Novel genes and sex differences in COVID-19 severity

Raquel Cruz1,2,3,4,5, Silvia Díez-de Almeida4,5, Miguel López de Heredia2, Inés Quintela1, Francisco C. Ceballos5, Guillermo Pita6, José M. Lorenzo-Salazar7, Rafaela González-Montelongo2, Manuela Gago-Domínguez3,8, Marta Sevilla Porras2,9, Jair Antonio Tenorio Cañasto9,10, Julian Nevado9,10, Jose Maria Aguado11,12,13,14, Carlos Aguilar15, Sergio Aguilera-Albesa16,17, Virginia Almadana18, Berta Almoguer19,2, Nuria Álvarez18, Álvaro Andreu-Bernabeu20,13, Eunate Arana-Arri21,22, Celso Arango20,23,13, María J. Arranz24, María-Jesus Artiga25, Raúl C. Baptista-Rosas26,27,28, María Barreda-Sánchez29,30, Monce Belfhassen-Garcia31,32, Joao F. Bezerra33, Marcos A.C. Bezerra34, Lucía Boix-Palop35, María Brion36,37, Ramón Brugada38,39,37,40, Matilde Bustos41, Enrique J. Calderón42,43,44, Cristina Carbonell45,46, Luis Castano41,46,47,48, Jose E. Castelao49, Rosa Conde-Vicente50, M. Lourdes Cordero-Lorenzana51, Jose L. Cortes-Sanchez52,53, Marta Corton19,2, M. Teresa Darnaude54, Alba De Martino-Rodriguez55,56, Víctor del Campo-Pérez27, Aranzazu Díaz de Bustamante54, Elena Domínguez-Garrido48, Andre D. Luchessi59, Rocío Eiros60, Gladys Mercedes Estigarribia Sanabria61, María Carmen Fariñas62,63,64, Uxía Fernández-Robelo65, Amanda Fernández-Rodriguez2,14, Tania Fernández-Villa66, Belén Gil-Fournier67, Javier Gómez-Arrue55,56, Beatriz González Álvarez55,56, Fernan Gonzalez Bernardo de Quiros68, Javier González-Peñas20,13,23, Juan F. Gutiérrez-Bautista69, María José Herrero70,71, Antonio Herrero-Gonzalez72, María A. Jimenez-Sousa14, Maria Claudia Lattig7,74, Anabel Liger Borja75, Rosario Lopez-Rodriguez19,2, Esther Mancheño6,77, Caridad Martín-López75, Vicente Martín8,43, Oscar Martinez-Nieto79,74, Iciar Martinez-Lopez80,81, Michel F. Martinez-Resendes52, Angel Martinez-Perez82, Juliana F. Mazzeu63,84,85, Eleutério Merayo Macías86, Pablo Mínguez19,2, Victor Moreno Cuerda87,88, Vivian N. Silbiger59, Silviene F. Oliveira83,89,90,91,92, Eva Ortega-Paino25, Mara Parellada20,23,23, Estela Paz-Artal76,77,79,91, Ney P.C. Santos94, Patricia Pérez-Matute95, Patricia Perez96, M. Elena Pérez-Tomás29, Teresa Peruchó87, Mel Lina Pinsach-Abuin38,37, Ericka N. Pompa-Mera98, Gloria L. Porras-Hurtado99, Aurora Pujo100,2,101, Soraya Ramiro León69, Salvador Resino5,14, Marianne R. Fernandes94,102, Emilio Rodriguez-Ruiz103,3, Fernando Rodríguez-Artalejo104,105,43,106, José A. Rodriguez-Garcia107, Francisco Ruiz Cabello69,108,109, Javier Ruiz-Hornillos101,110,111,112, Pablo Ryan113,114,115, José Manuel Soria82, Juan Carlos Souto116, Eduardo Tamayo117,118, Alvaro Tamayo-Velasco119, Juan Carlos Taracido-Fernandez72, Alejandro Teper120, Lilian Torres-Tobar121, Miguel Urioste122, Juan Valencia-Ramos123, Zuleima Vázquez124, Ruth Zarate125, Tomoko Nakanishi126,127,128,129,130, Sara Pigazzini131,132, Frauke Degenhardt133,134, Guillaume Butler-Laporte135,138, Douglas Maya-Miles136,137, Luis Bujanda138,137, Youssef Bouysran139, Adriana Palom140,141,142, David Ellinghaus143,144, Manuel Martinez-Bueno144, Selina Rolker145, Sara Amirian146, Luisa Roade137,140,147, Francesca Fava148,149,150, Christoph D. Spinner149,151, Daniele Prati152, David Bernard153,154, Federico García154,155, Guillermo Darias156,157, Israel Fernández-Cadenas158, Jan Cato Holter159,160, Jesus M. Banales161,163, Robert Frithiof162, Stefano Duga163,164, Rosanna Asselts165,166, Alexandre C. Pereira165, Manuel Romero-Gómez166, Beatriz Nafria-Jiménez166, Johannes R. Hou167,160,168, Isabelle Migeotte169,139, Alessandra Renieri170,171, Anna M. Planas170,171, Kerstin U. Ludwig172,144, Maria Buti173,140,147, Souad Rahmouni175, Marta E. Alarcón-Riquelme176,177, Eva C. Schulte173,174,175, Andre Franke133,134, Tom H. Karlsen167,160,176, Luca Valenti177,178, Hugo Zeberg179,180, Brent Richards181,128,182, Andrea Ganna132,183,3,137, Mercè Boada184,185, Itziar de Rojas184,185, Agustín Ruiz184,185, Pascual Sánchez-Juan186, Luis Miguel Rea187, SCOURGE Cohort Group*, HOSTAGE Cohort Group*, GRAÇE Cohort Group*, Encarna Guillen-Navarro29,188,189,190, Carmen Ayuso19,2, Ana González-Neira6, José A. Riancho62,63,64, Augusto Rojas-Martinez191,2, Carlos Flores1,172,193,194,1, Pablo Lapunzina2,9,10,11 and Angel Carracedo1,2,3,4,8,1.

1Centro Nacional de Genotipado (CEGEN), Universidade de Santiago de Compostela, 15706 Santiago de Compostela, Spain
2Centre for Biomedical Network Research on Rare Diseases (CIBERER), Instituto de Salud Carlos III, 28029 Madrid, Spain
3Instituto de Investigación Sanitaria de Santiago (IDIS), 15706 Santiago de Compostela, Spain

Received: February 15, 2022. Revised: May 16, 2022. Accepted: June 1, 2022
© The Author(s) 2022. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1093/hmg/ddac132
Advance access publication date 16 June 2022
Original Article
Here, we describe the results of a genome-wide study conducted in 11,939 coronavirus disease 2019 (COVID-19) positive cases with an extensive clinical information that were recruited from 34 hospitals across Spain (SCOURGE consortium). In sex-disaggregated genome-wide association studies for COVID-19 hospitalization, genome-wide significance \( (P < 5 \times 10^{-8}) \) was crossed for variants in 3p21.31 and 21q22.11 loci only among males \( (P = 1.3 \times 10^{-22}) \) and \( P = 8.1 \times 10^{-12} \), respectively, and for variants in 9q21.32 near TLE1 only among females \( (P = 4.4 \times 10^{-8}) \). In a second phase, results were combined with an independent Spanish cohort \((15,986 \text{ COVID-19 cases and } 1068 \text{ population controls})\), revealing in the overall analysis two novel risk loci in 9p13.3 and 19q13.12, with fine-mapping prioritized variants functionally associated with \( AQP3 (P = 2.7 \times 10^{-8}) \) and \( ARHGAP33 (P = 1.3 \times 10^{-8}) \), respectively. The meta-analysis of both phases with four European studies stratified by sex from the Host Genetics Initiative (HGI) confirmed the association of the 3p21.31 and 21q22.11 loci predominantly in males and replicated a recently reported variant in 11p13 \( (P = 4.1 \times 10^{-8}) \). Six of the COVID-19 HGI discovered loci were replicated and an HGI-based genetic risk score predicted the severity strata in SCOURGE. We also found more SNP-heritability and larger heritability differences by age (<60 or >60 years) among males than among females. Parallel genome-wide screening of inbreeding depression in SCOURGE also showed an effect of homozygosity in COVID-19 hospitalization and severity and this effect was stronger among older males. In summary, new candidate genes for COVID-19 severity and evidence supporting genetic disparities among sexes were identified. 

