Age-dependent impact of the major common genetic risk factor for COVID-19 on severity and mortality

Tomoko Nakanishi,[1,2,3,4,5] Sara Pigazzini,[1,6] Frauke Degenhardt,[7] Mattia Cordioli,[1] Guillaume Butler-Laporte,[3,8] Douglas Maya-Miles,[9,10] Luis Bujanda,[11] Youssef Bouysran,[12] Mari E.K. Niemi,[1] Adriana Palom,[13,14,15] David Ellinghaus,[7,16] Atlas Khan,[17] Manuel Martínez-Bueno,[18] Selina Rolker,[19] Sara Amitrano,[20] Luisa Roade Tato,[10,13,14] Francesca Fava,[20,21,22] FinnGen,[23] The COVID-19 Host Genetics Initiative (HGI),[24] Christoph D. Spinner,[25] Daniele Prati,[26] David Bernardo,[10,27] Federico Garcia,[28,29] Gilles Darcis,[30,31] Israel Fernández-Cadenas,[32] Jan Cato Holter,[33,34] Jesus M. Banales,[11,35] Robert Frithiof,[36] Krzysztof Kiryluk,[17] Stefano Duga,[37,38] Rosanna Asselta,[37,38] Alexandre C. Pereira,[39] Manuel Romero-Gómez,[9,10] Beatriz Nafría-Jiménez,[40] Johannes R. Hov,[33,41,42] Isabelle Migeotte,[12,43] Alessandra Renieri,[20,21,22] Anna M. Planas,[44,45] Kerstin U. Ludwig,[19] Maria Buti,[10,13,14] Souad Rahmouni,[29] Marta E. Alarcón-Riquelme,[18,46] Eva C. Schulte,[47,48,49] Andre Franke,[7,50] Tom H. Karlsen,[34,41,42] Luca Valenti,[51,52] Hugo Zeberg,[53,54] J. Brent Richards,[2,3,9,55] and Andrea Ganna[1,56]

