Establishment of the large-scale longitudinal multi-omics dataset in COVID-19 patients: data profile and biospecimen

Hye-Yeong Jo[1], Sang Cheol Kim[1], Do-hwan Ahn[1], Siyoung Lee[2], Se-Hyun Chang[1], So-Young Jung[3], Young-Jin Kim[4], Eugene Kim[3], Jung-Eun Kim[5], Yeon-Sook Kim[6], Woong-Yang Park[2,7], Nam-Hyuk Cho[8], Donghyun Park[2], Ju-Hee Lee[1] & Hyun-Young Park[9,*]

[1]Division of Healthcare and Artificial Intelligence, Department of Precision Medicine, Korea National Institute of Health, Korea Disease Control and Prevention Agency, Cheongju 28159, [2]Geninus Inc, Seoul 05836, [3]Division of Biobank, Department of Precision Medicine, Korea National Institute of Health, Korea Disease Control and Prevention Agency, Cheongju 28159, [4]Division of Genome Science, Department of Precision Medicine, Korea National Institute of Health, Korea Disease Control and Prevention Agency, Cheongju 28159, [5]Division of Bio Bigdata, Department of Precision Medicine, Korea National Institute of Health, Korea Disease Control and Prevention Agency, Cheongju 28159, [6]Division of Infectious Disease, Department of Internal Medicine, Chungnam National University School of Medicine, Daejeon 35015, [7]Samsung Genome Institute, Samsung Medical Center, Seoul 06351, [8]Department of Microbiology and Immunology, College of Medicine, Seoul National University, Seoul 08826, [9]Department of Precision Medicine, Korea National Institute of Health, Korea Disease Control and Prevention Agency, Cheongju 28159, Korea

Understanding and monitoring virus-mediated infections has gained importance since the global outbreak of the coronavirus disease 2019 (COVID-19) pandemic. Studies of high-throughput omics-based immune profiling of COVID-19 patients can help manage the current pandemic and future virus-mediated pandemics. Although COVID-19 is being studied since past 2 years, detailed mechanisms of the initial induction of dynamic immune responses or the molecular mechanisms that characterize disease progression remains unclear. This study involved comprehensively collected biospecimens and longitudinal multi-omics data of 300 COVID-19 patients and 120 healthy controls, including whole genome sequencing (WGS), single-cell RNA sequencing combined with T cell receptor (TCR) and B cell receptor (BCR) sequencing (scRNA(+scTCR/BCR)-seq), bulk BCR and TCR sequencing (bulk TCR/BCR-seq), and cytokine profiling. Clinical data were also collected from hospitalized COVID-19 patients, and HLA typing, laboratory characteristics, and COVID-19 viral genome sequencing were performed during the initial diagnosis. The entire set of biospecimens and multi-omics data generated in this project can be accessed by researchers from the National Biobank of Korea with prior approval. This distribution of large-scale multi-omics data of COVID-19 patients can facilitate the understanding of biological crosstalk involved in COVID-19 infection and contribute to the development of potential methodologies for its diagnosis and treatment. [BMB Reports 2022; 55(9): 465-471]

INTRODUCTION

In December 2019, the occurrence of a new infectious outbreak of severe respiratory illness was observed (1), which was named novel coronavirus disease 2019 (COVID-19) on January 7, 2020 by the World Health Organization (WHO), leading to the COVID-19 pandemic (2). As of March 2022, 451,611,588 cases of COVID-19 infection resulting in 6,022,199 deaths have been reported in 196 countries and territories worldwide (3, 4). For over 2 years, the COVID-19 pandemic has impacted political, social, economic, and healthcare issues worldwide. The establishment of healthcare strategies for COVID-19 patients requires a critical understanding of the pathophysiology of COVID infection.

The difficulties in predicting disease progression have made it challenging to set up healthcare services associated with quarantine, vaccination, and treatment of COVID-19 patients. Although several studies have reported the dynamics between the COVID-19 virus and infected host immune system (4-9), the molecular and cellular mechanisms resulting in arbitrary symptoms during COVID-19 infection remain poorly understood. Nations such as the countries of the European Union and Taiwan have emphasized on building research infrastructure and collecting data and biospecimens to devise multifaceted responses to the COVID-19 pandemic (10-12). However, there is a limited understanding of the complicated infection biology of COVID-19 owing to a lack of comprehensive and systematic multimodal datasets with matched clinical data and laboratory testing (13). Therefore, longitudinal multimodal data are needed to elucidate the viral infection dynamics and immune response to

COVID-19.

