Association Between the *LZTFL1* rs11385942 Polymorphism and COVID-19 Severity in Colombian Population

Mariana Angulo-Aguado [1†], David Corredor-Orlandelli [1†], Juan Camilo Carrillo-Martínez [1], Mónica Gonzalez-Cornejo [1], Eliana Pineda-Mateus [1], Carolina Rojas [1], Paula Triana-Fonseca [2], Nora Constanza Contreras Bravo [1], Adrien Morel [1], Katherine Parra Abaunza [3], Carlos M. Restrepo [1], Dora Janeth Fonseca-Mendoza [1‡] and Oscar Ortega-Recalde [1*‡]*

[1] *Center for Research in Genetics and Genomics – CIGGUR, GENIUROS Research Group, School of Medicine and Health Sciences, Universidad Del Rosario, Bogotá, Colombia,* [2] *Department of Molecular Diagnosis, Genética Molecular de Colombia SAS, Bogotá, Colombia,* [3] *Hospital Universitario Mayor – Méderi – Universidad del Rosario, Bogotá, Colombia*

Genetic and non-genetic factors are responsible for the high interindividual variability in the response to SARS-CoV-2. Although numerous genetic polymorphisms have been identified as risk factors for severe COVID-19, these remain understudied in Latin-American populations. This study evaluated the association of non-genetic factors and three polymorphisms: *ACE* rs4646994, *ACE2* rs2285666, and *LZTFL1* rs11385942, with COVID severity and long-term symptoms by using a case-control design. The control group was composed of asymptomatic/mild cases ($n = 61$) recruited from a private laboratory, while the case group was composed of severe/critical patients ($n = 63$) hospitalized in the Hospital Universitario Mayor-Méderi, both institutions located in Bogotá, Colombia. Clinical follow up and exhaustive revision of medical records allowed us to assess non-genetic factors. Genotypification of the polymorphism of interest was performed by amplicon size analysis and Sanger sequencing. In agreement with previous reports, we found a statistically significant association between age, male sex, and comorbidities, such as hypertension and type 2 diabetes mellitus (T2DM), and worst outcomes. We identified the polymorphism *LZTFL1* rs11385942 as an important risk factor for hospitalization ($p < 0.01$; OR = 5.73; 95% CI = 1.2–26.5, under the allelic test). Furthermore, long-term symptoms were common among the studied population and associated with disease severity. No association between the polymorphisms examined and long-term symptoms was found. Comparison of allelic frequencies with other populations revealed significant differences for the three polymorphisms investigated. Finally, we used the statistically significant genetic and non-genetic variables to develop a predictive logistic regression model, which was implemented in a Shiny web application. Model discrimination was assessed using the area under the receiver operating characteristic curve (AUC = 0.86; 95% confidence interval 0.79–0.93). These results suggest that *LZTFL1* rs11385942 may be a potential biomarker for COVID-19 severity in addition to conventional non-genetic risk factors.

A better understanding of the impact of these genetic risk factors may be useful to prioritize high-risk individuals and decrease the morbimortality caused by SARS-CoV2 and future pandemics.

INTRODUCTION

SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) is a novel coronavirus, first identified in China in late December 2019 (1). The disease caused by this virus, named COVID-19, rapidly spread across the globe being declared a pandemic by the WHO in March 2021 (2). Up to the first week of March 2022, more than 450 million confirmed cases and 6 million deaths were reported worldwide, from which ∼6 million confirmed cases and 139.000 deaths occurred in Colombia (3, 4). The clinical course and severity of COVID-19 disease are highly variable among individuals, ranging from asymptomatic cases to severe respiratory failure and death (5).

Different clinical risk factors, including aging, male sex and comorbidities such as cardiovascular disease, hypertension, diabetes mellitus, chronic obstructive pulmonary lung disease, immunosuppression and obesity have been linked to more severe courses of COVID-19 (6, 7). Importantly, numerous studies have shown that host genetic factors also play a critical role in SARS-CoV-2 disease progression and severity (8–10). Early works suggested a potential role of genes related to the renin-angiotensin-aldosterone system (RAAS) (*ACE1* and *ACE2*), the ABO blood group system and the human leukocyte antigen (HLA) (11–13). The RAAS pathway is a physiological system that plays an important role in the homeostatic control of blood pressure and body water-electrolyte balance (14). Angiotensin I converting enzyme and angiotensin converting enzyme 2, coded by the genes *ACE* and *ACE2*, respectively, are critical regulators of this pathway and may also contribute to multiple organ injuries in COVID-19. In lung vascular endothelium, ACE catalyzes Angiotensin I conversion into Angiotensin II, an active peptide that promotes vasoconstriction, inflammation and thrombosis (15). Conversely, ACE2 converts Angiotensin II into angiotensin-(1–7), molecules that counteract the effects of Angiotensin II, including vasodilatation and vascular protection (16). Polymorphisms that increase *ACE* expression have been associated with more severe COVID-19 infections. The *ACE* insertion(Ins)/deletion(Del) polymorphism (rs4646994) is of particular interest as the resulting decrease in ACE activity has been linked to a protective effect in Ins allele carriers (17). Moreover, ACE2 has a dual role as the SARS-CoV-2 receptor, allowing virus internalization, and as RAAS regulator, catalyzing angiotensin II degradation (16, 18). Whole exome studies (WES) have identified more than 30 variants in the *ACE2* gene, potentially interfering with protein structure, stabilization and expression, and contributing to the high interindividual variability and susceptibility to COVID-19 (19). Among these variants, NM_001371415.1:c.439+4G>A (rs2285666) polymorphism is related to an increase of 50% of ACE2 expression, compared to wild-type G/G genotype carriers,

and decreases the risk of severe SARS-CoV2 infection (20). In addition, two large genome-wide association studies, oriented to find genetic susceptibly locus, identified an association signal at chromosome 3p21.31 (rs11385942 and rs10490770) as the one with the most significant association with respiratory failure and mechanical ventilation requirement amongst severe COVID-19 patients (21, 22). This locus contains several genes related to cell signaling and solute transportation, including *CCR9*, *CXCR6*, *LZTFL1*, and *SLC6A20*. *LZTFL1* gene, the most promising candidate, codifies for a protein involved in the primary cilia function and the immunological synapse between T-cells and antigen-presenting cells (23).

Despite their relevance, genetic host factors related to COVID-19 severity remain understudied in Latin-American populations, limiting their potential use as predictive biomarkers and the development of predictive models. Furthermore, the study of these factors is particularly relevant considering that Latin-American countries have been severely affected by the COVID-19 pandemic. In this study, we performed an ambispective case-control analysis to evaluate the association between non-genetic factors and genetic factors, including the polymorphisms rs4646994 (*ACE*), rs2285666 (*ACE2*), and rs11385942 (*LZTFL1*), and COVID-19 severity and long-term symptoms in Colombian population. The results of this study support a positive association between the *LZTFL1* rs11385942 locus variant and an increased risk of severe SARS-CoV-2 infection. Furthermore, we developed a predictive model integrating non-genetic and genetic factors, potentially useful to identify high-risk individuals and prioritize prevention and mitigation efforts.

METHODS

Study Population and Sampling

This study enrolled 145 patients between 18 and 60 years with confirmed diagnosis of COVID-19 by positive RT-PCR (reverse transcriptase polymerase chain reaction), antigens or antibodies (IgG and/or IgM for SARS-CoV-2) tests. The control group consisted of 71 patients who were classified as asymptomatic or mild COVID-19, group non-hospitalized. The case group was composed of 74 patients with severe or critical disease, group hospitalized. Subcategorization of the case group was made with patients critically ill who required intensive care unit (ICU), group hospitalized-ICU. Clinical severity was determined according to national guidelines for COVID-19 by the Colombian Health Ministry (24). Cases were recruited among hospitalized patients at the Hospital Universitario Mayor-Méderi (Bogotá, Colombia). Controls were enrolled from a private laboratory (Genética Molecular de Colombia, Bogotá, Colombia). Cases and controls were invited to participate in this

study and those who accepted signed an informed consent and underwent buccal swap or peripheral blood sampling. Patients were enrolled between December 2020–July 2021 and all subjects were unvaccinated at the time of recruitment.

The sample size was calculated with a p (sample proportion) of 7% according to the minimum allele frequency (MAF) for the allele with the reported lowest frequency, in our case the polymorphism rs11385942, a confidence level of 95% (α = 0.05, z = 1.96), a margin of error (e) of 5%, and a population size N = 8,000,000 for Bogotá city. Using the formula n = $Nz^2*p(1\text{-}p)/\alpha^2(N\text{-}1)+z^2*p(1\text{-}p)$, implemented in the OpenEpi web-tool, we estimated that the minimum sample size was 101 (25). This value was approximated to 145 individuals considering possible clinical follow up lost. Given this is the first study to assess allele frequency for the polymorphisms of interest in Colombian population, MAF were obtained from the GnomAD database for Latino-American individuals (26). This study followed the guidelines of the Declaration of Helsinki and all experimental procedures were approved by the Ethics Committee of Universidad del Rosario (DVO005 1543-CV1334).

Clinical Data Collection and Follow Up

Data collection and clinical follow up were conducted through phone calls at least 21 days after the diagnosis. Data was obtained through a standardized format that included the following clinical and demographical information: sex, age, blood type, medical history, comorbidities, drugs use, symptoms, long-term symptoms, and any change in disease severity. Furthermore, we performed an exhaustive revision of clinical records of hospitalized patients to validate the information collected previously and verify the clinical classification and severity criteria according to the clinical guidelines mentioned before. One hundred and twenty four patients, 61 cases and 63 controls, completed the clinical follow up and continued in the study.

DNA Extraction and Genotyping

Total genomic DNA was obtained from buccal swab or blood samples using either the Quick-DNA™ Miniprep Plus Kit (Zymo Research) or the Buccal Swab DNA Kit (Promega). The buccal swab samples were collected in a cotton swab and the blood samples were collected in EDTA tubes, 5 mL for patient. Genomic DNA was quantified using a nanodrop spectrophotometer. All samples were aliquoted and stored at 4°C until analysis. Polymerase chain reaction (PCR) was used to amplify and genotype three polymorphisms of interest: *ACE* 289bp ALU Ins/Del (rs4646994), *ACE2* c.439+4G>A (rs2285666), and *LZTFL1* c.323+621dup (rs11385942). Primers were designed using PrimerBlast (27). Primers sequences and PCR conditions are listed in **Supplementary Table 1**. For *ACE* rs4646994 genotyping, PCR products were run on a 1% agarose gel stained by ethidium bromide and amplicon sizes were used to determine individual genotypes. Fragments obtained were 191 bp for the Del allele and 480 bp for the Ins allele. For *ACE2* rs2285666 and *LZTL1* rs11385942, PCR products were purified and sequenced through Sanger method. Sequences were analyzed with the software Geneious Prime v2021.2 (Biomatters) (28). Genotypes were assigned in batches of 20 samples by

two independent researchers. In case the results were in disagreement, a third researcher reassessed the results and a final consensus was achieved. These researchers were blind to the case-control status of the individuals. Genotypification was attempted in 125 individuals, being successful in 124 (99.2%).