[1]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. [2]Department of Human Genetics and [3]Centre for Clinical Epidemiology, Department of Medicine, Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Quebec, Canada. [4]Kyoto-McGill International Collaborative School in Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [5]Japan Society for the Promotion of Science, Tokyo, Japan. [6]University of Milano-Bicocca, Milano, Italy. [7]Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany. [8]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Quebec, Canada. [9]Digestive Diseases Unit, Virgen del Rocio University Hospital, Institute of Biomedicine of Seville, University of Seville, Seville, Spain. [10]Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto de Salud Carlos III (ISCIII), Madrid, Spain. [11]Department of Liver and Gastrointestinal Diseases, Biodonostia Health Research Institute, Donostia University Hospital, University of the Basque Country (UPV/EHU), CIBERehd, Ikerbasque, San Sebastian, Spain. [12]Centre de Génétique Humaine, Hôpital Erasme, Université Libre de Bruxelles (ULB), Brussels, Belgium. [13]Liver Unit, Department of Internal Medicine, Hospital Universitari Vall d'Hebron, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain. [14]Departament de Medicina: Bellaterra, Universitat Autònoma de Barcelona, Barcelona, Spain. [15]Liver Diseases, Vall d'Hebron Institut de Recerca (VHIR), Barcelona, Spain. [16]Novo Nordisk Foundation Center for Protein Research, Disease Systems Biology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [17]Division of Nephrology, Department of Medicine, Vagelos College of Physicians and Surgeons, Columbia University, New York, New York, USA. [18]GENYO, Centre for Genomics and Oncological Research: Pfizer/University of Granada/Andalusian Regional Government, Granada, Spain. [19]Institute of Human Genetics, University Hospital Bonn, Medical Faculty University of Bonn, Bonn, Germany. [20]Genetica Medica, Azienda Ospedaliero-Universitaria Senese, Siena, Italy. [21]Medical Genetics and [22]Med Biotech Hub and Competence Center, Department of Medical Biotechnologies, University of Siena, Italy. [23]FinnGen is detailed in Supplemental Acknowledgments. [24]The COVID-19 HGI is detailed in Supplemental Acknowledgments. [25]Technical University of Munich, School of Medicine, University Hospital Rechts der Isar, Department of Internal Medicine II, Munich, Germany. [26]Department of Transfusion Medicine and Hematology, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Università degli Studi di Milano, Milano, Italy. [27]Mucosal Immunology Lab, Unit of Excellence Institute of Biomedicine and Molecular Genetics (IBGM), University of Valladolid-CSIC, Valladolid, Spain. [28]Hospital Universitario Clinico San Cecilio, Granada, Spain. [29]Instituto de Investigación Ibs.Granada, Granada, Spain. [30]University of Liege, GIGA-Insitute, Liege, Belgium. [31]Liege University Hospital (CHU of Liege), Liege, Belgium. [32]Stroke Pharmacogenomics and Genetics Group, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain. [33]Department of Microbiology, Oslo University Hospital, Oslo, Norway. [34]Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [35]Department of Biochemistry and Genetics, School of Sciences, University of Navarra, Pamplona, Spain. [36]Department of Surgical Sciences, Anaesthesiology and Intensive Care Medicine, Uppsala University, Uppsala, Sweden. [37]Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy. [38]IRCCS Humanitas Clinical and Research Hospital, Rozzano, Milan, Italy. [39]Heart Institute (InCor)/University São Paulo Medical School, São Paulo, Brazil. [40]Osakidetza Basque Health Service, Donostialdea Integrated Health Organisation, Clinical Biochemistry Department, Sebastian, Spain. [41]Norwegian PSC Research Center and Section of Gastroenterology, Department of Transplantation Medicine and [42]Research Institute of Internal Medicine, Oslo University Hospital, Oslo, Norway. [43]Fonds de la Recherche Scientifique (FNRS), Brussels, Belgium. [44]Institute for Biomedical Research of Barcelona (IIBB), National Spanish Research Council (CSIC), Barcelona, Spain. [45]Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Barcelona, Spain. [46]Institute for Environmental Medicine, Karolinska Institutet, Solna, Sweden. [47]Institute of Virology, Technical University of Munich/Helmholtz Zentrum München, Munich, Germany. [48]Institute of Psychiatric Phenomics and Genomics and [49]Department of Psychiatry, University Hospital, LMU Munich University, Munich, Germany. [50]University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany. [51]Department of Pathophysiology and Transplantation, Università degli Studi di Milano, Milano, Italy. [52]Department of Transfusion Medicine and Hematology, Precision Medicine, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milano, Italy. [53]Department of Neuroscience, Karolinska Institutet, Sweden. [54]Max-Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [55]Department of Twin Research, King's College London, London, United Kingdom. [56]Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA.

▶ **Related Commentary: https://doi.org/10.1172/JCI155011**

BACKGROUND. There is considerable variability in COVID-19 outcomes among younger adults, and some of this variation may be due to genetic predisposition.

METHODS. We combined individual level data from 13,888 COVID-19 patients (*n* = 7185 hospitalized) from 17 cohorts in 9 countries to assess the association of the major common COVID-19 genetic risk factor (chromosome 3 locus tagged by rs10490770) with mortality, COVID-19-related complications, and laboratory values. We next performed metaanalyses using FinnGen and the Columbia University COVID-19 Biobank.

RESULTS. We found that rs10490770 risk allele carriers experienced an increased risk of all-cause mortality (HR, 1.4; 95% CI, 1.2–1.7). Risk allele carriers had increased odds of several COVID-19 complications: severe respiratory failure (OR, 2.1; 95% CI, 1.6–2.6), venous thromboembolism (OR, 1.7; 95% CI, 1.2–2.4), and hepatic injury (OR, 1.5; 95% CI, 1.2–2.0). Risk allele carriers age 60 years and younger had higher odds of death or severe respiratory failure (OR, 2.7; 95% CI, 1.8–3.9) compared with those of more than 60 years (OR, 1.5; 95% CI, 1.2–1.8; interaction, *P* = 0.038). Among individuals 60 years and younger who died or experienced severe respiratory failure, 32.3% were risk-variant carriers compared with 13.9% of those not experiencing these outcomes. This risk variant improved the prediction of death or severe respiratory failure similarly to, or better than, most established clinical risk factors.