In this study, we constructed a scalable multimodal dataset from 300 COVID-19 patients and 120 healthy controls. Comprehensively collected biospecimens and clinical data were deposited in the National Biobank of Korea (NBK). Laboratory testing, HLA typing, and whole-genome sequencing (WGS) of samples from the study participants were performed on admission. Additionally, longitudinal single-cell immune profiling comprising single-cell RNA sequencing combined with T cell receptor (TCR) and B cell receptor (BCR) sequencing (scRNA (+scTCR/BCR)-seq) and bulk BCR and TCR sequencing (bulk BCR/TCR-seq), as well as an profiling 192 cytokines were used to investigate dynamic infection progress in COVID-19 patients. Our dataset can potentially elucidate the understanding of COVID-19 susceptibility and severity and facilitate the development of target molecules and treatment strategies against COVID-19.

RESULTS

Participant enrollment and ethics statement

From October 2020 to June 2021, 120 healthy controls and 300 COVID-19 patients were enrolled at Chungnam National University Hospital, Seoul Medical Center, and Samsung Medical Center. The study protocol was reviewed and approved by these institutes and the Institutional Review Board of Korea National Institute of Health (CNUH 2020-12-002-008, SEOUL 2021-02-016, SMC-2021-03-160, and 2020-09-03-C-A). All the participants provided written informed consent.

Participants were enrolled when diagnosed with COVID-19 and biospecimens were collected at several time points based on each patient's disease progression. The clinical course of the patients was recorded daily until discharge. Moreover, we recruited 120 healthy controls who had no history of COVID-19 infection or were vaccinated for COVID-19.

Collection of biospecimens

The host response against COVID-19 at the molecular and cellular levels can be examined based on a multidimensional analysis of optimal biospecimens. Blood samples from participants collected using various sampling techniques (Supplementary Fig. 1) were used for complete blood count (CBC), laboratory testing, cytokine profiling, and multi-omics analyses such as WGS, HLA typing, scRNA(+scTCR/BCR)-seq, and bulk BCR/TCR-seq. Nasopharyngeal swabs (NPS) were collected for sequencing of the COVID-19 viral genome. Urine samples were collected and stored for use in future experiments, such as metabolic studies or trace determination of COVID-19. Detailed procedures for collecting of samples were described in Supplementary Methods.

The NBK website is the medium of distribution of the collected biospecimens and also provides biospecimen information (https://nih.go.kr/biobank). Researchers wishing to use these biospecimens can apply to the NBK for access approval based on research proposals.

Clinical data collection and baseline characteristics of study population

The severity of COVID-19 in patients was classified according to the WHO guidelines (14). Patients were classified as "severe" on presenting COVID-19 symptoms with pneumonia and one of the following: respiratory rate of more than 30 times/min, oxygen saturation of less than 93% at room temperature, less than 300 mmHg of PaO2/FiO2, or a rapid progression of infiltration ($>$ 50%) observed by chest computed tomography (CT) imaging.

Out of 300 COVID-19 patients, 243 (81.0%) were classified as moderate and 57 (19.0%) as severe (Fig. 1 and Table 1). Two patients with severe COVID-19, whose biospecimens were collected at three time points, died during hospitalization. Other patients were discharged after recovery. The average hospitalization period was 13 and 16 days for moderate and severe patients, respectively. Multi-omics data were obtained at 7 time points from 12 patients with severe COVID-19.

Clinical information collected daily using an electronic clinical data management system (https://icreat.nih.go.kr) and clinical characteristics of study participants are shown in Supplementary Table 1 and Table 1, respectively. The median age was 42.0 for healthy controls, and 52.0 and 64.0 for moderate and severe COVID-19 patients, respectively. Hypertension and diabetes
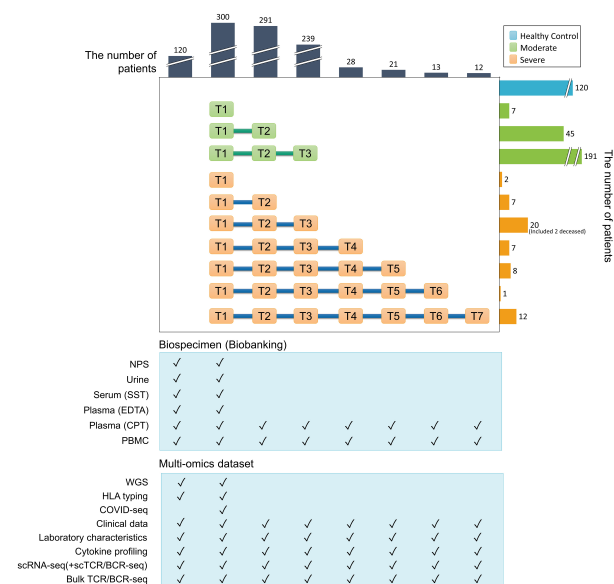
Fig. 1. Overview of multi-omics study of COVID-19. Blue, green, and orange bars indicate healthy controls and moderate and severe COVID-19 patients, respectively. T# indicates the time points of sample collection, and the number on the bars on the top right side indicates the number of patients. Sample collection was performed at a maximum of three and seven junctures for moderate and severe COVID-19 patients, respectively. The bottom panel shows six distinct types of biospecimens and eight distinct types of multi-omics data. COVID-seq indicates COVID-19 viral genome sequencing.

