scientific reports

OPEN

Presence of rare potential pathogenic variants in subjects under 65 years old with very severe or fatal COVID-19

Rosario López-Rodríguez[1,2,14], Marta Del Pozo-Valero[1,2], Marta Corton[1,2], Pablo Minguez [1,2,3], Javier Ruiz-Hornillos[4,5,6], María Elena Pérez-Tomás[7], María Barreda-Sánchez[7,8], Esther Mancebo[9,10], Cristina Villaverde[1,2], Gonzalo Núñez-Moreno[1,3], Raquel Romero[1,2], The STOP_Coronavirus Study Group, Estela Paz-Artal[9,10,11,12], Encarna Guillén-Navarro[2,7,13], Berta Almoguera[1,2] & Carmen Ayuso[1,2✉]

Rare variants affecting host defense against pathogens could be involved in COVID-19 severity and may help explain fatal outcomes in young and middle-aged patients. Our aim was to report the presence of rare genetic variants in certain genes, by using whole exome sequencing, in a selected group of COVID-19 patients under 65 years who required intubation or resulting in death (n = 44). To this end, different etiopathogenic mechanisms were explored using gene prioritization-based analysis in which genes involved in immune response, immunodeficiencies or blood coagulation were studied. We detected 44 different variants of interest, in 29 different patients (66%). Some of these variants were previously described as pathogenic and were located in genes mainly involved in immune response. A network analysis, including the 42 genes with candidate variants, showed three main components, consisting of 25 highly interconnected genes related to immune response and two additional networks composed by genes enriched in carbohydrate metabolism and in DNA metabolism and repair processes. In conclusion, we have detected candidate variants that may potentially influence COVID-19 outcome in our cohort of patients. Further studies are needed to confirm the ultimate role of the genetic variants described in the present study on COVID-19 severity.

Since the first outbreak of the Coronavirus disease in the 2019 (COVID-19) pandemic, over 243 million cases of COVID-19 and more than 6.1 million deaths have been confirmed (https://coronavirus.jhu.edu, *last accessed in March 2022*). SARS-CoV-2 infection displays high inter-individual clinical variability, ranging from asymptomatic to lethal outcomes[1]. The most important life-threatening factor is age, increasing the risk for critical

[1]Department of Genetics & Genomics, Instituto de Investigación Sanitaria-Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid (IIS-FJD, UAM), Madrid, Spain. [2]Center for Biomedical Network Research on Rare Diseases (CIBERER), Instituto de Salud Carlos III, 28029 Madrid, Spain. [3]Bioinformatics Unit, Instituto de Investigación Sanitaria-Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid (IIS-FJD, UAM), Madrid, Spain. [4]Allergy Unit, Hospital Infanta Elena, Valdemoro, Madrid, Spain. [5]Instituto de Investigación Sanitaria-Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid (IIS-FJD, UAM), Madrid, Spain. [6]Faculty of Medicine, Universidad Francisco de Vitoria, Pozuelo de Alarcón, Madrid, Spain. [7]Instituto Murciano de Investigación Biosanitaria Virgen de la Arrixaca (IMIB-Arrixaca), Murcia, Spain. [8]Health Sciences Faculty, Universidad Católica San Antonio de Murcia (UCAM), Murcia, Spain. [9]Department of Immunology, Hospital Universitario 12 de Octubre, Madrid, Spain. [10]Instituto de Investigación Sanitaria Hospital 12 de Octubre (imas12), Madrid, Spain. [11]Department of Immunology, Ophthalmology and ENT, Universidad Complutense de Madrid, Madrid, Spain. [12]Center for Biomedical Network Research on Infectious Diseases (CIBERINFEC), Instituto de Salud Carlos III, 28029 Madrid, Spain. [13]Medical Genetics Section, Pediatric Department, Virgen de la Arrixaca University Clinical Hospital, Faculty of Medicine, University of Murcia (UMU), Murcia, Spain. [14]Present address: Department of Pharmaceutical and Health Sciences, Faculty of Pharmacy, Universidad San Pablo-CEU, CEU Universities, Madrid, Spain. *A list of authors and their affiliations appears at the end of the paper. ✉email: cayuso@fjd.es

illness for individuals over 65 years of age[2]. Other known risk factors are being male and having comorbidities such as hypertension, diabetes and cardiovascular, renal or respiratory diseases[3,4]. However, these risk factors do not explain completely why apparently healthy young and middle-aged adults present severe COVID-19 with acute respiratory distress syndrome (ARDS) that cause a fulminant disease in some cases.

Genetic background has been proposed as a candidate factor to explain some of the inter-individual variability observed in COVID-19 severity. Recently, different genome-wide association studies (GWAS) have identified several loci associated with an increased susceptibility to SARS-CoV-2 infection and severe disease[5–7].These loci include genes involved in type I interferon (IFN) signaling pathway (*IFNAR2, DPP9 or OAS1-3*), autoimmunity (*TYK2*) or in lung function (*FOXP4*).However, top associated variants displayed low odd ratios to be considered predictive biomarkers of COVID-19 severity[5–7].

In addition to common variants detected in GWAS, rare variants affecting host defense against pathogens could be involved in COVID-19 severity and may help explain fatal outcomes in young and middle-aged patients. In fact, inborn errors of immunity producing increased infection susceptibility and/or infection recurrence (such as primary immunodeficiencies or N-glycosylation defects), may aggravate the course of SARS-CoV-2 infection[8–10]. Enrichment in loss-of-function (LoF) variants in 13 genes belonging to type I IFN signaling pathway has been reported in patients with life-threatening COVID-19 pneumonia[11], although this finding has not been replicated[12]. Moreover, LoF genetic variants in Toll-like receptor 7 (*TLR7*), which is critical in the recognition of single-stranded RNA viruses and fostering the antiviral responses, have been associated to more severe outcomes in young males without comorbidities[13–16].

Apart from ARDS, thrombosis and coagulopathy emerge as critical complications of SARS-CoV-2 infection[17–19]. In fact, elevation of the thrombotic related D-dimer is one of the most frequent laboratory findings, particularly in critically ill patients[20]. In addition, different studies have shown the influence of ABO blood groups on the risk of SARS-CoV-2 infection and/or the severity of the disease[21]. In this regard, the demonstrated impact of ABO phenotypes on vascular homeostasis and function[22], has been suggested as an explanation of the reported associations between COVID-19 severity and ABO blood groups. Thus, pathogenic genetic variants altering protein functionality of coagulation system may also impact on COVID-19 resolution.