**Abstract**

We describe the results of a genome-wide study conducted in 11,939 COVID-19 positive cases with an extensive clinical information that were recruited from 34 hospitals across Spain (SCOURGE consortium). In sex-disaggregated genome-wide association studies for COVID-19 hospitalization, genome-wide significance \( (P < 5 \times 10^{-8}) \) was crossed for variants in 3p21.31 and 21q22.11 loci only among males \( (P = 1.3 \times 10^{-22}) \) and \( P = 8.1 \times 10^{-12} \), respectively, and for variants in 9q21.32 near TLE1 only among females \( (P = 4.4 \times 10^{-8}) \). In a second phase, results were combined with an independent Spanish cohort \((15,986 \text{ COVID-19 cases and } 1068 \text{ population controls})\), revealing in the overall analysis two novel risk loci in 9p13.3 and 19q13.12, with fine-mapping prioritized variants functionally associated with \( AQP3 (P = 2.7 \times 10^{-8}) \) and \( ARHGAP33 (P = 1.3 \times 10^{-8}) \), respectively. The meta-analysis of both phases with four European studies stratified by sex from the Host Genetics Initiative (HGI) confirmed the association of the 3p21.31 and 21q22.11 loci predominantly in males and replicated a recently reported variant in 11p13 \( (P = 4.1 \times 10^{-8}) \). Six of the COVID-19 HGI discovered loci were replicated and an HGI-based genetic risk score predicted the severity strata in SCOURGE. We also found more SNP-heritability and larger heritability differences by age (<60 or >60 years) among males than among females. Parallel genome-wide screening of inbreeding depression in SCOURGE also showed an effect of homozygosity in COVID-19 hospitalization and severity and this effect was stronger among older males. In summary, new candidate genes for COVID-19 severity and evidence supporting genetic disparities among sexes were identified.
Introduction

Coronavirus disease 2019 (COVID-19)—caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)—develops with wide clinical variability, ranging from asymptomatic infection to a life-threatening condition (1). Advanced age and the presence of comorbidities are well-known major risk factors of COVID-19 severity (2,3). In addition, male sex is another risk factor associated with COVID-19 severity, regardless of comorbidities (4).

International genetic studies based on genome-wide association studies (GWAS) and/or comparative genome sequencing analyses have been conducted to identify genetic variants associated with COVID-19 severity (5,6). These studies have revealed the role of genes of the type-I interferon (IFN) signaling pathway as key players underlying disease severity (7–9). Besides, they have also identified other potential loci previously linked to lung function, respiratory diseases and autoimmunity (9). Regarding COVID-19 severity in males, sex-disaggregated genetic analyses have received limited attention despite the importance of sex disparities in clinical severity (10). Early studies suggested immune deficits presumably because of pre-existing neutralizing autoantibodies against type-I IFN in older male patients (11).

The effects of autozygosity, measured as the change of the mean value of a complex trait because of inbreeding, have been useful to identify alternative genetic risk explanations and effects that traditionally are not captured by GWAS (12). By analyzing the contribution of the inbreeding depression (ID) through the lens of the runs of homozygosity (ROH: genomic tracts where homozygous markers occur in an uninterrupted sequence), it is possible to assess the importance of directional dominance or overdominance in the genetic architecture of complex traits (13). Even though this is a relatively modern approach, different studies have shown the importance of homozygosity in a large range of complex phenotypes, including anthropometric, cardiometabolic and mental traits (14–16).

Through diverse nested sub-studies, the Spanish Coalition to Unlock Research on Host Genetics on COVID-19 (SCOURGE) consortium was launched in May 2020 aiming to find biomarkers of evolution and prognosis that can have an immediate impact on clinical management and therapeutic decisions in SARS-CoV-2 infections. This consortium has recruited patients from hospitals across Spain and Latin America in close collaboration with the STOP-Coronavirus initiative (https://www.scourge-covid.org). Here, we describe the results of the first SCOURGE genome-wide studies of COVID-19 conducted in patients recruited in Spain. This dataset has not been used in any previous GWAS of COVID-19 that has been published to date. To the best of our knowledge, this is the first time that the impact of homozygosity is considered in COVID-19 studies, serving as a complement to the traditional GWAS to assess both the additive and dominant components of the genetic architecture of COVID-19 severity. Likewise, the ID analysis could also help to explain the strong effect of age in COVID-19 severity.

Results

Discovery phase

In the SCOURGE study, 11 939 COVID-19 positive cases were recruited from 34 centers (Supplementary Material, Table S1) between March and December 2020. All diagnosed cases were classified in a five-level severity scale (Table 1). Two untested Spanish sample collections were used as general population controls in some analyses: 3437 samples from the Spanish DNA biobank (https://www.bancoadn.org) and 2506 samples from the GR@CE consortium (17). The discovery phase samples were genotyped with the Axiom Spain Biobank Array (Thermo Fisher Scientific). Details of quality control (QC), ancestry inference and imputation are shown in the Materials and Methods section. Individuals with inferred European ancestry were used for association testing. After post-imputation filtering, 15 045 individuals (9371 COVID-19 positive cases and 5674 population controls) and 8 933 154 genetic markers remained in the SCOURGE European study (discovery). Clinical and demographic characteristics of European patients from SCOURGE included in the analysis are shown in Table 2. Population controls were 46.3% females with a mean age of 55.5 years (standard deviation, SD = 16.2) and 53.7% males, with a mean age of 51 years (SD = 13.04).

The discovery phase of the GWAS was carried out with infection susceptibility and three severity outcomes (hospitalization, severe illness and critical illness), which were tested using three different control definitions (see Supplementary Material, Table S2).

- A1 analysis: COVID-19 positive not satisfying the case condition and control samples from the general population (COVID-19 untested).
- A2 analysis: control samples from the general population.
- C analysis: COVID-19 positive not satisfying the case condition.

The GWAS was carried on by fitting logistic mixed regression models adjusting for age, sex and the first 10 principal components (PCs; see Materials and Methods). Summary statistics can be accessed from https://github.com/CIBERER/Scourge-COVID19. The SCOURGE Board of Directors has agreed to aggregate the GWAS summaries with those from the COVID-19 Host Genetics Initiative (HGI) in the data freeze 7 that has not been used for any published article to date. Supplementary Material, Table S3 shows the independent significant associated loci for hospitalization, severity, critical illness and infection susceptibility risk, for global and sex-stratified analysis in the SCOURGE dataset. However, considering the overlap between the findings for these analyses, only the main results for the A1 analysis are presented.
Table 1. Five-level severity scale used to classify SCOURGE patients

| Level                  | Clinical findings                                                                 |
|------------------------|------------------------------------------------------------------------------------|
| Severity 0 (asymptomatic) | Asymptomatic                                                                        |
| Severity 1 (mild)      | With symptoms, but without pulmonary infiltrates or need of oxygen therapy         |
| Severity 2 (moderate)  | With pulmonary infiltrates affecting <50% of the lungs or need of supplemental oxygen therapy |
| Severity 3 (severe)    | Hospitalized with any of the following criteria:                                     |
|                        | • PaO₂ < 65 mmHg or SaO₂ < 90%                                                  |
|                        | • PaO₂/FiO₂ < 300                                                               |
|                        | • SaO₂/FiO₂ < 440                                                              |
|                        | • Dyspnea                                                                       |
|                        | • Respiratory frequency ≥ 22 bpm                                                  |
|                        | • Infiltrates affecting > 50% of the lungs                                        |
| Severity 4 (critical)  | Admission to the ICU or need of mechanical ventilation (invasive or non-invasive) |

Note: PaO₂, partial pressure of oxygen in arterial blood; SaO₂, saturation of oxygen in arterial blood; FiO₂, fraction of inspired oxygen.