Table 1. Clinical characteristics of COVID-19 patients

Clinical characteristics	Healthy control (n = 120)	COVID-19 moderate (n = 243)	COVID-19 severe (n = 57)	P-value
Age	42 (23-80)	52 (21-97)	64 (27-93)	-
Male (%)	59 (49.2%)	128 (52.7%)	33 (57.9%)	-
Current Smoker	15 (12.5%)	27 (11.1%)	0 (0.0%)	0.02
Comorbidity				
Hypertension	20 (16.7%)	63 (25.9%)	29 (50.9%)	<0.001
Diabetes	7 (5.8%)	34 (14.0%)	26 (45.6%)	<0.001
Coronary heart disease	3 (2.5%)	8 (3.3%)	6 (10.5%)	0.05
Stroke	1 (0.8%)	3 (1.2%)	2 (3.5%)	0.53
Malignant neoplasm	1 (0.8%)	4 (1.6%)	2 (3.5%)	0.71
Chronic hepatitis/liver cirrhosis	1 (0.8%)	4 (1.6%)	1 (1.8%)	<0.001
Symptom				
Cough	-	99 (40.7%)	28 (49.1%)	0.32
Dyspnea	-	9 (3.7%)	16 (28.1%)	<0.001
Fever	-	84 (34.6%)	27 (47.4%)	0.10
Sore throat	-	82 (33.7%)	14 (24.6)	0.24
Sputum production	-	46 (18.9%)	19 (33.3%)	0.03
Rhinorrhea	-	18 (7.4%)	2 (3.5%)	0.44
Myalgia	-	93 (38.3%)	22 (38.6%)	1.00
Malaise	-	56 (23.0%)	15 (26.3%)	0.73
Headache	-	56 (23.0%)	14 (24.6%)	0.94
Nausea	-	3 (1.2%)	3 (5.3%)	0.15
Diarrhea	-	4 (1.6%)	6 (10.5%)	0.00
Pneumonia during hospitalization	-	108 (44.4%)	53 (93.0%)	<0.001

The number and percentage of events in each group are shown for each clinical characteristic. P-values were calculated by chi-squared test for the categorical groups. Statistical significance was set at P < 0.05. Values in age are presented as median (minimum-maximum).

were the most common comorbidities, and pneumonia was observed in 44.4% and 93.0% of the moderate and severe patients, respectively. As the clinical information used in this study includes a broad range of categories recorded on the daily bases (Supplementary Table 1), we envisage that this result will provide insights into COVID-19 infection, such as studies on correlations between COVID-19 severity and demographics, comorbidity, or treatment and on changes in the clinical status during hospitalization for COVID-19 severity.

Blood chemistry and cytokine profiling

CBC and blood chemistry were performed on admission for the participants (Table 2). Among the 37 tested parameters, hsCRP levels [12.98 mg/L in moderate and 51.05 mg/L in severe] were significantly elevated in COVID-19 patients than in healthy controls [0.79 mg/L], which indicated an association between hsCRP level and the severity of COVID-19. This is consistent with previous studies reporting hsCRP as an indicator of COVID-19 progression (15, 16). Ferritin was also significantly elevated in COVID-19 patients [214.50 ng/ml in moderate and 490.09 ng/ml in severe] compared to that in healthy controls [119.56 ng/ml]. This finding agrees with that of previous reports (17, 18), suggesting that ferritin is relevant for distinguishing

disease severity in COVID-19 patients and leads to immune dysregulation and cytokine storms, especially in severe COVID-19 (19). Conversely, iron levels were lower in COVID-19 patients [71.73 μg/dl in moderate and 48.00 μg/dl in severe] than in healthy controls [108.01 μg/dl], consistent with previously reported results (20, 21). Interestingly, the WBC count was lower in moderate COVID-19 patients and elevated in severe COVID-19 patients compared to that in healthy controls. These findings would allow to understand the association between blood biochemical characteristics and COVID-19 severity, as well as that of the pathophysiology of COVID-19 progression.