Based on this evidence, our objective was to identify rare genetic variants related to COVID-19 severity. To this end, we selected a group of patients under 65 years who experienced a very severe outcome defined as requiring intubation or resulting in death and were subjected to whole exome sequencing. Different etiopathogenic mechanisms were explored using gene prioritization-based analysis in which more than 800 genes involved in immune response, immunodeficiencies or blood coagulation were studied.

Results

Clinical and demographic characteristics.
A total of 44 unrelated patients with very severe COVID-19) that required intubation, non-invasive ventilatory support or did not survive to SARS-CoV-2 infection were included in the present study. A total of 16 patients (36%) did not survive, most of them received invasive ventilatory support (intubation, 44%) and 31% received exclusively non-invasive ventilatory support (Table 1). The median age was 46 years, and patients were mainly males (70%, Table 1). Most patients were of European ancestry (86%), except for 7 Admixed Americans and one patient from another ethnicity. Main clinical characteristics of this cohort such as pre-existing diseases and COVID-19 management are summarized in Table 1. The most frequent comorbidities were obesity (defined as body mass index > 33 kg/m²; 39%), hypertension (18%), respiratory disease (16%) and oncohematological antecedents (14%). Detailed clinical data of each patient are provided in Supplementary Table 1.

Identification of candidate variants.
We have detected 44 different variants of interest located in 42 genes. These variants were identified in 29 patients, 11 of them carrying 2 or more candidate variants in different genes (Fig. 1A). Available data about pathogenic predictors, MAF and pathogenic annotation from public databases (*ClinVar* and *HGMD*) are summarized in Supplementary Table 2. A total of 12 (26%) variants were previously described as likely pathogenic or displayed strong evidence for being considered as likely pathogenic following ACMG criteria (Table 2).

Thirty eight percent of the identified variants (n = 17, 15 different variants and 1 variant found in 2 cases) was located in 15 genes involved in immune response (Fig. 1B). Eight of these candidate variants were found in genes with an autosomal dominant inheritance mode (such as *PLCG2, RNASEL, TLR4 or STAT3*), six variants were detected in genes with an autosomal recessive inheritance mode (*CFI, FCN3, HAVCR2 and IL17RC*) and three variants were identified in X-linked genes (*TLR7, TLR8 and G6PD*). The detected LoF variant in *TLR7* was found in a 30-year-old male who was included in a case-series recently reported[16]. The p.Arg378* variant on *IL17RC*, which was not described in public databases (*GnomAD*[23]*, ExAc*[24] or 1000 genome project[25]), was detected in two unrelated patients from our cohort (MAF of 2.3%). The patient carrying the variant p.Ser755Cys located in the *TLR8* gene was also hemizygous for a pathogenic variant in *G6PD* (both XL genes) and did not show pre-existing comorbidities or risk factors at the time of the SARS-CoV-2 infection (Supplementary Table 1). Besides, one of the deceased patients, clinically diagnosed with a primary immunodeficiency (FJD_0728); carried a variant on the *STAT3* gene (p.Asn175His), in addition to one pathogenic variant in the recessive *SRD5A3* gene (Table 2).

A total of 6 candidate variants (14%) were found in genes related to congenital disorders of glycosylation (*G6PC3, FKRP, MPI, MOGS, SDR5A3 and TMEM199*) and another four variants (9%) were detected in damaged DNA binding genes (*EP300, MSH6, POLD1 and RECQL4*). These 10 variants, identified in recessive genes, were carried in heterozygosis by 8 patients (Table 2).

Additionally, two variants were found in genes related to coagulation (*PLAU*) and cardiovascular risk (*LDLR*) in two different cases (Table 2). These two cases required intubation and developed severe complications during

	All (n = 44)	Survivors (n = 28)	Deceased (n = 16)
Males (n, %)	31, 70%	23, 82%	8, 50%
Age (median, range)	46, 24–62	42, 24–50	56, 45–62
Europeans (n, %)	36, 82%	20, 71%	16, 100%
Pre-existing conditions (n, %)			
Hypertension	8, 18%	5, 18%	3, 19%
Hypercholesterolemia	3, 7%	1, 4%	2, 13%
Type 2 diabetes	5, 11%	3, 11%	2, 13%
Obesity	17, 39%	9, 32%	8, 50%
Cardiovascular disease	4, 9%	2, 7%	2, 13%
Neurological disease	2, 4.5%	1, 4%	1, 6%
Respiratory disease	7, 16%	3, 11%	4, 25%
Digestive/liver disease	4, 9%	3, 11%	1, 6%
Oncohematological antecedents	6, 14%	1, 4%	5, 31%
Kidney disease	2, 4.5%	0	2, 13%
Management of COVID-19 disease			
ICU (n, %)	39, 89%	28, 100%	11, 69%
Invasive ventilation (intubation, n, %)	35, 80%	28, 100%	7, 44%
Non-invasive ventilation (CPAP/BiPAP/high flow nasal cannula)ª (n, %)	5, 11%%	0, 0%	5, 31%%
Without ventilatory support	4, 9%	0, 0%	4, 25%
Hospitalization (mean ± SD of days)	33.2 ± 24.8	38.8 ± 26.3	24.0 ± 19.3