Table 2. Baseline characteristics of European patients from SCOURGE included in the analysis

| Variable                        | Global N = 9371 | Males N = 4343 | Females N = 5028 |
|---------------------------------|-----------------|----------------|------------------|
| Age—mean years (SD)             | 62.6 (17.9)     | 64.3 (16.3)    | 61.1 (19.1)      |
| Severity—N (%)                  |                 |                |                  |
| 0—asymptomatic                  | 582 (6.6)       | 161 (3.9)      | 421 (8.9)        |
| 1—mild                          | 2689 (30.3)     | 748 (18.2)     | 1941 (40.8)      |
| 2—intermediate                  | 2099 (23.6)     | 1093 (26.5)    | 1006 (21.1)      |
| 3—severe                        | 2379 (26.8)     | 1300 (31.6)    | 1079 (22.7)      |
| 4—critical illness              | 1128 (12.7)     | 817 (19.8)     | 311 (6.5)        |
| Hospitalization—N (%)           | 5966 (63.8)     | 3436 (79.3)    | 2530 (50.5)      |
| Severe COVID—N (%)              | 3507 (39.2)     | 2117 (51.2)    | 1390 (28.9)      |
| Critical illness—N (%)          | 1128 (12.6)     | 817 (19.8)     | 311 (6.5)        |
| Comorbidities—N (%)             |                 |                |                  |
| Vascular/endoocrinal            | 4099 (43.7)     | 2207 (50.8)    | 1892 (37.6)      |
| Cardiac                         | 1057 (11.3)     | 634 (14.6)     | 423 (8.4)        |
| Nervous                         | 773 (8.3)       | 341 (7.9)      | 432 (8.6)        |
| Digestive                       | 264 (2.8)       | 153 (3.5)      | 111 (2.2)        |
| Onco-hematological             | 647 (6.9)       | 411 (9.5)      | 236 (4.7)        |
| Respiratory                     | 905 (9.7)       | 565 (13.0)     | 340 (6.8)        |

All analyses support the association of two known loci, i.e. 3p21.31 and 21q22.11. However, other suggestive associations were also found (Fig. 1 and Supplementary Material, Fig. S1). Strikingly, the leading signals found in the global (sex-aggregated) analysis were genome-wide significant in the analyses among males but not among females. Association in 3p21.31 was also found in the C analyses (rs10490770, \( P = 3.8 \times 10^{-12} \)) and once again, association was genome-wide significant only among males (\( P = 3.9 \times 10^{-9} \), females: \( P = 4.6 \times 10^{-5} \)). However, the leading variant of 9q21.32 (near TLE1 gene) reached genome-wide significance among females only (similarly, in the C analysis for females, rs140152223, \( P = 2.11 \times 10^{-9} \)). Several variants (rs17763742 near LZTFL1, rs2834164 in IFNAR2 and rs1826292621 near TLE1) showed a significant difference in effect sizes (SNP×sex interaction \( P < 0.0031 \), adjusted probability for 16 comparisons) linked not only to hospitalization, but also to critical illness and infection risk. The A2 and C analyses did not reveal any additional significant loci (Supplementary Material, Fig. S2). Although fine-mapping studies in 3p21.31 and 21q22.11 have led to gene and variant prioritization within these loci (Supplementary Material, Fig. S3), a Bayesian fine-mapping on the 9q21.32 did not allow to prioritize variants by their role as expression quantitative trait loci (eQTLs) or anticipate the function (Fig. 2). To assess if a higher prevalence of comorbidities in males may underlie these differential findings, we performed an additional C analysis in which the presence of comorbidities was tested for association within hospitalized patients. No significant association was found in either males or females (Supplementary Material, Fig. S4). Further explorations of the genetic associations with comorbidities are presented in the Supplementary Note.

This GWAS phase was also performed disaggregated by age (<60/>60 years old), and by age and sex simultaneously. Differences in effect sizes between both age groups were tested for the SNPs shown in the Supplementary Material, Table S3, in global and sex-specific analysis (Supplementary Material, Table S4). Significant findings were only found in the subgroup of males with <60 years old. This was also found in the C analysis for hospitalization where association in 3p21.31 was significant only in males <60 years old (\( P = 3.32 \times 10^{-9} \)). Differences in effect size (significant age-interaction) were significant
Figure 1. Association results of SCOURGE for global and sex-disaggregated A1 hospitalization analysis. (A) Manhattan plot of results from global analysis. A quantile–quantile plot of the global analysis is also shown as an inset. (B) Miami plot of results from sex-disaggregated analyses (top: males and bottom: females).

Figure 2. Regional plot of a novel association at 9q21.32 found among females from the SCOURGE study. The x axis reflects the chromosomal position, and the y axis the $-\log(P\text{-value})$. The sentinel variant is indicated by a diamond and all other variants are colour coded by their degree of LD with the sentinel variant in Europeans. Credible set for this signal is shown within a dashed square. The horizontal dotted blue line corresponds to the threshold for genome-wide significance ($P = 5 \times 10^{-8}$).
at 3p21.31 for severity and critical illness, and suggestive in hospitalization.

**Lookup of previously found COVID-19 host risk factors in the SCOURGE study**

Known significant loci for COVID-19 severity in 3p21.31 (near SLC6A20 and LZTFL1) and 21q22.11 (in IFNAR2) were clearly replicated at genome-wide significance in this study, specifically for risk of infection, hospitalization and severity. Three other loci, in 9q34.2 (in ABO), 12q24.13 (in OAS1) and 19p13.2 (near RAVER1 and TYK2), did not reach the genome-wide significance threshold but they were significant after correcting for the 390 tests performed in a lookup (13 SNPs and 30 analyses, significance threshold $P < 1.3 \times 10^{-4}$). In agreement with previous results, ABO was mainly associated with the risk of infection. However, other loci as those in 3q12.3 (near RPL24) and 19p13.3 (near DPP9), previously found associated with COVID-19 severity, were not replicated in the SCOURGE Europeans. The complete list of results for the list of COVID-19 HGI significant loci (9) is shown in Figure 3 and in the Supplementary Material, Table S5. Figure 3 also reinforces the clear sex differences found in this study.

**Genetic risk score and the COVID-19 severity scale**

We developed a genetic risk score (GRS) combining the 13 leading variants associated with infection risk, hospitalization or severity in the meta-analysis performed by the COVID-19 HGI (9). This GRS predicted the severity scale in SCOURGE but supporting the differentiation in three classes: (i) controls/asymptomatic/mild cases; (ii) moderate and severe cases and (iii) critical cases. (Supplementary Material, Fig. S5). Simultaneously disaggregating by age (<60/≥60 years old) and sex, we identify the three severity classes in the subgroup of males with <60 years old, supporting the importance of this group in the overall findings (Supplementary Material, Fig. S5). Details of this analysis can be found in Supplementary Note.

