



Article

Global and Sex-Stratified Genome-Wide Association Study of Long COVID Based on Patient-Driven Symptom Recall

Sara Polo-Alonso ^{1,2}, Álvaro Hernáez ^{1,3,4,*} , Irene R. Dégano ^{1,3,5,6}, Ruth Martí-Lluch ^{7,8} , Mel·lina Pinsach-Abuin ⁹, Roberto Elosua ^{3,5,10}, Isaac Subirana ^{3,10}, Marta Puigmulé ³, Alexandra Pérez ⁹ , Raquel Cruz ^{11,12} , Silvia Diz-de Almeida ¹¹ , Eulàlia Puigdecant ^{5,6}, Elisabet Selga ^{5,6} , Xavier Nogues ^{13,14,15,16} , Joan Ramon Masclans ^{17,18,19} , Roberto Güerri-Fernández ^{16,20,21}, Héctor Cubero-Gallego ^{22,23}, Helena Tizon-Marcos ^{3,22,23} , Beatriz Vaquerizo ^{3,16,22,23}, Ramon Brugada ^{3,9,24,25}, Rafel Ramos ^{7,8,25,26}, Anna Camps-Vilaró ^{1,3,5,*} and Jaume Marrugat ^{1,3,†}

- ¹ Registre Gironí del Cor (REGICOR) Study Group, Hospital del Mar Research Institute, 08003 Barcelona, Spain
- ² PhD Program in Biomedicine, Universitat Pompeu Fabra, 08003 Barcelona, Spain
- ³ Centro de Investigación Biomédica en Red de Enfermedades Cardiovasculares (CIBERCV), Instituto de Salud Carlos III, 28029 Madrid, Spain
- ⁴ Blanquerna School of Health Sciences, University Ramon Llull, 08022 Barcelona, Spain
- ⁵ Faculty of Medicine, University of Vic-Central University of Catalonia, 08500 Vic, Spain
- ⁶ Institute for Research and Innovation in Life Sciences and Health in Central Catalonia (IRIS-CC), 08500 Vic, Spain
- ⁷ Vascular Health Research Group, Institut Universitari per a la Recerca en Atenció Primària Jordi Gol i Gurina, 17002 Girona, Spain
- ⁸ Girona Biomedical Research Institute, 17190 Girona, Spain
- ⁹ Cardiovascular Genetics Center, Institut d'Investigació Biomèdica de Girona Dr. Josep Trueta (IdIBGi), 17190 Salt, Spain
- ¹⁰ Cardiovascular Epidemiology and Genetics Group, Hospital del Mar Research Institute, 08003 Barcelona, Spain
- ¹¹ Centro Singular de Investigación en Medicina Molecular y Enfermedades Crónicas (CIMUS), Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain
- ¹² Centro de Investigación Biomédica en Red de Enfermedades Raras, Instituto de Salud Carlos III, 28029 Madrid, Spain
- ¹³ Musculoskeletal Research Unit, Hospital del Mar Research Institute, 08003 Barcelona, Spain
- ¹⁴ Department of Internal Medicine, Hospital del Mar, 08003 Barcelona, Spain
- ¹⁵ Centro de Investigación Biomédica en Red de Fragilidad y Envejecimiento Saludable, Instituto de Salud Carlos III, 28029 Madrid, Spain
- ¹⁶ Faculty of Medicine, Universitat Autònoma de Barcelona (UAB), 08193 Bellaterra, Spain
- ¹⁷ Critical Illness Research Group (GREPAC), Hospital del Mar Research Institute, 08003 Barcelona, Spain
- ¹⁸ Department of Critical Care, Hospital del Mar, 08003 Barcelona, Spain
- ¹⁹ Faculty of Medicine and Life Sciences, Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain
- ²⁰ Centro de Investigación Biomédica en Red de Enfermedades Infecciosas, Instituto de Salud Carlos III, 28029 Madrid, Spain
- ²¹ Department of Infectious Diseases, Hospital del Mar Research Institute, 08003 Barcelona, Spain
- ²² Biomedical Research in Heart Diseases Group, Hospital del Mar Research Institute, 08003 Barcelona, Spain
- ²³ Department of Cardiology, Hospital del Mar, 08003 Barcelona, Spain
- ²⁴ Department of Cardiology, Hospital Josep Trueta, University of Girona, 17007 Girona, Spain
- ²⁵ Department of Medical Science, School of Medicine, University of Girona, 17071 Girona, Spain
- ²⁶ Primary Care Services, Catalan Institute of Health, 08007 Barcelona, Spain
- * Correspondence: ahernaez@researchmar.net (Á.H.); acamps@researchmar.net (A.C.-V.)
- † These authors contributed equally to this work.



Academic Editors: Manuela Rizzi and Pier Paolo Sainaghi

Received: 4 August 2025

Revised: 16 September 2025

Accepted: 22 September 2025

Published: 22 September 2025

Citation: Polo-Alonso, S.; Hernáez, Á.; Dégano, I.R.; Martí-Lluch, R.; Pinsach-Abuin, M.; Elosua, R.; Subirana, I.; Puigmulé, M.; Pérez, A.; Cruz, R.; et al. Global and Sex-Stratified Genome-Wide Association Study of Long COVID Based on Patient-Driven Symptom Recall. *Int. J. Mol. Sci.* **2025**, *26*, 9252. <https://doi.org/10.3390/ijms26189252>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract

We aimed to explore the global and sex-specific genetic variants associated with long COVID, as defined by patient-driven symptom recall. A 1-year cohort study of 2411 COVID-19 patients collected long COVID symptoms with an open-ended, non-directed questionnaire, and long COVID incidence was determined according to the World Health Organization definition. Global and sex-stratified genome-wide association analyses were conducted

by logistic regression models adjusted for age, sex (in the global analysis), and the first 10 principal components. We assessed sex-variant interactions and performed gene-based analyses, gene mapping, and gene-set enrichment analyses. When comparing the 1392 long COVID cases with the non-cases, we identified 23 lead variants from suggestive signals: 13 from the global analysis, 5 from females, and 5 from males. Five variants showed a significant interaction with sex (two in females, three in males). We mapped 15 protein-coding genes related to diseases of the immune and nervous systems and tumoral processes. Notably, *CD5* and *VPS37C*, linked to immune function, were significantly associated with long COVID in men. Our results suggest that persistent immune dysregulation may be involved in the development of precisely defined long COVID.

Keywords: post-acute COVID-19 syndrome; COVID-19; SARS-CoV-2; genome-wide association study; genetic polymorphism; sex characteristics

1. Introduction

Since the beginning of the COVID-19 pandemic, many patients have developed persistent symptoms lasting months after acute infection, a condition now recognized as long COVID or post-acute sequelae of SARS-CoV-2 infection [1,2]. It is estimated that between 6% and 70% of individuals infected with SARS-CoV-2 develop long COVID [3,4]. It includes a variety of symptoms such as fatigue, dyspnea, chest pain, anxiety, depression, fever, and hair loss, many of which limit quality of life [3,5–8].