Table 1. Clinical and demographic characteristics. ªPatients that received exclusively non-invasive ventilatory support.
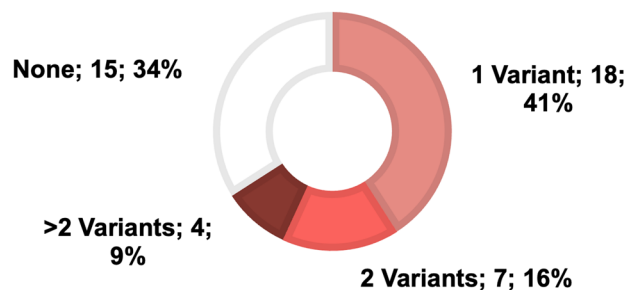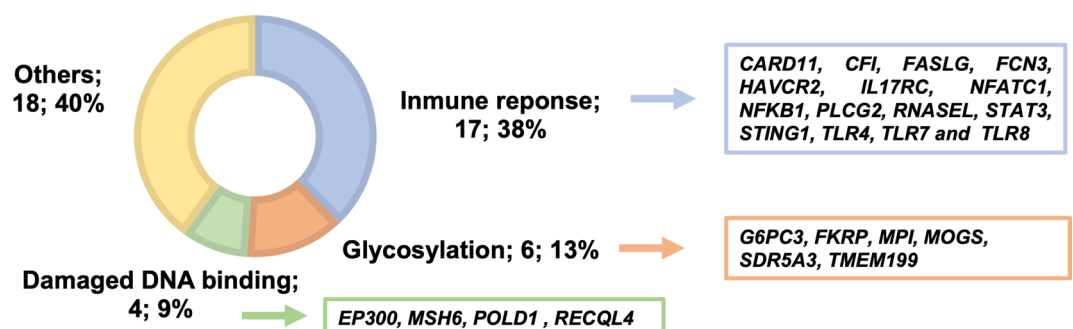

Figure 1. Identification of variants of interest in very severe COVID-19 patients: frequency and functional pathways involved. (**A**) Number and percentage of patients with none, one, two or more variants of interest, (**B**) Number and percentage of variants detected in each of the functional pathways.

Patient	Age	Sex	O	Gene	IM	GT	HGVSc	HGVSp	Varsome Class[a]	LP/P
FJD_0004	46	M	S	PLAU	AD	HT	NM_002658.3:c.970 + 1G>A	p.?	LP	No
FJD_0019	50	M	S	NPHS1	AR	HT	NM_004646.3:c.767G>A	p.Arg256Gln	LP	No
FJD_0044	49	M	S	G6PD	X-L	HM	NM_000402.4:c.934G>C	p.Asp312His	LP	Yes[b]
				NFATC1	AD	HT	NM_006162.5: c.230C>T	p.Pro77Leu	VUS	No
				TLR8	X-L	HM	NM_016610.3:c.2263A>T	p.Ser755Cys	VUS	No
FJD_0081	24	M	S	CARD11	AD	HT	NM_032415.7:c.572A>G	p.Asn191Ser	VUS	No
FJD_0325	57	M	S	PDGFRA	AD	HT	NM_006206.6:c.2464C>T	p.Arg847Cys	VUS	No
				SLC9A3	AR	HT	NM_004174.2:c.1144C>T	p.Arg382Trp	LP	No
FJD_0374	58	M	D	FANCD2	AR	HT	NM_033084.3:c.2444G>A	p.Arg815Gln	LP	Yes[b]
				FCN3	AR	HT	NM_003665.2:c.232 + 1G>A	p.?	LP	No[b]
				FKRP	AR	HT	NM_024301.4:c.265C>T	p.Pro89Ser	LP	No
				MYO5B	AR	HT	NM_001080467.2:c.3843G>C	p.Ala1281Ala	LP	No
FJD_0412	60	F	D	POLR3C	AD	HT	NM_001303456.1:c.706G>C	p.Asp236His	VUS	No
FJD_0626	59	F	D	FN1	AD	HT	NM_212482.2:c.4654C>G	p.Pro1552Ala	VUS	No
				MOGS	AR	HT	NM_006302.2:c.882del	p.Glu295Asnfs*10	LP	No[b]
				POLD1	AR	HT	NM_002691.4:c.378_394del	p.Ala127*	LP	No
FJD_0714	44	M	S	IL17RC	AR	HT	NM_153461.3:c.1132C>T	p.Arg378*	LP	No[b]
FJD_0728	62	F	D	SRD5A3	AR	HT	NM_024592.4:c.57G>A	p.Trp19*	LP	Yes[b]
				STAT3	AD	HT	NM_139276.3:c.523A>C	p.Asn270His	VUS	No
FJD_1417	38	M	S	PLCG2	AD	HT	NM_002661.4:c.3022C>T	p.Gln1008*	LP	No
FJD_1456	49	F	D	G6PC3	AR	HT	NM_138387.3:c.889del	p.Leu297Trpfs*27	LP	No[b]
				MPI	AR	HT	NM_002435.1:c.1123G>C	p.Gly375Arg	LP	No
				PIK3CD	AD	HT	NM_005026.5:c.2869C>T	p.Arg981Trp	VUS	No
FJD_1458	56	M	D	NFATC1	AD	HT	NM_006162.5:c.2311C>A	p.Leu771Ile	VUS	No
FJD_1462	51	F	D	IL17RC	AR	HT	NM_153461.3:c.1132C>T	p.Arg378 *	LP	No[b]
FJD_1532	46	M	S	TMEM199	AR	HT	NM_152464.1:c.92G>C	p.Arg31Pro	LP	Yes[b]
FJD_1645	59	M	D	RECQL4	AR	HT	NM_004260.3:c.2756G>A	:p.Ala919Thr	LP	No
FJD_2149	53	M	D	MSH6	AR	HT	NM_000179.2:c.3226C>T	p.Arg1076Cys	LP	Yes[b]
H12O_104	30	M	S	TLR7	X-L	HM	NM_016562.3:c.2050A>T	p.Lys684*	LP	Yes[b]
H12O_107	38	F	S	EP300	AD	HT	NM_001429.3:c.7180G>C	p.Gly2394Arg	VUS	No
				SLC37A4	AR	HT	NM_001164277.1:c.1016G>T	p.Gly339Cys	LP	Yes[b]
H12O_110	27	M	S	POLR3C	AD	HT	NM_001303456.1:c.1597A>T	p.Ile533Phe	VUS	No
H12O_111	45	F	D	NFKB1	AD	HT	NM_003998.4:c.233A>G	p.Asn86Ser	VUS	No
H12O_112	42	M	S	CFI	AR	HT	NM_000204.5:c.1643A>G	p.Glu556Gly	LP	No
				LDLR	AD	HT	NM_000527.4:c.1816G>A	p.Ala606Thr	LP	No
H12O_113	45	M	S	RNASEL	AD	HT	NM_021133.3:c.1450C>T	p.Gln484*	VUS	No
H12O_236	53	M	D	FASLG	AD	HT	NM_000639.1:c.596G>T	p.Gly199Val	VUS	No
				HAVCR2	AR	HT	NM_032782.4:c.291A>G	p.Ile97Met	LP	No
HVAM_083	48	M	S	SLX4	AR	HT	NM_032444.2:c.2340_2343del	p.Glu781Serfs*38	LP	No[b]
				TLR4	AD	HT	NM_138554.4:c.1976T>C	p.Met659Thr	VUS	No
HVAM_137	41	M	S	PTPN11	AD	HT	NM_002834.3:c.369G>T	p.Glu123Asp	VUS	No
HVAM_142	49	M	S	STING1	AD	HT	NM_198282.2:c.65C>A	p.Ala22Asp	VUS	No
HVAM_212	46	M	S	CHD7	AD	HT	NM_017780.3:c.8257A>G	p.Met2753Val	VUS	No
				TEK	AD	HT	NM_000459.3:c.2357A>G	p.Gln786Arg	VUS	No
HVAM_252	44	M	S	ITGB3	AD	HT	NM_000212.2:c.1658_1660del	p.Ser553del	VUS	No