Statistical Analysis and Predictive Model

A bivariate analysis was performed between clinical and demographic variables with the severe COVID-19 outcome (non-hospitalized vs. hospitalized, including UCI and non-UCI patients) or the presence of long-term COVID-19 symptoms using the χ^2, Mann-Whitney and OR statistics. All the analyses were conducted using this case-control definition unless otherwise stated. Significant thresholds were set as p < 0.05, and a 95% confidence interval for the OR. Long-term COVID-19 symptoms were defined as persistent symptoms beyond 3 weeks from initial symptoms onset (29). An extended analysis of long-term symptoms was performed grouping symptoms into the following categories: (1) frequent (fatigue, headache, attention deficit, alopecia, dyspnea), (2) organ system affected (neurological, psychiatric, osteomuscular, respiratory, and cardiovascular), and (3) others including the ones with low sample and literature prevalence (dysphagia, otorhinolaryngological, ophthalmological and cutaneous manifestations) according to Lopez-Leon et al. (30).

Population genetic statistics, including allelic frequencies, genotypic frequencies and Hardy–Weinberg equilibrium (HWE), were calculated using the SNPStats software (31). The deviation of the HWE was established using a χ^2 goodness-of-fit test with 1° of freedom (df) except for the SNP in *ACE2* rs2285666 located in the X chromosome, for which HWE was determined using the R package "HWadmiX" (32). Allelic frequencies obtained from the study were compared to other populations using the χ^2 and Fisher's exact test statistics (21, 26, 33–46). p-values <0.05 were considered statistically significant.

The bivariate association analysis between genetic polymorphisms and severity outcome or the presence of long-term symptoms was performed with the PLINK software (47). Different genetic models, including allelic, genotypic, dominant and recessive, were assessed with the Cochran-Armitage trend, genotypic (2df), dominant gene action (1df), and recessive gene (1df) tests. In addition, a subgroup analysis between control (non-hospitalized) and ICU-hospitalized patients (n = 26) was conducted under the allelic model. The clinical and genetic variables with a significant correlation were used to build a multivariate logistic regression model in order to develop a predictive risk model for severe disease. Different combinations of variables were tested to construct the models, and these were compared using the Akaike Information Criterion (AIC) and the Coefficient of Discrimination D (Tjur's R2) parameters. This last method, Tjur's R2, is used for binomial logistic models and a value approaching 1 indicates that there is a clear separation between the predicted values for the response outcomes (48). For the model construction we evaluated and handled the potentially cofounding and interacting variables. We assessed the variation inflation factor (VIF) to protect our model to be inflated by multicollinearity, all the variables included had

FIGURE 1 | Flowchart of the study participants.
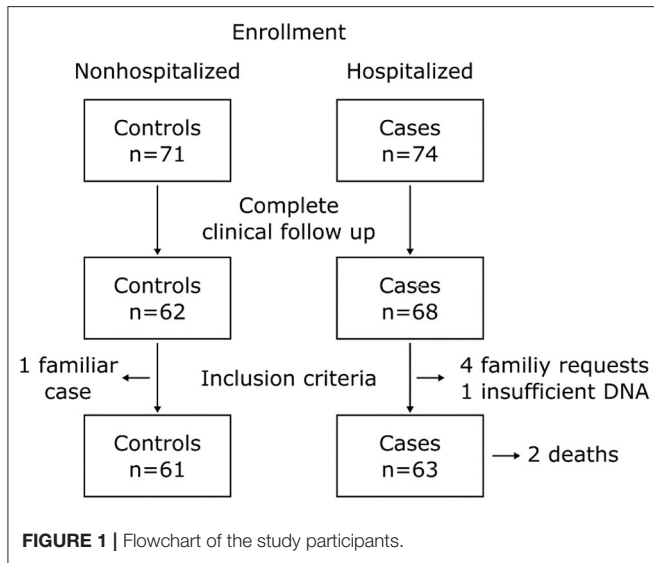
a VIF value of 1. Model comparison was assessed by calculating the area under the receiver operating characteristic curve (AUC), direct comparison between the scores obtained from the models, integrated discrimination improvement (IDI) and cross-validation parameters, including concordance, sensitivity, specificity, and net benefit at different cutoff probabilities. Concordance was defined as the correctly estimated outcomes using several cutoff values for the predicted affection probability. The IDI score and cross-validation parameters were calculated with the R packages PredictABEL and rmda, respectively (49, 50). Finally, an open-source and online application was developed for users to easily access and test the model. The predictive model was constructed using R v4.1.2 and the online application was built using the Shiny package for R (51).

RESULTS

Clinical and Demographic Data

In total 145 patients, 71 controls and 74 cases were enrolled in the study. Nine patients from the control group and six patients from the cases group were excluded from the study by loss to follow up. One control was excluded due to familial relationship, one case was excluded by insufficient DNA and four cases were excluded due to direct request from the family. The final number of patients included was 61 controls and 63 cases. Two patients from the cases group died due to COVID-19 complications; nevertheless, clinical follow up was completed with help of relatives. A summary of the study participants is presented in **Figure 1**. For the control group, 29.5% ($n = 18$) diagnoses were made by RT-PCR, 63.9% ($n = 39$) by antigen test, and 6.6% ($n = 4$) by antibodies. The sampling methods for this group were 67.2% ($n = 41$) by buccal swabs and 32.8% ($n = 20$) from peripheral blood. For the case group, 98.4% ($n = 62$) diagnoses were made by RT-PCR and 1.6% ($n = 1$) by antigen test and 100% samples were taken from peripheral blood.

Demographic and clinical characteristics of our study population are summarized in **Table 1**. The mean age for the control group was 36.6 ± 10.8 years and that for the case group was 47.3 ± 9.53 years. Men accounted 42.6% ($n = 26$) of controls and 65% ($n = 41$) of the case group. Among the most common comorbidities in our study population were type 2 diabetes mellitus (T2DM) 11.3% ($n = 14$), hypertension 16.1% ($n = 20$) and obesity 21.8% ($n = 27$). Most patients (56.5%, $n = 70$) presented no comorbidities, 27.4 ($n = 34$) patients had 1 comorbidity and 16.1% ($n = 20$) had two or more comorbidities. The different signs and symptoms observed in the patients are presented in **Table 2**. Respiratory symptoms were the most common, these included dyspnea 55.6% ($n = 69$) and cough 64.5% ($n = 80$), followed by systemic symptoms, including fever 52.4% ($n = 65$), fatigue 81.5% ($n = 101$) and osteomuscular pain 70.2% ($n = 87$). Long-term symptoms were frequent (57.3%, $n = 71$), these included common symptoms (39.5%, $n = 49$), respiratory (15.3%, $n = 19$), osteomuscular (8.9%, $n = 11$), neurologic (22.6%, $n = 28$) and psychiatric (19.4%, $n = 24$). Demographic and clinical characteristics in patients with and without long-term COVID-19 symptoms are presented in **Table 3**.

Clinical Association Analysis

Our study revealed a significant statistical correlation between SARS CoV-2 severity and multiple clinical variables reported previously, including age ($p < 0.01$), male sex ($p = 0.01$; OR = 2.51; 95% CI = 1.21–5.18), hypertension ($p < 0.01$; OR = 7.14; 95% CI = 1.97–25.88) and T2DM ($P < 0.01$; OR = 15.6; 95% CI = 1.97–123.42). Interestingly, other clinical variables, including blood group, cardiovascular, pulmonary, and other systemic diseases, such as cancer and obesity, were non-statistically significant in our sample ($p > 0.05$). Additionally, presence of no comorbidities was a protective factor ($p < 0.01$; OR = 0.17; 95% CI = 0.08–0.38) and presence of two or more comorbidities conferred an increased risk of severe disease ($p < 0.01$; OR = 11.8; 95% CI = 2.6–53.5). Symptoms who exhibited significant association with severe disease were mainly respiratory, systemic, and neurological, and included dyspnea ($p < 0.01$, OR = 29.54; 95% CI = 10.91–80.01), cough ($p < 0.01$; OR = 4.69; 95% IC=2.1–10.49) and fever ($p < 0.01$; OR = 4.41; 95% CI = 2.08–9.38) and mental status disturbance ($p = 0.04$; OR = 2.63; 95% CI = 1–6.93). In contrast, anosmia ($p < 0.01$; OR = 0.2; 95% CI = 0.09–0.42), ageusia ($p < 0.01$; OR = 0.30; 95% CI = 0.14–0.63), and headache ($p = 0.02$ OR = 0.42; 95% CI = 0.2–0.9) were more frequent in patients with mild disease (**Table 2**).

Presence of long-term symptoms was associated with disease severity ($p < 0.01$; OR = 3.37; 95% CI = 1.6–7.1). 42.6% patients in the control group developed these symptoms, in contrast to the 71.4% in the case group. Categories significantly different were common long-term symptoms ($p < 0.01$; OR = 6.91; 95% CI = 3.03–15.77), psychiatric ($p < 0.01$; OR = 34.5; 95% CI = 4.48–265.78) and respiratory ($p = 0.01$; OR = 4.45; 95% CI = 1.39–14.32), whereas cardiovascular long-term symptoms were present only in cases ($p < 0.01$). Multiple acute symptoms were associated with long-term symptoms, such as presence of fatigue

TABLE 1 | Demographic and clinical characteristics of the study population.

