J Chin Med Assoc

ORIGINAL ARTICLE

Predictive biosignatures for hospitalization in patients with virologically confirmed COVID-19

Kung-Hao Liang[a,b,c,d,*], Yu-Chun Chen[e,f,g], Chun-Yi Hsu[a,f], Zih-Kai Kao[f,h], Ping-Hsing Tsai[a,f], Hsin-Yi Huang[f,h], Yuan-Chia Chu[f,h,i], Hsiang-Ling Ho[j,k], Yi-Chu Liao[l,m,n], Yi-Chung Lee[l,m,n], Chi-Cheng Huang[o,p,q], Tzu-Chun Wei[r,s,t], Yi-Jia Liao[l], Yung-Hsiu Lu[u], Chen-Tsung Kuo[h], Shih-Hwa Chiou[a,v,w]

[a]Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [b]Biosafety Level 3 Laboratory, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [c]College of Medicine, Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, Taipei, Taiwan, ROC; [d]School of Pharmaceutical Sciences, Institute of Food Safety and Health Risk Assessment, National Yang Ming Chiao Tung University, Taipei, Taiwan, ROC; [e]Department of Family Medicine, Taipei Veterans General Hospital, Yuli Branch, Hualien, Taiwan, ROC; [f]Department of Medical Research, Big Data Center, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [g]School of Medicine, College of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan, ROC; [h]Department of Information Management, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [i]Department of Information Management, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan, ROC; [j]Department of Pathology and Laboratory Medicine, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [k]Department of Biotechnology and Laboratory Science in Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan, ROC; [l]Departments of Neurology, Taipei Veterans General Hospital, Taiwan, ROC; [m]Department of Neurology, School of Medicine, National Yang Ming Chao Tung University, Taipei, Taiwan, ROC; [n]College of Medicine, Brain Research Center, National Yang Ming Chao Tung University School of Medicine, Taipei, Taiwan, ROC; [o]Department of Surgery, Comprehensive Breast Health Center, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [p]Division of Breast Surgery, Department of Surgery, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [q]College of Public Health, Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan, ROC; [r]Department of Urology, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [s]Department of Urology, School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan, ROC; [t]College of Medicine, Shu-Tien Urological Institute, National Yang Ming Chiao Tung University, Taipei, Taiwan, ROC; [u]Department of Pediatrics, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; [v]Institute of Pharmacology, College of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan, ROC; [w]Department of Ophthalmology, Taipei Veterans General Hospital, Taipei, Taiwan, ROC

Abstract

Background: Coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus, presents with varying severity among individuals. Both viral and host factors can influence the severity of acute and chronic COVID-19, with chronic COVID-19 commonly referred to as long COVID. SARS-CoV-2 infection can be properly diagnosed by performing real-time reverse transcription polymerase chain reaction analysis of nasal swab samples. Pulse oximetry, chest X-ray, and complete blood count (CBC) analysis can be used to assess the condition of the patient to ensure that the appropriate medical care is delivered. This study aimed to develop biosignatures that can be used to distinguish between patients who are likely to develop severe disease and require hospitalization from patients who can be safely monitored in less intensive settings.

Methods: A retrospective investigation was conducted on 7897 adult patients with virologically confirmed SARS-CoV-2 infection between January 26, 2020, and November 30, 2023; all patients underwent comprehensive CBC testing at Taipei Veterans General Hospital. Among them, 1867 patients were independently recruited for a population study involving genome-wide genotyping of approximately 424 000 genomic variants. Therefore, the participants were divided into two patient cohorts, one with genomic data (n = 1867) and one without (n = 6030) for model validation and training, respectively.

Results: We constructed and validated a biosignature model by using a combination of CBC measurements to predict subsequent hospitalization events (hazard ratio = 3.38, 95% confidence interval: 3.07-3.73 for the training cohort and 3.03 [2.46-3.73] for the validation cohort; both $p < 10^{-8}$). The obtained scores were used to identify the top quartile of patients, who formed the "very high risk" group with a significantly higher cumulative incidence of hospitalization (log-rank $p < 10^{-8}$ in both the training and validation cohorts). The "very high risk" group exhibited a cumulative hospitalization rate of >60%, whereas the rate for the other patients was approximately 30% over a 1.5-year period, providing a binary classification of patients with distinct hospitalization risks. To investigate the genetic factors mediating this risk, we conducted a genome-wide association study. Specific regions in chromosomes 7 and 10 and the mitochondrial chromosome (M), harboring IKAROS family zinc finger 1 (*IKZF1*), actin binding LIM protein 1 (*ABLIM1*), and mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 3 (*MT-ND3*), exhibited prominent associations with binary risk classification. The identified exonic variants of *IKZF1* are linked to several autoimmune diseases. Notably, people with different genotypes of the leading variants (rs4132601, rs141492519, and Affx-120744614) exhibited varying cumulative hospitalization rates after infection.

Conclusion: We successfully developed and validated a biosignature model of COVID-19 severe disease in virologically confirmed patients. The identified genomic variants provide new insights for infectious disease research and medical care.

Keywords: COVID-19; Genome-wide association study; SARS-CoV-2

1. INTRODUCTION

Coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2 virus), emerged in late 2019 as an unprecedented infectious disease. To date, a considerable proportion of the global population has been infected. Although most individuals with COVID-19 recovered quickly, some experienced severe or fatal outcomes.[1] This virus is an enveloped, positive-strand RNA virus with numerous identified viral proteins, including spike, membrane, nucleocapsid, and envelope proteins.[2] The virus utilizes the angiotensin-converting enzyme 2 (ACE2) receptors on the human cell surface as gateways. After the virus enters the human cell, it replicates using its own RNA polymerase, producing viral particles that are subsequently released into the extracellular environment.[2] It can infect various types of human cells, including immune cells.[3]