**Second study phase and meta-analysis with the discovery**

Results for hospitalization risk were meta-analysed with a second Spanish cohort, the CNIO study (see Materials and Methods). This study was filtered following the same QC and imputation procedures. The final dataset of the CNIO study included 2446 European individuals (1378 COVID-19 positive cases and 1068 population controls) and 8 895 721 markers. Table 3 shows the results that were genome-wide significant either in global or sex-stratified meta-analysis with SCOURGE. Besides the widely replicated loci at 3p21.31 and 21q22.11, three additional signals were found: chr9:33426577:A:G (intergenic to AQP7 and AQP3), chr17:45422978:G:C (intronic to ARHGAP27) and chr19:35687796:G:A (intergenic to UPK1A and ZBTB32). Bayesian fine-mapping analysis around chr17:45422978:G:C failed to prioritize a credible set of variants, hindering functional links of the locus. Functional assessments of the prioritized variants by the Bayesian fine-mapping analysis in the other two regions supported that they were eQTLs of the AQP3 and ARHGAP33 genes, including whole blood and lung tissues (Fig. 4).

These variants were also associated with the three severity groups previously outlined in SCOURGE by the
Table 3. Genome-wide significant variants in global or sex-stratified meta-analysis between the SCOURGE and CNIO studies

| SNP          | chr:position | Meta-ALL | Meta-males | Meta-females | Nearest gene |
|--------------|--------------|----------|------------|--------------|--------------|
| rs115679256  | 3:45587795   | G        | A          | 0.43         | 0.08         | 1.1E–08    |
| rs17763742   | 3:45805277   | G        | A          | 0.60         | 0.05         | 4.1E–29    |
| rs35477280   | 3:45932600   | G        | A          | 0.39         | 0.05         | 1.4E–17    |
| rs14443214   | 3:46136372   | T        | C          | 0.25         | 0.04         | 9.0E–09    |
| rs115102354  | 3:46180545   | A        | G          | 0.41         | 0.07         | 1.6E–08    |
| rs10818937   | 9:33426577   | A        | G          | 0.18         | 0.03         | 2.7E–08    |
| rs1230082    | 17:4542978   | C        | G          | 0.16         | 0.03         | 2.1E–08    |
| rs77127536   | 19:35687796  | G        | A          | −0.22        | 0.04         | 1.3E–08    |
| rs17860169   | 21:33240996  | A        | G          | 0.19         | 0.03         | 2.3E–11    |

Note: Representitive SNPs were selected using the clump function of PLINK 1.9 (clumping parameters \( r^2 = 0.5 \), \( 5 \times 10^{-8} \) and \( p = 0.05 \)). EA, effect allele; NEA, non-effect allele; beta, effect coefficient; SE, standard error.

Table 4. Results of European meta-analysis for hospitalization risk

| SNP          | chr:position | Meta-all | Meta-males | Meta-females | Nearest gene |
|--------------|--------------|----------|------------|--------------|--------------|
| rs115679256  | 3:45587795   | G        | A          | 0.37         | 0.06         | 1.3E–08    |
| rs13078854   | 3:45820440   | G        | A          | 0.53         | 0.04         | 6.7E–34    |
| rs14289622   | 3:4593053    | T        | G          | 0.36         | 0.04         | 3.6E–21    |
| rs115102354  | 3:46180545   | A        | G          | 0.40         | 0.06         | 8.9E–12    |
| rs61882275   | 11:34482745  | G        | A          | −0.12        | 0.02         | 1.0E–06    |
| rs47607082   | 12:11292138  | A        | G          | −0.16        | 0.02         | 1.6E–10    |
| rs12609134   | 19:35687796  | G        | A          | −0.19        | 0.03         | 2.3E–08    |
| rs17860169   | 21:33240996  | A        | G          | 0.18         | 0.03         | 3.9E–12    |

Note: Summary statistics of both phases (SCOURGE and CNIO) were meta-analysed with four additional sex-disaggregated European studies from the COVID-19 HGI consortium. EA, effect allele; NEA, non-effect allele; beta, effect coefficient; SE, standard error.

GRS under a multinomial model (Supplementary Material, Table S6).

Meta-analysis in independent European studies

Hospitalization risk was meta-analysed with other European studies combining both Spanish cohorts (SCOURGE and CNIO) with other four sex-disaggregated studies from the COVID-19 HGI consortium, namely: BelCOVID, GenCOVID, Hostage-Spain and Hostage-Italy (Table 4). Once again, the most outstanding significant loci were found at 3p21.31 and 21q22.11 (in global and male-specific analyses), and three additional loci reached genome-wide significance in the meta-analysis of males: chr12:11292383:A-G (in OAS1), chr19:35687796:G-A (intergenic to UPK1A and ZBTB32) and chr11:34482745:G-A (in ELF5). The 3p21.31 variants reached genome-wide significance in females, although at significantly lower level than in males despite the similar sample sizes (Z = 3.33, P = 5 × 10^{-4}).

Significance of two interesting loci revealed in the Spanish studies was reduced in the meta-analysis with other European studies, although still showed suggestive associations: that of 9q21.32 near TLE1 previously found only in females (female meta-analysis P = 5.4 × 10^{-7}), and that of 9p13.3 near AQP3 (global meta-analysis, P = 1.23 × 10^{-7}).

Heritability of COVID-19 hospitalization

In the hospitalization risk analysis, we found that common variants (minor allele frequency, MAF > 1%) explain 27.1% (95% confidence interval, CI: 19.0–35.3%) of heritability on the observed scale (corresponding to 13.1% [95%CI: 9.2–17.0%] on the liability scale, assuming a prevalence of 0.5%; Fig 5). We observed less heritability among females than males (2.9% [95%CI: 0.00–10.6%] in females and 17.0% [95%CI: 9.2–24.9%] in males on the liability scale). In agreement with observations suggesting an accumulation of non-genetic risk factors with age, especially among males (11,18), we observed larger heritability differences by age groups among males (40.2% [95%CI: 22.8–57.5%] in <60 years versus 17.6% [95%CI: 0.00–38.0%] in ≥60 years on the liability scale) than among females (9.1% [0.00–31.3%] in <60 years versus 13.7% [0.00–29.6%] in ≥60 years on the liability scale).

ID and COVID-19 outcomes

ROH calling was performed in the European QC-ed genotyped dataset. Inbreeding depression analyses are described in Materials and Methods section and Supplemental Note.

The median genomic inbreeding coefficient, F_{ROH}, for the entire SCOURGE study was 0.0048 (IQR = 0.004). No differences were detected between males (F_{ROH} = 0.004, IQR = 0.0035) and females (F_{ROH} = 0.0056, IQR = 0.0038), or
Figure 4. Regional plots of novel association signals found from the meta-analysis between the SCOURGE and CNIO studies. Regional plots of novel association signals found in 9p13.3 (A–C), 17q21.31 (D–F) and 19q13.12 (G–I). The x axes reflect the chromosomal position, and the y axes the $-\log(P\text{-value})$ of the SCOURGE-CNIO meta-analysis. On A, D and G the sentinel variant is indicated by a diamond and all other variants are color coded by their degree of LD with the sentinel variant in Europeans. Whenever a concise set of variants was prioritized, a credible set is shown within a dashed square. The horizontal dotted blue line corresponds to the threshold for genome-wide significance ($P = 5 \times 10^{-8}$).