Understanding the biological mechanisms driving severe and long COVID, potentially linked to immune dysregulation and its consequences [9,10], is crucial for developing targeted therapeutic strategies. Genetics could provide valuable insights into these mechanisms [11]. The COVID-19 Host Genetics Initiative recently carried out a trans-ancestry GWAS meta-analysis, applying a strict case definition (test-verified infection) versus broad population controls, and identified a genome-wide significant association at the *FOXP4* locus in 3018 strict cases versus 994,582 controls [12]. However, by aggregating studies with variable definitions of long COVID, follow-up intervals, and limited sex stratification, such meta-analyses cannot uniformly capture the depth or temporal dynamics of long COVID symptoms. Moreover, its checklist-based phenotyping cannot distinguish persistent, patient-reported sequelae from prompted symptom endorsement, potentially diluting the long COVID case definition. Here, we present a prospectively recruited GWAS of long COVID in 2411 RT-PCR- or antigen-confirmed SARS-CoV-2-positive individuals from Catalonia, Spain, applying an open-ended, one-year follow-up questionnaire to capture 24 persistent symptoms with a homogeneous clinical definition. This open-ended, clinician-administered interview allows unaided symptom recall, thereby yielding a deeply phenotyped outcome less prone to misclassification. We aimed to analyze the genetic variants associated with the development of long COVID in female and male separately, as well as in both sexes combined, and to deepen the understanding of the biological mechanisms underlying the condition.

2. Results

2.1. Participants

2411 individuals were eligible for further analysis, 1392 suffered from long COVID, and 1019 did not (Figure 1). As seen in Table 1, individuals with long COVID were younger than those without long COVID. Females with long COVID showed a lower prevalence of

hypertension, while males with long COVID experienced a more severe acute phase of the infection and showed a different ancestry distribution.

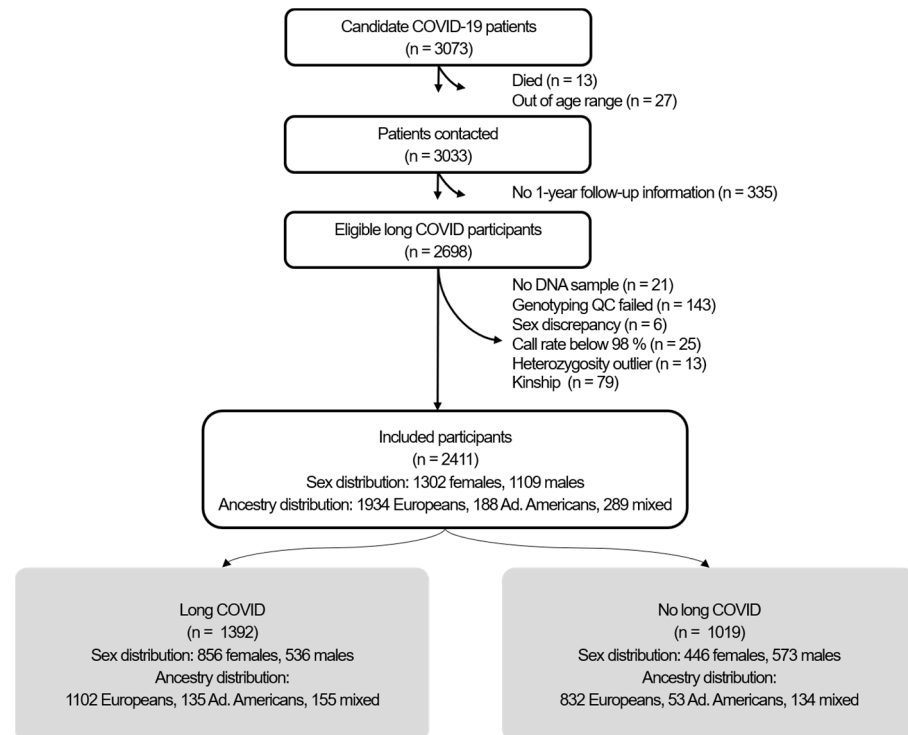


Figure 1. Flow chart of the GINA-COVID study.

Table 1. Participants with and without incident long COVID, stratified by sex.

	Females (n = 1302)			Males (n = 1109)		
	With Long COVID (n = 856)	Without Long COVID (n = 446)	p-Value	With Long COVID (n = 536)	Without Long COVID (n = 573)	p-Value
Age (mean ± SD)	56.4 ± 12.0	58.3 ± 12.6	0.009	57.8 ± 12.2	60.1 ± 12.7	0.002
Severe (hospitalized) acute phase (n, %)	196 (23.4%)	91 (20.5%)	0.261	242 (47.3%)	164 (29.6%)	<0.001
Current smokers (n, %)	57 (6.70%)	43 (9.71%)	0.070	49 (9.23%)	72 (12.7%)	0.086
Diabetes mellitus (n, %)	86 (10.0%)	44 (9.87%)	0.995	80 (14.9%)	99 (17.3%)	0.320
Hypertension (n, %)	202 (23.6%)	144 (32.3%)	0.001	209 (39.0%)	225 (39.3%)	0.956
Dyslipidemia (n, %)	201 (23.5%)	121 (27.1%)	0.175	179 (33.4%)	194 (34.0%)	0.888
BMI (categorized):			0.534			0.234
<18.5 kg/m ² (n, %)	6 (0.72%)	6 (1.40%)		2 (0.38%)	1 (0.18%)	
18.5–24.9 kg/m ² (n, %)	286 (34.3%)	155 (36.1%)		109 (20.7%)	137 (25.0%)	
25.0–29.9 kg/m ² (n, %)	268 (32.1%)	127 (29.6%)		266 (50.5%)	277 (50.5%)	
≥30.0 kg/m ² (n, %)	275 (32.9%)	141 (32.9%)		150 (28.5%)	133 (24.3%)	
Ancestry:			0.112			0.002
Admixed Americans (n, %)	89 (10.4%)	31 (6.95%)		46 (8.58%)	22 (3.84%)	
Europeans (n, %)	677 (79.1%)	370 (83.0%)		425 (79.3%)	462 (80.6%)	
Mixed (n, %)	90 (10.5%)	45 (10.1%)		65 (12.1%)	89 (15.5%)	

2.2. Genome-Wide Association Analysis and Variant–Sex Interaction

Several suggestive signals (p -value $< 5 \times 10^{-6}$) were identified (Figure 2). In the global analysis, 13 genomic loci were located on chromosomes 1, 2, 3, 4, 8, and 12. Sex-stratified

analyses revealed 5 genomic loci in females (chromosomes 4, 9, 17, and 19) and in males (chromosomes 1, 4, 11, 19, and 21). 23 lead genetic variants were identified, (13 in the global analysis, 5 in females, and 5 in males; Table 2), and none were shared across the analyses. The gene variant rs10888603, identified in males, exhibited the lowest p -value (5.2×10^{-8}). In the sensitivity analysis results (Figure S1), the three genomic risk loci described in males on chromosomes 4, 11, and 21 (Table S1) were confirmed. For females, no clear genetic signals were found.

