

## ORIGINAL ARTICLE

## Genomewide Association Study of Severe Covid-19 with Respiratory Failure

The Severe Covid-19 GWAS Group\*

## ABSTRACT

## BACKGROUND

There is considerable variation in disease behavior among patients infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes coronavirus disease 2019 (Covid-19). Genomewide association analysis may allow for the identification of potential genetic factors involved in the development of Covid-19.

## METHODS

We conducted a genomewide association study involving 1980 patients with Covid-19 and severe disease (defined as respiratory failure) at seven hospitals in the Italian and Spanish epicenters of the SARS-CoV-2 pandemic in Europe. After quality control and the exclusion of population outliers, 835 patients and 1255 control participants from Italy and 775 patients and 950 control participants from Spain were included in the final analysis. In total, we analyzed 8,582,968 single-nucleotide polymorphisms and conducted a meta-analysis of the two case-control panels.

## RESULTS

We detected cross-replicating associations with rs11385942 at locus 3p21.31 and with rs657152 at locus 9q34.2, which were significant at the genomewide level ( $P < 5 \times 10^{-8}$ ) in the meta-analysis of the two case-control panels (odds ratio, 1.77; 95% confidence interval [CI], 1.48 to 2.11;  $P = 1.15 \times 10^{-10}$ ; and odds ratio, 1.32; 95% CI, 1.20 to 1.47;  $P = 4.95 \times 10^{-8}$ , respectively). At locus 3p21.31, the association signal spanned the genes *SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6* and *XCR1*. The association signal at locus 9q34.2 coincided with the *ABO* blood group locus; in this cohort, a blood-group-specific analysis showed a higher risk in blood group A than in other blood groups (odds ratio, 1.45; 95% CI, 1.20 to 1.75;  $P = 1.48 \times 10^{-4}$ ) and a protective effect in blood group O as compared with other blood groups (odds ratio, 0.65; 95% CI, 0.53 to 0.79;  $P = 1.06 \times 10^{-5}$ ).

## CONCLUSIONS

We identified a 3p21.31 gene cluster as a genetic susceptibility locus in patients with Covid-19 with respiratory failure and confirmed a potential involvement of the *ABO* blood-group system. (Funded by Stein Erik Hagen and others.)

The authors' full names, academic degrees, and affiliations are listed in the Appendix. Address reprint requests to Dr. Franke at the Institute of Clinical Molecular Biology and University Hospital of Schleswig-Holstein, Christian-Albrechts-University, Rosalind-Franklin-Str. 12, D-24105 Kiel, Germany, or at a.franke@mucosa.de; or to Dr. Karlsen at the Division of Surgery, Inflammatory Diseases, and Transplantation, Oslo University Hospital Rikshospitalet and University of Oslo, Postboks 4950 Nydalen, N-0424 Oslo, Norway, or at t.h.karlsen@medisin.uio.no.

\*Dr. Franke serves as an author on behalf of the Covid-19 Host Genetics Initiative; members of the Initiative are listed in Supplementary Appendix 1, available at NEJM.org.

Dr. Ellinghaus and Ms. Degenhardt and Drs. Valenti, Franke, and Karlsen contributed equally to this article.

This article was published on June 17, 2020, at NEJM.org.

N Engl J Med 2020;383:1522-34.

DOI: 10.1056/NEJMoa2020283

Copyright © 2020 Massachusetts Medical Society.

SEVERE ACUTE RESPIRATORY SYNDROME coronavirus 2 (SARS-CoV-2) was discovered in Wuhan, China, in late 2019, and coronavirus disease 2019 (Covid-19), the disease caused by SARS-CoV-2, rapidly evolved into a global pandemic.<sup>1</sup> As of June 15, 2020, there were more than 8.03 million confirmed cases worldwide, with total deaths exceeding 436,900.<sup>2</sup> In Europe, Italy and Spain were severely affected early on, with epidemic peaks starting in the second half of February 2020 (Fig. 1) and 61,507 deaths reported by June 15, 2020. Covid-19 has varied manifestations,<sup>3</sup> with the large majority of infected persons having only mild symptoms or even no symptoms.<sup>4</sup> Mortality rates are driven predominantly by the subgroup of patients who have severe respiratory failure related to interstitial pneumonia in both lungs and acute respiratory distress syndrome.<sup>5</sup> Severe Covid-19 with respiratory failure requires early and prolonged support by mechanical ventilation.<sup>6</sup>

The pathogenesis of severe Covid-19 and the associated respiratory failure is poorly understood, but higher mortality is consistently associated with older age and male sex.<sup>7,8</sup> Clinical associations have also been reported for hypertension, diabetes, and other obesity-related and cardiovascular disease traits, but the relative role of clinical risk factors in determining the severity of Covid-19 has not yet been clarified.<sup>7-11</sup> Observational data on lymphocytic endotheliitis and diffuse microvascular and macrovascular thromboembolic complications suggest that Covid-19 is a systemic disease that involves injury to the vascular endothelium but provide little insight into the underlying pathogenesis.<sup>12-14</sup> At the peak of the epidemic in Italy and Spain in early 2020, we performed a genomewide association study (GWAS) in an attempt to delineate host genetic factors contributing to severe Covid-19 with respiratory failure. The relatively low disease burden of Covid-19 in Norway and Germany allowed for a complementary team to be set up, whereby genotyping and analysis could occur in parallel with the rapid recruitment of patients in the heavily affected Italian and Spanish epicenters.

