Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

PASD1 genetic variation is associated with COVID-19 clinical severity in an admixed Brazilian population: an exploratory genome-wide association study

Lauana Fogaca de Almeida
Universidade Estadual Paulista (Unesp)
Fabiana Barcelos Furtado
Universidade Estadual Paulista (Unesp)
Leonardo Nazário de Moraes
Universidade Estadual Paulista (Unesp)
Carlos Magno Castelo Branco Fortaleza
Universidade Estadual Paulista (Unesp)
Rejane Maria Tommasini Grotto
rejane.grotto@unesp.br

Universidade Estadual Paulista (Unesp)

Research Article

Keywords: Genome-wide association study (GWAS), Host genetics, PASD1, SARS-CoV-2, Brazilian admixed population, Disease severity

Posted Date: July 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10348565/v1

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

Additional Declarations: No competing interests reported.

Abstract

The marked interindividual variability in COVID-19 clinical manifestations suggests that host genetic variation contributes to disease severity. To investigate host genetic variants associated with COVID-19 severity, we conducted an exploratory genome-wide association study (GWAS) in an admixed Brazilian cohort. A subset of 96 individuals was selected from a larger clinical cohort, of whom 92 underwent high-density SNP genotyping. After genotype quality control and identity-by-descent filtering, 73 unrelated individuals remained for genome-wide association analyses. The strongest association was identified at the PASD1 locus (rs56035326), where the minor allele (T) was significantly more frequent in mild than in severe COVID-19 cases (MAF = 0.54 vs. 0.05; $p = 4.42 \times 10^{-8}$), remaining significant after false discovery rate correction. Additional suggestive associations were observed in DNAH9, ADRA1A, and C1ORF132 but did not remain significant after correction for multiple testing. To our knowledge, this study provides the first evidence implicating PASD1 genetic variation in COVID-19 severity, providing a basis for future replication and functional studies investigating the biological mechanisms underlying severe COVID-19, particularly in admixed populations.

1. Background

The clinical manifestations of coronavirus disease 2019 (COVID-19), caused by the SARS-CoV-2 virus, exhibit profound heterogeneity, ranging from asymptomatic infection to life-threatening acute respiratory distress syndrome (ARDS) and multiorgan failure (Cascella et al., 2020; Park, 2020). While advanced age and preexisting comorbidities, such as obesity, chronic kidney disease, and peripheral arterial disease, are established risk factors for critical illness, they do not fully account for the broad spectrum of clinical phenotypes observed (Wang et al., 2020; Zhou et al., 2020). Accumulating evidence indicates that host genetic variation, particularly single nucleotide polymorphisms (SNPs), plays a pivotal role in shaping the host immune response, influencing both susceptibility to infection and the subsequent severity of the disease (COVID-19 Host Genetics Initiative, 2021).

Previous genome-wide association studies (GWAS) have identified several genetic loci associated with COVID-19 severity, consistently implicating genes involved in interferon signaling, antiviral immunity, inflammatory responses, and viral entry mechanisms, including the HLA region and other immune-related loci (Ellinghaus et al., 2020; Pairo-Castineira et al., 2021; Zeberg et al., 2021). These findings have substantially advanced the understanding of the host genetic architecture underlying COVID-19. Despite these advances, most GWAS have been conducted in populations of predominantly European ancestry, limiting the generalizability of their findings to other genetically diverse populations. Because allele frequencies, linkage disequilibrium patterns, and genetic architecture vary across populations, the inclusion of admixed populations, such as the Brazilian population, is essential for identifying population-specific loci associated with COVID-19 severity and expanding the global understanding of the host genetic architecture underlying disease severity.

To address the limited representation of admixed populations in previous genomic studies, we conducted an exploratory genome-wide association study (GWAS) in a Brazilian admixed cohort including individuals with mild, moderate, and severe COVID-19. Our aim was to identify candidate loci associated with COVID-19 severity, generate biologically plausible hypotheses regarding the molecular mechanisms underlying COVID-19 progression, and provide a foundation for future genomic and functional studies aimed at elucidating the biological mechanisms underlying COVID-19 severity.