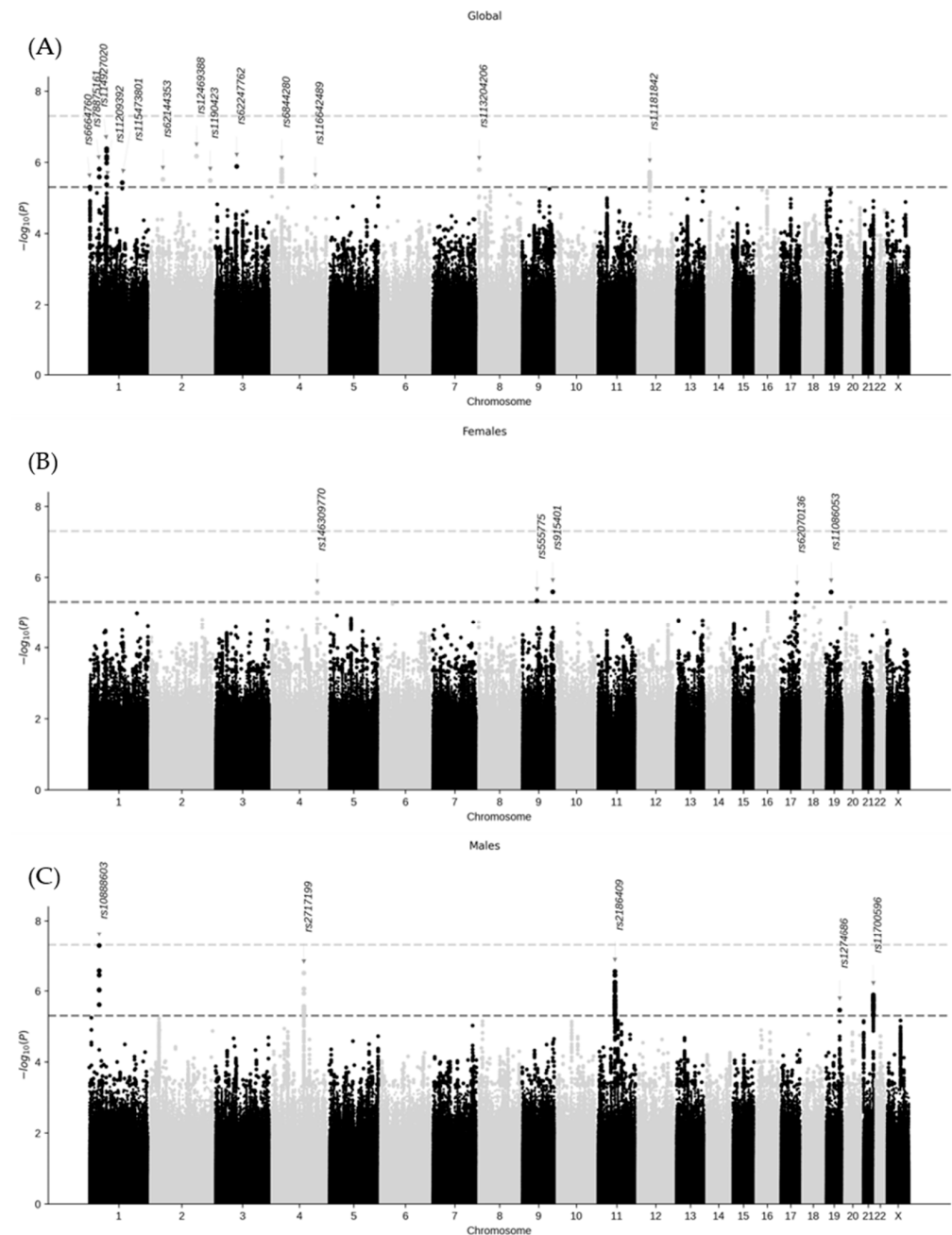


Figure 2. Manhattan plots of long COVID for the global (A) and sex-stratified ((B) females; (C) males) analyses. Lead variants for each association are indicated. Dark gray and light gray lines represent the p -value significance thresholds of 5×10^{-6} and 5×10^{-8} , respectively.

Table 2. List of the lead variants from global and sex-stratified analyses.

Gene Variant	Chrom.	Position (GRCh38)	Non-Effect Allele	Effect Allele	Effect Allele Frequency	β	Standard Error	GWAS p -Value	Gene Variant–Sex Interaction p -Value ¹
Females and males combined									
rs6664760	1	3,731,704	T	C	0.035	0.857	0.188	5.0×10^{-6}	1
rs78875161	1	39,627,261	C	T	0.034	−0.800	0.167	1.5×10^{-6}	1
rs114927020	1	68,967,153	G	A	0.099	−0.518	0.102	4.2×10^{-6}	1
rs11209392	1	69,122,167	T	G	0.016	−1.194	0.260	4.3×10^{-6}	1
rs115473801	1	157,613,190	G	A	0.027	−0.891	0.193	3.7×10^{-6}	1
rs62144353	2	42,755,360	C	A	0.111	0.467	0.100	3.0×10^{-6}	1
rs12469388	2	177,824,304	G	T	0.030	−0.908	0.183	6.7×10^{-6}	1
rs1190423	2	232,449,952	A	C	0.341	−0.298	0.064	3.3×10^{-6}	1
rs62247762	3	73,537,975	C	T	0.024	−0.989	0.204	1.3×10^{-6}	1
rs6844280	4	31,385,522	C	T	0.487	−0.293	0.061	1.6×10^{-6}	1
rs116642489	4	152,949,693	C	G	0.065	−0.558	0.122	4.8×10^{-6}	1
rs113204206	8	1,585,368	C	G	0.105	−0.474	0.099	1.6×10^{-6}	1
rs11181842	12	43,007,535	C	T	0.161	−0.387	0.081	1.9×10^{-6}	1
Only females									
rs146309770	4	160,217,229	T	TA	0.035	0.613	0.131	2.7×10^{-6}	1.6×10^{-2}
rs555775	9	75,445,184	C	T	0.034	−0.535	0.117	4.6×10^{-6}	1.0×10^{-1}
rs915401	9	131,013,384	A	G	0.099	0.419	0.089	2.6×10^{-6}	9.1×10^{-5}
rs62070136	17	72,195,002	T	G	0.016	−0.884	0.189	3.1×10^{-6}	2.1×10^{-1}
rs11086053	19	17,119,562	A	C	0.027	−0.693	0.148	2.6×10^{-6}	1
Only males									
rs10888603	1	38,972,945	A	C	0.035	−0.592	0.109	5.2×10^{-8}	5.5×10^{-2}
rs2717199	4	111,595,181	A	G	0.034	−0.478	0.094	3.1×10^{-7}	1.8×10^{-3}
rs2186409	11	61,127,020	T	G	0.099	−0.496	0.097	2.8×10^{-7}	7.8×10^{-4}
rs1274686	19	50,495,701	C	T	0.016	0.832	0.179	3.5×10^{-6}	1.0×10^{-5}
rs11700596	21	46,454,141	G	C	0.027	0.522	0.108	1.3×10^{-6}	1.2×10^{-1}

¹ Bonferroni-adjusted p -value for gene variant–sex interaction.

A statistically significant interaction with sex was observed in five genetic variants: two in females (rs146309770 and rs915401) and 3 in males (rs1274686, rs2186409, and rs2717199). The significance p -value threshold was established at a Bonferroni-adjusted p -value < 0.05.

2.3. Functional Annotation

Positional and expression quantitative trait locus (eQTL) gene-mapping of protein-coding genes was performed within the genomic regions defined by the five lead genetic variants that showed a statistically significant interaction with sex (Table 2). In females, rs915401 mapped to the gene Laminin Subunit Gamma 3 (*LAMC3*), and rs146309770 did not map to any protein-coding gene. In males, rs2186409 mapped to seven protein-coding genes: five by eQTL (Transmembrane Protein 109 (*TMEM109*), Pepsinogen A3 (*PGA3*), Pepsinogen A5 (*PGA5*), Von Willebrand Factor C and EGF Domains (*VWCE*), and Transmembrane Protein 216 (*TMEM216*)), and two by both position and eQTL (Vacuolar Protein Sorting-Associated Protein 37C (*VPS37C*) and *CD5*). rs1274686 mapped to seven protein-coding genes: five by position (Myosin Binding Protein C2 (*MYBPC2*), ER Membrane Protein Complex Subunit 10 (*EMC10*), Josephin Domain Containing 2 (*JOSD2*), Aspartate Dehydrogenase Domain Containing (*ASPDH*), and Leucine Rich Repeat Containing 4B (*LRRC4B*)), one by eQTL (Prostate Tumor Overexpressed 1 (*PTOV1*)), and one by both

position and eQTL (Family with sequence similarity 71 member E1 (*FAM71E1*)). rs2717199 was not mapped to any protein-coding gene.