## METHODS

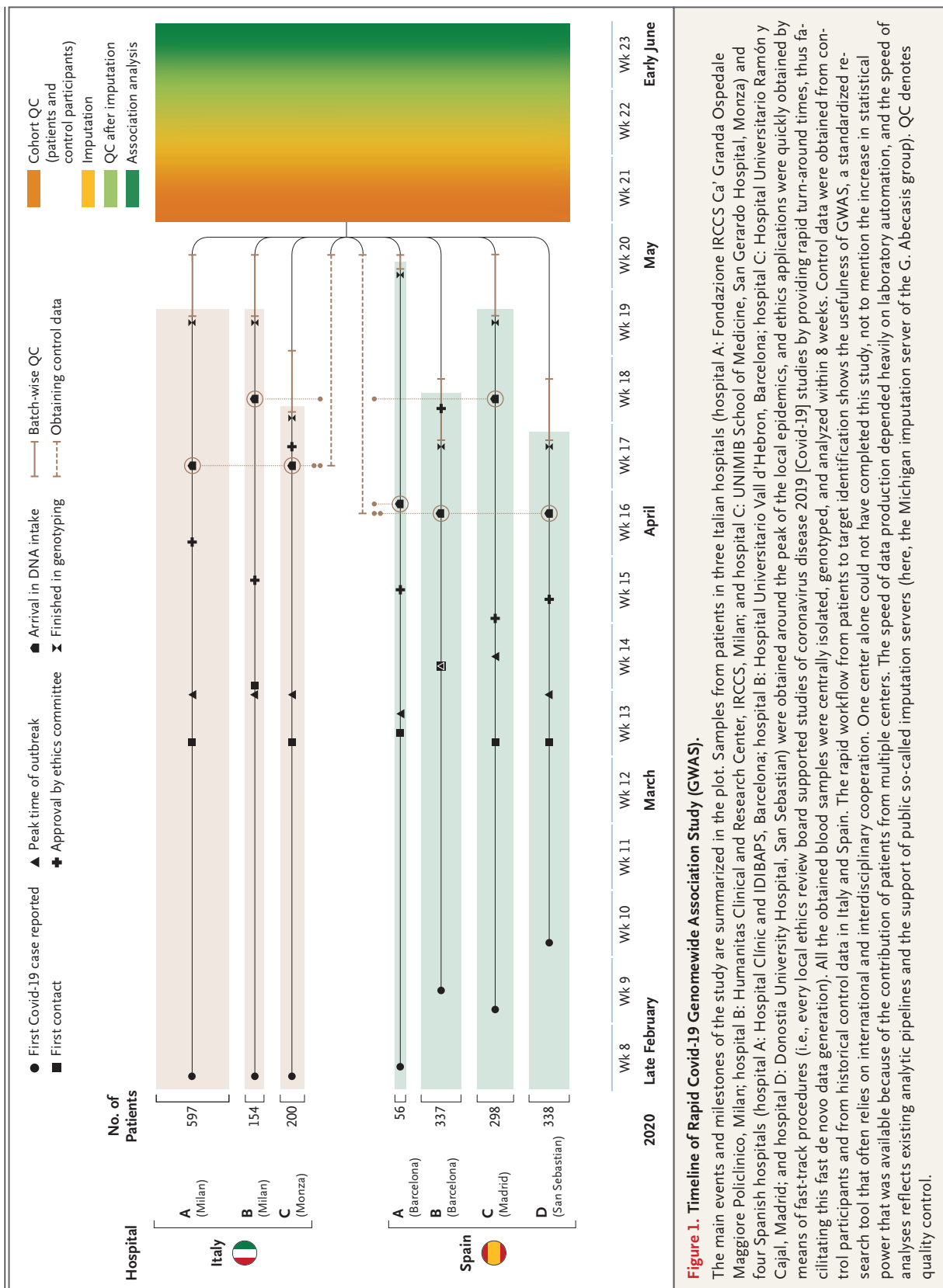
### STUDY PARTICIPANTS AND RECRUITMENT

We recruited 1980 patients with severe Covid-19, which was defined as hospitalization with respiratory failure and a confirmed SARS-CoV-2 viral

RNA polymerase-chain-reaction (PCR) test from nasopharyngeal swabs or other relevant biologic fluids, cross sectionally, from intensive care units and general wards at seven hospitals in four cities in the pandemic epicenters in Italy and Spain (Table S1A in Supplementary Appendix 1, available with the full text of this article at NEJM.org). The hospitals in Italy were the following: Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico in Milan (597 patients); Humanitas Clinical and Research Center, IRCCS, in Milan (154 patients); and UNIMIB (Università degli Studi di Milano–Bicocca) School of Medicine, San Gerardo Hospital, in Monza (a suburb of Milan) (200 patients). The hospitals in Spain were the following: Hospital Clínic and IDIBAPS (Instituto de Investigaciones Biomédicas August Pi i Sunyer) in Barcelona (56 patients), Hospital Universitario Vall d'Hebron in Barcelona (337 patients), Hospital Universitario Ramón y Cajal in Madrid (298 patients), and Donostia University Hospital in San Sebastian (338 patients).

Respiratory failure was defined in the simplest possible manner in order to ensure feasibility: the use of oxygen supplementation or mechanical ventilation, with severity graded according to the maximum respiratory support received at any point during hospitalization (supplemental oxygen therapy only, noninvasive ventilatory support, invasive ventilatory support, or extracorporeal membrane oxygenation). For severity assessments, severity was also dichotomized as no mechanical ventilation or mechanical ventilation. Whole-blood samples or buffy coats from diagnostic venipuncture were obtained for DNA extraction.

For comparison, we included 2381 control participants from Italy and Spain (Table S1B in Supplementary Appendix 1). We recruited 998 randomly selected blood donors at Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Milan, who underwent genotyping for the purpose of the present study. A total of 40 of these participants had evidence of the development of anti-SARS-CoV-2 antibodies, all of whom had mild or no Covid-19 symptoms. We also included two control panels with genotype data derived from previous studies and from persons with unknown SARS-CoV-2 infection status using the same genotyping array. The panels included 396 healthy volunteers, blood donors, and outpatients of gastroenterology departments in Italy<sup>15</sup> and 987 healthy blood donors in San Sebastian, Spain.



**Figure 1. Timeline of Rapid Covid-19 Genomewide Association Study (GWAS).**

The main events and milestones of the study are summarized in the plot. Samples from patients in three Italian hospitals (hospital A: Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan; hospital B: Humanitas Clinical and Research Center, IRCCS, Milan; and hospital C: UNIMIB School of Medicine, San Gerardo Hospital, Monza) and four Spanish hospitals (hospital A: Hospital Clinic and IDIBAPS, Barcelona; hospital B: Hospital Universitario Vall d'Hebron, Barcelona; hospital C: Hospital Universitario Ramón y Cajal, Madrid; and hospital D: Donostia University Hospital, San Sebastian) were obtained around the peak of the local epidemics, and ethics applications were quickly obtained by means of fast-track procedures (i.e., every local ethics review board supported studies of coronavirus disease 2019 [Covid-19] studies by providing rapid turn-around times, thus facilitating this fast de novo data generation). All the obtained blood samples were centrally isolated, genotyped, and analyzed within 8 weeks. Control data were obtained from control participants and from historical control data in Italy and Spain. The rapid workflow from patients to target identification shows the usefulness of GWAS, a standardized re-search tool that often relies on international and interdisciplinary cooperation. One center alone could not have completed this study, not to mention the increase in statistical power that was available because of the contribution of patients from multiple centers. The speed of data production depended heavily on laboratory automation, and the speed of analyses reflects existing analytic pipelines and the support of public so-called imputation servers (here, the Michigan imputation server of the G. Abecasis group). QC denotes quality control.

**ETHICS COMMITTEE APPROVAL**

The project protocol involved the rapid recruitment of patient-participants and no additional project-related procedures (we primarily used material from clinically indicated venipunctures) and afforded anonymity, owing to the minimal data set collected. Differences in recruitment and consent procedures among the centers arose because some centers integrated the project into larger Covid-19 biobanking efforts, whereas other centers did not, and because there were differences in how local ethics committees provided guidance on the handling of anonymization or deidentification of data as well as consent procedures. Written informed consent was obtained, sometimes in a delayed fashion, from the study patients at each center when possible. In some instances, informed consent was provided verbally or by the next of kin, depending on local ethics committee regulations and special policies issued for Covid-19 research. For some severely ill patients, an exemption from informed consent was obtained from a local ethics committee or according to local regulations in order to allow the use of completely anonymized surplus material from diagnostic venipuncture.

The following approvals of the project were obtained from the relevant ethics committees: Germany: Kiel (reference number, D464/20); Italy: Fondazione IRCCS Cá Granda Ospedale Maggiore Policlinico (reference numbers, 342\_2020 for patients and 334-2020 for control participants), Humanitas Clinical and Research Center, IRCCS (reference number, 316/20), the University of Milano-Bicocca School of Medicine, San Gerardo Hospital, Monza (the ethics committee of the National Institute of Infectious Diseases Lazzaro Spallanzani reference number, 84/2020); Norway: Regional Committee for Medical and Health Research Ethics in South-Eastern Norway (reference number, 132550); Spain: Hospital Clínic, Barcelona (reference number, HCB/2020/0405), Hospital Universitario Vall d'Hebron, Barcelona (reference number, PR[AG]244/2020), Hospital Universitario Ramón y Cajal, Madrid (reference number, 093/20) and Donostia University Hospital, San Sebastian (reference number, PI2020064).

**SAMPLE PROCESSING, GENOTYPING, AND IMPUTATION**

We performed DNA extraction using a Chemagic 360 (PerkinElmer) with the use of the low-volume kit CMG-1491 and the buffy-coat kit CMG-714

(Chemagen), respectively. For genotyping, we used the Global Screening Array (GSA), version 2.0 (Illumina), which contains 712,189 variants before quality control. Details on genotyping and quality-control procedures are provided in the Supplementary Methods section in Supplementary Appendix 1. To maximize genetic coverage, we performed single-nucleotide polymorphism (SNP) imputation on genome build GRCh38 using the Michigan Imputation Server and 194,512 haplotypes generated by the Trans-Omics for Precision Medicine (TOPMed) program (freeze 5).<sup>16</sup>

After the exclusion of samples during quality control (the majority of which were due to population outliers; see the Supplementary Methods section and Table S1B and S1C), the final case-control data sets comprised 835 patients and 1255 control participants from Italy and 775 patients and 950 control participants from Spain. A total of 8,965,091 SNPs were included in the Italian cohort and 9,140,716 SNPs in the Spanish cohort.