The longitudinal profiles for 191 cytokines in COVID-19 patients obtained by measuring cytokine levels at multiple time points from hospitalization to discharge can provide an in-depth understanding of dynamic cytokine patterns in the progression of COVID-19 over time (Supplementary Table 2). To comprehensively identify cytokine inflammation in COVID-19 patients, the expression levels of cytokines for the first time point were evaluated (Supplementary Fig. 2). Of the top 25 significant cytokines, expression levels of complement component C9, LRG1, CD14, CEACAM1-1, and IL-23 were elevated for severity, while the expression of serpin A4, properdin, fetuin A, and fibroblast activation protein alpha (FAP) were decreased.

Table 2. Laboratory characteristics of healthy controls and COVID-19 patients

Laboratory characteristics	Healthy control (n = 120)	COVID-19 moderate (n = 243)	COVID-19 severe (n = 57)
WBC (Thous/ul)	5.96 ± 1.55	4.68 ± 1.63*	7.02 ± 4.94[§, †]
RBC (Mil/ul)	4.60 ± 0.43	4.50 ± 0.59	4.37 ± 0.59[§]
Platelet (Thous/ul)	225.06 ± 48.98	197.28 ± 63.66*	178.95 ± 68.62[§]
Hemoglobin (g/dl)	14.06 ± 1.53	13.906 ± 1.79	13.48 ± 1.69[§]
Hematocrit (%)	41.82 ± 4.09	41.15 ± 5.19	39.96 ± 4.90[§]
hs-CRP (mg/L)	0.79 ± 1.11	12.98 ± 24.57*	51.05 ± 56.30[§, †]
Iron (ug/dl)	108.01 ± 42.47	71.73 ± 41.86*	48.00 ± 39.05[§, †]
Ferritin (ng/ml)	119.56 ± 110.05	214.50 ± 219.43*	490.09 ± 390.64[§, †]
UIBC (ug/dl)	217.55 ± 72.33	222.04 ± 60.26	196.40 ± 58.05[†]
Vitamin B12 (pg/ml)	631.26 ± 275.59	765.97 ± 468.82*	1045.74 ± 742.19[§, †]
Folate (ng/ml)	10.84 ± 6.12	12.73 ± 8.43	13.13 ± 10.00
Total protein (g/dl)	6.87 ± 0.35	6.68 ± 0.49*	6.23 ± 0.71[§, †]
Albumin (g/dl)	4.56 ± 0.23	4.34 ± 0.36*	3.91 ± 0.53[§, †]
Homocysteine (umol/L)	14.08 ± 3.80	14.32 ± 4.65	13.34 ± 4.95
ALT (U/L)	24.48 ± 16.38	28.88 ± 26.29	38.09 ± 31.52[§, †]
r-GTP (U/L)	24.11 ± 42.07	33.08 ± 67.44	45.18 ± 48.50[§]
AST (U/L)	24.48 ± 12.51	29.51 ± 16.10*	39.12 ± 20.16[§, †]
Total bilirubin (mg/dl)	0.83 ± 0.34	0.75 ± 0.35*	0.67 ± 0.36[§]
Direct bilirubin (mg/dl)	0.24 ± 0.11	0.24 ± 0.13	0.25 ± 0.17[§]
ALP (U/L)	62.26 ± 16.35	67.12 ± 21.75*	74.74 ± 27.10[§, †]
Calcium (mg/dl)	9.51 ± 0.33	9.04 ± 0.42*	8.59 ± 0.53[§, †]
Phosphorus (mg/dl)	94.58 ± 0.42	3.38 ± 0.68*	3.21 ± 0.58[§]
BUN (mg/dl)	14.54 ± 3.94	14.70 ± 11.00	19.79 ± 12.05[§, †]
Cystatin C (mg/L)	0.70 ± 0.18	0.96 ± 0.79*	1.24 ± 1.03[§, †]
Creatinine (mg/dl)	0.77 ± 0.18	0.88 ± 1.08	0.97 ± 0.95[§]
Uric acid (mg/dl)	5.24 ± 1.54	4.94 ± 1.57	4.50 ± 1.61[§]
Total CPK (U/L)	142.60 ± 212.26	96.51 ± 143.74*	122.89 ± 145.91[§]
Glucose (mg/dl)	94.58 ± 0.42	98.21 ± 35.96	142.09 ± 74.29[§, †]
HbA1c (%)	5.46 ± 0.74	5.89 ± 0.91*	7.29 ± 2.25[§, †]
Total cholesterol (mg/dl)	179.28 ± 33.34	167.35 ± 34.58*	144.28 ± 38.52[§, †]
Triglyceride (mg/dl)	130.72 ± 99.32	116.28 ± 58.64	138.46 ± 56.92[†]
HDL cholesterol (mg/dl)	56.71 ± 14.05	44.47 ± 12.91*	34.89 ± 8.31[†]
LDL cholesterol (mg/dl)	104.80 ± 30.57	105.18 ± 34.84	87.77 ± 37.73[§, †]
Apolipoprotein A-I (mg/dl)	154.93 ± 22.60	127.36 ± 26.50*	105.32 ± 20.23[§, †]
Apolipoprotein A-II (mg/dl)	31.55 ± 5.11	28.98 ± 5.86*	23.28 ± 6.06[§, †]
Apolipoprotein B (mg/dl)	88.46 ± 24.04	90.89 ± 23.52	85.91 ± 30.20
Lipoprotein(a) (mg/dl)	16.56 ± 12.03	18.41 ± 14.09	20.30 ± 19.64