Then, the gene set enrichment analysis was performed with the protein-mapped genes. In the case of females, it could not be assessed because only one gene was mapped and a minimum of two were needed for the analysis. In males, there was an overrepresentation of the target genes of the transcription factors protein inhibitor of activated STAT 4 (*FAM71E1*, *EMC10*, and *JOSD2*), zinc finger protein 524 (*VPS37C*, *PTOV1*, *FAM71E1*, and *EMC10*), and the gene set Nikolsky breast cancer 11q12-q14 amplicon (*TMEM109*, *CD5*, *VPS37C*, *PGA5*, *VWCE*, and *TMEM216*).

Finally, the gene-based analysis performed with MAGMA revealed two genes that were significantly associated with long COVID in males (Figure S2): *CD5* (p -value 2.16×10^{-7}) and *VPS37C* (p -value 2.97×10^{-7}). The gene-based analysis in females did not find any gene with statistical significance.

3. Discussion

In this study, by incorporating an open-ended long COVID symptoms questionnaire, we obtained highly accurate phenotypic data and performed sex-stratified analyses. Although no genetic variants reached genome-wide significant level, suggestive signals (p -value $< 5 \times 10^{-6}$) were found in the three GWAS, supporting the idea that genetic susceptibility to long COVID may be sex-specific.

Our gene-mapping analyses identified 15 genes associated with these suggestive variants (1 in females: *LAMC3*; and 14 in males). Of particular interest, *CD5* and *VPS37C*, were not only identified through positional and eQTL mapping, but also reached statistical significance in the MAGMA gene-based analysis. These genes are linked to immune system function. *CD5* encodes the CD5 protein, expressed in T cells, B1a cells, chronic lymphocytic leukemia cells, and dendritic cells [13]. Some studies have reported the development of chronic lymphocytic leukemia following SARS-CoV-2 infection [14–16]. A possible explanation is that mutant chronic lymphocytic leukemia cells may exist as small niches in healthy individuals and expand following SARS-CoV-2-induced cytokine dysregulation [15]. Similarly, *VPS37C* is included in the transcriptomic signature that differentiates multisystem inflammatory syndrome in children (occurring weeks after SARS-CoV-2 infection) from Kawasaki disease and other infections [17]. The involvement of *CD5* and *VPS37C* suggests that long COVID may be influenced by sustained immune dysregulation after the initial infection. Other mapped genes also support an immune-related mechanism. For instance, *FAM71E1* and *EMC10* have been associated with lupus erythematosus [18] and genes like *PGA3*, *VWCE*, and *PTOV1* are associated with immune infiltrates and immunity in various tumors [19–21]. Taken together, these findings suggest that persistent immune system alterations, possibly triggered by SARS-CoV-2, may play a critical role in the development of long COVID. This aligns with previous studies showing that interleukin-6 has been proposed as a predictive biomarker of long COVID risk [22]. Some of the identified genes may be involved in neurological aspects of long COVID. Variants in *EMC10* have been related to neurodevelopmental disorders and intellectual disability [23], and alterations in *LAMC3* have been linked to cortical development anomalies and epilepsy [24]. An association between central nervous system-related genes and long COVID may be compatible due to the frequent neurological symptoms seen in the disease [25], and neurocentric proteomic and microglial studies further support a link between viral persistence and long-term cognitive sequelae [26,27].

In addition, several genes mapped in this study are related to tumor progression and cancer prognosis. *JOSD2* influences the proliferation and progression of hepatocarcinoma, lung cancer, and esophageal squamous cell carcinoma [28–30] and modulates acute myeloid

leukemia progression [31]. *PGA3* is highly expressed in bone metastases of gastric cancer, indicating poor survival [32], and *PGA5* expression can induce changes related to tumor progression and epithelial–mesenchymal transition [33]. *VWCE* belongs to a 3-gene model proposed as a survival signature of uterine cancers [34]. *MYBPC2*, *TMEM109*, and *LACM3* have also been associated with various tumor types [35–37]. While the link between long COVID and cancer risk or progression remains speculative, these findings highlight pathways worth exploring further.

Additionally, in contrast to the recent meta-analysis by Lammi et al. [12], which reported a genome-wide significant association at the *FOXP4* locus, we did not observe this signal in our study. A key methodological difference is that Lammi et al. compared long COVID cases with controls drawn from the general population, which included both SARS-CoV-2-infected and non-infected individuals. Interestingly, when they restricted the analysis to controls with confirmed SARS-CoV-2 infection, the *FOXP4* association did not reach statistical significance. Our study design, which compared long COVID cases exclusively against SARS-CoV-2-infected non-cases, follows this latter approach and may explain the absence of association at the *FOXP4* locus in our results. This highlights how the choice of control group can influence the detection of genetic associations in long COVID research.

The main strengths of our study are, first, the use of an open-ended, clinician-administered interview that lets participants report symptoms freely, yielding a precise, checklist-free definition of long COVID; and second, our sex-stratified GWAS and gene variant-by-sex interaction analyses, which uncover genetic signals that differ between female and male. However, our study also has limitations. Firstly, long COVID characterization is recent and still evolving, encompassing a broad range of symptoms. This heterogeneity complicates the identification of specific genetic associations, as variability in symptom presentation reduces the strength of potential signals. Secondly, our limited sample size may reduce the statistical power to detect genetic signals involved in long COVID. Third, we were unable to conduct a meta-analysis with other available datasets due to differences in phenotype definitions. Fourth, differences in linkage disequilibrium patterns, allele frequencies, and individuals with and without long COVID distributions across ancestries posed additional challenges. Nevertheless, these differences were addressed by including the first 10 principal components as covariates in the association analysis. Fifth, SARS-CoV-2 reinfections were not recorded during follow-up. Reinfection events may exacerbate or re-trigger symptoms, potentially mimicking or amplifying the long COVID phenotype. Future studies should incorporate reinfection surveillance to better disentangle their role in the persistence of symptoms. Finally, our findings may not fully generalize to other populations. All participants were recruited in Catalonia, northeastern Spain, and were 35 to 84 years old, within a universal healthcare system. These factors differ from other settings and could influence how long COVID develops and persists.

4. Materials and Methods

4.1. Study Design and Participants

The GINA-COVID project is a prospective cohort study that comprised 3073 COVID-19 patients aged 35 to 84 years old. They were recruited from four centers in Catalonia, northeastern Spain [38]. All the participants had tested positive for SARS-CoV-2 using reverse transcription-polymerase chain reaction, rapid antigen, or IgG tests between February 2020 and December 2021. Participants were excluded if they had been vaccinated against SARS-CoV-2 before the diagnosis, if clinical data were not available in the electronic medical record of the Catalan health system (universal coverage), or if they did not participate in the 1-year follow-up for long COVID symptoms.