**STATISTICAL ANALYSIS**

To take imputation uncertainty into account, we tested for phenotypic associations with allele dosage data separately for both the Italian and Spanish case-control panels with the use of the PLINK logistic-regression framework for dosage data (PLINK, version 1.9).<sup>17</sup> We carried out two genomewide tests of association that included covariates from principal-component analyses, with adjustments to control for potential population stratification (main analysis) and potential population stratification and age and sex bias (analysis corrected for age and sex). A fixed-effects meta-analysis was conducted with the use of the meta-analysis tool METAL<sup>18</sup> on 8,582,968 variants that were common to both the Italian and Spanish data sets with the use of effect-size estimates and their standard errors from the study-specific association analyses.

For the genomewide meta-analysis, we used the commonly accepted threshold of  $5 \times 10^{-8}$  for joint P values to determine statistical significance. Bayesian fine-mapping analysis was performed for loci reaching genomewide significance (see the Supplementary Methods section). Genomewide summary statistics of our analyses are publicly available through our web browser ([www.c19-genetics.eu](http://www.c19-genetics.eu)) and have been submitted to the European Bioinformatics Institute ([www.ebi.ac](http://www.ebi.ac)



**Table 1. Overview of Patients Included in the Final Analysis.\***

Characteristic	Italian Hospitals			Spanish Hospitals			
	A (N=503)	B (N=140)	C (N=192)	A (N=45)	B (N=228)	C (N=201)	D (N=301)
Median age (IQR) — yr	64 (54–76)	67 (57–75)	66 (56–74)	69 (59–75)	65 (56–72)	69 (60–79)	67 (57–75)
Female sex — no. (%)	159 (32)	39 (28)	51 (27)	13 (29)	78 (34)	50 (25)	124 (41)
Respiratory support — no. (%)							
Supplemental oxygen only	0	70 (50)	67 (35)	7 (16)	105 (46)	106 (53)	255 (85)
Noninvasive ventilation	399 (79)	25 (18)	89 (46)	6 (13)	7 (3)	16 (8)	0
Ventilator	104 (21)	45 (32)	33 (17)	31 (69)	116 (51)	77 (38)	46 (15)
ECMO	0	0	3 (2)	1 (2)	0	2 (1)	0
Hypertension — no./total no. (%)	166/503 (33)	71/140 (51)	109/192 (57)	26/45 (58)	113/228 (50)	112/199 (56)	114/301 (38)
Coronary artery disease — no./total no. (%)	21/503 (4)	25/140 (18)	25/192 (13)	4/45 (9)	14/228 (6)	35/199 (18)	15/301 (5)
Diabetes — no./total no. (%)	63/503 (13)	18/140 (13)	34/192 (18)	10/45 (22)	50/228 (22)	57/199 (29)	65/301 (22)

\* In Italy, hospital A was Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan; hospital B Humanitas Clinical and Research Center, IRCCS, Milan; and hospital C UNIMIB School of Medicine, San Gerardo Hospital, Monza. In Spain, hospital A was Hospital Clínic and IDIBAPS, Barcelona; hospital B Hospital Universitario Vall d'Hebron, Barcelona; hospital C Hospital Universitario Ramón y Cajal, Madrid; and hospital D Donostia University Hospital, San Sebastián. The predominance of men among the patients and the advanced age (median, >63 years) were consistent across all the centers. The sample numbers provided are after quality control was conducted for the genome-wide association study (Table S1C in Supplementary Appendix 1). Allele distributions for detected risk variants at loci 3p21.31 and 9q34.2 in clinical subsets are shown in Table S2. ECMO denotes extracorporeal membrane oxygenation, and IQR interquartile range.

.uk/gwas; accession numbers, GCST90000255 and GCST90000256).

On the basis of the results from the TOPMed genotype imputation, we selected three ABO SNPs (rs8176747, rs41302905, and rs8176719)<sup>19,20</sup> to infer the ABO blood type and calculated odds ratios according to blood type (A vs. B, AB, or O; B vs. A, AB, or O; AB vs. A, B, or O; and O vs. A, AB, or B) (see the Supplementary Methods). To assess in detail the HLA complex at locus 6p21, we performed sequencing-based HLA typing of seven classical HLA loci (HLA-A, -C, -B, -DRB1, -DQA1, -DQB1, and -DPB1) in a subgroup of 835 patients and 891 control participants from Italy and 773 patients from Spain (see the Supplementary Methods). We also assessed allelic distribution according to no mechanical ventilation (supplemental oxygen only) as compared with mechanical ventilation of any type. A similar assessment was made for lead SNPs rs11385942 and rs657152 at loci 3p21.31 and 9q34.2, respectively.

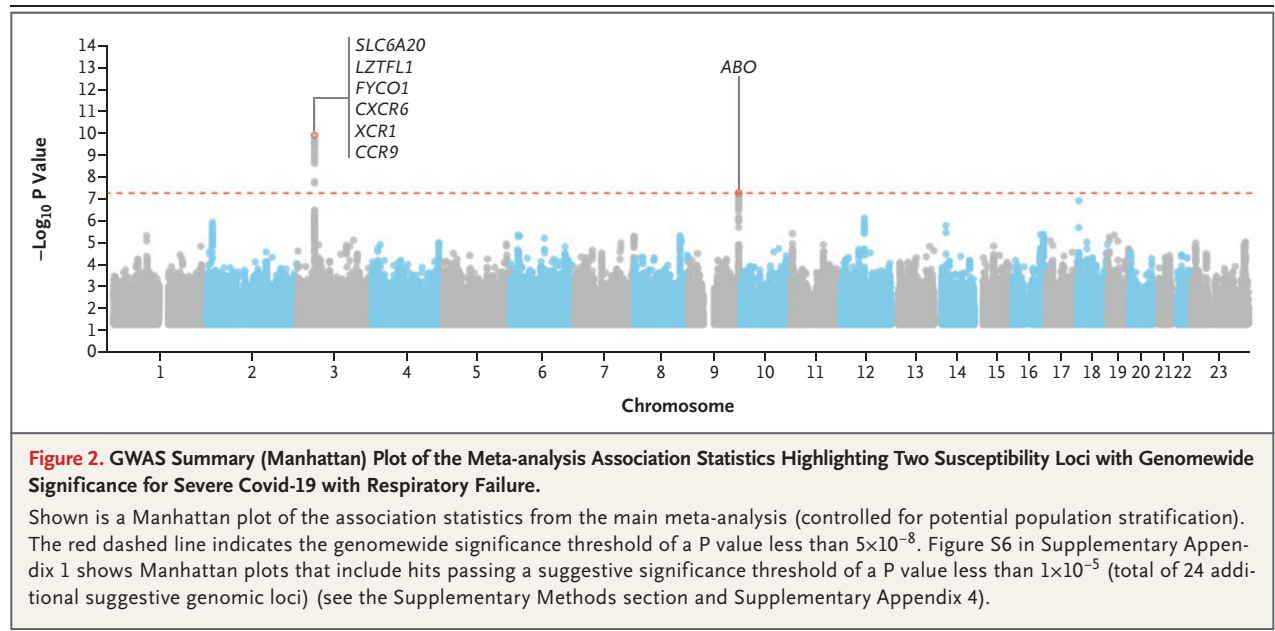
## RESULTS

### PATIENTS, GENOTYPING, AND QUALITY CONTROL

The milestones of the study in the context of the peak outbreaks in Italy and Spain are shown in Figure 1. Data on the age, sex, maximum respiratory support at any point during hospitalization, and relevant coexisting conditions (type 2 diabetes, hypertension, and coronary heart disease) in the patients who were included in the final analysis are shown in Table 1 and in Table S2 in Supplementary Appendix 1. Because we used the same genotyping platform (GSA) to obtain both data sets, we were able to perform a uniform quality control of the merged Italian and Spanish SNP data sets, thus reducing technical confounders to a minimum. A quantile–quantile (Q-Q) plot of the two meta-analyses (the main analysis and the analysis corrected for age and sex) showed significant associations in the tail of the distribution with minimal genomic inflation ( $\lambda_{GC}=1.015$  for main analysis and  $\lambda_{GC}=1.006$  for analysis corrected for age and sex) (Fig. S2 in Supplementary Appendix 1). We also carried out separate association analyses for the Italian and Spanish data sets (see the Supplementary Methods section and Fig. S3).