Variable	Controls (*n* = 61)	Cases (*n* = 63)	*p*-value	CI95%	OR
Age	36.6 (±10.8)	47.3 (±9.53)	<0.01*		
Male sex	26 (42.6%)	41 (65.0%)	0.01*	1.21–5.18	2.51
Blood group					
O	38 (62.3%)	42 (66.7%)	0.61	0.58–2.53	1.21
A	20 (32.8%)	12 (19.0%)	0.08	0.21–1.10	0.48
B	2 (3.3%)	2 (3.2%)	1	0.13–7.09	0.97
AB	1 (1.6%)	0 (0%)	0.98	–	–
Comorbidities					
Arrhythmia	0 (0%)	1 (1.58%)	1	–	–
Asthma	2 (3.27%)	1 (1.58%)	0.97	0.04–5.39	0.48
Autoimmune disease	0 (0%)	2 (3.17%)	0.49	–	
Cancer	1 (1.63%)	3 (4.76%)	0.63	0.30–29.66	3.00
Chronic kidney disease	5 (8.2%)	1 (1.58%)	0.22	0.59–45.63	5.17
COPD	0 (0%)	2 (3.2%)	0.49	–	–
Coronary disease	0 (0%)	2 (3.2%)	0.49	–	–
T2DM	1 (1.63%)	13 (20.6%)	<0.01*	1.97–123.42	15.6
Hypertension	3 (4.91%)	17 (26.9%)	<0.01*	1.97–25.88	7.14
HIV/Immunodeficiency	0 (0%)	2 (3.2%)	0.49	–	–
Obesity	9 (14.7%)	18 (28.5%)	0.09	0.95–5.65	2.31
No comorbidities	47 (77%)	23 (36.5%)	<0.01*	0.08–0.38	0.17
One comorbidity	12(19.7%)	22 (34.9%)	0.05	0.97–4.96	2.19
Two or more Comorbidities	2 (3.27%)	18 (28.5%)	<0.01*	2.60–53.50	11.80
Chronic use of steroids	1 (1.63%)	1 (1.58%)	1	0.06–15.83	0.97
Smoking history	28 (45.9%)	18 (28.5%)	0.05	0.23–1.02	0.48

Statistical significant, p-value < 0.05; COPD, Chronic obstructive pulmonary disease.

($p < 0.01$; OR = 5.12; 95% CI = 1.85–14.13), osteomuscular pain (0.04; OR = 2.26; 95% CI = 1.03–4.94), dyspnea ($p < 0.01$; OR = 4.96; 95% CI = 2.30–10.69), ageusia ($p = 0.01$; OR = 2.53; 95% CI = 1.22–5.27) and brain fog ($p = 0.02$; OR = 3.26; 95% CI 1.12–9.46).

Genetic Variants and Association Analysis

The *ACE* rs4646994 genotypic distribution in the total sample was 0.35 (43/124), 0.45 (56/124) and 0.2 (25/124) for Ins/Ins, Ins/Del and Del/Del, respectively. The allele frequency for the Del allele was 0.43 (106/248). For *ACE2* rs2285666, an X-linked SNP, the distribution was 0.5 (29/58), 0.4 (23/58) and 0.1 (6/58) for G/G and G/A and A/A genotypes, respectively, and 0.53 (35/66) and 0.47 (31/66) for G and A genotypes in hemizygous individuals, respectively. The allele frequency for the allele A was 0.36 (66/180). Finally, for *LZTFL1* rs11385942, the distribution was 0.9 (111/124) and 0.1 (13/124) for the genotypes WT/WT and WT/Ins, respectively. We did not observe homozygous individuals for the allele Ins. The allele frequency for this allele was 0.05 (13/235). Genotypic and allelic frequencies are presented in **Table 4**. All genotypes were found to be in HWE (*ACE* rs4646994 $p = 0.46$, *ACE2* rs2285666 $p = 0.25$ and *LZTFL1* rs11385942 $p = 1$). Genotype frequencies by clinical subgroups (controls, cases hospitalized

no ICU and cases hospitalized in ICU) are presented in **Supplementary Table 2**.

Bivariate analysis between the genetic polymorphisms and COVID-19 severity revealed a statistically significant association between the *LZTFL1* rs11385942 polymorphism with severe COVID-19 and severe COVID-19 requiring hospitalization in ICU ($p = 0.01$; OR = 5.73; 95% CI = 1.24–26.46 and $p = 0.02$; OR = 6.12; 95% CI = 1.14–32.63, respectively, under the allelic genetic model). No association was found between the *ACE* rs4646994 and *ACE2* rs2285666 polymorphisms, and COVID-19 severity under any of the models tested (**Table 5**). Nevertheless, an association between *ACE* rs4646994 Del and neurological long-term symptoms (e.g., ageusia, anosmia, and vertigo) was identified under the Cochran-Armitage test ($p < 0.01$; OR = 0.32; 95% CI = 0.16–0.63).

Population Genetic Analysis

Next, we compared the allelic frequencies obtained in this study with those of other datasets including populations of European, Asian, African, North American, and Latin-American ancestries (**Supplementary Table 3**). We found significant statistical differences for the three systems assessed. For *ACE* rs46469949, East Asia allelic frequencies were the only population with no statistical differences. For *ACE2* the rs2285666 allelic frequency found in our study was similar to those reported in Mexican

Variable	Controls (*n* = 61)	Cases (*n* = 63)	*p*-value	CI 95%	OR
Acute symptoms					
Ageusia	40 (65.5%)	23 (36.5%)	<0.01*	0.14–0.63	0.30
Anosmia	42 (68.8%)	19 (30.1%)	<0.01*	0.09–0.42	0.20
Cough	29 (47.5%)	51 (80.9%)	<0.01*	2.10–10.49	4.69
Diarrhoea	11 (18%)	20 (31.7%)	0.07	0.91–4.90	2.11
Dyspnea	13 (21.3%)	56 (88.8%)	<0.01*	10.91–80.01	29.54
Fatigue	42 (68.8%)	59 (93.6%)	<0.01*	2.12–21.04	6.67
Fever > 38°C	21 (34.4%)	44 (69.8%)	<0.01*	2.08–9.38	4.41
Haemoptysis	2 (3.2%)	6 (9.52%)	0.29	0.60–16.03	3.11
Headache	44 (72.1%)	33 (52.3%)	0.02*	0.20–0.90	0.42
Mental status disturbance	7 (11.4%)	16 (25.3%)	0.04*	1.00–6.93	2.63
Odynophagia	29 (47.5%)	29 (46%)	0.86	0.46–1.91	0.94
Osteomuscular pain	39 (63.9%)	48 (76.1%)	0.13	0.83–3.94	1.81
Rhinorrhea	33 (54%)	24 (38%)	0.07	0.26–1.07	0.52
Long-term symptoms					
Presence	26 (42.6%)	45 (71.4%)	<0.01*	1.60–7.09	3.37
Common	11 (18%)	38 (60.3%)	<0.01*	3.03–15.77	6.91
Cardiovascular	0 (0%)	8 (12.6%)	0.01*	–	–
Neurologic	16 (26.2%)	12 (19%)	0.33	0.28–1.55	0.66
Osteomuscular	4 (6.55%)	7 (11.1%)	0.56	0.49–6.42	1.78
Psychiatric	1 (1.63%)	23 (36.5%)	<0.01*	4.48–265.78	34.50
Respiratory	4 (6.55%)	15 (23.8%)	0.01*	1.39–14.32	4.45
Other long-term symptoms	0 (0%)	2 (3.2%)	0.49		

*Statistical significant, p-value < 0.05.

and American populations. Finally, for *LZTFL1* rs11385942, the comparison was made against COVID-19 patients obtained from a previous study. We found significant differences with Italian controls but not with Italian cases or Spanish population (**Table 6**).

Predictive Model and App Development

Genetic and non-genetic significant variables obtained from the previous analyses were entered into a logistic regression model. Different combinations of variables were tested, and the models obtained were compared by Akaike's Information Criterion (AIC) and Coefficient of Discrimination D (Tjur's R2). The best model had the lowest AIC and highest Tjur's R2 values. This model incorporated sex, age, number of comorbidities and the polymorphism *LZTFL1* rs11385942. The resulting predicting score that includes these variables was:

Adjusted score

$$= \frac{1}{1 + e^{-(-2.88 + (0.077*age) + 0.81(male) + (0.99*comorb) + 1.44(WT/Alt))}}$$

(1)

Where the adjusted score is a number between 0 and 1, "*age*" the age in years, "*male*" male sex, "*comorb*" represents the number

of comorbidities and "*WT/Alt*" the risk allele for the *LZTFL1* rs11385942 polymorphism.

Score distribution using this model for cases and controls is presented in **Figures 2A,B**. The model achieved good discrimination power (AUC = 0.857; 95% confidence interval 0.79–0.93) (**Figure 2C**) (**Supplementary Table 4**). Comparison between the clinical (Age + Sex + Comorbidities) and complete models (Age + Sex + Comorbidities + risk allele) showed a slight increase in the AUC, 0.846 vs. 0.857, respectively. Model comparison was assessed by three additional methods. First, direct comparison between the scores obtained from the clinical and complete model showed a high correlation, nevertheless, for several individuals, the risk scores changed noticeably when the risk allele is included in the model (**Figure 2D**). Next, we compared the models using the IDI score (52). This method is defined as the difference in the discrimination slopes between two models, the discrimination slopes are calculated as the difference of predicted probabilities for events and non-events (53). We obtained a positive IDI score (0.026; confidence interval 95% 0.001–0.051, *p*-value: 0.039) supporting a significant improvement for the complete model. Third, we calculated cross-validation parameters including concordance, sensitivity, specificity and net benefit for different probability cutoffs (54). Net benefit is a decision analytic measure, which puts benefits and harms on the same scale

TABLE 3 | Demographic and clinical characteristics in patients with and without long-term COVID-19 symptoms.

Variables	Patient with no long-term symptoms (*n* = 53)	Patients with long term symptoms (*n* = 71)	*p*-value	CI 95%	OR
Hospitalized	18 (34%)	45 (63.4)	0.00	1.60–7.10	3.37
Age	40 (±12.1)	43.5 (±10.8)	0.095		
Male sex	32 (60.3%)	35 (49.2%)	0.22	0.31–3.22	0.64
Blood group					
O	30 (56.6%)	50 (70.4%)	0.11	0.87–3.84	1.83
A	15 (28.3%)	17 (23.9%)	0.58	0.36–1.79	0.80
B	1 (1.88%)	3 (4.22%)	0.82	0.23–22.69	2.29
AB	0 (0%)	1 (1.40%)	1	–	–
Comorbidities					
Coronary disease	0 (0%)	2 (2.81%)	0.60	–	–
Arrhythmias	0 (0%)	1 (1.40%)	1	–	–
Hypertension	8 (15.0%)	12 (16.9%)	0.98	0.43–3.03	1.14
COPD	0 (0%)	2 (2.81%)	0.60	–	–
Asthma	1 (1.88%)	2 (2.81%)	0.97	0.04–5.39	0.48
T2DM	6 (11.3%)	8 (11.26%)	1	0.32–3.06	0.99
Chronic kidney disease	4 (7.54%)	2 (2.81%)	0.42	0.06–2.02	0.36
Cancer	1 (1.88%)	3 (4.22%)	0.82	0.23–22.69	2.29
Obesity	10 (18.8%)	17 (23.9%)	0.49	0.56–3.26	1.35
HIV/Immunodeficiency	1 (1.88%)	1 (1.40%)	1	0.05–12.15	0.74
Autoimmune disease	0 (0%)	2 (2.81%)	0.60	–	–
No comorbidities	34 (64.1%)	36 (50.7%)	0.13	0.28–1.19	0.57
One comorbidity	13(24.5%)	21(29.6%)	0.53	0.58–2.90	1.29
Two or more comorbidities	6 (11.3%)	14 (19.7%)	0.31	0.69–5.40	1.92
Chronic use of steroids	0 (0%)	2 (2.81%)	0.60	–	–
Smoking history	14 (26.4%)	32 (45.0%)	0.05	1.03–4.81	2.23
Acute symptoms					
Ageusia	20 (37.7%)	43 (60.5%)	0.01*	1.22–5.27	2.53
Anosmia	21 (39.6%)	40 (56.3%)	0.07	0.95–4.05	1.97
Cough	31 (58.4%)	49 (69.0%)	0.23	0.75–3.32	1.58
Diarrhoea	10 (18.8%)	21 (29.5%)	0.17	0.77–4.25	1.81
Dyspnoea	18 (33.9%)	51 (71.8%)	<0.01*	2.30–10.69	4.96
Fatigue	36 (67.9%)	65 (91%)	<0.01*	1.85–14.13	5.12
Fever > 38°C	21 (39.6%)	44 (61.9%)	0.01*	1.20–5.15	2.48
Haemoptysis	1 (1.88%)	7 (9.8%)	0.15	0.68–47.72	5.69
Headache	30 (56.6%)	47 (66.1%)	0.28	0.72–3.12	1.50
Odynophagia	23 (43.3%)	35 (49.2%)	0.51	0.62–2.59	1.27
Osteomuscular Pain	32 (60.3%)	55 (77.4%)	0.04*	1.03–4.94	2.26
Rhinorrhea	20 (37.7%)	37 (52.1%)	0.11	0.87–3.71	1.80
Brain fog	5 (9.43%)	18 (25.3%)	0.02*	1.12–9.46	3.26