Figure 5. Forest plot of the SNP-heritability estimates for the COVID-19 hospitalization risk analysis on the liability scale.

between younger and older individuals ($F_{\text{ROH}}$ individuals < 60 years old = 0.004, IQR = 0.0035; $F_{\text{ROH}}$ individuals ≥ 60 years old = 0.0052, IQR = 0.0047, respectively; Supplementary Material, Fig. S6). Regarding the ID in COVID-19 outcomes, we detected a positive association of the $F_{\text{ROH}}$ in COVID-19 hospitalization risk (Fig. 6), severity risk and risk for critical illness (Supplementary Material, Table S7). Our results showed that the hospitalization odds for COVID-19 patients with an $F_{\text{ROH}}$ = 0.0039 were 380% higher than individuals with $F_{\text{ROH}}$ = 0. No effect of the genomic relationship matrix ($F_{\text{GRM}}$) was found.

To assess whether ID in COVID-19 hospitalization in SCOURGE was different between sexes, we first tested the interaction between $F_{\text{ROH}}$ and biological sex. $F_{\text{ROH}}$, sex and the interaction of both ($F_{\text{ROH}}$:Sex) were significant ($F_{\text{ROH}} = 4.7 \times 10^{-3}$, Sex = 1.0 $\times 10^{-12}$, $F_{\text{ROH}}$:Sex = 1.2 $\times 10^{-3}$). This interaction was significant when comparing the hospitalized COVID-19 patients with different controls (A2 and C analyses, see Supplementary Material, Table S8). This interaction was also found significant with severity, but not with critical illness (Supplementary Material, Table S8). In sex-disaggregated analyses, we observed a sex-specific effect of inbreeding. $F_{\text{ROH}}$ was significant in hospitalized males but not in females (Fig. 6 and Supplementary Material, Table S8). This sex-specific effect was also observed with severity and for critical illness (Supplementary Material, Table S8). We then assessed whether ID in COVID-19 hospitalization was different by age. We detected a significant interaction between age and $F_{\text{ROH}}$ for the three outcomes considered (hospitalization, severity and critical illness) (Supplementary Material, Table S9). After disaggregating SCOURGE by sex and age (<60, ≥60), we found that the ID for hospitalization and severity were significant mainly in older males (Fig. 6 and Supplementary Material, Table S9). We detected significant ID for hospitalization and severity in males ≥ 60 years old, but it was marginally significant in females (Fig. 6 and...
Figure 6. Effect of the ID on COVID-19 hospitalization in the SCOURGE cohort. Forest plots are shown for global analyses as well as for sex and age-disaggregated analyses.

Supplementary Material, Table S9). Age and sex-specific effects in hospitalization and severity were robust across different experimental designs using different control groups (Supplementary Material, Fig. S7).

Finally, we then aimed to replicate the ID results with hospitalization, assessing sex and age-specific effects, in a 4418 case-enriched European cohort made of 16 studies from nine countries. Median $F_{\text{ROH}}$ in this other European cohort was slightly higher than that of SCOURGE, 0.05 (0.009–0.0577). A positive and significant ID in COVID-19 hospitalization was detected in this other European cohort when the entire cohort was considered ($F_{\text{ROH}}$ Beta = 18.22, $P = 3.33 \times 10^{-3}$). $F_{\text{GRM}}$ was not significant ($F_{\text{GRM}}$ Beta = $-7.34$, $P = 0.240$). ID was also detected in hospitalized COVID-19 males but not in females (Male $F_{\text{ROH}}$ Beta = 16.22, $P = 3.31 \times 10^{-3}$; Female $F_{\text{ROH}}$ Beta = 15.65, $P = 0.269$). $F_{\text{GRM}}$ was not significant in males or in female analyses. When disaggregating by age, it was possible to detect significant ID in hospitalization only in males $\geq$ 60 years old ($F_{\text{ROH}}$ Beta = 36.16, $P = 3.34 \times 10^{-3}$) (Supplementary Material, Table S10).

No evidence was found of major loci that may be exerting large effects. Rather, polygenicity seems to underlie the ID association. Different islands of ROH (ROHi) and regions of heterozygosity (RHZ) were found to be unique for hospitalized COVID-19 individuals (males and females) and non-hospitalized males, respectively (Supplementary Note, Supplementary Material, Table S11). Enrichment analysis of pathways based on the protein coding genes present in ROH islands were also different between sexes (Supplementary Note, Supplementary Material, Table S12), revealing links with coagulation and complement pathways in males.

Discussion

Here we report the replication of six COVID-19 loci across analyses and provide evidence supporting three additional loci, one of them specifically detected among females. Besides, our analyses provide new insights into disease severity disparities by sex and age and support the necessity of similarly stratified studies to increase the possibility of detecting additional risk variants. Our GWAS constitutes the largest study on COVID-19 genetic risk factors conducted in Spain, with an intrinsic design benefit that SCOURGE includes detailed clinical and genetic information gathered homogeneously across the country. Besides, the study included patients from the whole spectrum of COVID-19 severity covering from asymptomatic to life-threatening...
disease. To date, most research on COVID-19 disease has focused on respiratory failure. However, the inclusion of a severity scale provided a unique opportunity to assess whether previously reported loci combined into a GRS model were associated with differential risk by strata. We warn, however, that the GRS model findings should be interpreted with caution as sex and age-differential results in some of the severity strata needs appropriate replication. Association was tested for four main variables: infection, hospitalization, severe illness and critical illness, and using different definitions of controls to align with the COVID-19 HGI. Irrespective of the tested outcomes or the definition of controls, the results were very similar, supporting the use of population controls to increase power in these studies and the utility of using hospitalization as a proxy of severity. However, our results from the GRS analysis reported a need to maintain separated categories for medium–severe and critical illness.

We observed larger heritability differences by age groups among males than among females for COVID-19 hospitalization, which have diverse support from the literature. On the one hand, there is robust evidence supporting that the presence of X-linked deleterious variants in the TLR7 gene are causal for life-threatening COVID-19 only affecting males (19–21). Of note, most of these severe COVID-19 male patients were younger than 60 years (21). On the other hand, autoantibodies impairing type-I interferon signaling, which are supported to be strong determinants of critical COVID-19 pneumonia, are preferentially found among males older than 65 years (11,18). Taken together, this reconciles with the idea that non-genetic factors involved in severe COVID-19 are expected among older males.

We clearly replicated previously reported associations at 3p21.31 (near SLC6A20 and LZTFL1–FYCO1) (7,9,22,23) and 21q22.11 (in IFNAR2) (7,9), and other findings in ABO, OAS1, TYK2 and ARHGAP27. We also found a differential effect between males and females for SNPs in 3p21.31 and 21q22.11. Although in the meta-analysis with other European studies the leading variants of 3p21.31 reached genome-wide significance in females, there was still a difference in effect size that, considering its magnitude, should not be disregarded. It is important to remark that these association signals found in males were not associated with the presence of comorbidities (see Supplementary Material, Fig. S4). In fact, genetic effects were only found for younger males (<60 years old), consistent with other studies (24) and strongly supporting those comorbidities outweigh genetic effects in disease outcomes in the older patients.

Some novel genome-wide significant signals were found in our study, one in chromosome 19q13.12 (intergenic to UPK1A and ZBTB32, and also linked to the transcriptional regulation of ARHGAP33), and another in chromosome 9p13.3 (intergenic to AQP7 and AQP3). Interestingly, we also found two sex-specific signals: ELF5 in males and TLE1 in females. ELF5 has been recently reported as a new locus associated with critical illness in Europeans (25). Variants of this locus reached genome-wide significance in our male meta-analysis of European cohorts ($P = 4.1 \times 10^{-8}$). As regards of TLE1, this locus should be taken as speculative as the signal did not reach the standard genome-wide significance in the study. However, given that the meta-analysis involved a low number of studies (and the top marker was not imputed in one of them), this result should be taken with caution as further sex-specific studies will be needed to validate this finding.