4.2. Definition of Long COVID

We determined long COVID status in accordance with World Health Organization criteria [39]. Trained personnel conducted an open-ended questionnaire on participants' COVID-19 symptoms one year after diagnosis or one year after discharge from hospital for hospitalized patients. Participants were asked to describe spontaneously any symptoms they had experienced weeks or months after their infection. Trained personnel recorded the presence of symptoms with a predefined list of 24 symptoms previously reported in the bibliography for long COVID (>10% prevalence) [3,7,8,40]. Symptoms reported by participants that were not present in the predefined list were also recorded. For each symptom, participants were asked whether it was present before, during, or after their infection (less than 3 months, up to 3, 6, or 9 months, or up to 12 months or longer). A symptom was classified as persistent if it was absent before the infection but emerged during or after the infection and lasted for at least three months. Symptoms present prior to the infection but worsened during or after the infection for at least three months were also classified as persistent. Following the World Health Organization definition [39], long COVID included individuals who suffered from at least 1 persistent symptom from the 24 listed symptoms.

4.3. DNA Collection

DNA was obtained from both peripheral blood and saliva. Peripheral blood was obtained and placed in 4 mL EDTA Anti-Coagulant BD Vacutainer tubes. Saliva samples were collected using the DANASALIVA[®] Sample Collection kit (DANAGEN[®], Badalona, Spain). DNA was extracted with either a ChemagicTM[®] DNA Blood 7k Kit H12 (PerkinElmer[®], Barcelona, Spain) on a Chemagic MSM I instrument, or a FlexiGene[®] DNA Kit (Qiagen[®], Barcelona, Spain) [38].

4.4. Genotyping

DNA samples were genotyped with the Axiom Spain Biobank 1 Array (Thermo Fisher Scientific, Waltham, MA, USA), which contains 757,836 genetic variants, including rare and specific variants from the Spanish population. The genotyping process was performed at the Santiago de Compostela Node of the National Genotyping Center (CeGen-ISCI; <https://www.xenomica.eu/servicios/centro-nacional-de-genotipado/>, accessed on 14 January 2025) following the manufacturer's instructions. Genotyping calling and clustering were performed with Axiom Analysis Suite v5.3.0.45.

4.5. Quality Control of Genotype Data

Post-genotyping quality control was performed with R.4.3.2 [41] and PLINK v1.9 and v2.0 [42]. Variants with minor allele frequency < 1%, call rate < 98%, or those deviating from Hardy–Weinberg equilibrium (p -value < 10^{-6}) were excluded. For samples, those with call rate < 98% and a heterozygosity rate deviating by more than five standard deviations from the mean heterozygosity rate of the study were removed. Kinship and ancestry were assessed by pruning autosomal variants with minor allele frequency > 5% (window size of 1000 markers, step size of 80, and r^2 of 0.1), obtaining a subset of 72,546 genetic variants. High linkage disequilibrium regions previously described by Price et al. [43] were also removed. We performed identity-by-descent analysis with this set of variants and removed one individual, the one with the highest missingness rate, from each pair with first- and second-degree of kinship (PI_HAT > 0.25). Principal component analysis was conducted using the genetic variants from unrelated individuals to study population stratification and identify outliers for exclusion. Ancestry was analyzed by Admixture [44] on unrelated individuals using the 1000 Genomes Project data [45]. Participants were assigned to a

specific ancestry group if their probability of belonging to that group was $\geq 80\%$; otherwise, they were categorized as “mixed”. Overall, 2411 individuals (1392 with long COVID and 1019 without long COVID) and 584,906 genetic variants passed quality control.

4.6. Variant Imputation

Genetic variants were imputed using the TOPMed version r3 reference panel (GRCh38) through the TOPMed Imputation Server [46]. Post-imputation filtering was applied, retaining variants with an INFO score > 0.8 and a minor allele frequency $> 1\%$, resulting in 8,681,812 variants for further analysis.

4.7. Clinical Data

During admission or at the outpatient clinic, age, sex, weight, height, self-reported history of tobacco use (whether the participant was a current smoker or not), and history of diabetes, hypertension, and dyslipidemia were collected from electronic medical records. Body mass index was calculated by dividing an individual's weight (in kilograms) by the square of their height (in meters) with the following categories: underweight ($< 18.5 \text{ kg/m}^2$), normal weight (18.5 to $< 25.0 \text{ kg/m}^2$), overweight (25.0 to $< 30.0 \text{ kg/m}^2$), and obesity ($\geq 30.0 \text{ kg/m}^2$).

4.8. Statistical Analyses

Patient characteristics were presented as frequencies for categorical variables, while continuous variables were summarized as means with standard deviations for normally distributed data. Comparisons between people with and without long COVID were performed using *t*-tests for means and the chi-square test for categorical variables.

Global and sex-stratified genome-wide association tests were computed by fitting a logistic regression model in Plink 1.9, including age and sex in the global analysis, and the first 10 principal components as covariates. A sex-stratified sensitivity analysis was also performed in European population, including age and the first 10 European-specific principal components as covariates. The genetic variant association significance threshold was set at $p\text{-value} < 5 \times 10^{-8}$.

To formally evaluate effect modification by sex, we analyzed potential interactions between genetic variants and sex in relation to long COVID. To perform this analysis, we selected the lead variants from the global and the sex-stratified GWAS analyses. They were defined as a subset of significant independent variants ($p\text{-value} < 5 \times 10^{-6}$ and $r^2 < 0.6$) that were independent at each other at $r^2 < 0.1$. Once identified, a logistic regression model was fitted for each of them, in which the age, sex, the first 10 principal components, and the interaction term genetic variant*sex were included as covariates. Bonferroni correction of the *p*-values was performed for multiple testing, and the association significance threshold was established at Bonferroni-adjusted $p\text{-value} < 0.05$.

4.9. Functional Annotation

The functional annotation, gene mapping, and functional gene-set enrichment analyses were performed with the lead variants showing a significant interaction with sex, in the global or sex-stratified analyses, using SNP2GENE implemented in FUMA (v.1.5.2) [47].

Genomic risk loci included genetic variants ($p\text{-value} < 0.05$) in linkage disequilibrium ($r^2 < 0.6$) with the independent significant variants, referred to as candidate variants. The maximum distance between linkage disequilibrium blocks to be merged into a genomic locus was 250 kb. The 1000 Genomes Project, Phase 3, from all ancestries [45] was set as the reference panel population, as our sample was multi-ethnic.

Functional consequences of candidate variants on gene function were annotated with Annotate Variation [48], including the combined annotation dependent depletion score for

variant deleteriousness [49], Regulome database for potential regulatory functions [50], and gene expression effects were assessed using eQTL of genotype-tissue expression v.8 [51].

To further investigate the biological function of the candidate variants, they were mapped to protein-coding genes and prioritized by two procedures: positional mapping, in which variants were assigned to genes by physical distance (10 kb window), and eQTL mapping, in which genetic variants were mapped to a gene if they had significant effects on its expression, using the genotype-tissue expression v.8 database [51]. Significant variant–gene pairs were established at a false discovery rate < 0.05. Candidate genes were considered for the functional enrichment analysis.

An enrichment analysis of the candidate genes in pre-defined pathways was performed with the GENE2FUNC tool implemented in FUMA. The hypergeometric test was used to test for the overrepresentation of the candidate genes in any of the gene sets, which included Molecular Signatures Database [52,53], WikiPathways [54], and GWAS Catalog [55] information. For a gene set to be overrepresented in the candidate genes, a minimum of two overlapping genes and a Benjamini–Hochberg adjusted *p*-value < 0.05 were needed.

To test for the joint effect of genetic variants mapped to a protein-coding gene, sex-stratified gene-based analyses were performed with MAGMA [56], implemented in FUMA, under the gene variant-wide model with the 1000 Genomes Project, Phase 3 [45], from all the ancestries as the reference panel. For both females and males, genetic variants were mapped to a total of 19,394 protein-coding genes, and the gene-level significance *p*-value threshold was calculated using the Bonferroni method ($< 0.05/19,394 = 2.578 \times 10^{-6}$).