Statistical significant, p-value < 0.05; COPD Chronic obstructive pulmonary disease.

to be compared (55). The results of this analysis showed that the concordance and net benefit were better for most of the probability cutoffs tested (**Supplementary Table 5**). This improvement was particularly noticeable at probabilities between 0.3 and 0.4.

Finally, the complete model was used to design a web-based application using the R package Shiny. The application is open-access and is accessible through a shinyApp server (https://oscarortega.shinyapps.io/COVID19_UR_Shiny/). The source code of the shiny app is publicly available on Github at https://github.com/OscarOrt/COVID_19_risk.

DISCUSSION

During the last 2 years, the COVID-19 pandemic has caused vast disruptions in almost any sphere of human activity. Despite

TABLE 4 | Allelic and genotypic frequencies for cases and controls.

Gen	SNP	Allele frequency controls		Allele frequency cases		Genotype controls			Genotype cases			HWE
		WT	Alt	WT	Alt	WT/WT	WT/Alt	Alt/Alt	WT/WT	WT/Alt	Alt/Alt	
ACE	rs4646994	0.6	0.4	0.55	0.45	0.39	0.41	0.2	0.3	0.49	0.21	0.46
ACE2	rs2285666	0.63	0.37	0.65	0.35	0.47 ♀	0.39 ♀	0.14 ♀	0.55 ♀	0.41 ♀	0.04 ♀	0.25
						0.52 ♂	–	0.48 ♂	0.54 ♂	–	0.46 ♂	
LZTFL1	rs11385942	0.98	0.02	0.91	0.09	0.97	0.03	0	0.83	0.17	0	1

ACE WT allele (Ins), ACE2 WT allele (G), LZTFL WT allele (no dup); Alt, alternative; WT, Wild Type; ♀ Genotype frequencies in females; ♂ Genotype frequences in males (hemizygous).

TABLE 5 | Genetic association analysis for severe COVID-19.

SNP	Model	Genotypes/alleles in cases	Genotypes/alleles in controls	χ2	df	*p*-value	OR	IC 95%
ACE rs4646994	Genotypic (2 df) test	13/31/19	12/25/24	1.23	2	0.54	–	–
	Cochran-Armitage trend test	57/69	49/73	0.60	1	0.43	–	–
	Allelic	57/69	49/73	0.65	1	0.41	1.23	0.74–2.03
	Dominant	44/19	37/24	1.15	1	0.28	–	–
	Recessive	13/50	12/49	0.01	1	0.89	–	–
ACE2 rs2285666	Genotypic (2 df) test	1/9/12	5/14/17	–	–	–	–	–
	Cochran-Armitage trend test	11/33	24/48	0.85	1	0.35	–	–
	Allelic	11/33	24/48	0.90	1	0.34	0.92	0.50–1.69
	Dominant	10/12	19/17	–	–	–	–	–
	Recessive	1/21	5/31	–	–	–	–	–
LZTFL1 rs11385942	Genotypic (2 df) test	0/11/52	0/2/59	–	–	–	–	–
	Cochran-Armitage trend test	11/115	2/120	6.64	1	<0.01*	–	–
	Allelic	11/115	2/120	6.27	1	0.01*	5.73	1.24–26.46
	Dominant	11/52	2/59	–	–	–	–	–
	Recessive	0/63	0/61	–	–	–	–	–

*Statistical significant, p-value < 0.05; df degrees of freedom; Genotypic (2 df) test: Alt/Alt vs. WT/Alt vs. WT/WT; Cochran-Armitage trend test: Alt vs. WT; Allelic: Alt vs. WT; Dominant: Alt/Alt + WT/Alt vs. WT/WT; Recessive: Alt/Alt vs. WT/Alt + WT/WT; ACE WT allele (Ins), ACE2 WT allele (G), LZTFL WT allele (no dup); Alt, alternative; WT, Wild Type.

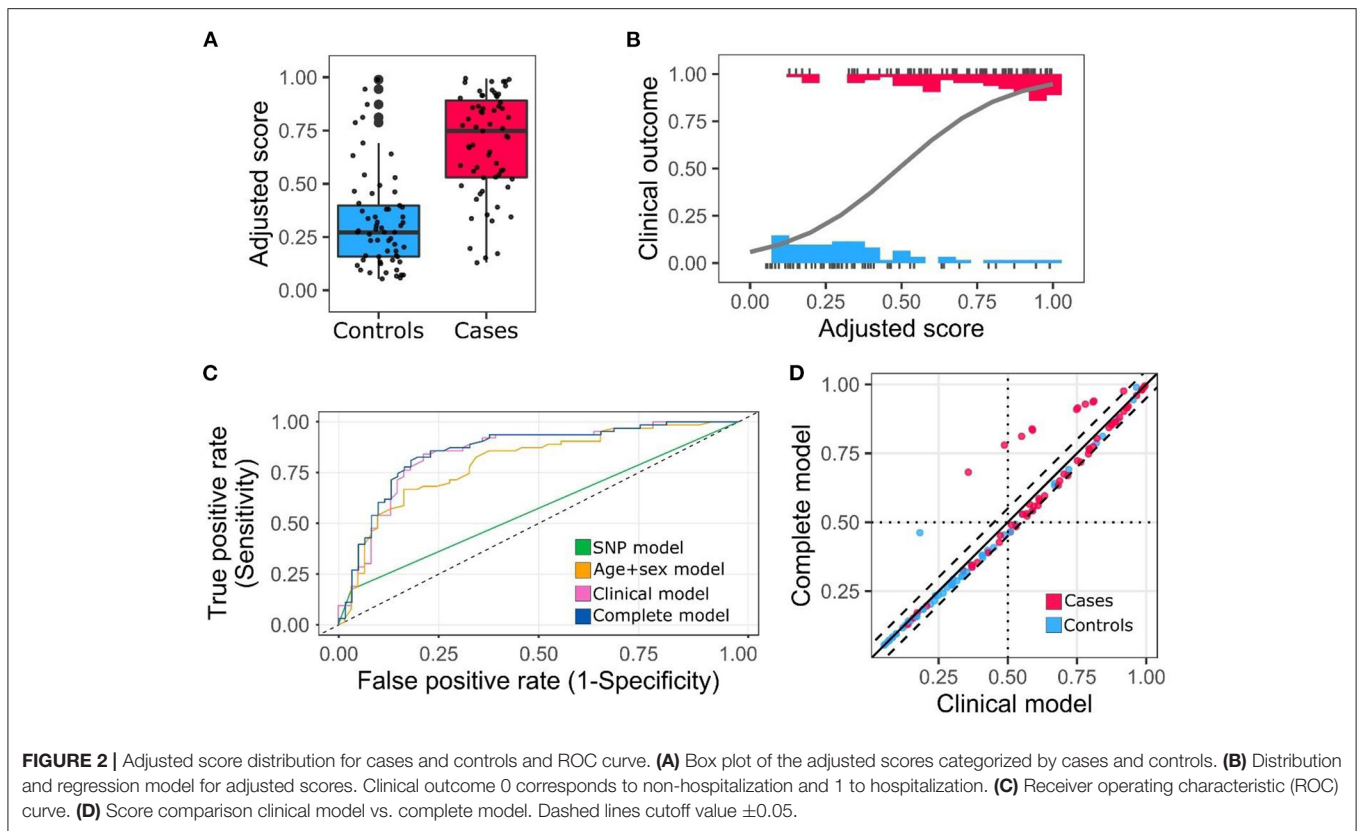**TABLE 6 |** Population case-control analysis of allele frequencies.

SNP	Region	Total case alleles	Cases WT alleles/AF	Cases Alt alleles/AF	*p*-value cases	Total controls alleles	Controls WT alleles/AF	Controls Alt alleles/AF	*p*-value controls	Source
LZTFL1 rs11385942	Present study	126	115/0.91	11/0.09		122	120/0.98	20/0.02		
	Italy	1,670	1,436/0.86	234/0.14	0.12	2,510	2,284/0.91	226/0.09	<0.01*	(21)
	Spain	1,550	1,410/0.91	140/0.09	0.96	1,900	1,805/0.95	9/0.05	0.14	(21)

*Statistical significant, p-value < 0.05; AF, Allele frequency; Alt, alternative; WT, Wild Type.

the growing knowledge about the biology and clinical features of this disease, many aspects of its physiopathology and clinical progression remain to be understood. Of particular interest in this process are host risk factors that could contribute to severe courses of COVID-19 and presence of long-term symptoms. These factors include non-genetic and genetic variables. In this study, we aimed to characterize the impact of these variables on COVID-19 outcomes in a sample of Colombian population. We identified several risk factors including the polymorphism *LZTFL1* rs11385942 and incorporated these variables into a predictive model. To the best of our knowledge, this is the first study to evaluate the association between genetic risk factors and

FIGURE 2 | Adjusted score distribution for cases and controls and ROC curve. **(A)** Box plot of the adjusted scores categorized by cases and controls. **(B)** Distribution and regression model for adjusted scores. Clinical outcome 0 corresponds to non-hospitalization and 1 to hospitalization. **(C)** Receiver operating characteristic (ROC) curve. **(D)** Score comparison clinical model vs. complete model. Dashed lines cutoff value ±0.05.