Table 2. List of filtered candidate variants identified in our cohort of COVID-19 patients. *M* male, *F* female, *O* Outcome, *D* deceased, *S* survivor, *IM* inheritance mode, *AD* autosomal dominant, *AR* autosomal recessive, *X-L* linked to X chromosome, *GT* genotype, *HT* heterozygous, *HM* Hemizygous, *Chr* chromosome, *HGVSc* HGVS coding sequence name, *HGVSp* HGVS protein sequence name, *LP* likely pathogenic, *P* pathogenic, *VUS* variant of unknown significance. [a]*Varsome* class following ACMG criteria; [b]variants previously described as likely pathogenic or displaying strong evidence for being considered as likely pathogenic.
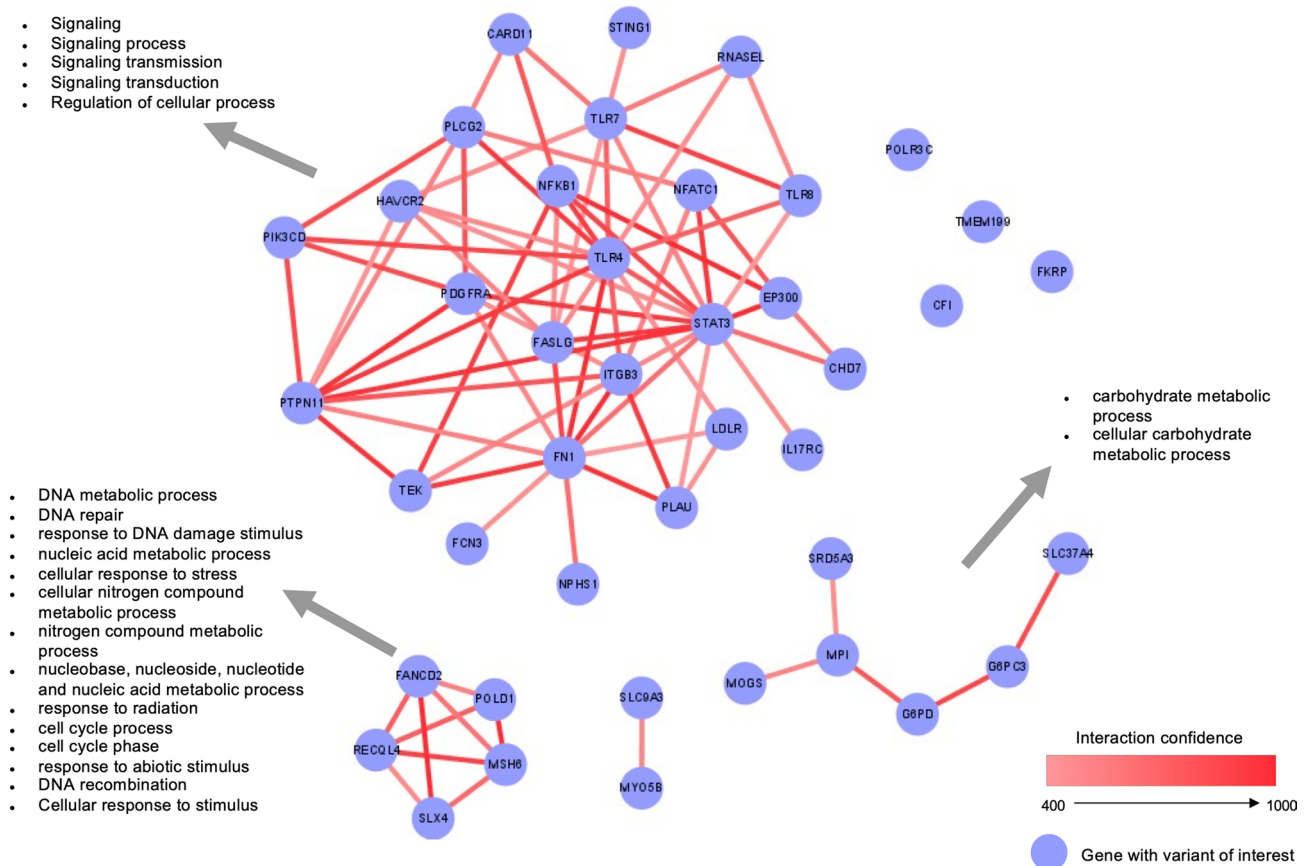
Figure 2. Network analysis of the genes with candidate variants.

SARS-COV-2 infection (Supplementary Table 2). The patient carrying the *LDLR* variant was also heterozygous for a variant in *CFI*.

Network analysis of genes with candidate variants. A network analysis, including genes with candidate variants (Supplementary Table 3), was performed to detect functional interactions among them. The network (Fig. 2) shows three main components. First component consists of 25 highly interconnected genes, 15 involved in immune response and enriched in cell signaling compared to the rest of the network (Fig. 2). Two additional network components were identified, one composed by 6 genes and enriched in carbohydrate metabolism and a third component with 5 genes enriched in DNA metabolism and repair processes, both compared to the rest of the network.