5. Conclusions

Our checklist-free, open-ended phenotyping and sex-stratified GWAS reveal immune-regulatory loci (notably CD5 and VPS37C in men) that point to persistent immune dysregulation as a driver of long COVID. By improving case definition and exposing sex-specific risk, our study helps public health efforts to identify individuals most likely to develop long-term sequelae of SARS-CoV-2 infection. Future work should replicate these signals in larger, harmonized cohorts (ideally applying similarly checklist-free phenotyping and mandatory sex stratification), ultimately guiding research for developing personalized strategies to prevent persistent immune dysregulation after COVID-19 infection.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms26189252/s1>.

Author Contributions: Conceptualization, Á.H., I.R.D., R.M.-L., M.P.-A., R.E., I.S., R.B., R.R., A.C.-V. and J.M.; Methodology, Á.H., I.R.D., R.M.-L., M.P.-A., R.E., I.S., R.C., R.B., R.R., A.C.-V. and J.M.; Software, S.P.-A. and I.S.; Validation, S.P.-A., Á.H. and I.R.D.; Formal Analysis, S.P.-A.; Investigation, S.P.-A., Á.H., I.R.D., R.M.-L., M.P.-A., R.E., I.S., M.P., A.P., R.C., S.D.-d.A., E.P., E.S., X.N., J.R.M., R.G.-F., J.M., H.C.-G., H.T.-M., B.V., R.B., R.R., A.C.-V. and J.M.; Resources, I.R.D., R.M.-L., R.B., R.R. and J.M.; Data Curation, S.P.-A., R.M.-L., M.P.-A., I.S. and A.C.-V.; Writing—Original Draft Preparation, S.P.-A.; Writing—Review and Editing, Á.H., I.R.D., R.M.-L., M.P.-A., R.E., I.S., M.P., A.P., R.C., S.D.-d.A., E.P., E.S., X.N., J.R.M., R.G.-F., J.M., H.C.-G., H.T.-M., B.V., R.B., R.R., A.C.-V. and J.M.; Visualization, S.P.-A.; Supervision, Á.H., I.R.D., R.M.-L., R.B., R.R., A.C.-V. and J.M.; Project Administration, J.M.; Funding Acquisition, I.R.D., R.M.-L., R.B., R.R. and J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by CIBER—Consortio Centro de Investigación Biomédica en Red—(CIBERCV CB16/11/00229 [J.M.] and CB16/11/00246 [R.E.]), Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación and Unión Europea—European Regional Development Fund; by a grant from the Government of Catalonia through the Agency for Management of University

and Research Grants (2021SGR144 [J.M.]); by the Crue-CSIC-Santander FONDO SUPERA COVID-19 [J.M., R.R., R.B., I.R.D.]; and by Fundació La Marató de TV3 (202119-30) [J.M., R.M., R.B., I.R.D.].

Institutional Review Board Statement: The study protocol of the GINA-COVID study complied with the Declaration of Helsinki for Medical Research Involving Human Subjects. Its protocol was approved by the Ethics Committee of Parc de Salut Mar (#2020/9297/I and #2020/9650/I), Doctor Josep Trueta University Hospital of Girona (#2020/058), Foundation University Institute for Primary Healthcare Research Jordi Gol i Gurina (IDIAPJGol) (#2021/084-PCV), and Fundació d'Osona per a la Recerca i l'Educació Sanitàries (FORES) (#2020152/PR283). The study followed all international regulations for biomedical research and data protection.

Informed Consent Statement: All participants were informed and gave their written consent to participate in the study.

Data Availability Statement: The data that support the findings of this study are available from the corresponding authors upon request.

Acknowledgments: We would like to thank Xavier Farré for his analytical contributions, and all the GINA-COVID participants and collaborators who participated in this study. A full roster of contributors can be found at: https://regicor.cat/cargencors_inv/ (accessed on 15 January 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

COVID-19	Coronavirus Disease 2019
eQTL	Expression Quantitative Trait Locus
GWAS	Genome-Wide Association Study
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2

References

1. Tsilingiris, D.; Vallianou, N.G.; Karampela, I.; Christodoulatos, G.S.; Papavasileiou, G.; Petropoulou, D.; Magkos, F.; Dalamaga, M. Laboratory Findings and Biomarkers in Long COVID: What Do We Know So Far? Insights into Epidemiology, Pathogenesis, Therapeutic Perspectives and Challenges. *Int. J. Mol. Sci.* **2023**, *24*, 10458. [\[CrossRef\]](#)
2. Greenhalgh, T.; Sivan, M.; Perlowski, A.; Nikolich, J.Ž. Long COVID: A Clinical Update. *Lancet* **2024**, *404*, 707–724. [\[CrossRef\]](#)
3. Robertson, M.M.; Qasmieh, S.A.; Kulkarni, S.G.; Teasdale, C.A.; Jones, H.E.; McNairy, M.; Borrell, L.N.; Nash, D. The Epidemiology of Long Coronavirus Disease in US Adults. *Clin. Infect. Dis.* **2023**, *76*, 1636–1645. [\[CrossRef\]](#)
4. Kim, Y.; Bae, S.; Chang, H.H.; Kim, S.W. Long COVID Prevalence and Impact on Quality of Life 2 Years after Acute COVID-19. *Sci. Rep.* **2023**, *13*, 11207; Erratum in *Sci. Rep.* **2023**, *13*, 11960.
5. O'Mahoney, L.L.; Routen, A.; Gillies, C.; Jenkins, S.A.; Almaqhawi, A.; Ayoubkhani, D.; Banerjee, A.; Brightling, C.; Calvert, M.; Cassambai, S.; et al. The Risk of Long Covid Symptoms: A Systematic Review and Meta-Analysis of Controlled Studies. *Nat. Commun.* **2025**, *16*, 4249. [\[CrossRef\]](#)
6. Constantinescu-Bercu, A.; Lobiuc, A.; Căliman-Sturdza, O.A.; Oîță, R.C.; Iavorschi, M.; Pavăl, N.-E.; Șoldănescu, I.; Dimian, M.; Covasa, M. Long COVID: Molecular Mechanisms and Detection Techniques. *Int. J. Mol. Sci.* **2023**, *25*, 408. [\[CrossRef\]](#)
7. Mateu, L.; Tebe, C.; Loste, C.; Santos, J.R.; Lladós, G.; López, C.; España-Cueto, S.; Toledo, R.; Font, M.; Chamorro, A.; et al. Determinants of the Onset and Prognosis of the Post-COVID-19 Condition: A 2-Year Prospective Observational Cohort Study. *Lancet Reg. Health Eur.* **2023**, *33*, 100724. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Dennis, A.; Wamil, M.; Alberts, J.; Oben, J.; Cuthbertson, D.J.; Wootton, D.; Crooks, M.; Gabbay, M.; Brady, M.; Hishmeh, L.; et al. Multiorgan Impairment in Low-Risk Individuals with Post-COVID-19 Syndrome: A Prospective, Community-Based Study. *BMJ Open* **2021**, *11*, e048391. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Hein, Z.M.; Thazin, Kumar, S.; Che Ramli, M.D.; Che Mohd Nassir, C.M.N. Immunomodulatory Mechanisms Underlying Neurological Manifestations in Long COVID: Implications for Immune-Mediated Neurodegeneration. *Int. J. Mol. Sci.* **2025**, *26*, 6214. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Gusev, E.; Sarapultsev, A. Exploring the Pathophysiology of Long COVID: The Central Role of Low-Grade Inflammation and Multisystem Involvement. *Int. J. Mol. Sci.* **2024**, *25*, 6389. [\[CrossRef\]](#)