### GENOMEWIDE ASSOCIATION ANALYSIS

We found two loci to be associated with Covid-19–induced respiratory failure with genome-wide sig-



nificance ( $P < 5 \times 10^{-8}$ ) in the main meta-analysis: the rs11385942 insertion–deletion GA or G variant at locus 3p21.31 (odds ratio for the GA allele, 1.77; 95% confidence interval [CI], 1.48 to 2.11;  $P = 1.15 \times 10^{-10}$ ) and the rs657152 A or C SNP at locus 9q34.2 (odds ratio for the A allele, 1.32; 95% CI, 1.20 to 1.47;  $P = 4.95 \times 10^{-8}$ ) (Fig. 2 and Table 2 and Supplementary Appendix 2, available at NEJM.org). Both loci showed nominally significant association in both the Spanish and Italian subanalyses (Table 2). The meta-analysis association results for recessive and heterozygous genetic models for the two meta-analyses (main analysis and the analysis corrected for age and sex) are provided in Supplementary Appendix 3, available at NEJM.org. The imputation quality of the associated markers was good (Table 2 and Supplementary Appendix 2), and manual inspection of genotype cluster plots of genotyped SNPs in these regions showed distinct genotype clouds for homozygous and heterozygous calls (Fig. S4 in Supplementary Appendix 1). Furthermore, the analyses that were corrected for age and sex corroborated the observations at both rs11385942 (meta-analysis odds ratio, 2.11; 95% CI, 1.70 to 2.61;  $P = 9.46 \times 10^{-12}$ ) and rs657152 (meta-analysis odds ratio, 1.39; 95% CI, 1.22 to 1.59;  $P = 5.35 \times 10^{-7}$ ) (Table 2 and Fig. S5 in Supplementary Appendix 1).

The allele frequencies in Spanish and Italian control data sets from previously published studies<sup>21–27</sup> are consistent with those we report here

(Supplementary Appendix 2). A further 24 different genomic loci showed suggestive evidence ( $P < 1 \times 10^{-5}$ ) for association with Covid-19–induced respiratory failure in the main analysis (Supplementary Appendix 4, available at NEJM.org, and Fig. S6 in Supplementary Appendix 1). Association signals at loci 3p21.31 and 9q34.2 were fine-mapped to 22 and 38 variants, respectively, with greater than 95% certainty (Fig. 3A and 3B and Supplementary Appendix 5, available at NEJM.org).

#### CHROMOSOME 3P21.31

The association signal at locus 3p21.31 comprised six genes (*SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6*, and *XCR1*) (Fig. 3A). The risk allele GA of rs11385942 is associated with reduced expression of *CXCR6* and increased expression of *SLC6A20*, and *LZTFL1* is strongly expressed in human lung cells (Fig. S7 and Supplementary Appendix 6, available at NEJM.org). We found that the frequency of the risk allele of the lead variant at 3p21.31 (rs11385942) was higher among patients who received mechanical ventilation than among those who received oxygen supplementation only in both the main meta-analysis (odds ratio, 1.70; 95% CI, 1.27 to 2.26;  $P = 3.30 \times 10^{-4}$ ) and the meta-analysis corrected for age and sex (odds ratio, 1.56; 95% CI, 1.17 to 2.01;  $P = 0.003$ ) (Supplementary Appendix 7, available at NEJM.org). Furthermore, the 19 patients who were homozygous for the rs11385942 risk allele were younger than

**Table 2. Susceptibility Loci Associated with Severe Covid-19 with Respiratory Failure.\***

Chromosome and Analysis	Meta-analysis				Italian Panel				Spanish Panel			
	P Value		Odds Ratio (95% CI)	P Value	Odds Ratio (95% CI)	Allele Frequency		P Value	Odds Ratio (95% CI)	Allele Frequency		
						patient	control			patient	control	
3p21.31†												
Main analysis	1.15×10 <sup>-10</sup>	1.77 (1.48–2.11)	1.98×10 <sup>-7</sup>	1.74 (1.27–2.38)	0.14	0.09	1.32×10 <sup>-4</sup>	1.85 (1.50–2.28)	0.09	0.05		
Analysis corrected for age and sex	9.46×10 <sup>-12</sup>	2.11 (1.70–2.61)	7.02×10 <sup>-8</sup>	1.95 (1.53–2.48)	0.14	0.09	1.17×10 <sup>-5</sup>	2.79 (1.76–4.42)	0.09	0.05		
9q34.2‡												
Main analysis	4.95×10 <sup>-8</sup>	1.32 (1.20–1.47)	2.90×10 <sup>-6</sup>	1.37 (1.20–1.57)	0.42	0.35	3.55×10 <sup>-3</sup>	1.26 (1.08–1.48)	0.42	0.35		
Analysis corrected for age and sex	5.35×10 <sup>-7</sup>	1.39 (1.22–1.59)	5.31×10 <sup>-5</sup>	1.37 (1.17–1.60)	0.42	0.35	2.81×10 <sup>-3</sup>	1.45 (1.13–1.84)	0.42	0.35		

\* The meta-analysis included 1610 patients and 2205 control participants; the Italian analysis, 835 and 1255, respectively; and the Spanish analysis, 775 and 950, respectively. Allele frequencies of the minor or risk allele (see below) are shown among the patients and the control participants. All the association test statistics were adjusted for the top 10 principal components from the principal-component analysis. Two analyses were performed: a main analysis, which was corrected for 10 principal components, and an analysis that was corrected for age and sex in addition to 10 principal components. In the analyses that were corrected for age and sex, 25 control participants were excluded from the Spanish analysis and the meta-analysis because of missing covariate data. The P values and corresponding odds ratios and 95% confidence intervals (CIs) are shown with respect to the minor allele. Association results for the recessive and heterozygous models for both meta-analyses (main and corrected for age and sex) are shown in Supplementary Appendix 3. Covid-19 denotes coronavirus disease 2019.

<sup>†</sup> For chromosome 3p21.31, the association boundaries for each index single-nucleotide polymorphism (SNP; see the Supplementary Methods section), with the genomic positions retrieved from genome build hg38, were chr3:45800446 through 46135604. The Single Nucleotide Polymorphism database (dbSNP) identifier was rs11385942 (the rs identifier from the National Center for Biotechnology Information, rs11385942, is annotated as chr3:45834968 through 45834969:AAA:AA in dbSNP, version 153, and as chr3:45834967:GA:G in the Trans-Omics for Precision Medicine [TOPMed] imputation reference panel). The SNP rs11385942 was imputed according to TOPMed with high confidence (TOPMed estimated imputation accuracy, R<sup>2</sup> = 0.94 and R<sup>2</sup> = 0.95 for the Italian and Spanish panels, respectively) (Supplementary Appendix 2). The minor or risk allele was GA, and the major allele was G. The key genes (i.e., the candidate genes in the region) were *SLC6A20*, *LZTFL1*, *FYCO1*, *CXCR6*, *XCR1*, and *CCR9*.

<sup>‡</sup> For chromosome 9q34.2, the association boundaries for each index SNP, with the genomic positions retrieved from genome build hg38, were chr9:133257521 through 133279871. The SNP rs657152 was genotyped according to the Global Screening Array (GSA) in the Italian and Spanish panels (Supplementary Appendix 2). The minor or risk allele was A, and the major allele was C. The key gene was *ABO*.

1591 patients who were heterozygous or homozygous for the nonrisk allele (median age, 59 years [interquartile range, 49 to 68] vs. 66 years [interquartile range, 56 to 75];  $P=0.005$ ). Available variant database entries suggest that the frequency of this risk allele varies among populations worldwide (Fig. S8 in Supplementary Appendix 1).