The infection can cause acute and chronic forms of COVID-19, presenting with diverse clinical manifestations and affecting various parts of the human body.[1,4–7] Acute respiratory distress syndrome (ARDS) and subsequent multiorgan failure are major causes of mortality, characterized by diffuse microthrombi in the pulmonary microcirculation that disrupt respiratory function and immune cell–mediated hypercoagulability.[8] Cytokine storms are a major cause of ARDS and mortality.[9] The severity and persistence of COVID-19 symptoms are associated with pre-existing conditions such as cardiovascular disease[10,11] and diabetes.[4] Endothelial disease, thromboembolic complications (eg, deep vein thrombosis and pulmonary embolism), and disseminated intravascular coagulation are potentially life-threatening complications associated with SARS-CoV-2 infections.[12–15] Notably, myocarditis, pulmonary embolism, and stroke pose considerable risks for heart failure or arrhythmias,[12,14–22] and SARS-CoV-2 infection–induced encephalitis can lead to neurological deficits and even death.[23] The infection can also cause chronic conditions collectively termed long COVID.[5,7,24–28]

Virological real-time reverse transcription polymerase chain reaction and antigen tests using nasal or throat swab samples are crucial for diagnosing SARS-CoV-2 infection. In addition, pulse oximetry, chest X-ray, and hematological complete blood count (CBC) examinations are often performed to provide valuable insights into patient health. CBC is a general evaluation of health status based on blood cell counts (eg, red blood cells, white blood cells, and platelets),[29] which are often altered by SARS-CoV-2 infections.[1] Neutrophilia, thrombocytopenia, and leukopenia are often observed in SARS-CoV-2 infections.[30] Developing a predictive model for COVID-19 outcomes can enable healthcare providers to identify high-risk patients early in their illness, thereby facilitating timely interventions to reduce disease severity and improve patient outcomes.[31] The present study investigated disease progression by using hospitalization for all causes as a surrogate indicator of severe illness. We focused on patients with confirmed COVID-19 to create a biosignature model for predicting hospitalization risk. Given the relationship between CBC and immune response to infection, we explored the associations between various CBC parameters and the course of SARS-CoV-2 infection. With the successful development of the model, we plan to investigate genetic variants associated with hospitalization risk.

2. METHODS

2.1. Patients

The present retrospective study was approved by the institutional review board of Taipei Veterans General Hospital, Taipei, Taiwan. We investigated a cohort of 7897 virologically confirmed patients, who were diagnosed with SARS-CoV-2 through the nasal swab real-time reverse transcription polymerase chain reaction test between January 26, 2020, and November 30, 2023, and underwent comprehensive CBC examinations at Taipei Veterans General Hospital (Fig. 1A). Among them, 1867 patients were also independently recruited for the Taipei Precision Medicine Initiative (TPMI) project during their visit to the hospital. Genome-wide genotype data were obtained from the TPMI participants by genotyping peripheral blood mononuclear cells using Affymetrix Axiom Genome-Wide TWB array plates. Therefore, the 7897 participants were divided into two patient populations, one with genotype data (n = 1867) and one without (n = 6790) for model validation and training, respectively (Fig. 1A). The clinical information of the patients is listed in Table 1.

2.2. Data and statistical analysis

This study aims to develop biosignatures to distinguish between patients likely to develop severe disease and require hospitalization and those who can be safely monitored in less intensive settings. To protect patients' privacy, we integrated deidentified blood biochemical test data, hospitalization records, and genome-wide genotypes (if available for research) by using computer-generated keys and lookup tables, without including any identifiable information such as patient name and chart numbers. Univariate analysis was performed using the Cox proportional hazards model, and multivariate biosignatures were generated using the generalized iterative modeling method.[32,33] A genome-wide association study (GWAS) was performed on the basis of the χ^2 test of allelic counts.

3. RESULTS

3.1. Predictive biosignature models for post SARS-CoV-2 infection hospitalization events

A univariate analysis was conducted on the associations of CBCs with the time to postinfection hospitalization events. The hazard ratios (HRs) with respect to CBC variables in the patients with training cohort are listed in Fig. 1B. Red cell distribution width (RDW) had a significant and positive association (HR: 1.045, confidence interval [CI]: 1.035-1.055) with time to postinfection hospitalization events, indicating that a higher RDW value is associated with increased risk. Furthermore, higher mean corpuscular hemoglobin (MCH; HR: 1.032, 95% CI: 1.019-1.044), monocyte count (MONO) (HR: 1.025, 95% CI: 1.015-1.036), mean corpuscular volume (MCV) (HR: 1.021, 95% CI: 1.016-1.026), band cell count (BAND) (HR: 1.014, 95% CI: 1.005-1.023), and segmented

*Address correspondence. Dr. Kung-Hao Liang, Department of Medical Research, Taipei Veterans General Hospital, 201, Section 2, Shi-Pai Road, Taipei 112, Taiwan, ROC. E-mail address: khliang@vghtpe.gov.tw (K.-H. Liang).

Author contributions: Dr. Kung-Hao Liang and Dr. Yu-Chun Chen contributed equally to this work.

Conflicts of interest: Dr. Yu-Chun Chen and Dr. Shih-Hwa Chiou, editorial board members at the Journal of the Chinese Medical Association, have no roles in the peer review process of or decision to publish this article. The other authors declare that they have no conflicts of interest related to the subject matter or materials discussed in this article.

Journal of Chinese Medical Association. (2025) 88: 246-252.