Continuous variables are presented as means ± standard deviations for each laboratory feature. P-values were calculated by two sample t-test after imputation using K-nearest neighbor (KNN) method. The number of neighbors (K) was 10 as a default setting. Statistical significance was set at P < 0.05. *, [§], and [†] indicate the significance between healthy controls versus COVID-19 moderate, healthy controls versus COVID-19 severe, and COVID-19 moderate versus severe patients, respectively. WBC, white blood cell; RBC, red blood cell; hsCRP, high-sensitivity C-reactive protein; UIBC, unsaturated iron binding capacity; ALT, alanine transaminase; AST, aspartate aminotransferase; BUN, blood urea nitrogen; CPK, creatine phosphokinase; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

This reveals an association between cytokines and COVID-19 severity and facilitates prediction of COVID-19 disease development (5, 22, 23). Additionally, integrated analysis with a matched transcriptome at the single-cell level can provide a better understanding of dynamic inflammatory responses under multi-layered control. Detailed procedures to perform blood biochemistry and cytokine profiling are described in the Supplementary Methods.

HLA genotyping

Here, we focused on genetic polymorphisms in *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DR*, *HLA-DP*, and *HLA-DQ* genes, which are associated with antiviral immune responses (Supplementary Table 3) (24). Association between these HLA gene polymorphisms and COVID-19 incidence and mortality have been previously

reported (25, 26). Detailed quality control (QC) metrices of all HLA typing data are given in Supplementary Table 4. *HLA-A* 02:01* and *HLA-A*24:02* were the most frequent types in both COVID-19 patients and healthy controls, followed by *HLA-A* 02:06* and *HLA-A*11:01*. Studies have reported these *HLA-A* types to possibly be associated with T-cell-mediated antiviral responses to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (26-28). However, we could not find a significant difference between the frequencies of these types in COVID-19 patients and healthy controls and could not, therefore, establish their association with COVID-19 severity in this study.

*HLA-B*15:01* and *HLA-B*46:01* alleles, previously reported to be associated with COVID-19 severity (27), were also frequently detected in this study, albeit without statistically significant differences between their levels in COVID-19 patients and healthy controls. *HLA-B*44:03*, also known to be frequent in the Korean population (29), had a frequency of approximately 9% in each group.

Additionally, we found relatively high frequencies of *HLA-C* 01:02*, *HLA-C*04:01*, and *HLA-C*07:02*, which were previously reported to be associated with COVID-19 occurrence and mortality (27). Particularly, *HLA-C*01:02*, which weakly binds to COVID-19 viral peptides (30), was the most frequent, albeit without significant differences between frequencies in COVID-19 patients and healthy controls. *HLA-C*03:02*, *HLA-C*03:03*, and *HLA-C*03:04* had relatively highly frequencies in moderate COVID-19 patients and healthy controls.

It has been reported HLA polymorphisms play a pivotal role in immune response for COVID-19 infection, especially resulting from pathogen-derived peptide presentation (31). Our HLA typing data along with matched clinic and other omics data would contribute to delineate detailed mechanisms by which HLA genotypes influence on susceptibility, progression, and severity of COVID-19 infection.

Viral genome sequencing and COVID-19 lineage analysis
In this study, 279 sequences of COVIDSeq data were collected, which excluded QC-failed data. The variants of COVID-19 sequences were analyzed to determine the variation in COVID-19 patients in terms of Pango lineage (32) and GISAID Clade (33) (https://www.gisaid.org) (Supplementary Fig. 3).