2. Methods

A total of 233 individuals diagnosed with COVID-19 by RT-qPCR were enrolled in the parent clinical study at the Hospital das Clínicas da Faculdade de Medicina de Botucatu (HCFMB), Brazil. For the present exploratory genome-wide association study (GWAS), a subset of 96 individuals was selected according to predefined inclusion criteria, ensuring balanced representation across the mild (n = 32), moderate (n = 33), and severe (n = 31) clinical groups. Clinical severity was assigned by infectious disease specialists according to the World Health Organization (WHO) and the Brazilian Ministry of Health guidelines for COVID-19 clinical management (WHO, 2021; BRASIL, 2021). Mild cases presented with symptoms such as fever, cough, and anosmia without hypoxia; moderate cases exhibited clinical signs of non-severe pneumonia; and severe cases were characterized by severe respiratory impairment, low oxygen saturation, or the need for advanced respiratory support.

Genomic DNA was isolated from peripheral blood leukocytes (buffy coat) using the Bio Gene Kit (Bioclin, Brazil). DNA purity and concentration were assessed using NanoDrop™ spectrophotometry and Qubit™ fluorometric quantification, respectively. Based on predefined inclusion criteria, samples with adequate DNA quality and concentration for microarray analysis were selected to ensure balanced representation across the three clinical severity groups (mild, moderate, and severe). A total of 92 samples fulfilled these technical requirements and were genotyped using the Axiom™ Human Genotyping SARS-CoV-2 Research Array (Thermo Fisher Scientific, CA, USA), comprising 859,833 single nucleotide polymorphism (SNP) markers enriched for variants associated with immune response and susceptibility to viral infections.

Genotype calling was performed using the Axiom Analysis Suite following the manufacturer's Best Practices Workflow, applying more stringent quality thresholds than the default recommendations. The following criteria were adopted: percentage of passing samples ≥ 95%, sample call rate ≥ 97%, average call rate for passing samples ≥ 98.5%, and CR-cutoff ≥ 99%. The resulting genotype dataset was subsequently subjected to additional quality-control filtering using PLINK v1.9 (Chang et al., 2015), excluding variants with more than 5% missing genotypes, significant

deviation from Hardy-Weinberg equilibrium (HWE; $p < 10^{-8}$), and ambiguous alleles. These procedures yielded a dataset comprising 78 samples and 597,313 SNPs, which was subsequently subjected to identity-by-descent analysis, population structure assessment, ancestry inference, and genome-wide association analyses.

Identity-by-descent (IBD) analysis was performed using the SNPRelate package implemented in R, applying both the method-of-moments and KING estimators. A kinship coefficient threshold of ≥ 0.125 was used to identify related individuals and assess the reproducibility of technical duplicate samples included as internal quality controls. One sample from each technical duplicate pair and one individual from each related pair were removed prior to the genome-wide association analyses, yielding a final dataset of 73 unrelated individuals.

Population structure was assessed by Principal Component Analysis (PCA) using the SNPRelate package (Zheng et al., 2012) implemented in the R statistical environment (R Core Team, 2019). Prior to PCA, linkage disequilibrium (LD) pruning was performed using an r^2 threshold of 0.2, resulting in 66,268 approximately independent SNPs used for population structure analysis.

For global ancestry inference, the study samples were merged with reference populations from the Human Genome Diversity Project (HGDP) (Bergström et al., 2020) and the Simons Genome Diversity Project (SGDP) (Mallick et al., 2016). An additional PCA was performed on the merged dataset to confirm the clustering of the reference populations and identify potential outlier samples prior to supervised ancestry inference. Subsequently, global ancestry was inferred using ADMIXTURE v1.3.0 (Alexander et al., 2009) in supervised mode. Ten independent ADMIXTURE runs were performed, and the consensus ancestry estimates were obtained using CLUMPP (Jakobsson & Rosenberg, 2007). The resulting ancestry proportions were used to characterize the genomic ancestry composition of the study population and to evaluate the potential influence of population structure on downstream association analyses.