Received July 5, 2024; accepted November 12, 2024.

doi: 10.1097/JCMA.0000000000001203

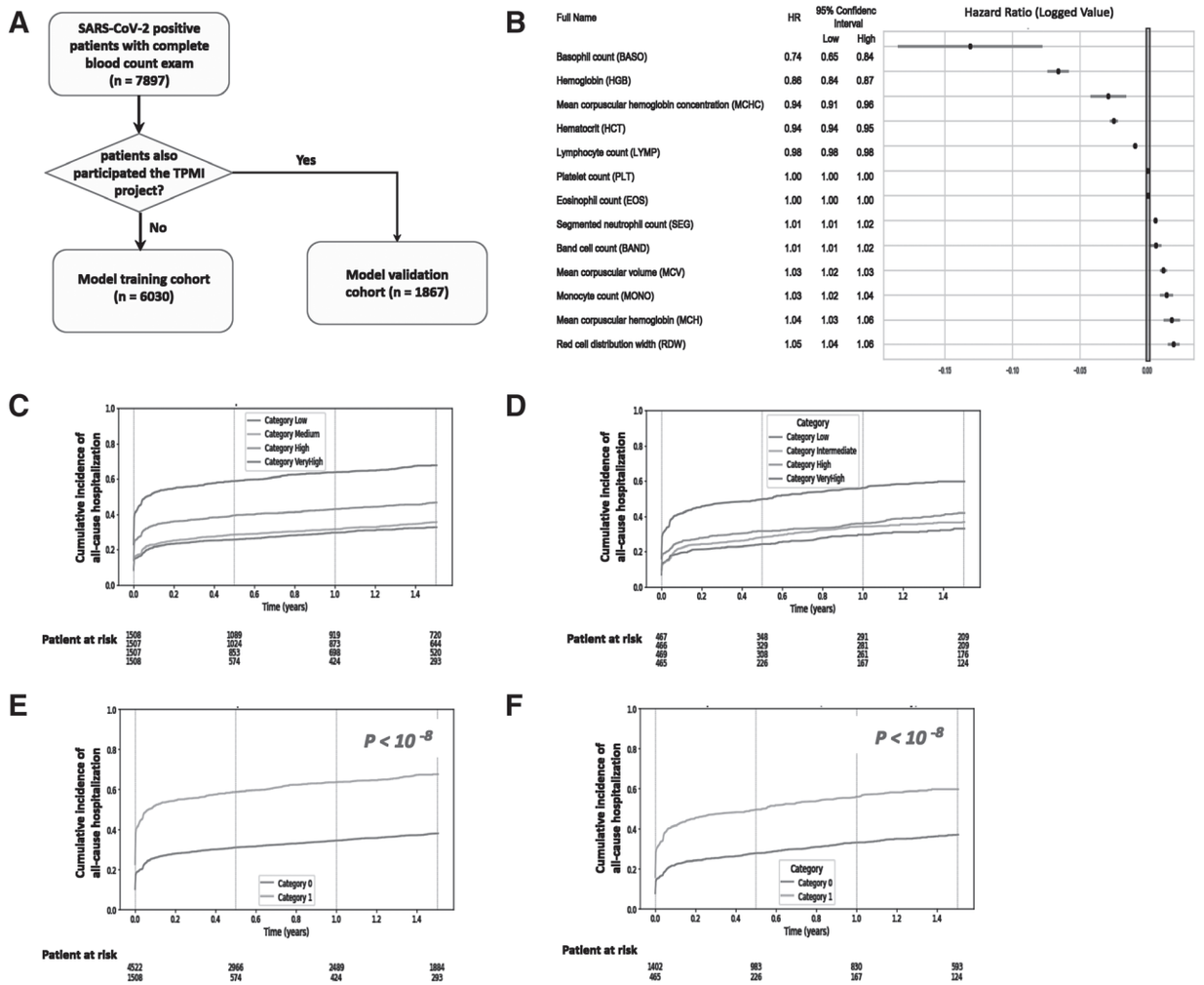

Fig. 1 Construction and validation of a clinical risk model for all-cause hospitalization events in patients diagnosed with SARS-CoV-2 infection through nasal swab RT-PCR tests. A, Participants were categorized into those with and without available genome-wide genotypes, forming two study cohorts for clinical model validation and training, respectively. B, Univariate analysis of complete blood count items associated with time to hospitalization events after infection. C, Longitudinal representation of cumulative incidence of patients in the training cohort, stratified into quartiles based on risk model values during a 1.5-year follow-up period postinfection. D, Longitudinal representation of patients in the validation cohort, stratified into risk quartiles. E, Longitudinal representation of patients in the training cohort, stratified into two groups (category 1: very high-risk; category 0: others). F, Longitudinal representation of patients in the validation cohort, stratified into two groups. RT-PCR = real-time reverse transcription polymerase chain reaction; SARS-CoV-2 = severe acute respiratory syndrome coronavirus 2.

neutrophil count (SEG) (HR: 1.012, 95% CI: 1.010-1.015) were all associated with increased hospitalization risk. By contrast, lymphocyte count (LYMP; HR: 0.982, 95% CI: 0.979-0.985), hematocrit (HCT) (HR: 0.947, 95% CI: 0.941-0.953), mean corpuscular hemoglobin concentration (MCHC; HR: 0.921, 95% CI: 0.896-0.948), hemoglobin (HGB; HR: 0.862, 95% CI: 0.847-0.877), platelet count (PLT) (HR: 1.000, 95% CI: 1.000-1.000), and basophil count (BASO) (HR: 0.796, 95% CI: 0.712-0.890) were all associated with decreased hospitalization risk. These findings suggest that these CBC variables can be used to generate a biosignature predict subsequent hospitalization events after infection, given that the combination of the variables correlated with the incidence of postinfection hospitalization events in both the training and validation cohorts. A multivariate biosignature model was subsequently developed. The biosignature model *f(x)* of individual variables *x* (CBC items in the present study) corresponds to the exponential term in the Cox regression equation:

$$h(t|x) = h_0(t)\, e^{f(x)}, \tag{1}$$

where $h(t|x)$ is the time (t)-dependent hazard function for an individual with given CBC values represented as a tensor x and $h_0(t)$ is the baseline hazard at time t. $f(x)$ is defined as

$$f(x) = \mathrm{MCHC} \times (-0.1821) + \mathrm{LYMP} \times (-0.0123) + \mathrm{HGB} \times (-0.1515) + \mathrm{MCHC} \times \mathrm{MCHC} \times \mathrm{MCH} \times (0.000060950).$$

The HR of $f(x)$ was 3.38 in the training cohort (95% CI: 3.07-3.73, $p < 10^{-8}$) and 3.03 in the validation cohort (95% CI: 2.46-3.73, $p < 10^{-8}$). $f(x)$ values were used to stratify the patients into quartiles (ie, very high–risk, high-risk, intermediate-risk, and low-risk quartiles) on the basis of their cumulative incidence of hospitalization (Fig. 1C). The $f(x)$ function in Equation (1) was then applied directly to the patients with validation cohort. Subsequently, the four equal-sized patient strata exhibited distinct cumulative incidences of hospitalization (Fig. 1D).