#### ABO LOCUS

At locus 9q34.2 the association signal coincided with the ABO blood group locus (Fig. 3B and Fig. S9 in Supplementary Appendix 1). Accordingly, the distribution of ABO blood groups (predicted from combinations of genotypes of three different SNPs) was skewed among patients with Covid-19 who had respiratory failure, as compared with the distribution among control participants. In the meta-analysis corrected for age and sex, we found a higher risk among persons with blood group A than among patients with other blood groups (odds ratio, 1.45; 95% CI, 1.20 to 1.75;  $P=1.48\times10^{-4}$ ) and a protective effect for blood group O as compared with the other blood groups (odds ratio, 0.65; 95% CI, 0.53 to 0.79;  $P=1.06\times10^{-5}$ ). Details are provided in Supplementary Appendix 8, available at NEJM.org. Both associations and effect directions were consistent in the separate Spanish and Italian case-control analyses. We found no significant difference in blood-group distribution between patients receiving supplemental oxygen only and those receiving mechanical ventilation of any kind. The ABO blood-group frequency distributions in public registries are provided for comparison in Supplementary Appendix 8, along with details of the results presented here, and corroborate our observations.

#### HLA ANALYSIS

Given its important role in several viral infections, we scrutinized the extended HLA region (chromosome 6, 25 through 34 Mb). There were no SNP association signals at the HLA complex that met even the significance threshold of suggestive association:  $P<1\times10^{-5}$  (Fig. S10 in Supplementary Appendix 1). Dedicated analysis of the classical HLA loci showed no significant allele associations with either Covid-19 or disease severity (oxygen supplementation only or mechanical ventilation of any kind), and further analysis of heterozygote and divergent allele advantage or predicted number of HLA-bound SARS-CoV-2 peptides did not show significant associations with

Covid-19 in this data set (see the HLA Analyses section in Supplementary Appendix 1 and Supplementary Appendix 9, available at NEJM.org).

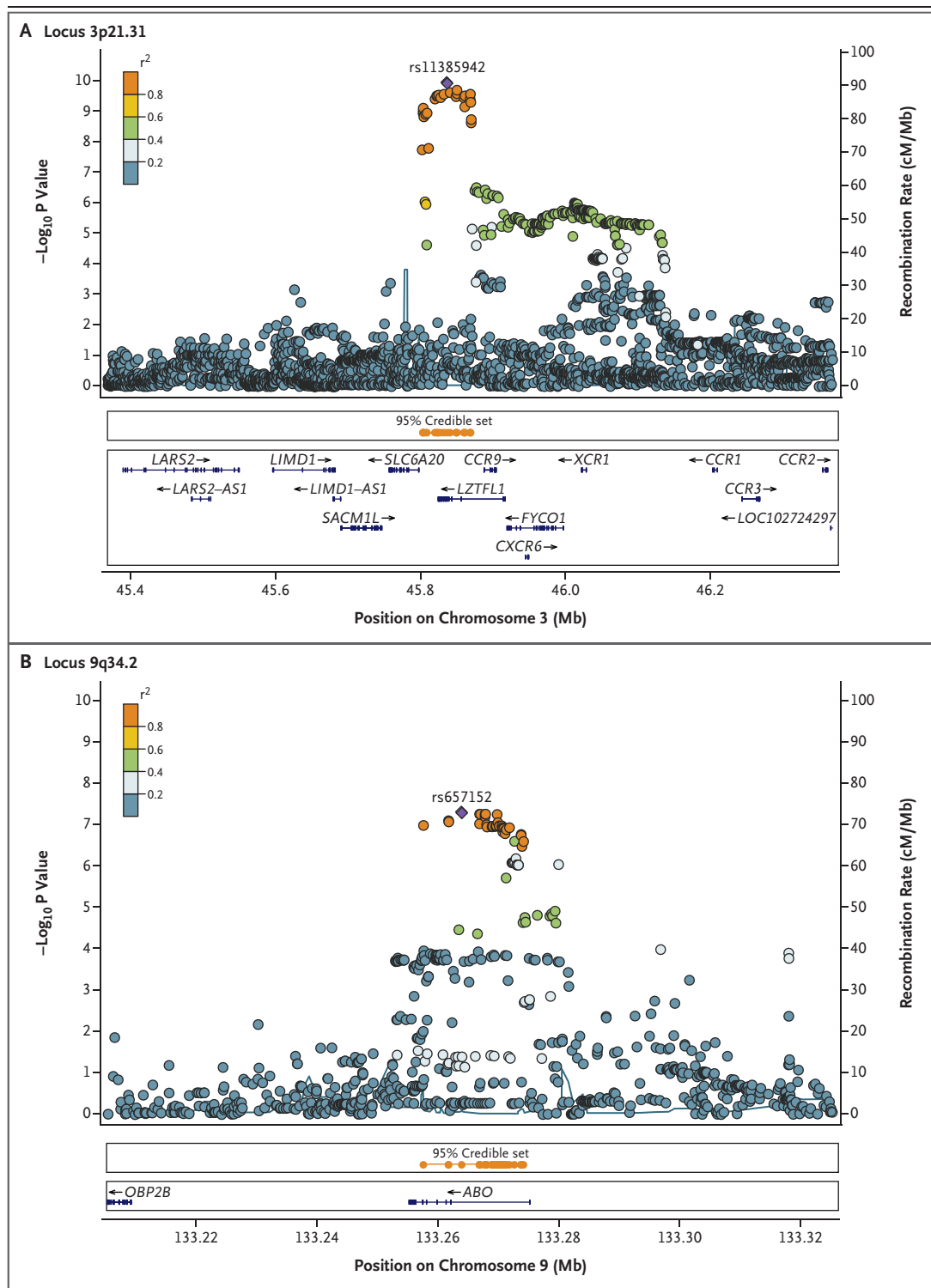
### DISCUSSION

Using a pragmatic approach with simplified inclusion criteria and a complementary team of clinicians at the European Covid-19 epicenters in Italy and Spain and scientists in the less-burdened countries of Germany and Norway, we performed a GWAS that included de novo genotyping for Covid-19 with respiratory failure in approximately 2 months. We detected a novel susceptibility locus at a chromosome 3p21.31 gene cluster and confirmed a potential involvement of the ABO blood-group system in Covid-19.

On chromosome 3p21.31, the peak association signal covered a cluster of six genes (*SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6*, and *XCR1*), several of which have functions that are potentially relevant to Covid-19. A causative gene cannot be reliably implicated by the present data. One candidate is *SLC6A20*, which encodes the sodium-imino acid (proline) transporter 1 (SIT1) and which functionally interacts with angiotensin-converting enzyme 2, the SARS-CoV-2 cell-surface receptor.<sup>28,29</sup> However, the locus also contains genes encoding chemokine receptors, including the CC motif chemokine receptor 9 (CCR9) and the C-X-C motif chemokine receptor 6 (CXCR6), the latter of which regulates the specific location of lung-resident memory CD8 T cells throughout the sustained immune response to airway pathogens, including influenza viruses.<sup>30</sup> Flanking genes (e.g., *CCR1* and *CCR2*) also have relevant functions,<sup>31</sup> and further studies will be needed to delineate the functional consequences of detected associations.