Discussion

Understanding inter-individual clinical variability in COVID-19 has important implications for the identification of high-risk patients, clinical decision-making and the development of individualized treatments. In the present study, a group of young and middle-aged patients with very severe COVID-19 were selected for a genetic study, in which 44 different variants of interest have been detected. As expected, most of the detected variants (40%) were encoded by genes directly related to immune response, as the gene panel used in this exploratory study was enriched in immune-related pathways and included additional immune genes than those reported in previous studies[11,13].

Innate immunity is crucial for early antiviral response; thus, LoF variants in related genes could affect the onset of the immune response and also alter the appropriate clearance of the infection by adaptive response[9,26,27]. In this sense, pathogenic rare variants in 13 candidate genes involved in TLR3- and IRF7-dependent type I IFN pathways showed a higher risk to severe SARS-CoV-2 infection, which could explain up to 3.5% of severe cases[11]. However, these initial findings were not replicated in subsequent studies[12,28]. In our study, we did not find any LoF variant in those 13 genes, but we have detected seven likely pathogenic variants in other genes directly related to immune response (IFN pathways, mainly).We have confirmed the presence of a *TLR7* variant in a male also participating in a recently published study[16], suffering from very severe COVID-19 and without relevant comorbidities or risk factors at the time of the infection. Therefore, our results support the genetic screening of *TLR7* variants in young men in absence of pre-existing conditions as a preventing biomarker that may help clinical management of this subset of patients. Even more, we have found two variants in other Toll-like receptor (TLRs) genes, *TLR4* (Chr9) and *TLR8* (ChrX), in two males under 50 years of age requiring intubation. Of

note, TLRs are crucial in innate response by recognizing pathogen-associated molecular patterns from different microorganisms[29], being *TLR3, 7* and 8 key sensors of RNA viruses[30].

Furthermore, nearly 40% of the variants detected in the present study were located in immune response genes, some of them with a high probability of intolerance to heterozygous LoF variation (pLI ≥ 0.9)[31];thus, a single LoF variant may lead to a severe clinical phenotype due to haploinsufficiency in genes such as *CARD11*, *STAT3* or *NFKB1*(pLI = 1, each). In addition to allelic dosage, subjects carrying the same genotypes can display variable expressivity and additional common or rare genetic variants may modify the penetrance of monogenic variants (polygenic risk)[32]. In this sense, 30% of our patients carried more than one variant of interest. Even more, other well-known COVID-19 risk factors, such as age, comorbidities, or environmental factors may affect monogenic variants penetrance to the final observed phenotype[33].

Additionally, we found 5 patients (11%) carrying heterozygous variants in genes related to glycosylation defects. Congenital defects of glycosylation (CDG) is a group of rare diseases caused mainly by recessive genes[34]. Clinical manifestations of CDG include neurological, cardiovascular, and hematologic involvement and recurrent infections, among others[35]. An increased risk of thrombotic events and bleeding complications have been related to abnormal glycosylation of coagulation factors[36] and thrombosis is one of the most important complications of COVID-19. Therefore, patients carrying a defective copy may experience a more severe course of SARS-CoV-2 infection due to the importance of glycosylation in immune response[35]. In contrast, ACE2 is a protein extensively glycosylated and previous studies showed that cellular SARS-CoV-2 entry is reduced by blocking the N-glycan and O-glycan formation[37]. Thus, it is difficult to conclude about the effect of these defective variants on the glycosylation status of the monoallelic carriers and the impact of those variants on SARS-CoV-2 clearance.

Moreover, four variants were detected in damaged DNA binding genes and a cluster including three of these genes (*POLD1*, *MSH6* and *REQL4*), in addition to *SLX4* and *FANCD2*, was detected in the network analysis. There is evidence that senescence is in part caused by accumulated DNA damage[38] and severity of some pathologies, as COVID-19, has been related to cell senescence, particularly in the elderly[39]. In addition, premature cellular senescence could be induced by viral infections[40]; therefore, COVID-19 patients with pathogenic variants in damaged DNA binding genes may be more likely to develop cellular senescence and severe COVID-19.

Interestingly, one of the candidate variants identified was in the canonical splice site of a key player of the coagulant pathway, *PLAU,* that has been previously related to bleeding disorders, tandem duplication of this gene is related to Quebec platelet disorder (MIM #601709) in a dominant model. Therefore, we could hypothesize that this variant may impair thrombosis resolution, as demonstrated previously in a knocked-out model[41] and inferred by the critical role of PLAU in the natural thrombus resolution by its fibrinolytic function[42]. Anticoagulant and fibrinolytic gene expression has been found dramatically down-regulated in the lung of COVID-19 patients compared to controls[43]. Thus, COVID-19 patients with loss of functions in the *PLAU* gene may be more likely to develop a thrombotic event. Moreover, we found a variant in the *LDRL* gene, classified as a variant of uncertain significance in relation to familial hypercholesterolemia[44]. Patients with impaired cholesterol metabolism could display a higher risk of COVID-19 severe outcomes due to the intimate relationship of hypercholesterolemia, metabolic syndrome, and heart disease[45]. Therefore, variants predisposing to hypercholesteremia such as *LDLR* pathogenic variants may confer a higher risk of suffering severe COVID-19 disease, even in the absence of other relevant comorbidities[46].

Our study has several limitations. First, we have a limited sample size. Despite recruiting more than 3500 in the Stop_Coronavirus cohort, the stringent cut-off for age (< 65 years) and outcome (only very severe COVID-19) led us to select those patients displaying an extreme phenotype in our cohort. Second, we have analyzed only the coding region; thus, we could have missed a second pathogenic allele (deep intronic regions or CNVs) in monoallelic patients that could help us explain the COVID-19 outcome. Besides, together with the effect of the detected genetic variants, it is necessary to consider the possible additional effect of pre-existing conditions related to COVID-19 severity in the patients on the outcome.

In conclusion, our descriptive study in very severe COVID-19 patients has reported the presence of rare variants in certain biological pathways such as immune response. Moreover, two additional signaling pathways have been detected including genes involved in carbohydrate metabolisms and DNA repair. Further studies are needed to confirm the ultimate role of the variants described in the present study on COVID-19 severity.