Table 1

The clinical information of participants in this study

Item	All participants	Training	Validation
Patient number	7897	6030	1867
Age, y	57.86 ± 20.16	57.94 ± 20.55	57.60 ± 18.86
Sex			
Male	3609 (45.7%)	2869 (47.6%)	740 (39.6%)
Female	4287 (54.3%)	3160 (52.4%)	1127 (60.4%)
Complete blood count			
Hemoglobin, g/dL	13.16 ± 1.93	13.17 ± 1.96	13.14 ± 1.84
Platelets, 1000/μL	236.77 ± 81.23	236.52 ± 81.16	237.55 ± 81.48
Mean corpuscular volume, femtoliters	89.66 ± 7.68	89.68 ± 7.68	89.59 ± 7.66
Hematocrit, %	39.59 ± 5.51	39.63 ± 5.55	39.45 ± 5.36
Mean corpuscular haemoglobin, pg/cell	29.76 ± 3.08	29.76 ± 3.08	29.77 ± 3.09
Red cell distribution width, %	13.96 ± 2.91	13.90 ± 2.68	14.16 ± 3.54
Mean corpuscular hemoglobin concentration, g/dL	33.14 ± 1.20	33.13 ± 1.21	33.18 ± 1.17
Eosinophils, %	3.60 ± 37.89	3.11 ± 29.45	5.16 ± 57.17
Lymphocytes, %	24.34 ± 11.40	23.85 ± 11.57	25.92 ± 10.68
Monocytes, %	7.44 ± 3.08	7.44 ± 3.13	7.47 ± 2.91
Basophils, %	0.55 ± 0.35	0.54 ± 0.34	0.61 ± 0.37
Bandemia, %	0.14 ± 2.37	0.15 ± 2.55	0.10 ± 1.65
Segmented neutrophils, %	65.52 ± 12.91	66.09 ± 13.12	63.68 ± 12.05

Because the patients in the first strata consistently exhibited a higher cumulative incidence relative to the other three strata, we compared this stratum (ie, the very high–risk group) with the other patients, for binary risk categorization. In the training cohort, patients in the two categories exhibited distinct cumulative incidences of hospitalization (Fig. 1E, $p < 10^{-8}$). A significant difference was also identified in the validation cohort (Fig. 1F, $p < 10^{-8}$).

3.2. Genetic determinants of post SARS-CoV-2 infection hospitalization risks

To gain genetic insights into hospitalization risks, we performed a GWAS, a case-control study where the case (n = 465) and control (n = 1402) patients were defined by binary risk categorization (Fig. 1F). The data points in the Manhattan plot (Fig. 2A) represent the strength of association for genomic variants, organized by genomic location. The P-P plot in Fig. 2B visually presents the results pertaining to the quality of this analysis, with an inflation factor of 1.024 suggesting good quality. Specific regions in chromosomes 7 and 10 and the mitochondria (M), encoding IKAROS family zinc finger 1 (*IKZF1*), actin binding LIM protein 1 (*ABLIM1*), and mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 3 (*MT-ND3*) were revealed to harbor genomic variants significantly associated with risk levels ($p < 5 \times 10^{-6}$). The identified variants in chromosome 7 resided in the *IKZF1* gene, as detailed in the zoomed-in plot in Figure 2C. The leading exonic variants, rs4132601 and rs11980379, were both

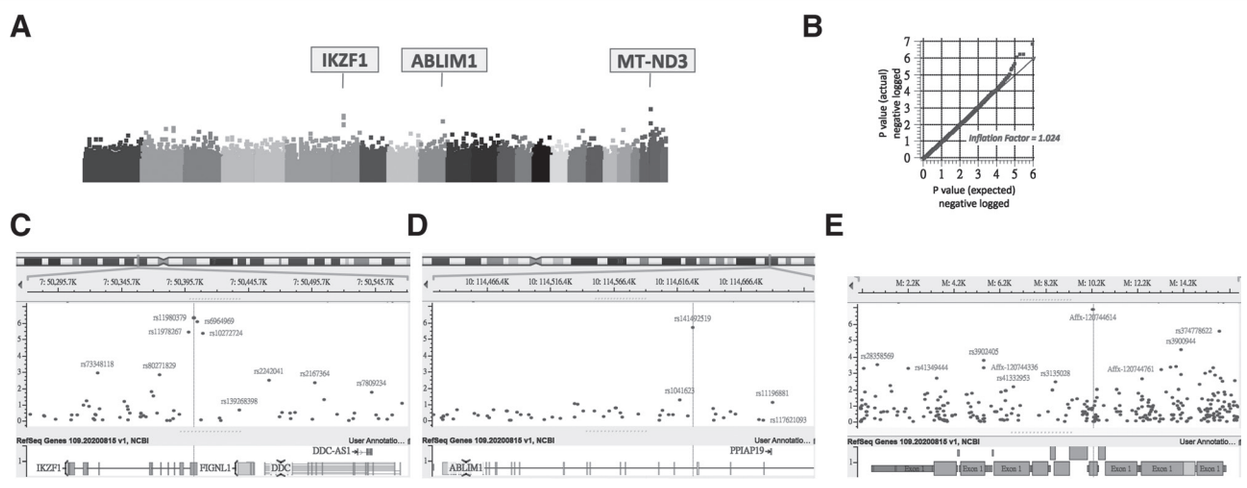

Fig. 2 Associations of genome-wide genomic variants with risk categories. A, Manhattan plot showing the significance of association (negative log-transformed *p* value in the vertical axis) for individual genetic variants, plotted according to their genomic locations (horizontal axis). Data from different chromosomes are represented in different colors. B, P-P plot comparing the observed *p* values from the GWAS (vertical axis) with the expected *p* values (horizontal axis) under the null hypothesis of no association, providing a visual assessment of analysis quality. For this analysis, the inflation factor was 1.024. C, Zoomed-in plot of chromosomes 7 (50271425-50571425). D, Zoomed-in plot of chromosome 10 (114423792-114904664). E, Zoomed-in plot of mitochondria genome (M), manifesting as a strong signal of association. ABLIM1 = actin binding LIM protein 1; GWAS = genome-wide association study; IKZF1 = IKAROS family zinc finger 1; MT-ND3 = mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 3.