The preliminary results from the Covid-19 Host Genetics Consortium<sup>32</sup> include suggestive associations within the same locus at chromosome 3p21.31, which lend considerable support to our findings (Fig. S11 in Supplementary Appendix 1). The consortium analysis also used population-based controls, but the patients included persons with mild Covid-19 and those with severe Covid-19. The parallel findings nevertheless underscore an important point about the ascertainment of patients and controls in genetic studies of Covid-19. Because the majority of patients with SARS-CoV-2 infection are asymp-





tomatic, any sample involving patients with a positive nasopharyngeal RNA test is likely to hold a bias toward some degree of symptomatic burden. Two of the identifiers for inclusion in the current study were a positive result for the presence of SARS-CoV-2 according to PCR testing and receipt of respiratory support (an extreme Covid-19 phenotype). As such, it seems

**Figure 3 (facing page). Regional Association Plots of Susceptibility Loci Associated with Severe Covid-19 with Respiratory Failure.**

Bayesian fine-mapping analysis prioritized 22 and 38 variants for loci 3p21.31 (Panel A) and 9q34.2 (Panel B), respectively, with greater than 95% certainty. The linkage disequilibrium values were calculated on the basis of genotypes of the merged Italian and Spanish data sets derived from TOPMed (Trans-Omics for Precision Medicine) imputation. The positions in the genome assembly hg38 are plotted. The recombination rate is shown in centimorgans (cM) per million base pairs (Mb). The plot shows the names and locations of the genes; the transcribed strand is indicated with an arrow. Genes are represented with intronic and exonic regions. The purple diamond in each panel represents the variant most strongly associated with severe Covid-19 and respiratory failure.

reasonable to conclude that the chromosome 3p21.31 locus is involved in Covid-19 susceptibility per se, with a possible enrichment in patients with severe disease. This latter interpretation is supported by the significantly higher frequency of the risk allele among patients who received mechanical ventilation than among those who received supplemental oxygen only as well as by the finding of younger age among patients who were homozygous for the risk allele than among patients who were heterozygous or homozygous for the nonrisk allele.

Nongenetic studies that were reported as preprints<sup>33,34</sup> have previously implicated the involvement of ABO blood groups in Covid-19 susceptibility, and ABO blood groups have also been implicated in susceptibility to SARS-CoV-1 infection.<sup>35</sup> Our genetic data confirm that blood group O is associated with a risk of acquiring Covid-19 that was lower than that in non-O blood groups, whereas blood group A was associated with a higher risk than non-A blood groups.<sup>33,34</sup> The biologic mechanisms undergirding these findings may have to do with the ABO group per se (e.g., with the development of neutralizing antibodies against protein-linked N-glycans)<sup>36</sup> or with other biologic effects of the identified variant,<sup>37-39</sup> including the stabilization of von Willebrand factor.<sup>40,41</sup> The ABO locus holds considerable risk for population stratification,<sup>42</sup> which is increased by the inclusion of randomly selected blood donors in the current study (for which there is an inherent risk of blood group O enrichment). Alignment of the allele frequencies at the ABO

locus in our control population with those in several non-blood-donor control populations would suggest that this is not a major bias, and at least one study<sup>34</sup> that tested for association with blood type used disease controls with no affiliation to blood donors.

The pragmatic aspects leading to the feasibility of this massive undertaking in a very short period of time during the extreme clinical circumstances of the pandemic imposed limitations that will be important to explore in follow-up studies. For example, to enable the recruitment of study participants, a bare minimum of clinical metadata was requested. For this reason, extensive genotype-phenotype elaboration of current findings could not be conducted, and adjustments for all potential sources of bias (e.g., underlying cardiovascular and metabolic factors relevant to Covid-19) could not be performed. Furthermore, we have limited information about the SARS-CoV-2 infection status in the control participants; this concern is mitigated by the fact that the presence of susceptible persons in the control group would only bias the tests toward the null. In addition, few restrictions were imposed during inclusion, which led to genotyped samples having to be excluded owing to differing ethnic groups (population outliers). Further exploration of current findings, both as to their usefulness in clinical risk profiling of patients with Covid-19 and toward a mechanistic understanding of the underlying pathophysiology, is warranted.

Supported by a philanthropic donation from Stein Erik Hagen and Canica; by a grant from the Deutsche Forschungsgemeinschaft Cluster of Excellence "Precision Medicine in Chronic Inflammation" (EXC2167); by a Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico Covid-19 Biobank grant (to Dr. Valenti); by grants from the Italian Ministry of Health (RF-2016-02364358, to Dr. Valenti) and Ministero dell'Istruzione, dell'Università e della Ricerca project "Dipartimenti di Eccellenza 2018–2022" (D15D18000410001 to the Department of Medical Sciences, University of Turin; by a grant from the Spanish Ministry of Science and Innovation JdC fellowship (IJC2018-035131-I, to Dr. Acosta-Herrera); and by the GCAT Cession Research Project PI-2020-01. HLA typing was performed and supported by the Stefan-Morsch-Stiftung.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank all the patients who consented to participate in this study, and we express our condolences to the families of patients who died from Covid-19. We also thank the entire clinical staff during the outbreak situation at the different centers who were able to work on this scientific study in parallel with their clinical duties; all the members of the Humanitas Covid-19 Task Force for contributions to the recruitment of patients (see the Supplementary Notes section in Supplementary Appendix 1); Sören Brunak and Karina Banasik for discussions on the ABO associa-

tion; Goncalo Abecasis and his team for providing the Michigan imputation server; Fabrizio Bossa and Francesca Tavano for contributions to control-sample acquisition; Maria Reig for help in the case-sample acquisition; the staff of the Basque Biobank in Spain for assistance in the acquisition of samples; the staff of GCAT|Genomes for Life, a cohort study of the Genomes of

Catalonia, Institute for Health Science Research Germans Trias i Pujol, for data contribution; Alexander Eck, Jenspeter Horst, and Jens Scholz for supporting the HLA typing in the project; and the members of the ethics commissions, review boards, and consortia who fast-track reviewed our applications and enabled this rapid genetic discovery study.

## APPENDIX

The authors' full names and academic degrees are as follows: David Ellinghaus, Ph.D., Frauke Degenhardt, M.Sc., Luis Bujanda, M.D., Ph.D., Maria Buti, M.D., Ph.D., Agustín Albillos, M.D., Ph.D., Pietro Invernizzi, M.D., Ph.D., Javier Fernández, M.D., Ph.D., Daniele Prati, M.D., Guido Baselli, Ph.D., Rosanna Asselta, Ph.D., Marit M. Grimsrud, M.D., Chiara Milani, Ph.D., Fátima Aziz, B.S., Jan Käsens, Ph.D., Sandra May, Ph.D., Mareike Wendorff, M.Sc., Lars Wienbrandt, Ph.D., Florian Uellendahl-Werth, M.Sc., Tenghao Zheng, M.D., Ph.D., Xiaoli Yi, Raúl de Pablo, M.D., Ph.D., Adolfo G. Chercoles, B.S., Adriana Palom, M.S., B.S., Alba-Estela Garcia-Fernandez, B.S., Francisco Rodríguez-Frias, M.S., Ph.D., Alberto Zanella, M.D., Alessandra Bandera, M.D., Ph.D., Alessandro Protti, M.D., Alessio Aghemo, M.D., Ph.D., Ana Lleo, M.D., Ph.D., Andrea Biondi, M.D., Andrea Caballero-Garralda, M.S., Ph.D., Andrea Gori, M.D., Anja Tanck, Anna Carreras Nolla, B.S., Anna Latiano, Ph.D., Anna Ludovica Fracanzani, M.D., Anna Peschuck, Antonio Julià, Ph.D., Antonio Pesenti, M.D., Antonio Voza, M.D., David Jiménez, M.D., Ph.D., Beatriz Mateos, M.D., Ph.D., Beatriz Nafria Jimenez, B.S., Carmen Quereda, M.D., Ph.D., Cinzia Paccapelo, M.Sc., Christoph Gassner, Ph.D., Claudio Angelini, M.D., Cristina Cea, B.S., Aurora Solier, M.D., David Pestaña, M.D., Ph.D., Eduardo Muñoz-Díaz, M.D., Ph.D., Elena Sandoval, M.D., Elvezia M. Paraboschi, Ph.D., Enrique Navas, M.D., Ph.D., Félix García Sánchez, Ph.D., Ferruccio Ceriotti, M.D., Filippo Martinelli-Boneschi, M.D., Ph.D., Flora Peyvandi, M.D., Ph.D., Francesco Blasi, M.D., Ph.D., Luis Téllez, M.D., Ph.D., Albert Blanco-Grau, B.S., M.S., Georg Hemmrich-Stanisak, Ph.D., Giacomo Grasselli, M.D., Giorgio Costantino, M.D., Giulia Cardamone, Ph.D., Giulia Cardamone, Ph.D., Serena Anelli, Ph.D., Hayato Kurihara, M.D., Hesham Elabd, M.Sc., Ilaria My, M.D., Iván Galván-Femenia, M.Sc., Javier Martín, M.D., Ph.D., Jeanette Erdmann, Ph.D., Jose Ferrusquía-Acosta, M.D., Koldo Garcia-Etxebarria, Ph.D., Laura Izquierdo-Sanchez, B.S., Laura R. Bettini, M.D., Lauro Sumoy, Ph.D., Leonardo Terranova, Ph.D., Leticia Moreira, M.D., Ph.D., Luigi Santoro, M.S., Luigia Scudeller, M.D., Francisco Mesonero, M.D., Luisa Roade, M.D., Malte C. Rühlemann, Ph.D., Marco Schaefer, Ph.D., Maria Carrabba, M.D., Ph.D., Mar Riveiro-Barciela, M.D., Ph.D., Maria E. Figuera Basso, Maria G. Valsecchi, Ph.D., María Hernandez-Tejero, M.D., Marialbert Acosta-Herrera, Ph.D., Mariella D'Angiò, M.D., Marina Baldini, M.D., Marina Cazzaniga, M.D., Martin Schulzky, M.A., Maurizio Cecconi, M.D., Ph.D., Michael Wittig, M.Sc., Michele Ciccarelli, M.D., Miguel Rodríguez-Gandía, M.D., Monica Boccione, M.D., Monica Miozzo, Ph.D., Nicola Montano, M.D., Ph.D., Nicole Braun, Nicoletta Sacchi, Ph.D., Nilda Martínez, M.D., Victor Moreno, Ph.D., Tobias L. Lenz, Ph.D., Tomas Pumarola, M.D., Ph.D., Valeria Rimoldi, Ph.D., Silvano Bosari, M.D., Wolfgang Albrecht, Wolfgang Peter, Ph.D., Manuel Romero-Gómez, M.D., Ph.D., Mauro D'Amato, Ph.D., Stefano Duga, Ph.D., Jesus M. Banales, Ph.D., Johannes R. Hov, M.D., Ph.D., Trine Folseraas, M.D., Ph.D., Luca Valenti, M.D., Andre Franke, Ph.D., and Tom H. Karlsen, M.D., Ph.D.