COVID-19 severity in a Latin-American population using a case-control design and illustrates the importance of host genetics in SARS-CoV-2 clinical outcomes.

Several non-genetic factors have been associated with poor COVID-19 prognosis, including age, male sex and comorbidities (56). In agreement with such reports, we found a significant association between age, male sex, hypertension and T2DM. Both hypertension and T2DM had been previously identified as independent risk factors for increased morbimortality in COVID-19 patients (57–61). The mechanism by which hypertension is a risk factor has been attributed to hyperactivation of the RAAS pathway, which increases the inflammatory response, cytokine storm, myocardial remodeling, acute lung injury, and endothelial damage (62). Similarly, it has been proposed that T2DM contributes to thromboembolic complications and organ damage through glucotoxicity, oxidative stress, and increased cytokine production (63). Interestingly, hyperglycemia in non-diabetic patients had a negative impact on patient outcomes (64), highlighting the importance of adequate metabolic control in the management of these patients. Other comorbidities analyzed did not show a statistically significant association individually, probably because the sample size was not large enough to detect such associations. Nonetheless, when grouped, the presence of two or more comorbidities conferred an increased risk of severe COVID-19, an effect possibly explained by the additive effect of risk factors to determine the clinical progression of the disease. The second point worth mentioning about clinical features in the studied population was the prevalence of acute symptoms. Among the most common symptoms reported in the literature are generalized weakness, dry cough, headache, dyspnea, and myalgias (65). In our sample, respiratory and systemic symptoms, including dyspnea, cough, fever and fatigue, were associated with severe disease, whereas flu-like symptoms, such as ageusia, anosmia and headache, were more frequent in patients with a mild form of the disease. Other studies, that included, populations have reported similar findings (66, 67). Lower respiratory tract symptoms are often related to severe COVID19, as they are a manifestation of underlying lung compromise.

Another element included in our analysis was the incidence of long-term COVID-19 symptoms, a phenomenon also reported in other viral infections including Spanish Flu SARS CoV-1 and MERS (68). Our findings are consistent with global literature, in which the most common long-term symptoms were fatigue (50–72.8%), joints pain (31.4%), headache (28.9%), chest pain (20–28.9%), dyspnea (28.2%) and palpitations (9%) (68–70). Remarkably, growing evidence suggests that psychiatric illness is an important COVID-19 sequel, affecting particularly specific populations such as Hispanic and African patients (71, 72). Psychiatric long-term symptoms were highly prevalent in hospitalized patients in our study (36.5%). Despite our study being limited by the absence of a standardized mental health scale for patient follow up, our data support these observations

(73). The mechanistic basis for these symptoms is attributed to the ability of the virus to infect the central nervous system via the blood-brain barrier and the olfactory bulb, affecting thereafter neurons on the hypothalamus, cortex and brainstem, which could explain many of the neuropsychiatric manifestations (71, 74). On the other hand, the absence of association between comorbidities and long-term symptoms has been also observed in the literature (75). Demographic variables such as sex are of much debate, as there is contradictory evidence of higher rates of long-term symptoms in female individuals (76). Finally, several acute symptoms associated with long-term compromise found in this study have been previously reported in the literature and include fatigue, dyspnea and osteomuscular pain and myalgias (77).

Regarding our genetic findings, our study identified the *LZTFL1* rs11385942 as a significant genetic factor associated with disease severity, conferring risk for severe/critical clinical outcomes. This polymorphism is located in the 3p21.31 locus, a region previously described as an important risk factor for severe respiratory disease in several studies (21, 78). There are six candidate genes in this locus potentially involved in the disease progression presumably by viral entry or clearance and immunological response, these are *SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6*, and *XCR1* (78). The rs11385942 polymorphism is located at intron 5 of *LZTFL1* and recent studies have assessed its functional significance in SARS-CoV-2 infection, suggesting a regulatory role. A CRISPRi analysis using lung epithelial cell lines showed that *LZTFL1* expression is severely affected by this polymorphism (79). LZTFL1 (leucine zipper transcription factor like 1) protein is highly expressed in lung cells and regulates airway cilia and epithelial-mesenchymal transition, a developmental process critical for the innate immune and inflammatory response. Remarkably, the rs11385942 polymorphism has been associated with higher levels of C5a and soluble terminal complement complex C5b-9 (SC5b-9) plasma levels during SARS-CoV-2 infection, suggesting that enhanced immune system and complement activation might be important pathways in the deleterious effect of this variant (80). Moreover, it has been described that complement activation and membrane attack complex (MAC) formation leads to upregulation of pro-inflammatory proteins and inflammasomes causing severe lung injury and, in parallel, endothelial cells death, platelet activation and induction of the coagulation cascade leading to thrombus formation, well-known physiopathological findings in severe COVID-19 (81, 82). The results of another recent study suggest that rs11385942 is in genetic linkage with the polymorphism rs17713054G>A, the gain-of-function risk A allele upregulates the expression of *LZTFL1* by generating a CCAAT/enhancer binding protein beta motif (23). Despite other molecular mechanisms cannot be discarded, this evidence supports *LZTFL1* as a candidate effector and provides further support to our findings. Additional studies have found supporting evidence for this association (79, 83). In line with these observations, genotypification of the risk allele in this gene could be useful as a molecular predictive biomarker for COVID-19 severe/critical clinical outcomes.

Since the beginning of the pandemic, numerous studies have explored the role of host genetic variability in COVID-19 severity and susceptibility. These studies have included genome-wide association studies (GWAS), which have identified multiple reproducible associations (21, 22, 84–86). Given the underrepresentation of Latin American population in these initiatives, our study allowed us to reproduce the association of the 3p21.32 locus in an ethnically different cohort and suggests that the variation in this region modulates the disease outcome (21). Importantly, detailed exploration of "expanded" phenotypes, other than clinical severity, including symptomatic/paucisymptomatic and Exposed_Positive/Exposed_Negative phenotypes have identified a much larger proportion of protective minor alleles (85). These results suggest that using additional phenotype definitions can identify protective associations. Our patients classified as asymptomatic-mild/severe-critical are more likely enriched for risk alleles conferred by loci such as those analyzed in our study.

It is important to highlight that case-control association studies are potentially influenced by population stratification due to undetected population substructure produced by differences in ancestry generating spurious associations (87). To avoid confounding due to population stratification, analysis using ancestry markers (AIMs) are useful to estimate variability between cases and controls (88). Although our study did not carry out this evaluation, we estimate that sampled population shares a similar gene pool without the influence of factors such as geographic isolation or non-random mating. Additionally, the individuals analyzed come from the Colombian Andean region, a geographical area where high inter-individual variation has not been identified (89), which supports the ethnic similarity of the cases and controls included. Here, LZTFL1 rs11385942 was identified as a significant genetic factor associated with severe COVID-19 ($p = 0.01$; OR = 5.73; 95% CI = 1.24–26.46) supporting an important genetic effect. Previously, it has been suggested a need for approaches such as family-based designs or genomic control when the identified genetic effects are very small (OR < 1.20) (90). Finally, although stratification may be less of a concern than originally anticipated and the evidence against a large effect of population stratification, hidden or otherwise, it is important to consider it in false positive or negative association arising from differences in local ancestry (87, 88, 91).

Two polymorphisms analyzed in our study, *ACE* rs4646994 and *ACE2* rs2285666, are important regulators of the RAAS pathway, a physiological system implicated in COVID-19 susceptibility and severity (92). Despite we did not find evidence of association between these polymorphisms and COVID-19 severity, numerous studies support a biological basis for such relationship (92–94). The *ACE2* rs2285666 T allele is associated with a significant increase in ACE2 expression (95). Interestingly, association studies of this polymorphism with COVID-19 severity have had contradictory results and similar findings to ours have been reported by several authors (43, 96, 97). Among these, next-generation sequencing analysis in patients hospitalized for COVID-19 indicated no association between *ACE2* variants and COVID-19 severity (97). Such discrepancies might be explained by population-specific differences, the

additive role of other genes interacting with risk alleles or other mechanisms not assessed such as epigenetic modifiers (98–100). Concerning *ACE* rs4646994 the Del allele has been associated with increased ACE expression, higher enzyme activity and elevated production of angiotensin II (101). Despite *ACE* Del/Del genotype and Del allele have been associated with increased COVID-19 patient severity (101–103), our results failed to replicate these findings in the Colombian population. In agreement with our results, other studies have reported no association between *ACE* rs4646994 and COVID-19 severity (43, 96). Collectively, current evidence contains conflicting results about the role of this polymorphism in SARS-CoV-2 infections. The reasons for these discrepancies are unclear and similar to the *ACE2* rs2285666 polymorphism require further exploration. Interestingly, a recent meta-analysis evaluating several polymorphisms related to COVID-19 outcomes found a significant association between the polymorphism ACE rs4646994 and COVID-19 severity (104). Results of individual association studies must be considered carefully and discrepancies in the findings may be the result of underpowered sample sizes, therefore replicates and more robust studies should be considered to validate these associations. On the other hand, we identified *ACE* rs4646994 Del allele as a protective factor for neurological long-term symptoms, we hypothesize this could be related to an increased catalytic activity resulting in vasoconstriction that counterbalances the intracerebral vasodilation and brain edema due to the anaerobic metabolism in cerebral cells in response to SARS-CoV-2 induced hypoxia (105, 106). Whereas, interesting, this hypothesis requires experimental and clinical validation.

Comparison of allelic frequencies obtained in our study with other populations revealed important differences. For *ACE* rs4646994, Asia was the only region with a similar allele frequency to our studied population (40). This may reflect the ancestral origin of Native American population in Colombia or the admixture between an ancestral population with a higher frequency and Europeans, where allele frequencies are considerably lower (107, 108). For the *ACE2* polymorphism, the allelic frequency was similar to Mexican population, probably due to a common ancestry and admixture history (44). For the variant *LZTFL1* rs11385942, no differences were found with Spanish, European and African populations. Remarkably, Zeberg and Pääbo (109) described that the 3p21.31 region, the locus where the variant is located, was inherited from Neanderthals. The mixture of native Americans and Europeans probably modified the ancestral genetic pool leading to the current allele frequencies. Additionally, it has been proposed that differences in allelic frequencies for the 3p21.31 risk haplotype are produced by natural selection in response to pathogens (109).

Another important determinant of COVID-19 severity is viral genetics (10). It has been identified that specific SARS-CoV-2 variants are associated with differences in severity and mortality, for example, the alpha and gamma variants are related to increased hospitalization, ICU admission and mortality risk (110–112). While our study did not assess variant differences in cases and controls, genomic surveillance studies conducted during the sample collection period (December 2020–July 2021) in Bogotá, showed that the predominant variants were B.1.621 (Mu) 57.3% (469/819), P.1 (Gamma) 14% (114/819) and B.1.1.7 (alpha) 2.8% (23/819) (113). The most common variant found in this interval of time, Mu, was classified as a variant being monitored (VBM) by the Centers for Disease Control and Prevention (CDC U.S.) without reported major effects on infectivity, transmissibility or severity (114). The coexistence of several variants during this period constitutes a source of variation and might reflect a more complex dynamics of host-pathogen interactions.