Genome-wide association analyses (GWAS) were performed using Fisher's exact test implemented in PLINK v1.9 (Chang et al., 2015), comparing allele frequencies among the severe versus mild, severe versus moderate, and moderate versus mild clinical groups. Raw *p*-values were adjusted for multiple testing using the Benjamini-Hochberg procedure implemented through the p.adjust function in the R statistical environment to control the False Discovery Rate (FDR). Association results were visualized using Manhattan plots, adopting $p = 1 \times 10^{-5}$ as the threshold for suggestive associations and $p = 5 \times 10^{-8}$ as the genome-wide significance threshold.

3. Results

Sample quality control and dataset composition

A total of 233 individuals were enrolled in parent clinical study. For the present exploratory GWAS, 96 individuals were selected according to predefined study design criteria to ensure balanced representation across clinical severity groups. Of these, 92 samples fulfilled the technical requirements for high-density genotyping. Following genotype calling and quality control filtering using the Axiom Analysis Suite and PLINK v1.9, 78 samples and 597,313 SNPs remained available for downstream analyses.

Identity by Descent (IBD) analysis confirmed five pairs of technical duplicates that had been intentionally included on the genotyping array as internal quality controls, demonstrating high reproducibility of genotype calling (**Supplementary Fig. 1**). Additionally, two pairs of individuals exhibited kinship coefficients compatible with first-degree relationships (kinship ≥ 0.125). To ensure independence among observations, one sample from each duplicate pair and one individual from each related pair were excluded from the association analyses, resulting in a final dataset of 73 unrelated individuals.

Population structure

Principal component analysis (PCA), performed using linkage disequilibrium-pruned SNPs, revealed no evident clustering according to COVID-19 clinical severity, indicating the absence of substantial population stratification among the mild, moderate, and severe groups (**Supplementary Fig. 2**).

Global ancestry inference demonstrated the expected admixed genetic background of the Brazilian population, with a predominance of European ancestry followed by African and Native American contributions. Similar ancestry distributions were observed across the three clinical groups, indicating that ancestry differences were unlikely to confound the subsequent association analyses (**Supplementary Fig. 3**).

Genome-wide association analysis

Genome-wide association analyses were performed through pairwise comparisons between the clinical groups (severe versus mild, severe versus moderate, and moderate versus mild) using Fisher's exact test followed by Benjamini–Hochberg false discovery rate (FDR) correction.

In the comparison between severe and mild COVID-19 cases, one variant remained significantly associated after FDR correction (Table 1). The strongest association was observed for the intronic variant rs56035326 within PASD1, in which the T allele frequency was markedly lower among severe than mild cases (5.6% vs. 54.0%; OR = 0.05; 95% CI: 0.01–0.18; $p = 4.43 \times 10^{-8}$; FDR = 0.022).

Table 1
Genetic variants associated with COVID-19 severity identified by genome-wide association analysis. Allele frequencies (AF), odds ratios (OR), 95% confidence intervals (95% CI), nominal *p*-values and Benjamini–Hochberg false discovery rate (FDR) adjusted *p*-values are shown for each comparison.

Comparison	SNP	Chr	Position (hg38)	Gene	Functional annotation	Effect allele	Case AF	Control AF	OR (95% CI)	*p*-value	FDR
Severe *vs* Mild	rs56035326	X	151,611,081	PASD1	Intronic	T	0.056	0.540	0.05 (0.01–0.18)	4.43×10^{-8}	**0.022**
Severe *vs* Mild	rs9902083	17	11,679,350	DNAH9	—	G	0.000	0.320	0.02 (0.001–0.34)	1.86×10^{-6}	> 0.05
Severe *vs* Moderate	rs13282836	8	26,820,689	ADRA1A	Intronic	T	0.037	0.500	0.04 (0.01–0.18)	9.30×10^{-8}	> 0.05
Severe *vs* Moderate	rs7055322	X	5,652,781	NLGN4X	Downstream	C	0.019	0.405	0.03 (0.003–0.22)	1.03×10^{-6}	> 0.05
Severe *vs* Moderate	rs2724373	1	207,825,855	C1orf132	—	T	0.074	0.500	0.08 (0.02–0.26)	3.25×10^{-6}	> 0.05
Moderate *vs* Mild	rs7620281	3	151,601,300	IGSF10	Upstream (~ 33 kb)	A	0.024	0.420	0.03 (0.004–0.26)	3.87×10^{-6}	> 0.05