The authors' affiliations are as follows: the Institute of Clinical Molecular Biology, Christian-Albrechts-University (D.E., F.D., J.K., S. May, M. Wendorff, L.W., F.U.-W., X.Y., A.T., A. Peschuck, C.G., G.H.-S., H.E.A., M.C.R., M.E.F.B., M. Schulzky, M. Wittig, N.B., S.J., T.W., W.A., M. D'Amato, A.F.), and University Hospital Schleswig-Holstein, Campus Kiel (N.B., A.F.), Kiel, the Institute for Cardiogenetics, University of Lübeck, Lübeck (J.E.), the German Research Center for Cardiovascular Research, partner site Hamburg–Lübeck–Kiel (J.E.), the University Heart Center Lübeck (J.E.), and the Institute of Transfusion Medicine, University Hospital Schleswig-Holstein (S.G.), Lübeck, Stefan-Morsch-Stiftung, Birkenfeld (M. Schaefer, W.P.), and the Research Group for Evolutionary Immunogenomics, Max Planck Institute for Evolutionary Biology, Plön (O.O., T.L.L.) — all in Germany; Novo Nordisk Foundation Center for Protein Research, Disease Systems Biology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen (D.E.); the Department of Liver and Gastrointestinal Diseases, Biodonostia Health Research Institute–Donostia University Hospital–University of the Basque Country (L.B., K.G.-E., L.I.-S., P.M.R., J.M.B.), Osakidetza Basque Health Service, Donostialdea Integrated Health Organization, Clinical Biochemistry Department (A.G.C., B.N.J.), and the Department of Liver and Gastrointestinal Diseases, Biodonostia Health Research Institute (M. D'Amato), San Sebastian, Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas, Instituto de Salud Carlos III (L.B., M. Buti, A. Albillos, A. Palom, F.R.-F., B.M., L. Téllez, K.G.-E., L.I.-S., F.M., L.R., M.R.-B., M. Rodríguez-Gandía, P.M.R., M. Romero-Gómez, J.M.B.), the Departments of Gastroenterology (A. Albillos, B.M., L. Téllez, F.M., M. Rodríguez-Gandía), Intensive Care (R.P., A.B.O.), Respiratory Diseases (D.J., A.S., R.N.), Infectious Diseases (C.Q., E.N.), and Anesthesiology (D. Pestaña, N. Martínez), Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria, University of Alcalá, and Histocompatibilidad y Biología Molecular, Centro de Transfusión de Madrid (F.G.S.), Madrid, the Liver Unit, Department of Internal Medicine, Hospital Universitari Vall d'Hebron, Vall d'Hebron Barcelona Hospital Campus (M. Buti, A. Palom, L.R., M.R.-B.), Hospital Clinic, University of Barcelona, and the August Pi i Sunyer Biomedical Research Institute (J.F., F.A., E.S., J.F.-A., L.M., M.H.-T., P.C.), the European Foundation for the Study of Chronic Liver Failure (J.F.), Vall d'Hebron Institut de Recerca (A. Palom, F.R.-F., A.J., S. Marsal), and the Departments of Biochemistry (A.-E.G.-F., F.R.-F., A.C.-G., C.C., A.B.-G.), Intensive Care (R.F.), and Microbiology (T.P.), University Hospital Vall d'Hebron, the Immunohematology Department, Banc de Sang i Teixits, Autonomous University of Barcelona (E.M.-D.), Catalan Institute of Oncology, Bellvitge Biomedical Research Institute, Consortium for Biomedical Research in Epidemiology and Public Health and University of Barcelona, l'Hospitalet (V. Moreno), and Autònoma University of Barcelona (T.P.), Barcelona, Universitat Autònoma de Barcelona, Bellaterra (M. Buti, F.R.-F., M.R.-B.), GenomesForLife–GCAT Lab Group, Germans Trias i Pujol Research Institute (A.C.N., I.G.-F., R.C.), and High Content Genomics and Bioinformatics Unit, Germans Trias i Pujol Research Institute (L. Sumoy), Badalona, Institute of Parasitology and Biomedicine Lopez-Neyra, Granada (J.M., M.A.-H.), the Digestive Diseases Unit, Virgen del Rocío University Hospital, Institute of Biomedicine of Seville, University of Seville, Seville (M. Romero-Gómez), and Ikerbasque, Basque Foundation for Science, Bilbao (M. D'Amato, J.M.B.) — all in Spain; the Division of Gastroenterology, Center for Autoimmune Liver Diseases, Department of Medicine and Surgery, University of Milan Bicocca (P.I., C.M.),

Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico (D. Prati, G.B., A.Z., A. Bandera, A.G., A.L.F., A. Pesenti, C.P., F.C., F.M.-B., F.P., F.B., G.G., G. Costantino, L. Terranova, L. Santoro, L. Scudeller, M. Carrabba, M. Baldini, M.M., N. Montano, R.G., S.P., S. Aliberti, V. Monzani, S. Bosari, L.V.), the Department of Biomedical Sciences, Humanitas University (R.A., A. Protti, A. Aghemo, A. Leo, E.M.P., G. Cardamone, M. Cecconi, V.R., S.D.), Humanitas Clinical and Research Center, IRCCS (R.A., A. Protti, A. Aghemo, A. Leo, A.V., C.A., E.M.P., H.K., I.M., M. Cecconi, M. Ciccirelli, M. Boccione, P.P., P.O., P.T., S. Badalamenti, S.D.), University of Milan (A.Z., A. Bandera, A.G., A.L.F., A. Pesenti, F.M.-B., F.P., F.B., G.G., G. Costantino, M.M., N. Montano, R.G., S.P., S. Aliberti, S. Bosari, L.V.), and the Center of Bioinformatics, Biostatistics, and Bioimaging (M.G.V.) and the Phase 1 Research Center (M. Cazzaniga), School of Medicine and Surgery, and the Departments of Emergency, Anesthesia, and Intensive Care (G.F.), Pneumologia (P.F.), and Infectious Diseases (P.B.); University of Milano-Bicocca, Milan, the European Reference Network on Hepatological Diseases (P.L., C.M.) and the Infectious Diseases Unit (P.B.), San Gerardo Hospital, Monza, the Pediatric Department and Centro Tetamanti-European Reference Network PaedCan, EuroBloodNet, MetabERN—University of Milano-Bicocca—Fondazione MBBM—Ospedale, San Gerardo (A. Biondi, L.R.B., M. D'Angiò), the Gastroenterology Unit, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (A. Latiano, O.P.), the Department of Medical Sciences, Università degli Studi di Torino, Turin (S. Aneli, G.M.), and the Italian Bone Marrow Donor Registry, E.O. Ospedali Galliera, Genoa (N.S.) — all in Italy; the Norwegian PSC Research Center, Department of Transplantation Medicine, Division of Surgery, Inflammatory Diseases, and Transplantation, and the Research Institute for Internal Medicine, Division of Surgery, Inflammatory Diseases, and Transplantation, Oslo University Hospital Rikshospitalet and University of Oslo (M.M.G., J.R.H., T.F., T.H.K.), and the Section for Gastroenterology, Department of Transplantation Medicine, Division for Cancer Medicine, Surgery, and Transplantation, Oslo University Hospital Rikshospitalet (J.R.H., T.F., T.H.K.), Oslo; the School of Biological Sciences, Monash University, Clayton, VIC, Australia (T.Z., M. D'Amato); Private University in the Principality of Liechtenstein (C.G.); the Institute of Biotechnology, Vilnius University, Vilnius, Lithuania (S.J.); and the Unit of Clinical Epidemiology, Department of Medicine Solna, Karolinska Institutet, Stockholm (M. D'Amato).

## REFERENCES

1. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020;382:727-33.
2. Dong E, Du H, Gardner L. An interactive Web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020;20:533-4.
3. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 2020;5:536-44.
4. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* 2020 February 24 (Epub ahead of print).
5. Berlin DA, Gulick RM, Martinez FJ. Severe Covid-19. *N Engl J Med*. DOI: 10.1056/NEJMcip2009575.
6. Marini JJ, Gattinoni L. Management of COVID-19 respiratory distress. *JAMA* 2020 April 24 (Epub ahead of print).
7. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395:1054-62.
8. Li X, Xu S, Yu M, et al. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. *J Allergy Clin Immunol* 2020 April 12 (Epub ahead of print).
9. Chen R, Liang W, Jiang M, et al. Risk factors of fatal outcome in hospitalized subjects with coronavirus disease 2019 from a nationwide analysis in China. *Chest* 2020 April 15 (Epub ahead of print).
10. Docherty AB, Harrison EM, Green CA, et al. Features of 20133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* 2020;369:m1985.
11. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. *JAMA* 2020;323(20):2052-9.
12. Levi M, Thachil J, Iba T, Levy JH. Coagulation abnormalities and thrombosis in patients with COVID-19. *Lancet Haematol* 2020;7(6):e438-e440.
13. Varga Z, Flammer AJ, Steiger P, et al. Endothelial cell infection and endothelitis in COVID-19. *Lancet* 2020;395:1417-8.
14. Ackermann M, Verleden SE, Kuehnel M, et al. Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in Covid-19. *N Engl J Med*. DOI: 10.1056/NEJMoa2015432.
15. Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010;42:1118-25.
16. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. March 6, 2019 (<https://www.biorxiv.org/content/10.1101/563866v1>). preprint.
17. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;4:7.
18. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* 2010;26:2190-1.
19. Bugert P, Rink G, Kemp K, Klüter H. Blood group ABO genotyping in paternity testing. *Transfus Med Hemother* 2012;39:182-6.
20. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA database. *Nucleic Acids Res* 2020;48(D1):D948-D955.
21. Dubois PC, Trynka G, Franke L, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 2010;42:295-302.
22. Benthall J, Morris DL, Graham DSC, et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet* 2015;47:1457-64.
23. Myocardial Infarction Genetics Consortium. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 2009;41:334-41.
24. Julià A, González I, Fernández-Nebro A, et al. A genome-wide association study identifies SLC8A3 as a susceptibility locus for ACPA-positive rheumatoid arthritis. *Rheumatology (Oxford)* 2016;55:1106-11.
25. López-Isac E, Acosta-Herrera M, Kerick M, et al. GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nat Commun* 2019;10:4955.
26. Obón-Santacana M, Vilardell M, Carreras A, et al. GCAT[Genomes for life: a prospective cohort study of the genomes of Catalonia. *BMJ Open* 2018;8(3):e018324.
27. Galván-Femenía I, Obón-Santacana M, Piñeyro D, et al. Multitrait genome association analysis identifies new susceptibility genes for human anthropometric variation in the GCAT cohort. *J Med Genet* 2018;55:765-78.
28. Vuille-dit-Bille RN, Camargo SM, Emmenegger L, et al. Human intestine luminal ACE2 and amino acid transporter



- expression increased by ACE-inhibitors. *Amino Acids* 2015;47:693-705.
29. Kuba K, Imai Y, Ohto-Nakanishi T, Penninger JM. Trilogy of ACE2: a peptidase in the renin-angiotensin system, a SARS receptor, and a partner for amino acid transporters. *Pharmacol Ther* 2010; 128:119-28.
  30. Wein AN, McMaster SR, Takamura S, et al. CXCR6 regulates localization of tissue-resident memory CD8 T cells to the airways. *J Exp Med* 2019;216:2748-62.
  31. Hickey MJ, Held KS, Baum E, Gao JL, Murphy PM, Lane TE. CCR1 deficiency increases susceptibility to fatal coronavirus infection of the central nervous system. *Viral Immunol* 2007;20:599-608.
  32. COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur J Hum Genet* 2020;28:715-8.
  33. Zhao J, Yang Y, Huang H, et al. Relationship between the ABO blood group and the COVID-19 susceptibility. March 27, 2020 (<https://www.medrxiv.org/content/10.1101/2020.03.11.20031096v2>). preprint.
  34. Zietz M, Tatonetti NP. Testing the association between blood type and COVID-19 infection, intubation, and death. April 11, 2020 (<https://www.medrxiv.org/content/10.1101/2020.04.08.20058073v1>). preprint.
  35. Cheng Y, Cheng G, Chui CH, et al. ABO blood group and susceptibility to severe acute respiratory syndrome. *JAMA* 2005;293:1450-1.
  36. Breiman A, Ruvën-Clouet N, Le Pendu J. Harnessing the natural anti-glycan immune response to limit the transmission of enveloped viruses such as SARS-CoV-2. *PLoS Pathog* 2020;16(5):e1008556.
  37. Comuzzie AG, Cole SA, Laston SL, et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS One* 2012; 7(12):e51954.
  38. Aziz M, Fatima R, Assaly R. Elevated interleukin-6 and severe COVID-19: a meta-analysis. *J Med Virol* 2020 April 28 (Epub ahead of print).
  39. Naitza S, Porcu E, Steri M, et al. A genome-wide association scan on the levels of markers of inflammation in Sardinians reveals associations that underpin its complex regulation. *PLoS Genet* 2012; 8(1):e1002480.
  40. Franchini M, Crestani S, Frattini E, Sissa C, Bonfanti C. ABO blood group and von Willebrand factor: biological implications. *Clin Chem Lab Med* 2014;52:1273-6.
  41. Murray GP, Post SR, Post GR. ABO blood group is a determinant of von Willebrand factor protein levels in human pulmonary endothelial cells. *J Clin Pathol* 2020;73:347-9.
  42. Thomson G, Bodmer WF. Letter: population stratification as an explanation of IQ and ABO association. *Nature* 1975; 254:363-4.

Copyright © 2020 Massachusetts Medical Society.

RECEIVE IMMEDIATE NOTIFICATION WHEN AN ARTICLE  
IS PUBLISHED ONLINE FIRST

To be notified by email when *Journal* articles  
are published online first, sign up at NEJM.org.