TLE1 encodes for the transducin-like enhancer of split 1, a co-repressor of other transcription factors and signaling pathways. Besides repressing the transcriptional activity of FOXA2 and of the Wnt signaling, TLE1 has been shown to negatively regulate NF-$\kappa$B, which is fundamental in controlling inflammation and the immune response. The deficiency of TLE1 activity in mice results in enhancement of the NF-$\kappa$B-mediated inflammatory response in diverse tissues including the lung (26). Interestingly, TLE1 is one of the 332 high-confidence SARS-CoV-2 protein–human protein interactions identified so far (27). Taken together, SARS-CoV-2 would be directly targeting the innate immunity and inflammation signaling pathways by interfering with the NF-$\kappa$B activity. Thus, it is not surprising that TLE1 is a top-ranking regulator of inflammation that allows to transcriptionally distinguish mild from severe COVID-19 (28).

In the 19q13.12 locus, the most biologically plausible genes are ARHGAP33 (also showing the best functional support based on the fine-mapping variants) and ZBTB32. ARHGAP33 is transcriptionally regulated by IRF1—a prominent antiviral effector and IFN-stimulated gene (29). It also harbors NF-$\kappa$B binding site that modifies its expression in human lymphoblastoid cell lines by the presence of genetic variants in the binding site linked to many inflammatory and immune-related diseases including sepsis, and bacterial and viral infection (30). Its expression is also altered in human induced pluripotent stem cells-derived pancreatic cultures in response to SARS-CoV-2 infection (31). ARHGAP33 was identified in an unbiased genome-wide CRISPR-based knockout screen in human Huh7.5.1 hepatoma cells infected by coronaviruses including SARS-CoV-2 and further interactome studies (32). With respect to the transcription factor ZBTB32, it has been shown to impair antiviral immune responses by affecting cytokine production and the proliferation of natural killer and T cells, and the generation of memory cells (33). In single cell studies, transcripts of ZBTB32 were enriched in T follicular helper cells and were also expressed at significantly higher levels in hospitalized COVID-19 patients (34).

AQP3 is expressed most strongly in the kidney collecting duct, the gastrointestinal tract, large airways (in basal epithelial cells and the nasopharynx), skin and the urinary bladder; whereas AQP7 is expressed primarily in the testis, fat cells and, to a lesser extent in a subsegment of the kidney proximal tubule (35). In
addition, AQP3 is upregulated in the lung tissues during viral or bacterial-induced diffuse alveolar damage (36). Based on this, in the fact that SARS-CoV-2 interacts with host proteins with the highest expression in lung tissues (27), and the functional evidence linking the fine-mapped variants with eQTLs in lung tissues, our data support AQP3 as the most likely 9p13.3 gene driving the association with COVID-19 hospitalization. Many patients develop acute respiratory distress syndrome (ARDS) during severe COVID-19 (37), and one of the hallmarks of ARDS is the increase of fluid volume (edema) in the airspaces of the lung because of an increase in the alveolo-capillary membrane permeability. Some of the aquaporins, including AQP3, essentially function as water transport pores between the airways and the pulmonary capillaries (38), are key in lung fluid clearance and the formation of this lung edema as a consequence of the lung injury (35). In fact, the use of aquaporin modulators in lung inflammation and edema has been proposed for potential use for the treatment of COVID-19 respiratory comorbidity (39).

We have also shown for the first time that COVID-19 severity risk suffers from ID, where individuals with higher levels of homozygosity associate with higher risk of being hospitalized and of developing severe COVID-19. Our results also suggested that autozygous rare recessive variants found in ROH across the genome, rather than homozygous common variants in strong LD, are underlying the ID. Furthermore, the ID is stronger in males than in females suffering from COVID-19 hospitalizations, especially in males ≥60 years old. Although these results may be found counterintuitive with the rest of findings, they are supported by the mutation accumulation senescence theory. Under this theory, alleles with detrimental effects that act in late life are expected to accumulate and cause senescence, thus increasing the ID (40). We detected further sex-specific effects of homozygosity through ROHi. In hospitalized males, coagulation and complement pathways, which have been previously associated with severe COVID-19 (41), were enriched among the protein coding genes located in ROHi, highlighting the role of homozygosity whereas the Lectin pathway of complement activation is reported to be involved in the response to SARS-CoV-2 infection (42–44). In hospitalized females, PI3K-Akt signaling genes were found to be enriched in ROH islands, whose networks are affected by a great variety of viruses (45).

Given that the effect of the genetic variants in SARS-CoV-2 severity is clearer among males and previous evidence on this regard, we elucidate on the role of androgens in COVID-19 severity. Androgenic hormones have been suggested to be responsible of the excess male mortality observed in COVID-19 patients (46), and several lines of evidence suggest that the androgen receptor (AR) pathway is involved in the severity of SARS-CoV-2 infection: (i) A higher mortality rate among men has been established (47); (ii) A substantial proportion of individuals, both males and females, hospitalized for severe COVID-19 have androgenetic alopecia [AGA; (47)] and (iii) Most of the genes on COVID-19 severity in this study have been identified in male-only analyses, and these genes have been shown to interact with the AR. The following lines of evidence suggest the AR pathway is a mechanism responsible for some identified genes-COVID-19 severity relationship: (i) FYCO1 is regulated by the AR (48), and binding sites between the sex hormone receptor AR and FYCO1 have been demonstrated (48,49); (ii) There is a cross-talk between the IFN pathways and the androgen signaling pathways (50), and androgens are regulated by IFNs in human prostate cells (51); (iii) TMPRSS2, another gene associated with COVID-19 severity in other studies, is induced by androgens through a distal AR binding enhancer (52); (iv) AR induces the expression of chemokine receptors such as CCR1; (v) Variants of LZTFL1 gene are likely pathogenic for male reproductive system diseases (53) and (vi) Genetic polymorphisms in the AR (long polyQ alleles ≥23) and higher testosterone levels in subjects with AR long-polyQ appear to predispose some men to develop more severe disease (54). Thus, it is not unexpected to find that androgen treatments are under the focus of treatment options and prophylaxis of severe COVID-19 (47) and that randomized controlled clinical trials with bicalutamide (NCT04374279), degarelix (NCT04397718) and spironolactone (NCT04345887) are currently underway.

**Materials and Methods**

**Recruitment of cases and phenotype definitions for the discovery phase**

In Spain, 11,939 COVID-19 positive cases were recruited as part of SCOURGE study from 34 centers in 25 cities between March and December 2020. The complete list of hospitals or research centers and the number of samples that each contributed to the study is shown in Supplementary Material, Table S1. Study samples and data were collected by the participating centers, through their respective biobanks after informed consent, with the approval of the respectiveEthic and Scientific Committees. The whole project was approved by the Galician Ethical Committee Ref 2020/197. All samples and data were processed following normalized procedures. Study data were collected and managed using REDCap electronic data capture tools hosted at Centro de Investigación Biomédica en Red [CIBER; (55,56); Supplementary Material, Supplemental Note]. Individuals were diagnosed as COVID-19 positive through a PCR-based test (81.7% of cases) or according to local clinical (3.4%) and laboratory procedures (antibody test: 14.2%; other microbiological tests: 0.7%). All cases were classified in a five-level severity scale (Table 1).