Another variant, rs9902083, annotated to DNAH9, also exhibited a marked difference in allele frequency, with the G allele being absent among severe cases and present at a frequency of 32.0% among mild cases (OR = 0.02; 95% CI: 0.001–0.34; $p = 1.86 \times 10^{-6}$). However, this association did not remain significant after correction for multiple testing.

In the comparison between severe and moderate cases, three variants exhibited significant differences in allele frequencies before correction for multiple testing (Table 1). The intronic variant rs13282836 within ADRA1A showed T allele frequencies of 3.7% among severe cases and 50.0% among moderate cases (OR = 0.04; 95% CI: 0.01–0.18; $p = 9.30 \times 10^{-8}$). The downstream variant rs7055322, located near NLGN4X, exhibited C allele frequencies of 1.9% and 40.5%, respectively (OR = 0.03; 95% CI: 0.003–0.22; $p = 1.03 \times 10^{-6}$). The variant rs2724373, annotated to C1orf132, also showed a significant difference in allele frequencies between severe and moderate cases (7.4% vs. 50.0%; OR = 0.08; 95% CI: 0.02–0.26; $p = 3.25 \times 10^{-6}$). Nevertheless, none of these associations remained significant after FDR correction.

Finally, in the comparison between moderate and mild cases, the upstream variant rs7620281, located approximately 33 kb from IGSF10, exhibited an A allele frequency of 2.4% among moderate cases and 42.0% among mild cases (OR = 0.03; 95% CI: 0.004–0.26; $p = 3.87 \times 10^{-6}$). However, this association did not remain significant after FDR correction (Table 1).

4. Discussion

Unraveling the host genetic architecture is paramount to understanding why SARS-CoV-2 infection induces asymptomatic presentation in some individuals while others progress to severe clinical outcomes (Blanco-Melo et al., 2020). We performed an exploratory genome-wide association study (GWAS) in an admixed Brazilian cohort to identify host genetic variants associated with different degrees of COVID-19 severity.

Cryptic relatedness may introduce bias in genetic association studies, particularly when related individuals are retained in downstream analyses. In the present study, identity-by-descent (IBD) analysis confirmed five pairs of technical duplicates intentionally included as internal quality controls and identified two pairs of individuals with kinship coefficients consistent with first-degree relationships. The high concordance observed among technical duplicates supported the reproducibility of genotype calling, whereas the exclusion of one sample from each duplicate pair and one individual from each related pair ensured the statistical independence of the 73 unrelated individuals included in the final genome-wide association analyses. This quality-control procedure is widely recommended in GWAS to minimize confounding and reduce the risk of spurious genetic associations (Astle & Balding, 2009; COVID-19 Host Genetics Initiative, 2021; Namkoong et al., 2021).

Beyond sample relatedness, characterization of the ancestry composition of the study population was an essential step to evaluate the potential influence of population structure on the association analyses. Global ancestry inference confirmed the highly admixed genetic background

expected for the Brazilian population, with a predominance of European ancestry accompanied by African and Native American contributions, consistent with previous studies of populations from Southeastern Brazil (Pena et al., 2011; Kehdy et al., 2015). Principal component analysis further demonstrated no evident clustering according to COVID-19 clinical severity, suggesting that major population stratification was unlikely to account for the observed association signals. These findings support the robustness of the downstream analyses and are consistent with previous genomic studies of COVID-19 that emphasized the importance of evaluating population structure in genetically admixed cohorts (Ellinghaus et al., 2020; Shelton et al., 2021).