Fourteen COVID-19 lineages were identified in this study (Supplementary Fig. 3A). B.1.497 was the most common lineage found in our study, followed by B.1.619. The COVID-19 lineages mostly showed a change from B.1.497 to B.1.619 in biospecimens collected from January to June 2021 (Supplementary Fig. 3B). From the virus samples collected between January and March, 89.3% were of the B.1.497 lineage, which was mainly prevalent in South Korea (32). The proportions of B.1.619 lineage in the virus samples collected in April, May, and June were 8.3%, 49.0%, and 50.0%, respectively. Additionally, the B.1.620 lineage was present in virus samples collected in May and June at proportions of 15.6% and 25.0%, respectively.

Longitudinal multi-omics profiling based on COVID-19 infection
Multi-omics datasets were generated for longitudinal time points to investigate genomic and immunogenic dynamics in COVID-19 patients (Fig. 1). The omics datasets collected at three time points for moderate patients and seven time points for severe patients included scRNA-seq(+scTCR/BCR-seq), bulk TCR/BCR-seq, and cytokine profiling.

For scRNAseq data, 4,525 cells on average were captured and 24,530 genes on average were detected, and a total of 11,921,825 cells were obtained for 483 samples. A total of 184,549 B cells with complete heavy and light chain sequences and 386,553 T cells with complete TCR alpha and beta chain sequences were obtained (Supplementary Tables 5-7). Quality control data for bulk TCR/BCR-seq were also recorded in Supplementary Tables 8 and 9. In addition, WGS were also performed for the samples at the first time points to study genome-wide genetic variants of the participants. Detailed QC metrices of all WGS data are given in Supplementary Table 10.

The entire set of multi-omics data are available through the Clinical and Omics Data Archive (CODA) to researchers with approval. And an in-depth analysis of these datasets is also available on an analytical platform on CODA (https://coda. nih.go.kr). Detailed procedures to perform multi-omics datasets were described in Supplementary Methods.

Strength and limitations
This study presents a comprehensive collection of longitudinal multi-modal data of Korean COVID-19 patients and healthy controls based on state-of-the-art high-resolution technologies. The multi-omics data generated in this study includes matched clinical and laboratory testing data, thus acting as a powerful data source to delineate dysregulated immune biology and the underlying mechanisms against novel viruses.

There are several international resources for COVID-19 research. Global Initiative for Sharing All Influenza Data (GISAID) share over 11 million genomes to track the evolution of COVID-19 viruses and to study the worldwide genomic epidemiology of SARS-CoV-2 in real-time (33). The UK biobank also provides health record data on a regular basis to facilitate rapid research of COVID-19 and to obtain epidemiological insights into molecular characteristics of COVID-19 (34, 35), combined with the extensive data previously collected on genetic factors on UK biobank participants.

Compared to these international resources of COVID-19, our multi-omics study has tremendous strengths to study COVID-19 pathogenesis. The time-resolved set of multi-omics data and biospecimens collected from each patient allow to in-depth study of the dynamic progression of COVID-19. Moreover, the entire set of multi-omics, including single-cell omics, bulk TCR/BCRseq and cytokine profiling, was performed simultaneously for each time point, enabling the investigation of the relationship between dynamic cytokine levels and peripheral immune response and revealing patient-specific immune

responses. Considering that clinical data for each patient were collected daily, our dataset can also provide an in-depth understanding of disease progression after COVID-19 infection from the clinic point of view. Additionally, our results include a dataset for the early response to COVID-19 infection in moderate and severe patients that can be used to develop risk prediction models via multi-omics integration methodologies and deep learning. The dataset of multi-omics associated with specific COVID-19 lineages (Supplementary Fig. 3) can facilitate the study of coordinated immune responses strongly correlated with COVID-19 pathogenesis (36).

We expect that the combined single-cell omics, WGS, HLA typing, cytokine profiles, laboratory testing, and matched clinical data in our study will contribute to future integrative meta-analyses of COVID-19 to help investigate potential functions or mechanisms driving COVID-19 infection (37).

This COVID-19 project has several limitations. First, as the participants were recruited from October 2020 to June 2021, new variants of COVID-19, such as delta or omicron, were not included in this project. We are collecting the same set of resources for additional COVID-19 patients and vaccinated participants for a subsequent project, expecting to obtain biospecimens for novel COVID-19 variants and study antibody responses against COVID-19 vaccines (38). This can enable exploration of the immunologic landscape based on COVID-19 lineages or of complications after COVID-19 infection. Second, the biospecimens were collected after COVID-19 diagnosis, thereby having a possibility of variations in the exact time point of COVID-19 infection. Therefore, the very early immunological responses could not be included in this dataset. Lastly, as the datasets in this study were produced based on analysis of blood samples, local immune responses in infected tissues such as lungs or organs could not be accounted for (39, 40).