CONCLUSIONS. The major common COVID-19 genetic risk factor is associated with increased risks of morbidity and mortality, which are more pronounced among individuals 60 years or younger. The effect was similar in magnitude and more common than most established clinical risk factors, suggesting potential implications for future clinical risk management.

Introduction

The COVID-19 pandemic has led to the deaths of millions of individuals and the largest economic contraction since the Great Depression (1). The clinical outcomes of COVID-19 are remarkably variable, such that some individuals remain asymptomatic (2), while others develop severe COVID-19 with systemic inflammation, respiratory failure, or death. This variability in outcome creates difficulties in clinical management when estimating who is at risk of severe disease and may develop a need for intensive care. Furthermore, recent guidelines suggest risk stratification should be considered when deciding upon prophylactic treatment (3–5).

Some of this variation in COVID-19 behavior has been attributed to risk factors such as age, sex (6), comorbidities (7), socioeconomic factors (8), and genetic variants in the SARS-CoV-2 genome (9). While the main risk factor for severe outcomes is age, the impact of which increases exponentially after age 60 (7), some younger individuals experience severe COVID-19 outcomes and death. The early onset of several common diseases, such as breast cancers, myocardial infarction, and Alzheimer's disease, is disproportionally influenced by human genetic factors (10–13), and this may also be the case for COVID-19. Several GWAS have identified multiple loci in the human genome associated with severity of COVID-19 (14–17). Among GWAS findings, a genetic risk locus on chromosome 3 is the strongest and most consistent signal (16). This genetic risk locus harbors a cluster of genes on chromosome 3; however, the true causal variant is still unknown. The fact that the risk allele sits on a long haplotype inherited from Neanderthals (18) makes the identification of the causal allele and the gene or genes involved challenging. The SNP rs10490770 serves as a marker for this genetic risk factor (as well as other SNPs on the same haplotype; ref. 19), and approximately 15% of individuals of European ancestry carry the C risk allele (19). However, the clinical relevance of this locus and its potential age-dependent impact are unknown.

We therefore assembled individual-level COVID-19 clinical and human genomic data in a large international consortium of 17 cohorts in 9 countries (Belgium, Brazil, Canada, Germany, Italy, Norway, Spain, Sweden, and the United Kingdom) to assess the relationship between the chromosome 3 SNP rs10490770 and COVID-19 severity, complications, and mortality, focusing on age-dependent effects. Finally, in order to assess the relative importance of this locus, we compared its ability to predict COVID-19 outcomes to that of a polygenic risk score (PRS), which aggregates information from common genetic variants across the genome, and other established clinical risk factors.

Results

Study participants. We collected and harmonized individual-level clinical and genomic data from 13,888 COVID-19 patients diagnosed with COVID-19 from February 5, 2020, to February 7, 2021. Table 1 illustrates the participants' demographic and clinical characteristics. By genetically inferring the ancestry using 1000G genetic superpopulations (20) as a reference, the majority of participants were of European descent (12,091; 87.1%). However, considerable numbers of individuals who were not of European were also included in metaanalyses: 389 (2.8%) were of South Asian ancestry, and 602 (4.3%) were of admixed American ancestry. Of these patients, 7185 were hospitalized, among whom 1695 (24.3%) were admitted to the intensive care unit (ICU); 1264 (10.0%) died following COVID-19 diagnosis, and 1704 (14.6%) met the criteria for severe respiratory failure (noninvasive ventilation, high-flow oxygen therapy, or intubation); their mean age was 62.9 years, and 31.2% were females. Clinical information was obtained with different degrees of completeness across studies. A detailed description of study-specific demographics, clinical characteristics, and their missingness rates is provided in Supplemental Figure 1 and Supplemental