11. Varillas-Delgado, D.; Jimenez-Antona, C.; Lizcano-Alvarez, A.; Cano-de-la-Cuerda, R.; Molero-Sanchez, A.; Laguarda-Val, S. Predictive Factors and ACE-2 Gene Polymorphisms in Susceptibility to Long COVID-19 Syndrome. *Int. J. Mol. Sci.* **2023**, *24*, 16717. [\[CrossRef\]](#)
12. Lammi, V.; Nakanishi, T.; Jones, S.E.; Andrews, S.J.; Karjalainen, J.; Cortés, B.; O'Brien, H.E.; Ochoa-Guzman, A.; Fulton-Howard, B.E.; Broberg, M.; et al. Genome-Wide Association Study of Long COVID. *Nat. Genet.* **2025**, *57*, 1402–1417. [\[CrossRef\]](#)
13. Soldevila, G.; Raman, C.; Lozano, F. The Immunomodulatory Properties of the CD5 Lymphocyte Receptor in Health and Disease. *Curr. Opin. Immunol.* **2011**, *23*, 310–318. [\[CrossRef\]](#)
14. Ali, E.; Badawi, M.; Abdelmahuod, E.; Kohla, S.; Yassin, M.A. Chronic Lymphocytic Leukemia Concomitant with COVID 19: A Case Report. *Am. J. Case Rep.* **2020**, *21*, e926062. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Saluja, P.; Gautam, N.; Amisha, F.; Safar, M.; Bartter, T. Emergence of Chronic Lymphocytic Leukemia During Admission for COVID-19: Cause or Coincidence? *Cureus* **2022**, *14*, e23470. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Lanza, L.; Korovesi, B.; Barducci, F.; Lorenzo, A.; Venturino, E.; Cappelli, E.; Lillo, F.; Bain, B.J. A New Diagnosis of Monoclonal B-cell Lymphocytosis with Cytoplasmic Inclusions in a Patient with COVID-19. *Am. J. Hematol.* **2022**, *97*, 1372–1373. [\[CrossRef\]](#)
17. Jackson, H.R.; Miglietta, L.; Habgood-Coote, D.; D'Souza, G.; Shah, P.; Nichols, S.; Vito, O.; Powell, O.; Davidson, M.S.; Shimizu, C.; et al. Diagnosis of Multisystem Inflammatory Syndrome in Children by a Whole-Blood Transcriptional Signature. *J. Pediatr. Infect. Dis. Soc.* **2023**, *12*, 322–331. [\[CrossRef\]](#)
18. Delgado-Vega, A.M.; Martínez-Bueno, M.; Oparina, N.Y.; López Herráez, D.; Kristjansdottir, H.; Steinsson, K.; Kozyrev, S.V.; Alarcón-Riquelme, M.E. Whole Exome Sequencing of Patients from Multicase Families with Systemic Lupus Erythematosus Identifies Multiple Rare Variants. *Sci. Rep.* **2018**, *8*, 8775. [\[CrossRef\]](#)
19. Shen, S.; Li, H.; Liu, J.; Sun, L.; Yuan, Y. The Panoramic Picture of Pepsinogen Gene Family with Pan-Cancer. *Cancer Med.* **2020**, *9*, 9064–9080. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Huo, Q.; Li, Z.; Chen, S.; Wang, J.; Li, J.; Xie, N. VWCE as a Potential Biomarker Associated with Immune Infiltrates in Breast Cancer. *Cancer Cell Int.* **2021**, *21*, 272. [\[CrossRef\]](#)
21. Xie, S.A.; Zhang, W.; Du, F.; Liu, S.; Ning, T.T.; Zhang, N.; Zhang, S.-T.; Zhu, S.-T. PTOV1 Facilitates Colorectal Cancer Cell Proliferation through Activating AKT1 Signaling Pathway. *Heliyon* **2024**, *10*, e36017. [\[CrossRef\]](#)
22. Giannitrapani, L.; Mirarchi, L.; Amodeo, S.; Licata, A.; Soresi, M.; Cavaleri, F.; Casalicchio, S.; Ciulla, G.; Ciuppa, M.E.; Cervello, M.; et al. Can Baseline IL-6 Levels Predict Long COVID in Subjects Hospitalized for SARS-CoV-2 Disease? *Int. J. Mol. Sci.* **2023**, *24*, 1731. [\[CrossRef\]](#)
23. Kaiyrzhanov, R.; Rocca, C.; Suri, M.; Gulieva, S.; Zaki, M.S.; Henig, N.Z.; Siquier, K.; Guliyeva, U.; Mounir, S.M.; Marom, D.; et al. Biallelic Loss of EMC10 Leads to Mild to Severe Intellectual Disability. *Ann. Clin. Transl. Neurol.* **2022**, *9*, 1080–1089. [\[CrossRef\]](#)
24. Saarela, A.; Timonen, O.; Kirjavainen, J.; Liu, Y.; Silvennoinen, K.; Mervaala, E.; Kälviäinen, R. Novel LAMC3 Pathogenic Variant Enriched in Finnish Population Causes Malformations of Cortical Development and Severe Epilepsy. *Epileptic Disord.* **2024**, *26*, 498–509. [\[CrossRef\]](#)
25. Premraj, L.; Kannapadi, N.V.; Briggs, J.; Seal, S.M.; Battaglini, D.; Fanning, J.; Suen, J.; Robba, C.; Fraser, J.; Cho, S.-M. Mid and Long-Term Neurological and Neuropsychiatric Manifestations of Post-COVID-19 Syndrome: A Meta-Analysis. *J. Neurol. Sci.* **2022**, *434*, 120162. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Pulliam, L.; Sun, B.; McCafferty, E.; Soper, S.A.; Witek, M.A.; Hu, M.; Ford, J.M.; Song, S.; Kapogiannis, D.; Glesby, M.J.; et al. Microfluidic Isolation of Neuronal-Enriched Extracellular Vesicles Shows Distinct and Common Neurological Proteins in Long COVID, HIV Infection and Alzheimer's Disease. *Int. J. Mol. Sci.* **2024**, *25*, 3830. [\[CrossRef\]](#)
27. Chagas, L.D.S.; Serfaty, C.A. The Influence of Microglia on Neuroplasticity and Long-Term Cognitive Sequelae in Long COVID: Impacts on Brain Development and Beyond. *Int. J. Mol. Sci.* **2024**, *25*, 3819. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Yuan, T.; Zeng, C.; Liu, J.; Zhao, C.; Ge, F.; Li, Y.; Qian, M.; Du, J.; Wang, W.; Li, Y.; et al. Josephin Domain Containing 2 (JOSD2) Promotes Lung Cancer by Inhibiting LKB1 (Liver Kinase B1) Activity. *Signal Transduct. Target. Ther.* **2024**, *9*, 11. [\[CrossRef\]](#)
29. Wang, W.P.; Shi, D.; Yun, D.; Hu, J.; Wang, J.F.; Liu, J.; Yang, Y.-P.; Li, M.-R.; Wang, J.-F.; Kong, D.-L. Role of Deubiquitinase JOSD2 in the Pathogenesis of Esophageal Squamous Cell Carcinoma. *World J. Gastroenterol.* **2024**, *30*, 565–578. [\[CrossRef\]](#)
30. Huang, Y.; Zeng, J.; Liu, T.; Xu, Q.; Song, X.; Zeng, J. Deubiquitinating Enzyme JOSD2 Promotes Hepatocellular Carcinoma Progression through Interacting with and Inhibiting CTNNB1 Degradation. *Cell Biol. Int.* **2022**, *46*, 1089–1097. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Lei, H.; Yang, L.; Wang, Y.; Zou, Z.; Liu, M.; Xu, H.; Wu, Y. JOSD2 Regulates PKM2 Nuclear Translocation and Reduces Acute Myeloid Leukemia Progression. *Exp. Hematol. Oncol.* **2022**, *11*, 42. [\[CrossRef\]](#)
32. Oh, S.; Nam, S.K.; Lee, K.W.; Lee, H.S.; Park, Y.; Kwak, Y.; Lee, K.S.; Kim, J.-W.; Kim, J.W.; Kang, M.; et al. Genomic and Transcriptomic Characterization of Gastric Cancer with Bone Metastasis. *Cancer Res. Treat.* **2024**, *56*, 219–237. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Stabenau, K.A.; Samuels, T.L.; Lam, T.K.; Mathison, A.J.; Wells, C.; Altman, K.W.; Battle, M.A.; Johnston, N. Pepsinogen/Proton Pump Co-Expression in Barrett's Esophageal Cells Induces Cancer-Associated Changes. *Laryngoscope* **2023**, *133*, 59–69. [\[CrossRef\]](#) [\[PubMed\]](#)