CONCLUSION

Our dataset generated based on biospecimens from COVID-19 patients can facilitate a better understanding of the dynamic peripheral immune responses during COVID-19 infection and be used to develop predictive models for estimating the severity or newly emerging viruses. This study can potentially uncover the genetic and biological basis of COVID-19 by combining their relationship with the clinical phenotypes. We anticipate that our data will provide a valuable resource for future COVID-19 studies and integrative meta-analyses of multi-omics datasets of COVID-19 patients worldwide.

DATA AVAILABILITY

All data and biospecimens from this project are available from the National Biobank of Korea (NBK) on prior approval (https:// nih.go.kr/biobank). Fundamental datasets, including clinical, laboratory testing and cytokine profiling data, were deposited at NBK, while the multi-omics dataset was deposited at the Clinical and Omics Data Archive (CODA; https://coda.nih.go.kr).

The accession numbers for the clinical data, laboratory characteristics, cytokine profiling, WGS, HLA typing, bulk TCRseq, bulfk BCRseq, and scRNA(+scTCR/BCR)-seq reported in this study are CODA-000034, CODA-000035, CODA-000036, CODA-000037, CODA-000038, CODA-000039, CODA-000040, and CODA-000041, respectively.

Any additional information required to analyze the dataset collected in this project is available from the lead contact upon request.

ACKNOWLEDGEMENTS

CONFLICTS OF INTEREST

The authors have no conflicting interests.

MATERIALS AND METHODS

Further detailed information is provided in the Supplementary Information.

REFERENCES

1. Coronaviridae Study Group of the International Committee on Taxonomy of V (2020) The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 5, 536-544
2. World Health Organization (January 2020) Surveillance case definitions for human infection with novel coronavirus (nCoV): interim guidance v1. (https://apps.who.int/iris/handle/10665/330376.
3. Ministry of Health and Welfare (10 March 2022); https://ncov.mohw.go.kr/.
4. Stephenson E, Reynolds G, Botting RA et al (2021) Single-cell multi-omics analysis of the immune response in COVID-19. Nat Med 27, 904-916
5. Bernardes JP, Mishra N, Tran F et al (2020) Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19. Immunity 53, 1296-1314 e1299
6. Wang S, Yao X, Ma S et al (2021) A single-cell transcriptomic landscape of the lungs of patients with COVID-19. Nat Cell Biol 23, 1314-1328

7. Delorey TM, Ziegler CGK, Heimberg G et al (2021) COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. Nature 595, 107-113

8. Overmyer KA, Shishkova E, Miller IJ et al (2021) Large-scale multi-omic analysis of COVID-19 severity. Cell Syst 12, 23-40 e27

9. Wu P, Chen D, Ding W et al (2021) The trans-omics landscape of COVID-19. Nat Commun 12, 4543

10. Huang SF, Huang YC, Chang FY et al (2020) Rapid establishment of a COVID-19 biobank in NHRI by National Biobank Consortium of Taiwan. Biomed J 43, 314-317

11. Holub P, Kozera L, Florindi F et al (2020) BBMRI-ERIC's contributions to research and knowledge exchange on COVID-19. Eur J Hum Genet 28, 728-731

12. Feldstein LR, Rose EB, Horwitz SM et al (2020) Multisystem inflammatory syndrome in U.S. children and adolescents. N Engl J Med 383, 334-346

13. Iyer M, Jayaramayya K, Subramaniam MD et al (2020) COVID-19: an update on diagnostic and therapeutic approaches. BMB Rep 53, 191-205

14. National Institute of Health (NIH) (March 2022) Clinical spectrum of SARS-CoV-2 infection : https://www.covid19 treatmentguidelines.nih.gov/overview/clinical-spectrum/

15. An H, Zhang J, Li T et al (2022) Inflammation/coagulopathy/immunology responsive index predicts poor COVID-19 prognosis. Front Cell Infect Microbiol 12, 807332

16. He C, Liu C, Yang J et al (2022) Prognostic significance of day-by-day in-hospital blood pressure variability in COVID-19 patients with hypertension. J Clin Hypertens (Greenwich) 24, 224-233