Table 1. Patient characteristics

	Hospitalized	Total
	(n = 7185)	(n = 13,888)
Female	2866 (39.9%)	6549 (47.2%)
Age (years)[A]	64.8 (14.7)	63.7 (12.8)
Ancestry		
European	6054 (84.3%)	12,091 (87.1%)
South Asian	113 (1.6%)	389 (2.8%)
African	234 (3.3%)	421 (3.0%)
Others	187 (2.6%)	276 (2.0%)
East Asian	64 (0.9%)	109 (0.8%)
Admixed American	533 (7.4%)	602 (4.3%)
ICU admission	1695 (24.3%)	1695 (12.5%)
Death status		
Survived	4887 (79.3%)	11,369 (90.0%)
Deceased	1264 (20.5%)	1264 (10.0%)
Respiratory failure		
Severe respiratory failure	1704 (30.2%)	1704 (14.6%)
Oxygen supplementation	2051 (36.4%)	2051 (17.6%)
Hepatic injury	532 (10.8%)	536 (4.7%)
Cardiovascular complications	1017 (19.6%)	1040 (9.3%)
Kidney injury	1172 (21.8%)	1182 (10.0%)
VTE	288 (6.9%)	289 (2.7%)

[A]Mean (SD); percentage was calculated among those with complete information. The missing rates for each study are listed in Supplemental Table 1. Others in ancestry included remaining individuals who were not assigned as either of European, South Asian, African, East Asian, or admixed American descent.

Table 1 (supplemental material available online with this article; https://doi.org/10.1172/JCI152386DS1).

Chromosome 3 genetic risk and a PRS. In order to tag the chromosome 3 locus, we selected the SNP rs10490770, which was most significantly associated with hospitalization in the COVID-19 GWAS from the COVID-19 HGI, since this is the largest GWAS metaanalysis of COVID-19 severity (16) (cases/controls = 12,888/1,295,966). We then compared the predictive performance of rs10490770 and a PRS. Using the COVID-19 HGI GWAS release 6 (https://www.covid19hg.org/results/r6/), we first metaanalyzed GWAS results from cohorts that were not included in our study (Supplemental Table 2) and calculated PRSs using a pruning and thresholding method. A PRS of $P = 5 \times 10^{-4}$ and $r = 0.7$ had the maximum accuracy in prediction for death or severe respiratory failure and was more significantly associated with death or severe respiratory failure than the chromosome 3 SNP only (OR, 1.7 vs. 1.2 per 1 SD increase in PRS and rs10490770, respectively; Supplemental Tables 3 and 4). Nevertheless, we focused on exploring the clinical implications of rs10490770, given that a single variant can be more easily tested in a clinical context, requires fewer computational resources than a PRS, and is less influenced by limitations, such as the poor transferability of PRSs across different ancestry groups.

Risk allele frequency. We applied a dominant model by grouping participants into 2 groups according to their genotype at rs10490770; C is the allele associated with COVID-19 severity. Those with TC genotype or CC genotype were labeled as carriers, and those with TT genotype were labeled as noncarriers. According to the population frequencies in gnomAD (19), we estimate that 14.4% of individuals of European descent carry at least 1 rs10490770 C allele, as well as 9.5% of admixed Americans, 2.4% of Africans, 47.1% of South Asians, and 0.4% of East Asians. The carrier frequency was 16.2% among individuals of European descent in our cohort.

Association with mortality. We first estimated the HR for all-cause mortality and COVID-19–related death. All analyses were performed separately for each ancestry group. Because the sample size in non-Europeans was limited, we reported the results from individuals of European descent as the main analyses, but the results from non-European ancestry individuals are presented in Supplemental Figures 4–7. All analyses were based on mixed-effects model adjusted for age, sex, and the first 5 genetic principal components (PCs) as fixed effects. Study groups were also included as random effects to account for the study variability.

Risk allele carriers at rs10490770 had a higher HR for all-cause mortality compared with noncarriers (HR, 1.4; 95% CI, 1.2–1.7, $P = 4.5 \times 10^{-5}$, dead/alive = 870/8829) over a median follow-up duration of 43 days (IQR, 17.5–69 days; Figure 1A). A competing risk model to estimate the HR for COVID-19–related death while accounting for non-COVID-19–related deaths estimated a similar HR for COVID-19–related mortality (HR, 1.6; 95% CI, 1.3–1.8, $P = 4.5 \times 10^{-7}$, dead/alive = 750/8829; Figure 1B). The association with mortality was reduced, but still significant, when the analysis was restricted to hospitalized individuals (HR for all-cause mortality, 1.2; 95% CI, 1.0–1.4, $P = 0.03$, dead/alive = 870/3206, and HR for COVID-19 related mortality, 1.3; 95% CI, 1.1–1.6, $P = 1.1 \times 10^{-3}$, dead/alive = 750/3206), indicating that the effect of rs10490770 on mortality was not simply explained by the higher hospitalization rate among the carriers.