Following genome-wide association analysis and correction for multiple testing, rs56035326 within the PASD1 (PAS domain-containing repressor 1) locus emerged as the only variant that remained significantly associated with COVID-19 clinical severity after FDR adjustment. The minor allele (T) was substantially more frequent among patients with mild disease than among those with severe COVID-19, indicating that this allele is associated with a lower probability of severe disease in this cohort. To our knowledge, this is the first report describing an association between genetic variation in PASD1 and COVID-19 clinical severity, thereby expanding the spectrum of candidate loci potentially involved in the host response to SARS-CoV-2 infection.

The identification of a novel candidate locus is consistent with evidence from large-scale GWAS and whole-genome sequencing studies demonstrating that host genetic determinants involved in antiviral immunity, interferon signaling, epithelial barrier integrity, and inflammatory responses contribute substantially to interindividual variability in COVID-19 clinical manifestations (COVID-19 Host Genetics Initiative, 2021; Pairo-Castineira et al., 2021; Kousathanas et al., 2022; Zhang et al., 2020; Van der Made et al., 2022). Although PASD1 has not previously been reported among the loci associated with susceptibility to SARS-CoV-2 infection or severe COVID-19, the identification of rs56035326 broadens the spectrum of candidate genes potentially involved in disease progression. Furthermore, this finding highlights the importance of conducting genomic studies in historically underrepresented populations, where population-specific variants may reveal biological mechanisms that remain undetected in cohorts of predominantly European ancestry.

The associated variant, rs56035326, is located within an intronic region of PASD1, indicating that its biological effect is unlikely to result from direct alterations in protein sequence. Instead, the observed association may reflect linkage disequilibrium with one or more functional regulatory variants influencing PASD1 expression or the activity of nearby regulatory elements. This interpretation is consistent with growing evidence indicating that most susceptibility variants identified by GWAS are located within noncoding genomic regions and frequently exert their biological effects by modulating enhancer activity, chromatin accessibility, promoter interactions, or gene expression programs (Visscher et al., 2017; Tam et al., 2019). More recently, the integration of GWAS with expression quantitative trait locus (eQTL) analyses, chromatin conformation data, and single-cell transcriptomics has substantially improved the functional interpretation of disease-associated loci, including those involved in severe COVID-19 (Schmiedel et al., 2021; Pahl et al., 2022). Therefore, future studies investigating the regulatory landscape surrounding PASD1 will be important to determine whether rs56035326 directly influences gene expression or represents a marker in linkage disequilibrium with the true functional variant.

Biologically, PASD1 encodes a transcriptional regulatory protein containing a PAS domain, a structural motif involved in cellular signaling and environmental sensing (Michael et al., 2015). Although the biological role of PASD1 in viral infections remains largely unexplored, transcriptional regulation has emerged as a central component of the host response to SARS-CoV-2 infection. Dysregulated interferon signaling, exaggerated inflammatory responses, and widespread alterations in host transcriptional programs have been consistently associated with severe COVID-19 (Zhang et al., 2020; Pairo-Castineira et al., 2021; Rosa et al., 2021; Pahl et al., 2022; Schmiedel et al., 2021). In this context, it is plausible that genetic variation within the PASD1 locus may influence regulatory pathways involved in the host response to infection. The higher frequency of the rs56035326 minor allele among patients with mild disease is consistent with this hypothesis, although the molecular mechanisms underlying this association remain to be elucidated.

Although the present findings provide evidence supporting PASD1 as a novel candidate locus associated with COVID-19 severity, the exploratory nature of this study and the limited sample size remaining after stringent quality-control filtering warrant cautious interpretation. The observed association does not permit direct inference regarding the molecular consequences of rs56035326, and functional studies investigating gene expression, epigenetic regulation, chromatin organization, and immune cell-specific transcriptional responses will be necessary to determine whether genetic variation at the PASD1 locus contributes to pathways involved in antiviral immunity or other mechanisms influencing COVID-19 severity.