Table 2

The identified genomic variants associated with post SARS-CoV-2 infection hospitalization

Genomic variant	Chr	Position	Gene	OR	OR 95% CI low	OR 95% CI high	Minor allele	Major allele	MAF case group	MAF control group
rs11978267	7	50398606	*IKZF1*	1.60	1.31	1.96	G	A	0.182	0.122
rs11980379	7	50402283	*IKZF1*	1.68	1.37	2.05	C	T	0.181	0.116
rs4132601	7	50402906	*IKZF1*	1.68	1.37	2.05	G	T	0.181	0.116
rs6964969	7	50405553	*IKZF1*	1.66	1.36	2.04	G	A	0.180	0.116
rs10272724	7	50409515	*IKZF1*	1.60	1.31	1.97	C	T	0.178	0.119
rs141492519	10	114628407	*ABLIM1*	3.30	1.96	5.55	T	C	0.032	0.010
Affx-120744614	M	10211	*MT-ND3*	18.36	4.10	82.17	T	C	0.013	0.001
rs374778622	M	15712	*MT-ND3*	15.20	3.32	69.49	G	A	0.011	0.001

ABLIM1 = actin binding LIM protein 1; Chr = chromosome; CI = confidence interval; *IKZF1* = IKAROS family zinc finger 1; MAF = minor allele frequencies; *MT-ND3* = mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 3; OR = odds ratio; SARS-CoV-2 = severe acute respiratory syndrome coronavirus 2.

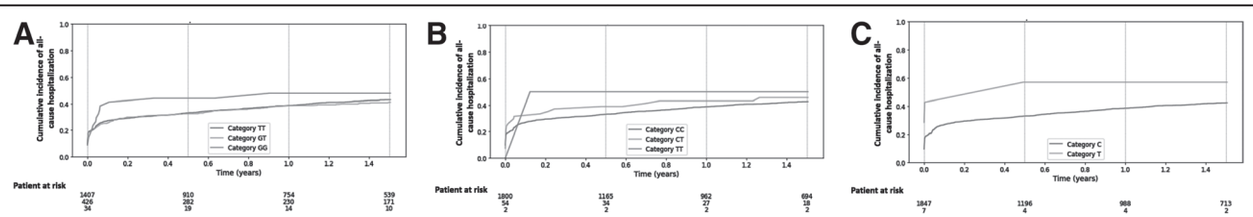

Fig. 3 Genotype-specific cumulative incidence of postinfection hospitalization. A, Cumulative incidence of postinfection hospitalization of patients, stratified by genotypes of rs4132601 on chromosome 7. B, Cumulative incidence of hospitalization of patients, stratified by genotype rs141492519 on chromosome 10. C, Cumulative incidence of hospitalization of patients stratified by genotype Affx-120744614 on mitochondrial genome.

located in the 3′-untranslated region (3′-UTR) of the *IKZF1* gene (Fig. 2C). The minor allele "G" of the leading variant rs4132601 (T/G) conferred an odds ratio of 1.68 (95% CI: 1.37-2.05) for the very high–risk group. The genotypes of this variant did not deviate significantly from the Hardy-Weinberg equilibrium in both the case and control groups (p = 0.610 and 0.730, respectively). We identified five nearby variants in linkage disequilibrium, all reaching statistical significance ($p < 5 \times 10^{-5}$; Table 2). Chromosome 10 harbored *ABLIM1*, where the leading variant rs141492519 was located in the intronic region (Fig. 2D). The minor allele "T" of the leading variant rs141492519 (C/T) conferred an odds ratio of 3.30 (95% CI: 1.96-5.55) to the very-high-risk group. Similarly, no significant deviation from the Hardy-Weinberg equilibrium was identified in both the case and control groups (p = 0.081 and 0.931, respectively). The mitochondrial genome harbored *MT-ND3* and the leading variants Affx-120744614 and rs374778622 (Fig. 2E). The minor allele frequencies (MAF) and relevant information of the leading variants are summarized in Table 2.

We then investigated the cumulative incidence of all-cause hospitalization among the patients, stratified by the genotypes of the leading variants rs4132601, rs141492519, and Affx-120744614. For the rs4132601 on chromosome 7, the data indicated higher hospitalization rates in patients with the GG (homozygous minor allele) genotype than in those with the GT and TT genotypes (Fig. 3A). For the rs141492519 on chromosome 10, the cumulative incidences of hospitalization exhibited discernible genotype-related differences (Fig. 3B), with the TT (homozygous minor allele) genotype exhibiting higher risks. For the Affx-120744614 on the mitochondrial genome, differences in hospitalization rates were identified between patients with C and T alleles (Fig. 3C). These findings suggest that germline variants influence the risk of severe disease outcomes in patients with COVID-19.

4. DISCUSSION

When COVID-19 initially emerged in Taiwan, nearly all affected patients were hospitalized, regardless of the severity of their symptoms. This universal hospitalization strategy was born out of an abundance of caution, which can be attributed to the limited clinical understanding of COVID-19 severity and progression at that time. Although the goal was to monitor and manage potential complications, this strategy led to the hospitalization of numerous patients who might have required only observation. The present study examines this context, highlighting the need for more refined risk stratification tools that can help distinguish between patients likely to develop severe disease and require hospitalization and those who can be safely monitored in less intensive settings.