Our clinical and genetic association analysis allowed us to identify several risk factors related to disease severity. These factors were incorporated into a predictive risk model using a multivariate logistic regression including demographic, clinical, and genetic traits. To date, ∼50 prediction models and scoring systems, have been published (115). These models are useful tools to facilitate decision-making in healthcare services and rely mostly on clinical features such as age, sex, number of comorbidities, hypertension, T2DM, chronic obstructive lung disease, cancer, cardiovascular disease. However, it is noteworthy that COVID-19 severity is influenced by viral and host genetic factors (10). Recent models, which like ours incorporate a multifactorial approach (genetic and non-genetic factors), included several single nucleotide variants (SNVs) (116). These models have achieved good results in discriminating COVID-19 severity groups and highlighted the role of integrated approaches to predict clinical outcomes. Furthermore, other models aiming to predict adverse outcomes are based on detailed clinical features during diagnosis, admission and hospitalization have been developed, nevertheless, its accessibility and clinical implementation have been limited (117, 118). We propose our model as a useful tool to estimate *a priori* severe or critical illness risk. Notably, despite the minor increase in the AUC when the clinical and complete models were compared, detailed analysis of the discrimination performance and cross-validation parameters suggest that the incorporation of the risk allele improves the risk prediction model. Further studies involving larger sample sizes might be useful to validate these findings. Likewise, the implementation of our model into a web application might facilitate its usage by healthcare providers in limited-resource settings during the current SARS-CoV-2 pandemics and future health emergencies caused by similar pathogens.

In summary, our study explores the relation between non-genetic and genetic factors, with COVID-19 outcomes in Colombian population, demonstrating a positive association between the *LZTFL1* rs11385942 polymorphism and severe disease. By establishing such association, we point up the importance of genetic host factors in SARS-CoV-2 infection. In addition, our work identified previously known non-genetic factors and developed a predictive model which was implemented in a web application, providing a useful tool for risk prediction. Integrative approaches, like ours, may be helpful to better understand COVID-19 clinical progression, refine healthcare efforts and reduce the morbimortality of patients with this disease.

Study Limitations

Our study has potential limitations. First, the sample size was calculated in order to have 80% statistical power based on previous association reports for the variant with the lower allele frequency (*LZTF1* rs11385942.), nevertheless, it could have been limited to detect potential small effect sizes for the rs4646994 and rs228566 SNPs in our population. Second, some clinical variables assessed in the clinical follow-up interview were self-reported. Even though most of this information was confirmed in the clinical record, this could have been a potential source of bias. Third, we did not match the case-control groups by age or sex for the statistical analysis. Considering these variables are known risk factors, we aimed to assess their impact on COVID-19 outcome. Fourth, as previously mentioned, analysis of potential population stratification was not performed. In addition, COVID-19 severity is a multifactorial trait and other important variables, including environmental factors, SARS-CoV-2 variants, and additional host genetic polymorphisms, described as risk or protective factors were not evaluated. Assessment of such variables in future studies could help to improve discriminative models and medical risk assessment. Finally, we should highlight that our proposed risk model constitutes a proof-of-concept of the feasibility of this integrative approach and further studies with larger sample sizes and independent replications are required to validate the model.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Universidad del Rosario (DVO005 1543-CV1334). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MA-A, DC-O, MG-C, EP-M, CR, DF-M, and OO-R contributed to conception and design of the study. MA-A, PT-F, NC, AM, DF-M, and OO-R performed DNA extraction and genetic analysis. MA-A, DC-O, JC-M, EP-M, CR, PT-F, KP, DF-M, and OO-R collected clinical data and organized the database. MA-A, JC-M, PT-F, and OO-R performed the statistical analysis. MA-A, DC-O, JC-M, PT-F, AM, DF-M, and OO-R wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2022.910098/full#supplementary-material

REFERENCES

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med.* (2020) 382:727–33. doi: 10.1056/NEJMoa2001017
2. Hu B, Guo H, Zhou P, Shi Z-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol.* (2021) 19:141–54. doi: 10.1038/s41579-020-00459-7
3. Center for Systems Science and Engineering (CSSE) by the Johns Hopkins University (JHU). *COVID-19 Dashboard.* (2022). Available online at: https://coronavirus.jhu.edu/map.html
4. Ministerio de Salud y Protección Social C. *Situación Actual Coronavirus (COVID-19) Colombia.* (2022). Available online at: https://www.minsalud.gov.co/salud/publica/PET/Paginas/Covid-19_copia.aspx
5. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis.* (2020) 20:669–77. doi: 10.1016/S1473-3099(20)30243-7
6. Deng G, Yin M, Chen X, Zeng F. Clinical determinants for fatality of 44,672 patients with COVID-19. *Crit Care.* (2020) 24:179. doi: 10.1186/s13054-020-02902-w
7. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature.* (2020) 584:430–36. doi: 10.1038/s41586-020-2521-4
8. Velavan TP, Pallerla SR, Rüter J, Augustin Y, Kremsner PG, Krishna S, et al. Host genetic factors determining COVID-19 susceptibility and severity. *eBioMedicine.* (2021) 72:103629. doi: 10.1016/j.ebiom.2021.103629
9. Fricke-Galindo I, Falfán-Valencia R. Genetics insight for COVID-19 susceptibility and severity: a review. *Front Immunol.* (2021) 12:622176. doi: 10.3389/fimmu.2021.622176
10. Ovsyannikova IG, Haralambieva IH, Crooke SN, Poland GA, Kennedy RB. The role of host genetics in the immune response to SARS-CoV-2 and COVID-19 susceptibility and severity. *Immunol Rev.* (2020) 296:205–19. doi: 10.1111/imr.12897
11. Kachuri L, Francis SS, Morrison ML, Wendt GA, Bossé Y, Cavazos TB, et al. The landscape of host genetic factors involved in immune response to common viral infections. *Genome Med.* (2020) 12:93. doi: 10.1186/s13073-020-00790-x
12. Gemmati D, Bramanti B, Serino ML, Secchiero P, Zauli G, Tisato V. COVID-19 and individual genetic susceptibility/receptivity: role of ACE1/ACE2 genes, immunity, inflammation and coagulation. Might the double X-chromosome in females be protective against SARS-CoV-2 compared to the single X-chromosome in males? *Int J Mol Sci.* (2020) 21:3474. doi: 10.3390/ijms21103474
13. Wang F, Huang S, Gao R, Zhou Y, Lai C, Li ZZ, et al. Initial whole-genome sequencing and analysis of the host genetic contribution to COVID-19 severity and susceptibility. *Cell Discov.* (2020) 6:83. doi: 10.1038/s41421-020-00231-4
14. Gelen V, Kükürt A, Sengül E. "Role of the Renin-Angiotensin-Aldosterone System in Various Disease Processes: An Overview". *Renin-Angiotensin Aldosterone System.* London: IntechOpen (2021). doi: 10.5772/intechopen.97354
15. Saad H, Jabotian K, Sakr C, Mahfouz R, Akl IB, Zgheib NK. The role of angiotensin converting enzyme 1 insertion/deletion genetic polymorphism in the risk and severity of COVID-19 infection. *Front Med.* (2021) 8:798571. doi: 10.3389/fmed.2021.798571
16. Ni W, Yang X, Yang D, Bao J, Li R, Xiao Y, et al. Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19. *Crit Care.* (2020) 24:422. doi: 10.1186/s13054-020-03120-0

17. Hatami N, Ahi S, Sadeghinikoo A, Foroughian M, Javdani F, Kalani N, et al. Worldwide ACE (I/D) polymorphism may affect COVID-19 recovery rate: an ecological meta-regression. *Endocrine.* (2020) 68:479–84. doi: 10.1007/s12020-020-02381-7

18. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell.* (2020) 181:271–80.e8. doi: 10.1016/j.cell.2020.02.052

19. Benetti E, Tita R, Spiga O, Ciolfi A, Birolo G, Bruselles A, et al. ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. *Eur J Hum Genet.* (2020) 28:1602–14. doi: 10.1038/s41431-020-0691-z

20. Srivastava A, Bandopadhyay A, Das D, Pandey RK, Singh V, Khanam N, et al. Genetic association of ACE2 rs2285666 polymorphism with COVID-19 spatial distribution in India. *Front Genet.* (2020) 11:564741. doi: 10.3389/fgene.2020.564741

21. Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, et al. Genomewide association study of severe Covid-19 with respiratory failure. *N Engl J Med.* (2020) 383:1522–34. doi: 10.1056/NEJMoa2020283

22. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature.* (2021) 600:472–7. doi: 10.1038/s41586-021-03767-x

23. Downes DJ, Cross AR, Hua P, Roberts N, Schwessinger R, Cutler AJ, et al. Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus. *Nat Genet.* (2021) 53:1606–15. doi: 10.1038/s41588-021-00955-3

24. Ministerio de Salud y Protección Social de Colombia. *Lineamientos Para el Manejo Clínico de Pacientes con Infección por Nuevo Coronavirus COVID-19.* (2020). Available online at: https://www.minsalud.gov.co/Ministerio/Institucional/Procesos y procedimientos/PSSS03.pdf

25. Dean AG. *OpenEpi: Open Source Epidemiologic Statistics for Public Health.* (2007). Available online at: http://www.OpenEpi.com; http://ci.nii.ac.jp/naid/10025863040/en/ (accessed March 10, 2022).

26. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* (2020) 581:434–43. doi: 10.1530/ey.17.14.3

27. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics.* (2012) 13:134. doi: 10.1186/1471-2105-13-134

28. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* (2012) 28:1647–9. doi: 10.1093/bioinformatics/bts199

29. Raveendran AV, Jayadevan R, Sashidharan S. Long COVID: an overview. *Diabetes Metab Syndr Clin Res Rev.* (2021) 15:869–75. doi: 10.1016/j.dsx.2021.04.007

30. Lopez-Leon S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo PA, Cuapio A, et al. More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. *Sci Rep.* (2021) 11:16144. doi: 10.1038/s41598-021-95565-8

31. Solé X, Guinó E, Valls J, Iniesta R, Moreno V. SNPStats: a web tool for the analysis of association studies. *Bioinformatics.* (2006) 22:1928–9. doi: 10.1093/bioinformatics/btl268

32. Backenroth D, Carmi S. A test for deviations from expected genotype frequencies on the X chromosome for sex-biased admixed populations. *Heredity.* (2019) 123:470–78. doi: 10.1038/s41437-019-0233-z

33. Al-Harbi EM, Farid EM, Gumaa KA, Singh J. Genotypes and allele frequencies of angiotensin-converting enzyme (ACE) insertion/deletion polymorphism among Bahraini population with type 2 diabetes mellitus and related diseases. *Mol Cell Biochem.* (2012) 362:219–23. doi: 10.1007/s11010-011-1146-1

34. Salem A, Batzer MA. High frequency of the D allele of the angiotensin-converting enzyme gene in Arabic populations. *BMC Res Notes.* (2009) 2:99. doi: 10.1186/1756-0500-2-99

35. Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, et al. Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet.* (2000) 107:312–319. doi: 10.1007/s004390000370

36. Barley J, Blackwood A, Carter ND, Crews DE, Cruickshank JK, Jeffery S, et al. Angiotensin converting enzyme insertion/deletion polymorphism: association with ethnic origin. *J Hypertens.* (1994) 12:955–7.