Associations with COVID-19 severity. We next examined the effect of risk allele carrier status at rs10490770 for COVID-19 severity. We confirmed that risk allele carrier status at rs10490770 was significantly associated with hospitalization (OR, 1.5; 95% CI, 1.3–1.7, $P = 1.2 \times 10^{-9}$, cases/controls = 6054/6004). A stronger effect was observed for ICU admission (OR, 2.5; 95% CI, 1.9–3.2, $P = 1.6 \times 10^{-12}$, cases/controls = 1234/6004) and death or severe respiratory failure (OR, 1.7; 95% CI, 1.5–2.1, $P = 9.0 \times 10^{-10}$, cases/controls = 2005/7047; Figure 2 and Supplemental Table 5). Restricting analyses to hospitalized individuals, we observed consistent results, some of which were with diminished effect sizes (Figure 2 and Supplemental Table 5). For instance, a significant reduction in effect size was observed in OR for ICU admission (OR, 1.6; 95% CI, 1.3–1.8, $P = 3.5 \times 10^{-8}$, cases/controls = 1234/4820).

We next explored the association of the rs10490770 risk allele with laboratory values that are known to be associated with the severity of COVID-19 (21–25). rs10490770 risk allele carrier status was associated with the worst value for each of these laboratory values at hospital (e.g., lactate dehydrogenase: 0.23 SD increase, $P = 3.5 \times 10^{-7}$, D-dimer: 0.14 SD increase, $P = 2.1 \times 10^{-3}$; IL-6: 0.16 SD increase, $P = 8.7 \times 10^{-3}$; Supplemental Table 6 and Supplemental Figures 2 and 3).

Associations with COVID-19 complications. Risk allele carrier status at rs10490770 was associated with multiple COVID-19–related severe complications (Figure 2). These included severe
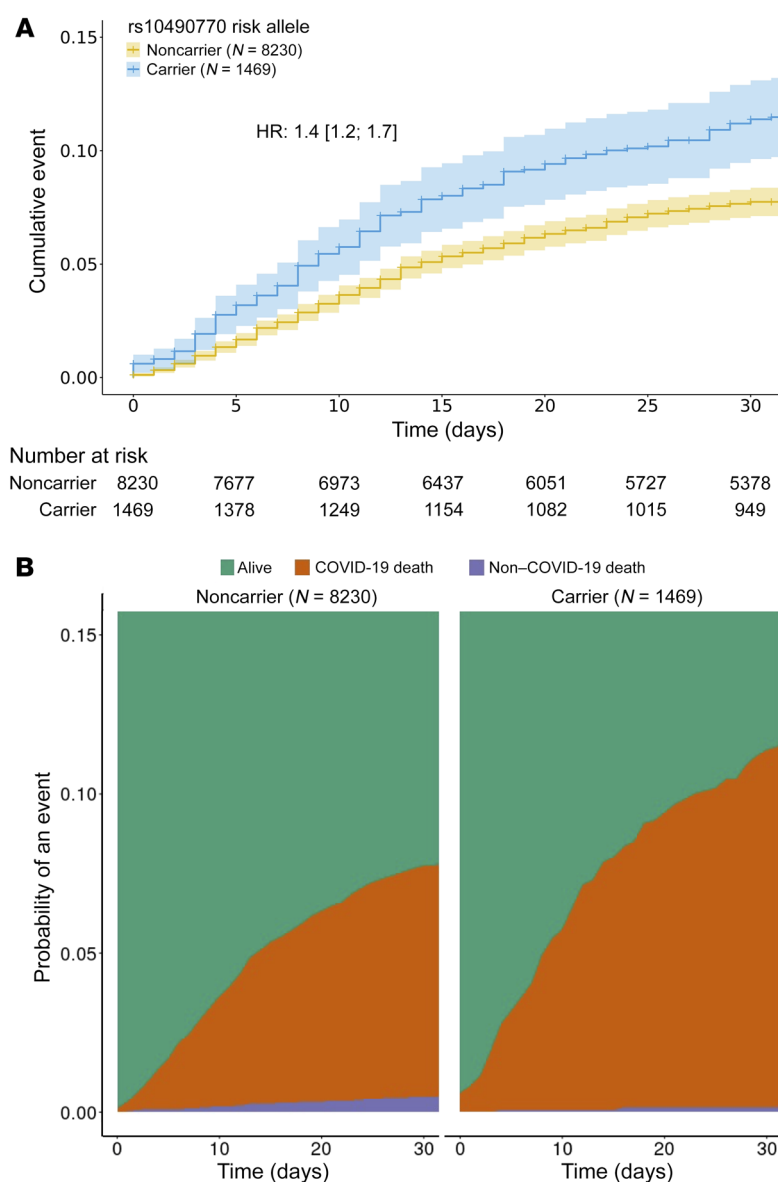
Figure 1. Associations with mortality. The results described here were restricted to 9699 COVID-19 patients of European ancestry with available follow-up and cause of death information. (**A**) Survival analysis using Cox's proportional hazard model. Kaplan-Meier curves stratified by rs10490770 risk allele carrier status. (carriers: n = 1469 vs. noncarriers: n = 8,230). HRs were calculated by adjusting for age, sex, and genetic PCs 1 to 5 as fixed effects and a dummy variable representing the participating studies as random effects. (**B**) Cumulative incidence curves for COVID-19–related death and COVID-19–unrelated death among the same individuals as described in **A**.