Patients and methods

Subjects and clinical data.
A case series study was performed by selecting a subgroup of patients (n = 44) from the Spanish STOP_Coronavirus[47] cohort, which comprises more than 3,500 COVID-19 patients, from 4 hospitals (three from Madrid and one from Murcia). Extreme phenotypes were selected from our STOP_Coronavirus cohort using a similar design to previous case series studies[13]. Inclusion criteria were young and middle-aged patients (age under 65 years) with a confirmatory test of SARS-CoV-2 infection that presented ARDS (survivors or deceased). More information about the Spanish STOP_Coronavirus cohort is provided in Supplementary methodology. Cases were retrospectively and prospectively enrolled from March to May 2020 and followed-up until December 2020. SARS-CoV-2 infection was confirmed by a positive PCR (n = 41) and/or serological test (n = 3, IgG and IgM both positives). COVID-19 patients were recruited from four hospitals in Spain: Hospital Universitario Fundación Jiménez Díaz (HUFJD), Hospital Universitario Infanta Elena (HUIE) and Hospital Universitario 12 de Octubre (H12O) in Madrid, and Hospital Clínico Universitario Virgen de la Arrixaca in Murcia (HVAM).

Clinical data obtained in HUFJD and HUIE were extracted from the patients' electronic medical records using batch-based complex queries and then reviewed and refined manually by two clinicians and two clinician researchers. At H12O and HVAM, clinical data were manually collected by researchers from electronic medical records. Clinical information included primary demographic data, comorbidities, COVID-19 symptoms,

laboratory findings, treatments, related complications from COVID-19, ICU admissions, and outcomes (Supplementary Table 1). Descriptive statistics (mean and SD) were calculated for main clinical and demographic data (Table 1).

This study was approved by the research ethics committees of HUFJD, HVAM and H12O. Wherever was possible, patients provided written or verbal informed consent to participate in this study. Due to the health emergency, the research ethics committees of each center waived the requirement for informed consent for the STOP_Coronavirus cohort. All samples were de-identified (pseudonymized) and clinical data were managed in accordance with national legislation and institutional requirements.

Ancestry inference. Principal component analysis (PCA) based on the variance-standardized relationship matrix was used to infer the ancestry of each patient and classify them as one of the selected ancestry groups (European, African, admixed American, and East Asian) using a set of 1000 genome samples (phase 3) as a reference population. For PCA, we used previously collected genetic data from our cohort (unpublished) obtained with the Applied Biosystems™ Axiom™ Spain Biobank Array (COL32017 1217, Thermo Fisher Scientific Inc.), which contains 758,740 variants. PCA was performed using Plink software version 1.9[48].

Kinship test. To assess kinship, we used previously collected genetic data from our cohort[2] obtained with the Applied Biosystems™ Axiom™ Spain Biobank Array (COL32017 1217, Thermo Fisher Scientific Inc.), which contains 758,740 variants. Autosomal SNPs (MAF > 5%) were pruned with PLINK[3] using a window of 1000 markers, a step size of 80 and a r^2 of 0.1. A subset of 131,937 independent SNPs was used to evaluate kinship (IBD estimation) in PLINK[3]. Only one individual from each pair of individuals with PI_HAT > 0.25 (second-degree relatives) that showed a Z0, Z1, and Z2 coherent pattern (according to theoretically expected values for each relatedness level), was removed.

Whole exome sequencing analysis. DNA was isolated from EDTA-collected peripheral blood samples using an automated DNA extractor (BioRobot EZ1, QIAGEN GmbH). DNA samples were subjected to library construction using SureSelect Human All Exon V6 (Agilent Technologies, Santa Clara, CA, USA) and sequenced on a Novaseq 6000 instrument (Illumina, San Diego, CA, USA), following the manufacturer's protocol. Paired-end reads of 2×150 bp were generated per sample to provide an on-target coverage of minimum 100 ×, with a total coverage of 12 GB/sample.

For WES analysis we applied an in-house maintained bioinformatics pipeline using bwa v0.7.17[49] for mapping to the GRCh37/hg19 human genome assembly, gatk v4.2.0 HaplotypeCaller[50] for single nucleotide variants calling and hard filtering (SNP_filter: QD (Quality of Depth) < 2.0, MQ (Mapping Quality) < 40.0, MQRankSum < − 12.5, and ReadPosRankSum < − 8.0, and INDEL_filter: QD < 2.0, and ReadPosRankSum < − 20.0). Annotations were performed using VEP r103[51]. More details can be seen at the github repository https://github.com/TBLabFJD/VariantCallingFJD and application of the same tool in Romero et al.[52].

Single variant analysis. To search for candidate variants involved in the pathophysiology of severe COVID-19, we used a candidate virtual gene panel summarized in Supplementary Table 4. Candidate gene panel included 330 genes mainly involved in type I IFN immunity, primary immunodeficiencies, and genes related to coagulation (panel 1). Moreover, 234 additional genes were selected by using the COVID-19 severity and susceptibility panel published in PanelApp[53], by selecting only green-labelled genes (panel 2)**.** Besides, other functionally related genes were included by using our GLOWgenes prioritization method (www.glowgenes.org) using the 564 genes from panels 1 and 2 as a seed set. Top 300 prioritized genes were selected and included as panel 3 (Supplementary methodology). Thus, a total of 864 genes (564 candidates and 300 selected by GLOWgenes) were included in the final panel.

The PriorR v.2.1 package (https://github.com/TBLabFJD/PriorR) was used for variant filtering and prioritization**.** Variants were filtered according to a minor allele frequency (MAF) < 0.01 in population databases [the 1000 genomes project[25], the Exome Aggregation Consortium (*ExAc*)[24], and the Genome Aggregation Database[23] (*GnomAD*)]. Synonymous, intronic and non-coding variants were excluded from the analysis. *ClinVar* (ncbi.nlm.nih.gov/clinvar/) and the *Human Gene Mutation Database*[54] (HGMD) were used to identify variants previously reported as pathogenic and those described as likely benign/benign variants were discarded. The impact of missense variants was assessed using several predictor tools (DANN[55], FATHMM[56], GERP++[57], LRT[58], M-CAP[59], CADD[60], MutationTaster[61], MutationAssessor[62], PhyloP[63], Polyphen2_HDIV[64], Polyphen2_HVAR[64], PROVEAN[65], RadiaISVM[66], SIFT[67], SiPhy[68], among others). Canonical and noncanonical splicing variants were assessed using 5 predictors (MaxEntScan[69], Human Splicing Finder[70], Splice Site Finder-like[71], NNSPLICE[72], and GeneSplicer[73]) using the Alamut software (Interactive Biosoftware, Rouen, France). The potential pathogenicity of prioritized variants was assessed using the *Varsome* tool[74] following ACGM criteria[75].