The present study developed and validated a biosignature model on the basis of CBC values and genetic factors of virologically confirmed patients to predict the risk of severe COVID-19 progression. The model was validated in a large cohort, allowing for the early identification of patients at elevated risk of severe outcomes.

CBC is closely related to the physiological and immunological conditions of patients, which may affect disease outcomes. Anemia, which has various causes, is a medical indication identifiable through CBC, specifically low HGB and HCT levels. An increase in the number of white blood cells, particularly neutrophils, can indicate inflammation in the body. In addition, abnormalities in red blood cells, white blood cells, and PLTs can indicate bone marrow disorders such as leukemia, lymphoma, and myelodysplastic syndrome. An increase in HCT levels can indicate dehydration. Abnormalities in white blood cell counts can be suggestive of autoimmune disorders, including lupus and rheumatoid arthritis. Abnormal PLTs can indicate platelet disorders such as thrombocytopenia, which may be related to liver fibrosis or thrombocytosis.

The proposed model incorporates MCHC, LYMP, HGB, and an interaction term that combines MCHC and MCH, enabling the effective identification of "very high risk" patients, namely those with a significantly higher risk of hospitalization. The incorporation of MCH and MCHC suggests the importance of HGB levels and concentrations in red blood cells for the efficient transport of oxygen under the

pathogenic influence of the virus. MCH has been identified as an independent predictor of a higher risk of long-term major adverse cardiovascular events.[34] In patients with heart failure with preserved ejection fraction, low MCHC was independently associated with worse clinical outcomes, such as any-cause hospitalization, in those with renal dysfunction.[35] In addition, viral infection may trigger an active host immune response, such as lymphocytosis, resulting in high LYMPs. In this context, lymphocytopenia may develop before and during the progression into severe disease. By identifying patients at higher risk of severe outcomes, the model can inform clinical decisions and enhance patient care.

A GWAS was conducted to identify genetic variants significantly associated with the risks of severe diseases. The genes in these associated genomic regions—*IKZF1* on chromosome 7, *ABLIM1* on chromosome 10, and *MT-ND3* in mitochondrial DNA—play key roles in viral–host interactions in the context of COVID-19 progression. The *IKZF1* gene, located on chromosome 7 at 7p12.2, encodes the Ikaros protein, a zinc finger transcription factor that is crucial for the development and function of lymphoid cells such as B cells.[36] The Ikaros protein also plays a pivotal role in the differentiation of several T helper cells (eg, T helper 1 [Th1], Th2, Th17, T follicular [Tfh], and T regulatory [Treg] cells) and the regulation of various immune responses.[37] *IKZF1* plays a crucial role in regulating cytokine expression and immune cell differentiation, and its proper function is essential for maintaining a balanced immune response. The leading exonic variants rs4132601, rs6964969, and their nearby variants in *IKZF1* have been implicated in B-cell deficiency,[38] autoimmune disorders such as Sjogren's syndrome,[39] systemic lupus erythematosus (SLE),[40] SLE nephritis,[41] and acute lymphoblastic leukemia,[42,43] all of which are associated with the dysregulation of immune pathways. One other exonic variant of *IKZF1* has been implicated in impairing the humoral immune response to SARS-CoV-2 vaccination, highlighting its potential role in modulating vaccine efficacy and adaptive immunity.[44] The present study is the first to suggest *IKZF1* as a potential contributor to aberrant immune response relevant to the severity and progression of COVID-19. The exonic variants in the 3′-UTR may affect the host's ability to mount a precise and effective immune response against SARS-CoV-2, or inducing immune overactivation that can cause cytokine storms and severe COVID-19 outcomes. The *ABLIM1* on chromosome 10 influences cytoskeletal organization and cellular signaling. Its potential role in viral infections may stem from its recently discovered function as a ubiquitin E3 ligase, which specifically targets IκBα for degradation, thereby activating the NF-κB/CCL-20 signaling pathway, a key player in the body's response to infection and inflammation.[45,46] Finally, the *MT-ND3* gene, part of the mitochondrial genome, encodes a subunit of the NADH dehydrogenase (complex I), which is involved in the mitochondrial respiratory chain. Variants in *MT-ND3* can affect cellular energy metabolism and oxidative stress responses, which are crucial in the context of viral infections where mitochondrial function is often disrupted.[47] Understanding the interplay between these genetic factors and viral pathogenesis can provide deeper insights into individual susceptibilities to severe disease and inform personalized therapeutic strategies. The results in Fig. 3 suggest that these variants can facilitate the stratification of patients to reflect their different levels of hospitalization risk. These genetic insights can lead to more targeted and effective interventions, ultimately improving patient care and management.

The findings of present study can help refine hospitalization decisions, reducing unnecessary admissions, and focusing on the patients most likely to benefit from hospitalization or even intensive care. Overall, integrating genetic and hematological indicators into predictive models for COVID-19 severity provides new insights into host–pathogen interactions and enhances the precision of clinical triage and resource allocation in managing infectious diseases.

In conclusion, we developed and successfully validated a biosignature for postinfection severe events in patients with virologically confirmed COVID-19. The identified genomic variants provide novel insights into both viral pathogenesis and patient care for SARS-CoV-2 infection. The results underscore the complex nature of integrating clinical and genetic data in the context of emerging infectious diseases such as COVID-19. This comprehensive model enhances our understanding of disease mechanisms and supports the development of targeted interventions for improving patient outcomes.

ACKNOWLEDGMENTS

This study was funded by Taipei Veterans General Hospital (V113E-002-4 and V114E-006-3), Academia Sinica (40-05-GMM, AS-GC-110-MD02, and 236e-1100202), and National Development Fund, Executive Yuan (NSTC 111-3114-Y-001-001). Data for analysis were provided by the Big Data Center and the Information Department, Taipei Veterans General Hospital, which is really appreciated. The interpretations and conclusions contained herein do not represent the position of Taipei Veterans General Hospital. This manuscript was edited by Wallace Academic Editing.