17. Kappert K, Jahic A and Tauber R (2020) Assessment of serum ferritin as a biomarker in COVID-19: bystander or participant? Insights by comparison with other infectious and non-infectious diseases. Biomarkers 25, 616-625

18. Carubbi F, Salvati L, Alunno A et al (2021) Ferritin is associated with the severity of lung involvement but not with worse prognosis in patients with COVID-19: data from two Italian COVID-19 units. Sci Rep 11, 4863

19. Alunno A, Carubbi F and Rodriguez-Carrio J (2020) Storm, typhoon, cyclone or hurricane in patients with COVID-19? Beware of the same storm that has a different origin. RMD Open 6, e001295

20. Chakurkar V, Rajapurkar M, Lele S et al (2021) Increased serum catalytic iron may mediate tissue injury and death in patients with COVID-19. Sci Rep 11, 19618

21. Biamonte F, Botta C, Mazzitelli M et al (2021) Combined lymphocyte/monocyte count, D-dimer and iron status predict COVID-19 course and outcome in a long-term care facility. J Transl Med 19, 79

22. Han H, Ma Q, Li C et al (2020) Profiling serum cytokines in COVID-19 patients reveals IL-6 and IL-10 are disease severity predictors. Emerg Microbes Infect 9, 1123-1130

23. Ling L, Chen Z, Lui G et al (2021) Longitudinal cytokine profile in patients with mild to critical COVID-19. Front Immunol 12, 763292

24. Kwok AJ, Mentzer A and Knight JC (2021) Host genetics and infectious disease: new tools, insights and translational opportunities. Nat Rev Genet 22, 137-153

25. Sakuraba A, Haider H and Sato T (2020) Population difference in allele frequency of HLA-C*05 and its correlation with COVID-19 mortality. Viruses 12, 1333

26. Tomita Y, Ikeda T, Sato R and Sakagami T (2020) Association between HLA gene polymorphisms and mortality of COVID-19: an in silico analysis. Immun Inflamm Dis 8, 684-694

27. Deng H, Yan X and Yuan L (2021) Human genetic basis of coronavirus disease 2019. Signal Transduct Target Ther 6, 344

28. Nesterenko PA, McLaughlin J, Tsai BL et al (2021) HLA-A(*)02:01 restricted T cell receptors against the highly conserved SARS-CoV-2 polymerase cross-react with human coronaviruses. Cell Rep 37, 110167

29. Park HJ, Kim YJ, Kim DH et al (2016) HLA allele frequencies in 5802 Koreans: varied allele types associated with SJS/TEN according to culprit drugs. Yonsei Med J 57, 118-126

30. Barquera R, Collen E, Di D et al (2020) Binding affinities of 438 HLA proteins to complete proteomes of seven pandemic viruses and distributions of strongest and weakest HLA peptide binders in populations worldwide. HLA 96, 277-298

31. Migliorini F, Torsiello E, Spiezia F, Oliva F, Tingart M and Maffulli N (2021) Association between HLA genotypes and COVID-19 susceptibility, severity and progression: a comprehensive review of the literature. Eur J Med Res 26, 84

32. Rambaut A, Holmes EC, O'Toole A et al (2020) A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 5, 1403-1407

33. Khare S, Gurry C, Freitas L et al (2021) GISAID's role in pandemic response. China CDC Wkly 3, 1049-1051

34. Kolin DA, Kulm S and Elemento O (2020) Clinical and genetic characteristics of COVID-19 patients from UK Biobank. medRxiv 2020.05.05.20075507

35. Kolin DA, Kulm S, Christos PJ and Elemento O (2020) Clinical, regional, and genetic characteristics of Covid-19 patients from UK Biobank. PLoS One 15, e0241264

36. Choi H and Shin EC (2022) Hyper-inflammatory responses in COVID-19 and anti-inflammatory therapeutic approaches. BMB Rep 55, 11-19

37. Tian Y, Carpp LN, Miller HER, Zager M, Newell EW and Gottardo R (2022) Single-cell immunology of SARS-CoV-2 infection. Nat Biotechnol 40, 30-41

38. Wei J, Stoesser N, Matthews PC et al (2021) Antibody responses to SARS-CoV-2 vaccines in 45,965 adults from the general population of the United Kingdom. Nat Microbiol 6, 1140-1149

39. Kayaaslan B and Guner R (2021) COVID-19 and the liver: a brief and core review. World J Hepatol 13, 2013-2023

40. Maiese A, Manetti AC, La Russa R et al (2021) Autopsy findings in COVID-19-related deaths: a literature review. Forensic Sci Med Pathol 17, 279-296