34. Almorox, L.; Antequera, L.; Rojas, I.; Herrera, L.J.; Ortuño, F.M. Gene Expression Analysis for Uterine Cervix and Corpus Cancer Characterization. *Genes* **2024**, *15*, 312. [\[CrossRef\]](#)
35. Kunitomi, H.; Kobayashi, Y.; Wu, R.C.; Takeda, T.; Tominaga, E.; Banno, K.; Aoki, D. LAMC1 Is a Prognostic Factor and a Potential Therapeutic Target in Endometrial Cancer. *J. Gynecol. Oncol.* **2020**, *31*, e11. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Cai, J.; Zhang, X.; Xie, W.; Li, Z.; Liu, W.; Liu, A. Identification of a Basement Membrane-Related Gene Signature for Predicting Prognosis and Estimating the Tumor Immune Microenvironment in Breast Cancer. *Front. Endocrinol.* **2022**, *13*, 1065530. [\[CrossRef\]](#)
37. Takenami, T.; Maeda, S.; Karasawa, H.; Suzuki, T.; Furukawa, T.; Morikawa, T.; Takadate, T.; Hayashi, H.; Nakagawa, K.; Motoi, F.; et al. Novel Biomarkers Distinguishing Pancreatic Head Cancer from Distal Cholangiocarcinoma Based on Proteomic Analysis. *BMC Cancer* **2019**, *19*, 318. [\[CrossRef\]](#)
38. Camps-Vilaró, A.; Pinsach-Abuin, M.L.; Degano, I.R.; Ramos, R.; Martí-Lluch, R.; Elosua, R.; Subirana, I.; Solà-Richarte, C.; Puigmulé, M.; Pérez, A.; et al. Genetic Characteristics Involved in COVID-19 Severity. The CARGENCORS Case-Control Study and Meta-Analysis. *J. Med. Virol.* **2024**, *96*, e29404. [\[CrossRef\]](#)
39. World Health Organization Post COVID-19 Condition (Long COVID). Available online: <https://www.who.int/europe/news-room/fact-sheets/item/post-covid-19-condition> (accessed on 18 December 2024).
40. Ceban, F.; Ling, S.; Lui, L.M.W.; Lee, Y.; Gill, H.; Teopiz, K.M.; Rodrigues, N.B.; Subramaniapillai, M.; Di Vincenzo, J.D.; Cao, B.; et al. Fatigue and Cognitive Impairment in Post-COVID-19 Syndrome: A Systematic Review and Meta-Analysis. *Brain Behav. Immun.* **2022**, *101*, 93–135. [\[CrossRef\]](#)
41. R Core Team. *R Core Team R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023.
42. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [\[CrossRef\]](#)
43. Price, A.L.; Weale, M.E.; Patterson, N.; Myers, S.R.; Need, A.C.; Shianna, K.V.; Ge, D.; Rotter, J.I.; Torres, E.; Taylor, K.D.; et al. Long-Range LD Can Confound Genome Scans in Admixed Populations. *Am. J. Hum. Genet.* **2008**, *83*, 132–138. [\[CrossRef\]](#)
44. Alexander, D.H.; Novembre, J.; Lange, K. Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Res.* **2009**, *19*, 1655–1664. [\[CrossRef\]](#)
45. 1000 Genomes Project Consortium. A Global Reference for Human Genetic Variation. *Nature* **2015**, *526*, 68–74. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Taliun, D.; Harris, D.N.; Kessler, M.D.; Carlson, J.; Szpiech, Z.A.; Torres, R.; Taliun, S.A.G.; Corvelo, A.; Gogarten, S.M.; Kang, H.M.; et al. Sequencing of 53,831 Diverse Genomes from the NHLBI TOPMed Program. *Nature* **2021**, *590*, 290–299. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Watanabe, K.; Taskesen, E.; van Bochoven, A.; Posthuma, D. Functional Mapping and Annotation of Genetic Associations with FUMA. *Nat. Commun.* **2017**, *8*, 1826. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* **2010**, *38*, e164. [\[CrossRef\]](#)
49. Kircher, M.; Witten, D.M.; Jain, P.; O’Roak, B.J.; Cooper, G.M.; Shendure, J. A General Framework for Estimating the Relative Pathogenicity of Human Genetic Variants. *Nat. Genet.* **2014**, *46*, 310–315. [\[CrossRef\]](#)
50. Boyle, A.P.; Hong, E.L.; Hariharan, M.; Cheng, Y.; Schaub, M.A.; Kasowski, M.; Karczewski, K.J.; Park, J.; Hitz, B.C.; Weng, S.; et al. Annotation of Functional Variation in Personal Genomes Using RegulomeDB. *Genome Res.* **2012**, *22*, 1790–1797. [\[CrossRef\]](#)
51. GTEx Consortium. The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues. *Science* **2020**, *369*, 1318–1330. [\[CrossRef\]](#)
52. Castanza, A.S.; Recla, J.M.; Eby, D.; Thorvaldsdóttir, H.; Bult, C.J.; Mesirov, J.P. Extending Support for Mouse Data in the Molecular Signatures Database (MSigDB). *Nat. Methods* **2023**, *20*, 1619–1620. [\[CrossRef\]](#)
53. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [\[CrossRef\]](#)
54. Agrawal, A.; Balci, H.; Hanspers, K.; Coort, S.L.; Martens, M.; Slenter, D.N.; Ehrhart, F.; Digles, D.; Waagmeester, A.; Wassink, I.; et al. WikiPathways 2024: Next Generation Pathway Database. *Nucleic Acids Res.* **2024**, *52*, D679–D689. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Sollis, E.; Mosaku, A.; Abid, A.; Buniello, A.; Cerezo, M.; Gil, L.; Groza, T.; Güneş, O.; Hall, P.; Hayhurst, J.; et al. The NHGRI-EBI GWAS Catalog: Knowledgebase and Deposition Resource. *Nucleic Acids Res.* **2023**, *51*, D977–D985. [\[CrossRef\]](#) [\[PubMed\]](#)
56. de Leeuw, C.A.; Mooij, J.M.; Heskes, T.; Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **2015**, *11*, e1004219. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.