Network analysis of genes with candidate variants. Genes carrying at least one of the candidate variants (Supplementary Table 3) were submitted to the STRING database v11.5[76] and interactions with a STRING combined score ≥ 400 were downloaded as a file (.tsv) in short tabular text output format from the Exports tab. Cytoscape[77] version 3.4.0 was used for visualization. Clusters were defined as subgraphs with any two nodes (genes) connected to each other by edges, and not connected to other nodes in the graph, this normally called network components and the most extreme version of a cluster. We applied BINGO[78] Cytoscape app for the enrichment analysis extracting over-represented Gene Ontology (GO) biological processes terms comparing their annotation in every cluster to the rest of the network including genes not grouped in clusters. In the net-

work representation, the STRING combined score, which represents the interaction confidence, is used to characterize edges between genes. Functions enriched for every cluster were selected as having an FDR < 0.05.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

References

1. Casanova, J.-L., Su, H. C., COVID Human Genetic Effort. A global effort to define the human genetics of protective immunity to SARS-CoV-2 infection. *Cell* **181**, 1194–1199 (2020).
2. O'Driscoll, M. *et al.* Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature* **590**, 140–145 (2021).
3. Yang, J. *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.* **94**, 91–95 (2020).
4. Debnath, M., Banerjee, M. & Berk, M. Genetic gateways to COVID-19 infection: Implications for risk, severity, and outcomes. *FASEB J.* **34**, 8787–8795 (2020).
5. Severe Covid-19 GWAS Group *et al.* Genomewide association study of severe covid-19 with respiratory failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).
6. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature* https://doi.org/10.1038/s41586-021-03767-x (2021).
7. Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, 92–98 (2021).
8. Meyts, I. *et al.* Coronavirus disease 2019 in patients with inborn errors of immunity: An international study. *J. Allergy Clin. Immunol.* **147**, 520–531 (2021).
9. Kwok, A. J., Mentzer, A. & Knight, J. C. Host genetics and infectious disease: New tools, insights and translational opportunities. *Nat. Rev. Genet.* **22**, 137–153 (2021).
10. Delavari, S. *et al.* Impact of SARS-CoV-2 pandemic on patients with primary immunodeficiency. *J. Clin. Immunol.* **41**, 345–355 (2021).
11. Zhang, Q. *et al.* Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **370**, 1–10 (2020).
12. Kosmicki, J. A. *et al.* Pan-ancestry exome-wide association analyses of COVID-19 outcomes in 586,157 individuals. *Am. J. Hum. Genet.* **108**, 1350–1355 (2021).
13. van der Made, C. I. *et al.* Presence of genetic variants among young men with severe COVID-19. *JAMA* **324**, 663–673 (2020).
14. Fallerini, C. *et al.* Association of Toll-like receptor 7 variants with life-threatening COVID-19 disease in males: Findings from a nested case-control study. *Elife* **10**, 1–10 (2021).
15. Solanich, X. *et al.* Genetic screening for TLR7 variants in young and previously healthy men with severe COVID-19. *Front. Immunol.* **12**, 719115 (2021).
16. Asano, T. *et al.* X-linked recessive TLR7 deficiency in ~1% of men under 60 years old with life-threatening COVID-19. *Sci. Immunol.* **6**, 10 (2021).
17. Bussani, R. *et al.* Persistence of viral RNA, pneumocyte syncytia and thrombosis are hallmarks of advanced COVID-19 pathology. *EBioMedicine* **61**, 103104 (2020).
18. Katneni, U. K. *et al.* Coagulopathy and thrombosis as a result of severe COVID-19 infection: A microvascular focus. *Thromb. Haemost.* **120**, 1668–1679 (2020).
19. Sriram, K. & Insel, P. A. Inflammation and thrombosis in COVID-19 pathophysiology: Proteinase-activated and purinergic receptors as drivers and candidate therapeutic targets. *Physiol. Rev.* **101**, 545–567 (2021).
20. Richardson, S. *et al.* Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York city area. *JAMA* **323**, 2052–2059 (2020).
21. Pendu, J. L., Breiman, A., Rocher, J., Dion, M. & Ruvoën-Clouet, N. ABO Blood Types and COVID-19: Spurious, anecdotal, or truly important relationships? A reasoned review of available data. *Viruses* **13**, 1–10 (2021).
22. Stowell, S. R. & Stowell, C. P. Biologic roles of the ABH and Lewis histo-blood group antigens part II: Thrombosis, cardiovascular disease and metabolism. *Vox Sang.* **114**, 535–552 (2019).
23. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
24. Karczewski, K. J. *et al.* The ExAC browser: Displaying reference data information from over 60,000 exomes. *Nucleic Acids Res.* **45**, D840–D845 (2017).
25. Clarke, L. The 1000 genomes project. *Powerpoint* https://doi.org/10.1001/jama.299.7.755-d (2013).
26. Sette, A. & Crotty, S. Adaptive immunity to SARS-CoV-2 and COVID-19. *Cell* **184**, 861–880 (2021).
27. Hanan, N., Doud, R. L., Park, I.-W., Jones, H. P. & Mathew, S. O. The many faces of innate immunity in SARS-CoV-2 infection. *Vaccines* **9**, 596 (2021).
28. Povysil, G. *et al.* Rare loss-of-function variants in type I IFN immunity genes are not associated with severe COVID-19. *J. Clin. Invest.* **131**, 15 (2021).
29. Kawasaki, T. & Kawai, T. Toll-like receptor signaling pathways. *Front. Immunol.* **5**, 461 (2014).
30. Nazmi, A., Dutta, K., Hazra, B. & Basu, A. Role of pattern recognition receptors in flavivirus infections. *Virus Res.* **185**, 32–40 (2014).
31. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
32. Fahed, A. C. *et al.* Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat. Commun.* **11**, 3635 (2020).
33. Goodrich, J. K. *et al.* Determinants of penetrance and variable expressivity in monogenic metabolic conditions across 77,184 exomes. *Nat. Commun.* **12**, 3505 (2021).
34. Péanne, R. *et al.* Congenital disorders of glycosylation (CDG): Quo vadis?. *Eur. J. Med. Genet.* **61**, 643–663 (2018).
35. Verheijen, J., Tahata, S., Kozicz, T., Witters, P. & Morava, E. Therapeutic approaches in congenital disorders of glycosylation (CDG) involving N-linked glycosylation: an update. *Genet. Med.* **22**, 268–279 (2020).
36. Funke, S. *et al.* Perinatal and early infantile symptoms in congenital disorders of glycosylation. *Am. J. Med. Genet. A* **161A**, 578–584 (2013).
37. Yang, Q. *et al.* Inhibition of SARS-CoV-2 viral entry upon blocking N- and O-glycan elaboration. *Elife* **9**, 61552 (2020).
38. Schumacher, B., Garinis, G. A. & Hoeijmakers, J. H. J. Age to survive: DNA damage and aging. *Trends Genet.* **24**, 77–85 (2008).
39. Nehme, J., Borghesan, M., Mackedenski, S., Bird, T. G. & Demaria, M. Cellular senescence as a potential mediator of COVID-19 severity in the elderly. *Aging Cell* **19**, e13237 (2020).