Two Spanish sample collections with unknown COVID-19 status were included as general population controls in some analyses: 3437 samples from the Spanish DNA biobank (https://www.bancoadn.org) and 2506
samples from the GR@CE consortium (17). General population controls were collected from branches of the National Blood Service from adult unrelated individuals self-reporting Spanish origin and absence of personal and familial history of diseases including infectious, cancerous, blood and circulatory, endocrine, mental or behavioral, nervous, vision, hearing, respiratory, immunological, bone, congenital, skin and digestive.

**Second phase: the CNIO study**

A total of 1598 COVID-19 cases from six different Spanish Biobanks (Biobanco CNIO, Biobanco Vasco, Biobanco Hospital Ramón y Cajal, Biobanco Hospital Puerta de Hierro, Biobanco Hospital San Carlos, and Banco Nacional de ADN) were obtained according to the ethical committee approval CEI P1 34_2020-v2. In addition, 1068 individuals from Spanish DNA biobank with unknown COVID-19 status were included as healthy controls in the analysis whenever necessary. Classification as healthy was based on self-reported absence of cardiovascular, renal, pulmonary, hepatic, hematological illnesses or any other chronic conditions, which require continuous treatment, hepatitis B, C infections or acquired immunodeficiency syndrome (AIDS). No clinical characterization was performed on any subject, no information from medical record was incorporated and no medical testing was performed on these individuals. We will refer to these cases and controls as the Centro Nacional de Investigaciones Oncológicas (CNIO) study.

**Genotyping**

The discovery phase samples were genotyped with the Axiom Spain Biobank Array (Thermo Fisher Scientific) following the manufacturer’s instructions in the Santiago de Compostela Node of the National Genotyping Center (CeGen-IS CIII; http://www.usc.es/cegen). This array contains 757 836 markers, including rare variants selected in the Spanish population. Genomic DNA was obtained from peripheral blood and isolated using the Chemagic DNA Blood100 kit (PerkinElmer Chemagen Technologies GmbH), following the manufacturer’s recommendations.

For the second phase study samples, a total of 250 ng of DNA was processed according to the Infinium HTS assay Protocol (Part # 15045738 Rev A, Illumina), including amplification, fragmentation and hybridization using the Global Screening Array Multi-disease v3.0. This array contains a total of 730 059 markers and was scanned on an iScan platform (Illumina, Inc.). Clustering and genotype calling were performed using Genome Studio v2.0.4 (Illumina, Inc.).

**Quality control**

A QC procedure was carried out for the SCOURGE study samples and control datasets. First, a list of probe sets was removed based on poor cluster separation or excessive MAF difference from The 1000 Genomes Project data (1KGP) (57). Then, the following QC steps were applied using PLINK 1.9 (58) and a custom R script. We excluded variants with MAF < 1%, call rate < 98%, a difference in missing rate between cases and controls >0.02, or deviating from Hardy–Weinberg equilibrium (HWE) expectations [\(P < 1 \times 10^{-6}\) in controls, \(P < 1 \times 10^{-10}\) in cases, with a mid-\(P\) adjustment (59)]. Samples with a call rate < 98% and those in which heterozygosity rate deviated > 5 SD from the mean heterozygosity of the study were also removed.

To assess kinship and assign ancestries, autosomal SNPs (MAF > 5%) were pruned with PLINK using a window of 1000 markers, a step size of 80 and a \(r^2\) of 0.1. In addition, high-linkage disequilibrium (LD) regions described in Price et al. (60) were also excluded. A subset of 131 937 independent SNPs was used to evaluate kinship (IBD estimation) in PLINK. Given the possible confusion between relatedness and ancestry, we removed 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 theoretical expected values for each relatedness level). The unrelated SCOURGE individuals were merged with samples from 1KGP and the common SNPs were LD-pruned as previously indicated. Ancestry was then inferred with Admixture (61) using the defined 1KGP superpopulations. Those individuals with an estimated probability > 80% of pertaining to European ancestry were defined as European (\(N = 15 571\)).

Genomic PCs were also computed using a LD-pruned \((r^2 < 0.1\) with a window size of 1000 markers) subset of genotyped SNPs passing quality check for controlling the population structure in the GWAS.

The CNIO study data was filtered following the same QC procedures, where 220 individuals were identified as non-European and, therefore, were excluded from further analysis.

**Variant imputation**

Imputation was conducted based on the TOPMed version r2 reference panel [GRCh38; (62)] in the TOPMed Imputation Server. After post-imputation filtering (Rsq > 0.3, HWE \(P > 1 \times 10^{-6}\), MAF > 1%), 15 045 individuals (9371 COVID-19 positive cases and 5674 population controls) and 8 933 154 genetic markers remained in the SCOURGE European study (discovery). The final dataset of the CNIO study (replication) included 2446 individuals (1378 COVID-19 positive cases and 1068 population controls) and 8 895 721 markers. For directly genotyped variants, the original genotype was maintained in place of the imputed data.

**Statistical analysis**

Association testing was computed by fitting logistic mixed regression models adjusting for age, sex and the first 10 ancestry-specific PCs. SNPRelate (63) was used for prior LD-pruning and data management. Association analyses were performed in SAIGEgds (64), which implements the SAIGE (65) two-step mixed model.
methodology and the SPA test while using more efficient objects for genotype storage. A null model was fitted in the first step using the LD-pruned genotyped variants (MAF > 0.5%, missing rate < 98%) to estimate variance components and the genetic relationship matrix. Then, in a second step, association analyses were performed for both genotyped and imputed SNPs. Significance was established at P < 5 × 10−8 after meta-analysis of the discovery and the second study phases.

To align the results with those from the COVID-19 HGI, three outcomes were evaluated in relation to severity: hospitalization, severe COVID-19 (severity ≥ 3) and very severe COVID-19 (severity = 4, critical illness). For each comparison, three control definitions (A1, A2 and C) were used (Supplementary Material, Table S2).

In addition, the risk to COVID-19 infection was also analysed by comparing all COVID-19 positive cases with control samples from the general population.

All analyses were conducted for each complete dataset and stratified by sex and age (<60 years, ≥60 years). The SNP-sex and SNP-age-interaction terms were tested for each SNP in the subset of clumped signals, adjusting the models for the same covariates.

Then, the main results of both Spanish cohorts (SCOURGE and CNIO) for the overall and sex-stratified analyses were meta-analysed assuming a fixed-effects model using METAL (66).

Because of the similarity of both the SCOURGE and CNIO studies in the clinical variables recorded and, more importantly, in the definition of the severity scale, the leading variants from the significant and validated loci in the hospitalization analysis were also analysed under a multinomial model (Supplementary Material, Supplemen
tal Note).

### Meta-analysis in independent European studies

In order to validate the findings in independent study samples of European ancestry, a meta-analysis of hospitalization risk was performed for the overall and sex-stratified summary statistics of both Spanish cohorts (SCOURGE and CNIO) and other four sex-stratified Europeans studies from the COVID-19 HGI consortium (BelCOVID, GenCOVID, Hostage-Spain and Hostage-Italy).

### Bayesian fine-mapping of GWAS findings

Credible sets were calculated for the GWAS loci to identify a subset of variants most likely containing the causal variant at 95% confidence level, assuming that there is a single causal variant and that it has been tested. We used corrcoversion for R (67) to calculate the posterior probabilities of the variant being causal for all variants with r² > 0.1 with the leading SNP and within 1 Mb. Variants were added to the credible set until the sum of the posterior probabilities was ≥0.95. VEP (https://www.ensembl.org/info/docs/tools/vep/index.html) and the V2G aggregate scoring from Open Targets Genetics (https://genetics.opentargets.org) were used to annotate the most prominent biological effects of the variants in the credible sets.