Although none of the remaining variants retained statistical significance after FDR correction, several suggestive association signals merit further investigation based on the biological functions of the genes involved. Among these, rs9902083 within the DNAH9 locus is particularly noteworthy because DNAH9 encodes a dynein axonemal heavy chain that is essential for ciliary motility and respiratory epithelial defense. Efficient mucociliary clearance constitutes one of the primary innate defense mechanisms responsible for removing inhaled pathogens from the respiratory tract (Ferkol & Leigh, 2012; Horani et al., 2018). Since SARS-CoV-2 primarily infects ciliated airway epithelial cells, genetic variation affecting ciliary structure or function could plausibly influence viral clearance and, consequently, disease progression. Consistent with this hypothesis, experimental and histopathological studies have demonstrated that SARS-CoV-2 infection causes extensive damage to ciliated epithelial cells, impairing mucociliary transport and delaying viral clearance from the respiratory tract (Robinot et al., 2021). Therefore, although

rs9902083 did not remain statistically significant after FDR correction, the established biological role of DNAH9 supports prioritizing this locus for replication in independent cohorts and subsequent functional studies.

Another suggestive association was observed for rs13282836 within the ADRA1A locus. ADRA1A encodes the α1A-adrenergic receptor, which plays an important role in vascular tone regulation and has also been implicated in the modulation of inflammatory responses. Transcriptomic studies have reported altered activity of adrenergic signaling pathways, including RNA hyper-editing of ADRA1A, during SARS-CoV-2 infection (Picardi et al., 2021). In addition, severe COVID-19 is characterized by endothelial dysfunction, vascular injury, and dysregulated inflammatory responses (Del Valle et al., 2020; Mehta et al., 2020), suggesting that genes involved in adrenergic signaling may contribute to interindividual variability in disease severity. Experimental evidence further indicates that adrenergic pathways participate in immune regulation by influencing cytokine production, leukocyte migration, endothelial activation, and tissue perfusion (Pairo-Castineira et al., 2021). Therefore, although rs13282836 did not remain statistically significant after FDR correction, the biological functions of ADRA1A support further evaluation of this locus in larger cohorts and functional studies.

Additional suggestive association signals were identified within or near the C1ORF132, NLGN4X, and IGSF10 loci. Although the biological relevance of C1ORF132 and NLGN4X to COVID-19 remains poorly understood, these findings broaden the range of candidate loci that may warrant investigation in future studies. By contrast, IGSF10, a member of the immunoglobulin superfamily, has been implicated in cell migration and signaling processes involved in immune regulation (Wojtowicz et al., 2020), making it a biologically plausible candidate for further investigation. Nevertheless, none of these variants remained statistically significant after FDR correction and should therefore be regarded as exploratory findings. Replication in independent cohorts, particularly in admixed Latin American populations, together with functional characterization studies, will be essential to determine whether these loci represent true genetic determinants of COVID-19 severity or reflect chance findings resulting from the limited statistical power of the present study.

The present study has several limitations that should be considered when interpreting these findings. First, the relatively limited sample size, particularly after the exclusion of related individuals during stringent quality-control procedures, reduced the statistical power to detect variants with smaller effect sizes. Consequently, the associations reported here require replication in larger and ethnically diverse independent cohorts (Chamnanphon et al., 2022). Importantly, despite the limited sample size, rs56035326 within the PASD1 locus remained significantly associated with COVID-19 severity after FDR correction, supporting its potential relevance as a candidate genetic determinant of disease severity. Nevertheless, the rigorous quality-control procedures, comprehensive assessment of population structure, and conservative statistical approach adopted in this study strengthen the robustness of the reported associations and provide a solid framework for future replication studies and meta-analyses investigating the host genetic architecture of COVID-19.

5. Conclusions

In conclusion, this exploratory GWAS identified rs56035326 within the PASD1 locus as the only variant that remained significantly associated with COVID-19 clinical severity after FDR correction. To our knowledge, this is the first report implicating PASD1 as a candidate locus associated with COVID-19 severity, expanding the spectrum of host genetic factors that may contribute to interindividual differences in disease outcome. Although these findings require replication in larger and ethnically diverse cohorts, they reinforce the importance of investigating genetically underrepresented populations and integrating GWAS with functional genomic approaches, including transcriptomics, epigenomics, expression quantitative trait locus analyses, and experimental validation, to elucidate the biological mechanisms underlying COVID-19 severity.