40. Martínez, I. *et al.* Induction of DNA double-strand breaks and cellular senescence by human respiratory syncytial virus. *Virulence* **7**, 427–442 (2016).
41. Singh, I. *et al.* Failure of thrombus to resolve in urokinase-type plasminogen activator gene-knockout mice: Rescue by normal bone marrow-derived cells. *Circulation* **107**, 869–875 (2003).
42. D'Alonzo, D., De Fenza, M. & Pavone, V. COVID-19 and pneumonia: a role for the uPA/uPAR system. *Drug Discov. Today* **25**, 1528–1534 (2020).
43. Mast, A. E. *et al.* SARS-CoV-2 suppresses anticoagulant and fibrinolytic gene expression in the lung. *Elife* **10**, 330 (2021).
44. Alonso, R. *et al.* Genetic diagnosis of familial hypercholesterolemia using a DNA-array based platform. *Clin. Biochem.* **42**, 899–903 (2009).
45. Grasselli, G. *et al.* Risk factors associated with mortality among patients with COVID-19 in intensive care units in Lombardy, Italy. *JAMA Intern. Med.* **180**, 1345–1355 (2020).
46. Vuorio, A., Raal, F., Kaste, M. & Kovanen, P. T. Familial hypercholesterolaemia and COVID-19: A two-hit scenario for endothelial dysfunction amenable to treatment. *Atherosclerosis* **320**, 53–60 (2021).
47. Lopez-Rodriguez, R. *et al.* Androgen receptor polyQ alleles and COVID-19 severity in men: A replication study. *MedRvix* https://doi.org/10.1101/2022.03.25.22271678 (2022).
48. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
49. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
50. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv* (2018).
51. McLaren, W. *et al.* The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
52. Romero, R. *et al.* An evaluation of pipelines for DNA variant detection can guide a reanalysis protocol to increase the diagnostic ratio of genetic diseases. *NPJ Genomic Med.* **7**, 7 (2022).
53. Martin, A. R. *et al.* PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat. Genet.* **51**, 1560–1565 (2019).
54. Stenson, P. D. *et al.* The human gene mutation database: Towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* **136**, 665–677 (2017).
55. Quang, D., Chen, Y. & Xie, X. DANN: A deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* **31**, 761–763 (2015).
56. Shihab, H. A. *et al.* Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.* **34**, 57–65 (2013).
57. Davydov, E. V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* **6**, e1001025 (2010).
58. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).
59. Jagadeesh, K. A. *et al.* M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* **48**, 1581–1586 (2016).
60. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
61. Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. MutationTaster2: Mutation prediction for the deep-sequencing age. *Nat. Methods* **11**, 361–362 (2014).
62. Reva, B., Antipin, Y. & Sander, C. Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biol.* **8**, R232 (2007).
63. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
64. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
65. Choi, Y. & Chan, A. P. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747 (2015).
66. Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **24**, 2125–2137 (2015).
67. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
68. Garber, M. *et al.* Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* **25**, i54-62 (2009).
69. Eng, L. *et al.* Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: maximum entropy estimates of splice junction strengths. *Hum. Mutat.* **23**, 67–76 (2004).
70. Desmet, F.-O. *et al.* Human splicing finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, e67–e67 (2009).
71. Houdayer, C. *et al.* Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum. Mutat.* **33**, 1228–1238 (2012).
72. Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved splice site detection in genie. *J. Comput. Biol.* **4**, 311–323 (1997).
73. Pertea, M. GeneSplicer: A new computational method for splice site prediction. *Nucleic Acids Res.* **29**, 1185–1190 (2001).
74. Kopanos, C. *et al.* VarSome: The human genomic variant search engine. *Bioinformatics* **35**, 1978–1980 (2019).
75. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
76. Szklarczyk, D. *et al.* The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**, 1–10 (2011).
77. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
78. Maere, S., Heymans, K. & Kuiper, M. BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449 (2005).

Acknowledgements

Author contributions

R.L.-R., M.C. and C.A. contributed to the conception of the study. J.R.-H., M.E.P.-T., M.B.-S., E.M., E.P.-A., E.G.-N. and C.A. provided samples and clinical data. R.L.-R., J.R.-H., M.C., B.A., P.M. and C.A. reviewed the clinical data and classify patients in severity categories. P.M., G.N.-M. and R.R. performed the bioinformatic analysis.

R.L-.R., M.C., M.d.P.V., B.A.and C.V. contributed to the acquisition and interpretation of the genetic data. All authors contributed to the review and approval of the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-14035-x.

Correspondence and requests for materials should be addressed to C.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.