The authors thank all the participants and investigators from Taiwan Precision Medicine Initiative.

REFERENCES

1. Machhi J, Herskovitz J, Senan AM, Dutta D, Nath B, Oleynikov MD, et al. The natural history, pathobiology, and clinical manifestations of SARS-CoV-2 infections. *J Neuroimmune Pharmacol* 2020;15:359–86.
2. Liang KH, Teng YC, Liao YT, Yarmishyn AA, Chiang SH, Hung WC, et al. The natural history of SARS-CoV-2-incurred disease: from infection to long COVID. *J Transl Med* 2024;4:72–86.
3. Matveeva O, Nechipurenko Y, Lagutkin D, Yegorov YE, Kzhyshkowska J. SARS-CoV-2 infection of phagocytic immune cells and COVID-19 pathology: antibody-dependent as well as independent cell entry. *Front Immunol* 2022;13:1050478.
4. Su Y, Yuan D, Chen DG, Ng RH, Wang K, Choi J, et al; ISB-Swedish COVID-19 Biobanking Unit. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell* 2022;185:881–95.e20.
5. Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. *Nat Rev Microbiol* 2023;21:133–46.
6. dos Santos WG. Natural history of COVID-19 and current knowledge on treatment therapeutic options. *Biomed Pharmacother* 2020;129:110493.
7. Hastie CE, Lowe DJ, McAuley A, Mills NL, Winter AJ, Black C, et al. Natural history of long-COVID in a nationwide, population cohort study. *Nat Commun* 2023;14:3504.
8. Selickman J, Vrettou CS, Mentzelopoulos SD, Marini JJ. COVID-19-related ARDS: key mechanistic features and treatments. *J Clin Med* 2022;11:4896.
9. Montazersaheb S, Hosseiniyan Khatibi SM, Hejazi MS, Tarhriz V, Farjami A, Ghasemian Sorbeni F, et al. COVID-19 infection: an overview on cytokine storm and related interventions. *Virol J* 2022;19:92.
10. Vasbinder A, Meloche C, Azam TU, Anderson E, Catalan T, Shadid H, et al; STOP-COVID Investigators‡. Relationship between preexisting cardiovascular disease and death and cardiovascular outcomes in critically ill patients with COVID-19. *Circ Cardiovasc Qual Outcomes* 2022;15:e008942.
11. Yamakawa M, Kuno T, Mikami T, Takagi H, Gronseth G. Clinical characteristics of stroke with COVID-19: a systematic review and meta-analysis. *J Stroke Cerebrovasc Dis* 2020;29:105288.
12. Giordo R, Paliogiannis P, Mangoni AA, Pintus G. SARS-CoV-2 and endothelial cell interaction in COVID-19: molecular perspectives. *Vasc Biol* 2021;3:R15–23.

