may not have been peer reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by the Hong Kong Academy of Medicine and the Hong Kong Medical Association.

References
47. Daamen AR, Bachali P, Owen KA, et al. Comprehensive Comprehensive