### Genetic risk score

A GRS was created for the SCOURGE cohort individuals and population controls using the list of SNPs associated with hospitalization, severity or risk in the meta-analysis performed by the COVID-19 HGI [see Supplementary Material, Table S2 in (9)] to appraise its prediction power of the severity scale in SCOURGE. Details of this analysis can be found in Supplementary Note.

### SNP-heritability of COVID-19 severity

We relied on GCTA-GREML 1.93.2beta (68) to assess the heritability of severe COVID-19 symptoms among SCOURGE patients, excluding those with cryptic relatedness or with missing genotypes above 0.5% and assuming a prevalence of COVID-19 hospitalization of 0.5%. This analysis considered all patients (modelling for age, sex, sex-age and the 10 first PCs), and males and females separately (modelling for age and the 10 first PCs). We also partitioned the variance to assess the heritability changes among the patients <60 or ≥60 years old. We focused on the 547,206 autosomal variants with MAF > 1% and missingness < 0.5%. Assuming 0.5% of prevalence of severe COVID-19, and at least 1500 cases and 1500 controls per stratum, we estimate > 97.9% power to detect a heritability > 0.2.

### ROH calling

The ROH segments longer than 300 Kb were called in SCOURGE using PLINK 1.9 in the European QC-ed genotyped dataset (before imputation) with the following parameters: homoyzg-snp 30, homoyzg-kb 300, homoyzg-density 30, homoyzg-window-sn 30, homoyzg-gap 1000, homoyzg-window-het 1, homoyzg-window-missing 5 and homoyzg-window-threshold 0.05. No LD pruning was performed.

### Calculating genomic inbreeding coefficients

Different genomic inbreeding coefficients were calculated (69):

\[ F_{\text{ROH}} = \frac{\sum_{i=1}^{n} \text{ROH} > 1.5 \text{ Mb}}{3 \text{ Gb}} \]

\[ F_{\text{GRM}} \]

is an alternative genomic inbreeding coefficient that was obtained using PLINK’s parameter -ibc (Fat3). This coefficient is described by Yang et al. (68) where N is the number of SNPs, \( p_i \) is the reference allele frequency of the ith SNP, and \( x_i \) is the number of copies of the reference allele. The reference allele frequencies were site-specific.
and included only variants with MAF > 0.05.

\[
F_{\text{GRM}} = \frac{1}{N} \sum_{i} \left( x_i^2 - (1 + 2p_i)x_i + 2p_i^2 \right) \frac{1}{2p_i(1 - p_i)}
\]

Testing and replicating the ID

Inbreeding depression is defined as the change in the mean phenotypic value in a population because of inbreeding (12,13). The ID was modelled in SCOURGE by a multiple logistic regression. The covariables used in this study were sex, age and the first 10 PCs.

The results were replicated in a cohort gathered by Nakanishi et al. (24). This consists of clinical and genomic data from 4418 individuals of European ancestry (3946 hospitalized COVID-19 cases and 422 controls): 2597 males (1072 males < 60 years old, 1525 males ≥ 60 years old) and 1821 females (808 females < 60 years old, 1013 females ≥ 60 years old). The cohort was built by harmonizing individual-level data from 16 different studies (24). The \( F_{\text{GRM}} \) coefficients were obtained following the procedure explained previously. The model described previously with the same covariables (age, sex and the first then PCs) was applied in this cohort.

Genome-specific effects on COVID-19 severity and hospitalization were tested through ID in genomic windows, ROH islands (ROH) and regions of heterozygosity (RHZ) (Supplementary Material, Supplemental Note).

Supplementary Material

Supplementary Material is available at HMGJ online.

Acknowledgements

R.L.-R. is granted by Cátedra de Medicina Genómica IIS-Fundación Jiménez-Díaz-UAM, M.B. by Nextgeneration EU funds. M.C., P.M., M.A.J.S., A.F.R. are granted by the Miguel Servet Programme (CP17/00006, CP16/00116, CPII20CIII/0001, CPII20CIII/0001 respectively) and B.A. by the Juan Rodés Programme (JR17/00020), all of them from Instituto de Salud Carlos III, and cofunded by European Union (ERDF) ‘A way of making Europe’. Fundación Amancio Ortega, Banco de Santander (to A.C.), Estrella de Levante S.A. and Colabora Mujer Association (to E.G.-N.) and Obra Social La Caixa (to R.B.); Agencia Estatal de Investigación (RTC-2017-6471-1 to C.F.), Cabildo Insular de Tenerife (CGIEU0000219140 ‘Apuestas científicas del ITER para colaborar en la lucha contra la COVID-19’ to C.F.) and Fundación Canaria Instituto de Investigación Sanitaria de Canarias (PIIFI20/57 to C.F.).

Conflict of Interest statement: The authors declare no competing interests.

Funding

Instituto de Salud Carlos III (COV20_00622 to A.C., COV20/00792 to M.B., COV20_00181 to C.A., COV20_1144 to M.A.J.S., P120/00876 to C.F.); European Union (ERDF) ‘A way of making Europe’. Fundación Amancio Ortega, Banco de Santander (to A.C.), Estrella de Levante S.A. and Colabora Mujer Association (to E.G.-N.) and Obra Social La Caixa (to R.B.); Agencia Estatal de Investigación (RTC-2017-6471-1 to C.F.), Cabildo Insular de Tenerife (CGIEU0000219140 ‘Apuestas científicas del ITER para colaborar en la lucha contra la COVID-19’ to C.F.) and Fundación Canaria Instituto de Investigación Sanitaria de Canarias (PIIFI20/57 to C.F.).

References

1. Tang, D., Komish, P. and Kang, R. (2020) The hallmarks of COVID-19 disease. PLoS Pathog., 16, e1008536. https://doi.org/10.1371/journal.ppat.1008536.
2. Goyal, P., Choi, J., Pinheiro, L., Schenck, E., Chen, R., Jabri, A., Satlin, M., Campion, R., Nahid, M., Ringel, J. et al. (2020) Clinical characteristics of Covid-19 in new york City. N. Engl. J. Med., 382, 2372–2374.
3. Richardson, S., Hirsch, J., Narasimhan, M., Crawford, J., McGinn, T., Davidson, K. and the Northwell COVID-19 Research Consortium (2020) Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA, 323, 2052–2059.
4. Vahidy, F., Pan, A., Ahnstedt, H., Munshi, Y., Choi, H., Tiruneh, Y., Nasir, K., Kash, B., Andrieni, J. and McCullough, L. (2021) Sex differences in susceptibility, severity, and outcomes of coronavirus disease 2019: cross-sectional analysis from a diverse US metropolitan area. PLoS One, 16, e0245556. https://doi.org/10.1371/journal.pone.0245556.
5. The COVID-19 Host Genetics Initiative (2020) 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., 28, 715–718.
6. Casanova, C., Su, H. and COV19 Human Genetic Effort (2020) A global effort to define the human genetics of protective immunity to SARS-CoV-2 infection. Cell, 181, 1194–1199.
7. Pairo-Castinea, E., Clohisey, S., Klaric, L., Bretherick, A., Rawlik, K., Pasko, D., Walker, S., Parkinson, N., Fourman, M., Russell, C. et al. (2021) Genetic mechanisms of critical illness in COVID-19. Nature, 591, 92–98.
8. Zhang, Q., Bastard, B., Liu, Z., Le Pen, J., Moncada-Velez, M., Chen, J., Ogishi, M., Sabli, I., Hodeib, S., Korol, C. et al. (2020) Inborn errors in type I IFN immunity in patients. Science, 370, eabd4570. https://doi.org/10.1126/science.abb4570.
9. COVID-19 Host Genetics Initiative (2021) Mapping the human genetic architecture of COVID-19. Nature, 600, 472–477.