Abbreviations

AF
Allele Frequency
ARDS
Acute Respiratory Distress Syndrome
CAAE
Certificate of Presentation for Ethical Consideration
CAPES
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CI
Confidence Interval
CNPq
National Council for Scientific and Technological Development
COVID-19
Coronavirus Disease 2019
DNA
Deoxyribonucleic Acid

eQTL
Expression Quantitative Trait Locus
FAPESP
São Paulo Research Foundation
FDR
False Discovery Rate
GWAS
Genome-Wide Association Study
HCFMB
Hospital das Clínicas da Faculdade de Medicina de Botucatu
HGDP
Human Genome Diversity Project
HLA
Human Leukocyte Antigen
HWE
Hardy–Weinberg Equilibrium
IBD
Identity by Descent
IFN
Interferon
IgSF
Immunoglobulin Superfamily
LD
Linkage Disequilibrium
MAF
Minor Allele Frequency
OR
Odds Ratio
PAS
PER-ARNT-SIM
PCA
Principal Component Analysis
RT-qPCR
Reverse Transcription Quantitative Polymerase Chain Reaction
SARS-CoV-2
Severe Acute Respiratory Syndrome Coronavirus 2
SGDP
Simons Genome Diversity Project
SNP
Single Nucleotide Polymorphism
UNESP
São Paulo State University
WHO
World Health Organization.

Declarations

7. Ethics approval and consent to participate

The study was conducted in accordance with the principles of the Declaration of Helsinki and approved by the Research Ethics Committee of the Hospital das Clínicas, Faculdade de Medicina de Botucatu, São Paulo State University (UNESP) (CAAE: 31557120.3.0000.5411). Written informed consent was obtained from all participants or, when applicable, from their legally authorized representatives prior to enrollment.

8. Consent for publication

Not applicable.

10. Clinical trial number

Not applicable.

11. Competing interests

The authors declare that they have no competing interests.

12. Funding

This work was supported by the São Paulo Research Foundation (FAPESP; grant #2022/02261-5), the National Council for Scientific and Technological Development (CNPq; grant #312765/2023-1), and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. The funding agencies had no role in the study design; collection, analysis, or interpretation of data; writing of the manuscript; or the decision to submit the manuscript for publication.

Author Contribution

LFA contributed to conceptualization, study design, experimental execution, data analysis, and drafting of the manuscript. FBF contributed to experimental execution, scientific writing, and English language revision. LNM contributed to study design, experimental setup, assay standardization, data interpretation, and manuscript writing. CMCBF contributed to conceptualization, intellectual supervision, project development, and funding acquisition. RMTG conceived the initial project, supervised the study, provided resources, acquired funding, and contributed to writing and manuscript revision. All authors read and approved the final manuscript.

Acknowledgement

The authors sincerely thank all patients who participated in this study. We also acknowledge the Hospital das Clínicas da Faculdade de Medicina de Botucatu (HCFMB) and all healthcare professionals involved in the care of patients with COVID-19 for their dedication, commitment, and invaluable contributions throughout the pandemic.