13. Rabaan AA, Al-Ahmed SH, Muhammad J, Khan A, Sule AA, Tirupathi R, et al. Role of inflammatory cytokines in COVID-19 patients: a review on molecular mechanisms, immune functions, immunopathology and immunomodulatory drugs to counter cytokine storm. *Vaccines (Basel)* 2021;9:436.
14. Vrints CJM, Krychtiuk KA, Van Craenenbroeck EM, Segers VF, Price S, Heidbuchel H. Endothelialitis plays a central role in the pathophysiology of severe COVID-19 and its cardiovascular complications. *Acta Cardiol* 2021;76:109–24.
15. Won T, Wood MK, Hughes DM, Talor MV, Ma Z, Schneider J, et al. Endothelial thrombomodulin downregulation caused by hypoxia contributes to severe infiltration and coagulopathy in COVID-19 patient lungs. *EBioMedicine* 2022;75:103812.
16. Ackermann M, Verleden SE, Kuehnel M, Haverich A, Welte T, Laenger F, et al. Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in Covid-19. *N Engl J Med* 2020;383:120–8.
17. Artifoni M, Danic G, Gautier G, Gicquel P, Boutoille D, Raffi F, et al. Systematic assessment of venous thromboembolism in COVID-19 patients receiving thromboprophylaxis: incidence and role of D-dimer as predictive factors. *J Thromb Thrombolysis* 2020;50:211–6.
18. Ambrosino P, Calcaterra IL, Mosella M, Formisano R, D'Anna SE, Bachetti T, et al. Endothelial dysfunction in COVID-19: a unifying mechanism and a potential therapeutic target. *Biomedicines* 2022;10:812.
19. Garnier Y, Claude L, Hermand P, Sachou E, Claes A, Desplan K, et al. Plasma microparticles of intubated COVID-19 patients cause endothelial cell death, neutrophil adhesion and netosis, in a phosphatidylserine-dependent manner. *Br J Haematol* 2022;196:1159–69.
20. Gavriilaki E, Anyfanti P, Gavriilaki M, Lazaridis A, Douma S, Gkaliagkousi E. Endothelial dysfunction in COVID-19: lessons learned from coronaviruses. *Curr Hypertens Rep* 2020;22:63.
21. Smadja DM, Guerin CL, Chocron R, Yatim N, Boussier J, Gendron N, et al. Angiopoietin-2 as a marker of endothelial activation is a good predictor factor for intensive care unit admission of COVID-19 patients. *Angiogenesis* 2020;23:611–20.
22. Siddiqi HK, Libby P, Ridker PM. COVID-19—a vascular disease. *Trends Cardiovasc Med* 2021;31:1–5.
23. Tyagi K, Rai P, Gautam A, Kaur H, Kapoor S, Suttee A, et al. Neurological manifestations of SARS-CoV-2: complexity, mechanism and associated disorders. *Eur J Med Res* 2023;28:307.
24. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV; WHO Clinical Case Definition Working Group on Post-COVID-19 Condition. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis* 2022;22:e102–7.
25. Ayoubkhani D, Bermingham C, Pouwels KB, Glickman M, Nafilyan V, Zaccardi F, et al. Trajectory of long covid symptoms after covid-19 vaccination: community based cohort study. *BMJ* 2022;377:e069676.
26. Caze AB, Cerqueira-Silva T, Bomfim AP, de Souza GL, Azevedo AC, Brasil MQ, et al. Prevalence and risk factors for long COVID after mild disease: a cohort study with a symptomatic control group. *J Glob Health* 2023;13:06015.
27. Dagher H, Chaftari AM, Subbiah IM, Malek AE, Jiang Y, Lamie P, et al. Long COVID in cancer patients: preponderance of symptoms in majority of patients over long time period. *eLife* 2023;12:e81182.
28. Tana C, Bentivegna E, Cho SJ, Harriott AM, Garcia-Azorin D, Labastida-Ramirez A, et al. Long COVID headache. *J Headache Pain* 2022;23:93.
29. Seo IH, Lee YJ. Usefulness of complete blood count (CBC) to assess cardiovascular and metabolic diseases in clinical settings: a comprehensive literature review. *Biomedicines* 2022;10:2697.
30. Violetis OA, Chasouraki AM, Giannou AM, Baraboutis IG. COVID-19 infection and haematological involvement: a review of epidemiology, pathophysiology and prognosis of full blood count findings. *SN Compr Clin Med* 2020;2:1089–93.
31. Berenguer J, Borobia AM, Ryan P, Rodríguez-Baño J, Bellón JM, Jarrín I, et al; COVID-19@Spain and COVID@HULP Study Groups. Development and validation of a prediction model for 30-day mortality in hospitalised patients with COVID-19: the COVID-19 SEIMC score. *Thorax* 2021;76:920–9.
32. Huang YH, Liang KH, Chien RN, et al. A circulating microRNA signature capable of assessing the risk of hepatocellular carcinoma in cirrhotic patients. *Sci Rep* 2017;7:523.
33. Liang KH, Zhang P, Lin CL, Wang SC, Hu TH, Yeh CT, et al. Morphomic signatures derived from computed tomography predict hepatocellular carcinoma occurrence in cirrhotic patients. *Dig Dis Sci* 2020;65:2130–9.
34. Zhang Z, Gao S, Dong M, Luo J, Xu C, Wen W, et al. Relationship between red blood cell indices (MCV, MCH, and MCHC) and major adverse cardiovascular events in anemic and nonanemic patients with acute coronary syndrome. *Dis Markers* 2022;2022:2193343.
35. Choy M, Zhen Z, Dong B, Chen C, Dong Y, Liu C, et al. Mean corpuscular haemoglobin concentration and outcomes in heart failure with preserved ejection fraction. *ESC Heart Fail* 2023;10:1214–21.
36. Kirstetter P, Thomas M, Dierich A, Kastner P, Chan S. Ikaros is critical for B cell differentiation and function. *Eur J Immunol* 2002;32:720–30.
37. Powell MD, Read KA, Sreekumar BK, Oestreich KJ. Ikaros zinc finger transcription factors: regulators of cytokine signaling pathways and CD4(+) T helper cell differentiation. *Front Immunol* 2019;10:1299.
38. Nunes-Santos CJ, Kuehn HS, Rosenzweig SD. IKAROS family zinc finger 1–associated diseases in primary immunodeficiency patients. *Immunol Allergy Clin North Am* 2020;40:461–70.
39. Qu S, Du Y, Chang S, Guo L, Fang K, Li Y, et al. Common variants near IKZF1 are associated with primary Sjogren's syndrome in Han Chinese. *PLoS One* 2017;12:e0177320.
40. Chen L, Niu Q, Huang Z, Yang B, Wu Y, Zhang J. IKZF1 polymorphisms are associated with susceptibility, cytokine levels, and clinical features in systemic lupus erythematosus. *Medicine (Baltim)* 2020;99:e22607.
41. Mosaad YM, Hammad A, Shouma A, Darwish M, Hammad EM, Sallam RA, et al. IKZF1 rs4132601 and rs11978267 gene polymorphisms and paediatric systemic lupus erythematosus; relation to lupus nephritis. *Int J Immunogenet* 2024;51:173–82.
42. Mahjoub S, Chayeb V, Zitouni H, Ghali RM, Regaieg H, Almawi WY, et al. IKZF1 genetic variants rs4132601 and rs11978267 and acute lymphoblastic leukemia risk in Tunisian children: a case-control study. *BMC Med Genet* 2019;20:159.
43. Evans TJ, Milne E, Anderson D, de Klerk NH, Jamieson SE, Talseth-Palmer BA, et al. Confirmation of childhood acute lymphoblastic leukemia variants, ARID5B and IKZF1, and interaction with parental environmental exposures. *PLoS One* 2014;9:e110255.
44. Diaz-Alberola I, Espuch-Oliver A, Garcia-Aznar JM, Ganoza-Gallardo C, Aguilera-Franco M, Sampedro A, et al. Common variable immunodeficiency associated with a de novo IKZF1 variant and a low humoral immune response to the SARS-CoV-2 vaccine. *J Clin Med* 2022;11:2303.
45. He Y, Shi Q, Ling Y, Guo H, Fei Y, Wu R, et al. ABLIM1, a novel ubiquitin E3 ligase, promotes growth and metastasis of colorectal cancer through targeting IkBα ubiquitination and activating NF-κB signaling. *Cell Death Differ* 2024;31:203–16.
46. Jin SH, Kim H, Gu DR, Park KH, Lee YR, Choi Y, et al. Actin-binding LIM protein 1 regulates receptor activator of NF-κB ligand-mediated osteoclast differentiation and motility. *BMB Rep* 2018;51:356–61.
47. Purandare N, Ghosalkar E, Grossman LI, Aras S. Mitochondrial oxidative phosphorylation in viral infections. *Viruses* 2023;15:2380.