37. Rutledge DR, Browe CS, Ross EA. Frequencies of the angiotensinogen gene and angiotensin I converting enzyme (ACE) gene polymorphisms in African Americans. *Biochem Mol Biol Int.* (1994) 34:1271–5.

38. Lee E. Population genetics of the angiotensin-converting enzyme in Chinese. *Br J Clin Pharmacol.* (1994) 37:212–14. doi: 10.1111/j.1365-2125.1994.tb04264.x

39. Saab YB, Gard PR, Overall ADJ. The geographic distribution of the ACE II genotype: a novel finding. *Genet Res.* (2007) 89:259–67. doi: 10.1017/S0016672307009019

40. Sarangarajan R, Winn R, Kiebish MA, Bountra C, Granger E, Narain NR. Ethnic prevalence of angiotensin-converting enzyme deletion (D) polymorphism and COVID-19 risk: rationale for use of angiotensin-converting enzyme inhibitors/angiotensin receptor blockers. *J Racial Ethn Heal Disparities.* (2021) 8:973–80. doi: 10.1007/s40615-020-00853-0

41. Strafella C, Caputo V, Termine A, Barati S, Gambardella S, Borgiani P, et al. Analysis of ACE2 genetic variability among populations highlights a possible link with COVID-19-related neurological complications. *Genes.* (2020) 11:741. doi: 10.3390/genes11070741

42. Clarke L, Zheng-Bradley X, Smith R, Kulesha E, Xiao C, Toneva I, et al. The 1000 genomes project: data management and community access. *Nat Methods.* (2012) 9:459–62. doi: 10.1038/nmeth.1974

43. Gómez J, Albaiceta GM, García-Clemente M, López-Larrea C, Amado-Rodríguez L, Lopez-Alonso I, et al. Angiotensin-converting enzymes (ACE, ACE2) gene variants and COVID-19 outcome. *Gene.* (2020) 762:145102. doi: 10.1016/j.gene.2020.145102

44. Lozano-Gonzalez K, Padilla-Rodríguez E, Texis T, Gutiérrez MN, Rodríguez-Dorantes M, Cuevas-Córdoba B, et al. Allele frequency of ACE2 intron variants and its association with blood pressure. *DNA Cell Biol.* (2020) 39:2095–101. doi: 10.1089/dna.2020.5804

45. Phan L, Jin Y, Zhang H, Qiang W, Shekhtman E, Shao D, et al. *ALFA: Allele Frequency Aggregator.* National Center for Biotechnology Information, US National Library of Medicine (2020).

46. Balanovsky O, Petrushenko V, Mirzaev K, Abdullaev S, Gorin I, Chernevskiy D, et al. Variation of genomic sites associated with severe Covid-19 across populations: global and national patterns. *Pharmgenomics Pers Med.* (2021) 14:1391–402. doi: 10.2147/PGPM.S320609

47. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* (2007) 81:559–75. doi: 10.1086/519795

48. Tjur T. Coefficients of determination in logistic regression models—a new proposal: the coefficient of discrimination. *Am Stat.* (2009) 63:366–72. doi: 10.1198/tast.2009.08210

49. Kundu S, Aulchenko YS, van Duijn CM, Janssens ACJW. PredictABEL: an R package for the assessment of risk prediction models. *Eur J Epidemiol.* (2011) 26:261–4. doi: 10.1007/s10654-011-9567-4

50. Brown M. *Package 'rmda'* (2018). Available online at: https://cran.r-project.org/web/packages/rmda/rmda.pdf (accessed May 25, 2022).

51. R Core Team. *R: A Language and Environment for Statistical Computing* (2020).

52. Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med.* (2008) 27:157–72. doi: 10.1002/sim.2929

53. Pencina MJ, D'Agostino RB, Demler OV. Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. *Stat Med.* (2012) 31:101–13. doi: 10.1002/sim.4348

54. Kesselmeier M, Legrand C, Peil B, Kabisch M, Fischer C, Hamann U, et al. Practical investigation of the performance of robust logistic regression to predict the genetic risk of hypertension. *BMC Proc.* (2014) 8:S65. doi: 10.1186/1753-6561-8-S1-S65

55. Zhang Z, Rousson V, Lee W-C, Ferdynus C, Chen M, Qian X, et al. Decision curve analysis: a technical note. *Ann Transl Med.* (2018) 6:308. doi: 10.21037/atm.2018.07.02

56. Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, et al. Risk factors of critical & mortal COVID-19 cases: a systematic literature review and meta-analysis. *J Infect.* (2020) 81:e16–25. doi: 10.1016/j.jinf.2020.04.021

57. Mancia G, Rea F, Ludergnani M, Apolone G, Corrao G. Renin–Angiotensin–Aldosterone System Blockers and the Risk of Covid-19. *N Engl J Med.* (2020) 382:2431–40. doi: 10.1056/NEJMoa2006923

58. Liang X, Shi L, Wang Y, Xiao W, Duan G, Yang H, et al. The association of hypertension with the severity and mortality of COVID-19 patients: evidence based on adjusted effect estimates. *J Infect.* (2020) 81:e44–7. doi: 10.1016/j.jinf.2020.06.060

59. Gao C, Cai Y, Zhang K, Zhou L, Zhang Y, Zhang X, et al. Association of hypertension and antihypertensive treatment with COVID-19 mortality: a retrospective observational study. *Eur Heart J.* (2020) 41:2058–66. doi: 10.1093/eurheartj/ehaa433

60. Holman N, Knighton P, Kar P, O'Keefe J, Curley M, Weaver A, et al. Risk factors for COVID-19-related mortality in people with type 1 and type 2 diabetes in England: a population-based cohort study. *Lancet Diabetes Endocrinol.* (2020) 8:823–33. doi: 10.1016/S2213-8587(20)30271-0

61. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *Int J Infect Dis.* (2020) 94:91–5. doi: 10.1016/j.ijid.2020.03.017

62. Paz Ocaranza M, Riquelme JA, García L, Jalil JE, Chiong M, Santos RAS, et al. Counter-regulatory renin–angiotensin system in cardiovascular disease. *Nat Rev Cardiol.* (2020) 17:116–29. doi: 10.1038/s41569-019-0244-8

63. Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *J Thromb Haemost.* (2020) 18:844–7. doi: 10.1111/jth.14768

64. Wang S, Ma P, Zhang S, Song S, Wang Z, Ma Y, et al. Fasting blood glucose at admission is an independent predictor for 28-day mortality in patients with COVID-19 without previous diagnosis of diabetes: a multi-centre retrospective study. *Diabetologia.* (2020) 63:2102–11. doi: 10.1007/s00125-020-05209-1

65. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med.* (2020) 382:1708–20. doi: 10.1056/NEJMoa2002032

66. Iroungou BA, Mangouka LG, Bivigou-Mboumba B, Moussavou-Boundzanga P, Obame-Nkoghe J, Nzigou Boucka F, et al. Demographic and clinical characteristics associated with severity, clinical outcomes, and mortality of COVID-19 infection in gabon. *JAMA Netw Open.* (2021) 4:e2124190. doi: 10.1001/jamanetworkopen.2021.24190

67. Mehta OP, Bhandari P, Raut A, Kacimi SEO, Huy NT. Coronavirus disease (COVID-19): comprehensive review of clinical presentation. *Front Public Heal.* (2021) 8:582932. doi: 10.3389/fpubh.2020.582932

68. Islam MF, Cotler J, Jason LA. Post-viral fatigue and COVID-19: lessons from past epidemics. *Fatigue Biomed Heal Behav.* (2020) 8:61–9. doi: 10.1080/21641846.2020.1778227

69. Carod Artal FJ. Síndrome post-COVID-19: epidemiología, criterios diagnósticos y mecanismos patogénicos implicados. *Rev Neurol.* (2021) 72:384. doi: 10.33588/rn.7211.2021230

70. Kamal M, Abo Omirah M, Hussein A, Saeed H. Assessment and characterisation of post-COVID-19 manifestations. *Int J Clin Pract.* (2021) 75:e13746. doi: 10.1111/ijcp.13746

71. Nakamura ZM, Nash RP, Laughon SL, Rosenstein DL. Neuropsychiatric complications of COVID-19. *Curr Psychiatry Rep.* (2021) 23:25. doi: 10.1007/s11920-021-01237-9

72. Czeisler MÉ, Wiley JF, Facer-Childs ER, Robbins R, Weaver MD, Barger LK, et al. Mental health, substance use, and suicidal ideation during a prolonged COVID-19-related lockdown in a region with low SARS-CoV-2 prevalence. *J Psychiatr Res.* (2021) 140:533–44. doi: 10.1016/j.jpsychires.2021.05.080

73. Parker C, Shalev D, Hsu I, Shenoy A, Cheung S, Nash S, et al. Depression, anxiety, and acute stress disorder among patients hospitalized with COVID-19: a prospective cohort study. *J Acad Consult Psychiatry.* (2021) 62:211–19. doi: 10.1016/j.psym.2020.10.001

74. Moldofsky H, Patcai J. Chronic widespread musculoskeletal pain, fatigue, depression and disordered sleep in chronic post-SARS syndrome; a case-controlled study. *BMC Neurol.* (2011) 11:37. doi: 10.1186/1471-2377-11-37

75. Moreno-Pérez O, Merino E, Leon-Ramirez J-M, Andres M, Ramos JM, Arenas-Jiménez J, et al. Post-acute COVID-19 syndrome. Incidence and risk factors: a Mediterranean cohort study. *J Infect.* (2021) 82:378–83. doi: 10.1016/j.jinf.2021.01.004

76. Yong SJ. Long COVID or post-COVID-19 syndrome: putative pathophysiology, risk factors, and treatments. *Infect Dis.* (2021) 53:737–54. doi: 10.1080/23744235.2021.1924397

77. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and predictors of long COVID. *Nat Med.* (2021) 27:626–31. doi: 10.1038/s41591-021-01292-y

78. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, et al. Genetic mechanisms of critical illness in Covid-19. *Nature.* (2020) 591:92–8. doi: 10.1101/2020.09.24.20200048

79. Fink-Baldauf IM, Stuart WD, Brewington JJ, Guo M, Maeda Y. CRISPRi links COVID-19 GWAS loci to LZTFL1 and RAVER1. *eBioMedicine.* (2022) 75:103806. doi: 10.1016/j.ebiom.2021.103806

80. Valenti L, Griffini S, Lamorte G, Grovetti E, Uceda Renteria SC, Malvestiti F, et al. Chromosome 3 cluster rs11385942 variant links complement activation with severe COVID-19. *J Autoimmun.* (2021) 117:102595. doi: 10.1016/j.jaut.2021.102595

81. Xie CB, Jane-Wit D, Pober JS. Complement membrane attack complex. *Am J Pathol.* (2020) 190:1138–50. doi: 10.1016/j.ajpath.2020.02.006

82. Malas MB, Naazie IN, Elsayed N, Mathlouthi A, Marmor R, Clary B. Thromboembolism risk of COVID-19 is high and associated with a higher risk of mortality: a systematic review and meta-analysis. *EClinicalMedicine.* (2020) 29–30:100639. doi: 10.1016/j.eclinm.2020.100639

83. Nakanishi T, Pigazzini S, Degenhardt F, Cordioli M, Butler-Laporte G, Maya-Miles D, et al. Age-dependent impact of the major common genetic risk factor for COVID-19 on severity and mortality. *J Clin Invest.* (2021) 131:e152386. doi: 10.1101/2021.03.07.21252875

84. Horowitz JE, Kosmicki JA, Damask A, Sharma D, Roberts GHL, Justice AE, et al., Yadav A, Leader JB, et al. Genome-wide analysis provides genetic evidence that ACE2 influences COVID-19 risk and yields risk scores associated with severe disease. *Nat Genet.* (2022) 54:382–92. doi: 10.1038/s41588-021-01006-7

85. Roberts GHL, Partha R, Rhead B, Knight SC, Park DS, Coignet M V., et al. Expanded COVID-19 phenotype definitions reveal distinct patterns of genetic association and protective effects. *Nat Genet.* (2022) 54:374–81. doi: 10.1038/s41588-022-01042-x

86. Shelton JF, Shastri AJ, Ye C, Weldon CH, Filshtein-Sonmez T, Coker D, et al. Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat Genet.* (2021) 53:801–8. doi: 10.1038/s41588-021-00854-7

87. Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, et al. Assessing the impact of population stratification on genetic association studies. *Nat Genet.* (2004) 36:388–93. doi: 10.1038/ng1333

88. Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL. Population stratification in genetic association studies. *Curr Protoc Hum Genet.* (2017) 95:1.22.1–1.22.23. doi: 10.1002/cphg.48

89. Ossa H, Aquino J, Pereira R, Ibarra A, Ossa RH, Pérez LA, et al. Outlining the ancestry landscape of colombian admixed populations. *PLoS ONE.* (2016) 11:e0164414. doi: 10.1371/journal.pone.0164414

90. Little J, Higgins JP., Ioannidis JP., Moher D, Gagnon F, von Elm E, et al. STrengthening the REporting of Genetic Association Studies (STREGA)— an extension of the STROBE statement. *PLoS Med.* (2009) 6:e1000022. doi: 10.1371/journal.pmed.1000022

91. Ioannidis JPA, Ntzani EE, Trikalinos TA. "Racial" differences in genetic effects for complex diseases. *Nat Genet.* (2004) 36:1312–18. doi: 10.1038/ng1474

92. Saengsiwaritt W, Jittikoon J, Chaikledkaew U, Udomsinprasert W. Genetic polymorphisms of ACE1, ACE2, and TMPRSS2 associated with COVID-19 severity: a systematic review with meta-analysis. *Rev Med Virol.* (2022) e2323. doi: 10.1002/rmv.2323

93. Coto E, Avanzas P, Gómez J. The renin-angiotensin-aldosterone system and coronavirus disease 2019. *Eur Cardiol.* (2021) 16:e07. doi: 10.15420/ecr.2020.30

94. Zipeto D, Palmeira J da F, Argañaraz GA, Argañaraz ER. ACE2/ADAM17/TMPRSS2 interplay may be the

main risk factor for COVID-19. *Front Immunol.* (2020) 11:576745. doi: 10.3389/fimmu.2020.576745

95. Khayat AS, de Assumpção PP, Meireles Khayat BC, Thomaz Araújo TM, Batista-Gomes JA, Imbiriba LC, et al. ACE2 polymorphisms as potential players in COVID-19 outcome. *PLoS ONE.* (2020) 15:e0243887. doi: 10.1371/journal.pone.0243887

96. Karakaş Çelik S, Çakmak Genç G, Pişkin N, Açikgöz B, Altinsoy B, Kurucu Işsiz B, Dursun A. Polymorphisms of ACE (I/D) and ACE2 receptor gene (Rs2106809, Rs2285666) are not related to the clinical course of COVID-19: a case study. *J Med Virol.* (2021) 93:5947–52. doi: 10.1002/jmv.27160

97. Novelli A, Biancolella M, Borgiani P, Cocciadiferro D, Colona VL, D'Apice MR, et al. Analysis of ACE2 genetic variants in 131 Italian SARS-CoV-2-positive patients. *Hum Genomics.* (2020) 14:29. doi: 10.1186/s40246-020-00279-z

98. Li Y, Li H, Zhou L. EZH2-mediated H3K27me3 inhibits ACE2 expression. *Biochem Biophys Res Commun.* (2020) 526:947–52. doi: 10.1016/j.bbrc.2020.04.010

99. Lambert DW, Clarke NE, Hooper NM, Turner AJ. Calmodulin interacts with angiotensin-converting enzyme-2 (ACE2) and inhibits shedding of its ectodomain. *FEBS Lett.* (2008) 582:385–90. doi: 10.1016/j.febslet.2007.11.085

100. Saponaro F, Rutigliano G, Sestito S, Bandini L, Storti B, Bizzarri R, et al. ACE2 in the era of SARS-CoV-2: controversies and novel perspectives. *Front Mol Biosci.* (2020) 7:588618. doi: 10.3389/fmolb.2020.588618

101. Mir MM, Mir R, Alghamdi MAA, Alsayed BA, Wani JI, Alharthi MH, et al. Strong association of angiotensin converting enzyme-2 gene insertion/deletion polymorphism with susceptibility to SARS-CoV-2, hypertension, coronary artery disease and COVID-19 disease mortality. *J Pers Med.* (2021) 11:1098. doi: 10.3390/jpm11111098

102. Gunal O, Sezer O, Ustun GU, Ozturk CE, Sen A, Yigit S, et al. Angiotensin-converting enzyme-1 gene insertion/deletion polymorphism may be associated with COVID-19 clinical severity: a prospective cohort study. *Ann Saudi Med.* (2021) 41:141–146. doi: 10.5144/0256-4947.2021.141

103. Verma S, Abbas M, Verma S, Khan FH, Raza ST, Siddiqi Z, et al. Impact of I/D polymorphism of angiotensin-converting enzyme 1 (ACE1) gene on the severity of COVID-19 patients. *Infect Genet Evol.* (2021) 91:104801. doi: 10.1016/j.meegid.2021.104801

104. de Araújo JLF, Menezes D, de Aguiar RS, de Souza RP. IFITM3, FURIN, ACE1, and TNF-α genetic association with COVID-19 outcomes: systematic review and meta-analysis. *Front Genet.* (2022) 13:775246. doi: 10.3389/fgene.2022.775246

105. Soltani Zangbar H, Gorji A, Ghadiri T. A review on the neurological manifestations of COVID-19 infection: a mechanistic view. *Mol Neurobiol.* (2021) 58:536–49. doi: 10.1007/s12035-020-02149-0

106. Purwaningroom DL, Saifurrohman M, Widodo N, Putri JF, Lukitasari M. Alteration of splicing pattern on angiotensin converting enzyme gene due to the insertion of ALU elements. *Int J Comput Biol.* (2015) 4:53. doi: 10.34040/IJCB.4.2.2015.61

107. Silva WA, Bonatto SL, Holanda AJ, Ribeiro-Dos-Santos AK, Paixão BM, Goldman GH, et al. Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. *Am J Hum Genet.* (2002) 71:187–92. doi: 10.1086/341358

108. Rishishwar L, Conley AB, Wigington CH, Wang L, Valderrama-Aguirre A, Jordan IK. Ancestry, admixture and fitness in Colombian genomes. *Sci Rep.* (2015) 5:12376. doi: 10.1038/srep12376

109. Zeberg H, Pääbo S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature.* (2020) 587:610–12. doi: 10.1038/s41586-020-2818-3

110. Funk T, Pharris A, Spiteri G, Bundle N, Melidou A, Carr M, et al. Characteristics of SARS-CoV-2 variants of concern B.1.1.7, B.1.351 or P.1: data from seven EU/EEA countries, weeks 38/2020 to 10/2021. *Eurosurveillance.* (2021) 26:2100348. doi: 10.2807/1560-7917.ES.2021.26.16.2100348

111. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature.* (2021) 593:266–9. doi: 10.1038/s41586-021-03470-x

112. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido D da S, Mishra S, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science.* (2021) 372:815–21. doi: 10.1126/science.abh2644

113. Instituto Nacional de Salud Colombia. *COVID-19 en Colombia.* (2021). Available online at: https://www.ins.gov.co/Noticias/Paginas/coronavirus-casos.aspx (accessed February 1, 2021).

114. Centers for Disease Control and Prevention (U.S.). *SARS-CoV-2 Variant Classifications and Definitions.* (2022). Available online at: https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html

115. Chen Y, Zhou X, Yan H, Huang H, Li S, Jiang Z, et al. CANPT score: a tool to predict severe COVID-19 on admission. *Front Med.* (2021) 8:608107. doi: 10.3389/fmed.2021.608107

116. Dite GS, Murphy NM, Allman R. An integrated clinical and genetic model for predicting risk of severe COVID-19: a population-based case-control study. *PLoS ONE.* (2021) 16:e0247205. doi: 10.1371/journal.pone.0247205

117. Jimenez-Solem E, Petersen TS, Hansen C, Hansen C, Lioma C, Igel C, et al. Developing and validating COVID-19 adverse outcome risk prediction models from a bi-national European cohort of 5594 patients. *Sci Rep.* (2021) 11:3246. doi: 10.1038/s41598-021-81844-x

118. Gao Y, Cai G-Y, Fang W, Li H-Y, Wang S-Y, Chen L, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nat Commun.* (2020) 11:5033. doi: 10.1038/s41467-020-18684-2