9. Availability of data and materials

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

References

1. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64.
2. Astle W, Balding DJ. Population structure and cryptic relatedness in genetic association studies. Stat Sci. 2009;24:451–71.
3. Bergström A, et al. Insights into human genetic variation and population history from 929 diverse genomes. Science. 2020;367:aay5012.
4. Blanco-Melo D, et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell. 2020;181:1036–e459.
5. Brazil. Ministry of Health. COVID-19: clinical signs, symptoms, and severity classification. Brasília: Ministry of Health; 2021.
6. Cascella M, et al. In: StatPearls, editor. Features, evaluation, and treatment of coronavirus. Treasure Island (FL): StatPearls Publishing; 2020.
7. Chamnanphon M, et al. Host genetic factors of COVID-19 susceptibility and disease severity in a Thai population. J Hum Genet. 2022;67:377–86.
8. Chang CC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.
9. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. Nature. 2021;600:472–7.
10. Del Valle DM, et al. An inflammatory cytokine signature predicts COVID-19 severity and survival. Nat Med. 2020;26:1636–43.
11. Ellinghaus D, et al. Genomewide association study of severe Covid-19 with respiratory failure. N Engl J Med. 2020;383:1522–34.
12. Ferkol TW, Leigh MW. Ciliopathies: the central role of cilia in a spectrum of pediatric disorders. J Pediatr. 2012;160:366–71.
13. Horani A, Genetics and biology of primary ciliary dyskinesia. Paediatr Respir Rev., Jakobsson M, Rosenberg NA et al. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007;23:1801-6.

14. Kehdy FSG, et al. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. Proc Natl Acad Sci U S A. 2015;112:8696–701.
15. Kousathanas A, Pairo-Castineira E, Rawlik K, et al. Whole-genome sequencing reveals host factors underlying critical COVID-19. Nature. 2022;607:97–103.
16. Mallick S, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature. 2016;538:201–6.
17. Mehta P, et al. COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet. 2020;395:1033–4.
18. Michael AK, et al. Cancer/Testis antigen PASD1 silences the circadian clock. Mol Cell. 2015;60:203–18.
19. Namkoong H, et al. DOCK2 is involved in the host genetics and biology of severe COVID-19. Nature. 2022;609:754–60.
20. Pahl MC, et al. Implicating effector genes at COVID-19 GWAS loci using promoter-focused Capture-C in disease-relevant immune cell types. Genome Biol. 2022;23:125.
21. Pairo-Castineira E, et al. Genetic mechanisms of critical illness in COVID-19. Nature. 2021;591:92–8.
22. Park SE. Epidemiology, virology, and clinical features of severe acute respiratory syndrome coronavirus 2. Pediatr Infect Vaccine. 2020;27:1–10.
23. Pena SDJ, et al. The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. PLoS ONE. 2011;6:e17063.
24. Picardi E, Mansi L, Pesole G. Detection of A-to-I RNA editing in SARS-CoV-2. Genes (Basel). 2021;12:1225.
25. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019.
26. Robinot R, Hubert M, de Melo GD, et al. SARS-CoV-2 infection damages airway motile cilia and impairs mucociliary clearance. Nat Commun. 2021;12:4354.
27. Rosa BA, et al. IFN signaling and neutrophil degranulation transcriptional signatures are induced during SARS-CoV-2 infection. Commun Biol. 2021;4:290.
28. Schmiedel BJ, et al. COVID-19 genetic risk variants are associated with expression of multiple genes in diverse immune cell types. Nat Commun. 2021;12:6767.
29. Shelton JF, et al. Trans-ethnic analysis reveals genetic and non-genetic associations with COVID-19 susceptibility and severity. Nat Genet. 2021;53:801–8.
30. Tam V, et al. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019;20:467–84.
31. van der Made CI, Netea MG, van der Veerdonk FL, Hoischen A. Clinical implications of host genetic variation and susceptibility to severe or critical COVID-19. Genome Med. 2022;14:96.
32. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet. 2017;101:5–22.
33. World Health Organization. Living guidance for clinical management of COVID-19. Geneva: World Health Organization; 2021.
34. Wojtowicz WM, et al. A human IgSF cell-surface interactome reveals a complex network of protein-protein interactions. Cell. 2020;182:1027–43.
35. Zeberg H, Pääbo S. A genomic region associated with protection against severe COVID-19 is inherited from Neandertals. Proc Natl Acad Sci U S A. 2021;118:e2026309118.
36. Zhang Q, et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. Science. 2020;370:eabd4570.
37. Zheng X, et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28:3326–8.
38. Zhou Y, et al. Obesity and diabetes as high-risk factors for severe coronavirus disease 2019. Diabetes Metab Res Rev. 2021;37:e3377.

Supplementary Files

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

- Supplementaryinfo.docx