respiratory failure (OR, 2.1; 95% CI, 1.6–2.6, P = 2.3 × 10⁻¹⁰, cases/controls = 1284/7047), venous thromboembolism (VTE) (OR, 1.7; 95% CI, 1.2–2.4, P = 1.1 × 10⁻³, cases/controls = 208/8,936), and hepatic injury (OR, 1.5; 95% CI, 1.2–2.0, P = 1.4 × 10⁻³, cases/controls = 352/9541). No significant effect was observed for cardiovascular complications (OR, 1.2; 95% CI, 1.0–1.5, P = 0.10, cases/controls = 854/8890), although this might be due to lack of statistical power to detect such effects. Similar results were observed when restricting the analyses o hospitalized patients (Figure 2 and Supplemental Table 5).

Age-dependent associations with COVID-19 severity. We explored how the effects of rs10490770 risk allele carrier status on severe COVID-19 outcomes in individuals of European descent varied by age. Among severe patients who died or had severe respiratory failure, rs10490770 risk allele carriers were on average 2.3 (95% CI, 1.1–3.5) years younger than noncarriers (P = 1.6 × 10⁻⁴, n = 2005; Figure 3A and Supplemental Table 5). Stratifying by age, we found that among those who were 60 years or younger,

risk allele carrier status had markedly increased odds of death or severe respiratory failure (OR, 2.7 95% CI, 1.8–3.9), whereas risk allele carrier status had more modest effects among those older than 60 years with an OR of 1.5 (95% CI, 1.2–1.9, P value interaction = 0.038; Figure 3B and Supplemental Tables 5 and 7). Among all participants 60 years or younger who died or experienced a severe respiratory COVID-19 outcome, we found that 32.3% (95% CI, 28.3%–36.7%) were rs10490770 risk variant carriers, compared with 13.9% (95% CI, 12.6%–15.2%) of those who did not experience severe disease (Table 2). When considering other severity phenotypes, such as hospitalization and ICU admission, we observed that risk allele carriers tended to be younger than noncarriers. However, we did not detect a different effect in the association between rs10490770 risk allele carriers and these additional severity phenotypes among those who were 60 or younger versus more than 60 years old. This could be attributed to the heterogeneity of the criteria of hospitalization or ICU admission or case-control imbalance in some participating studies.
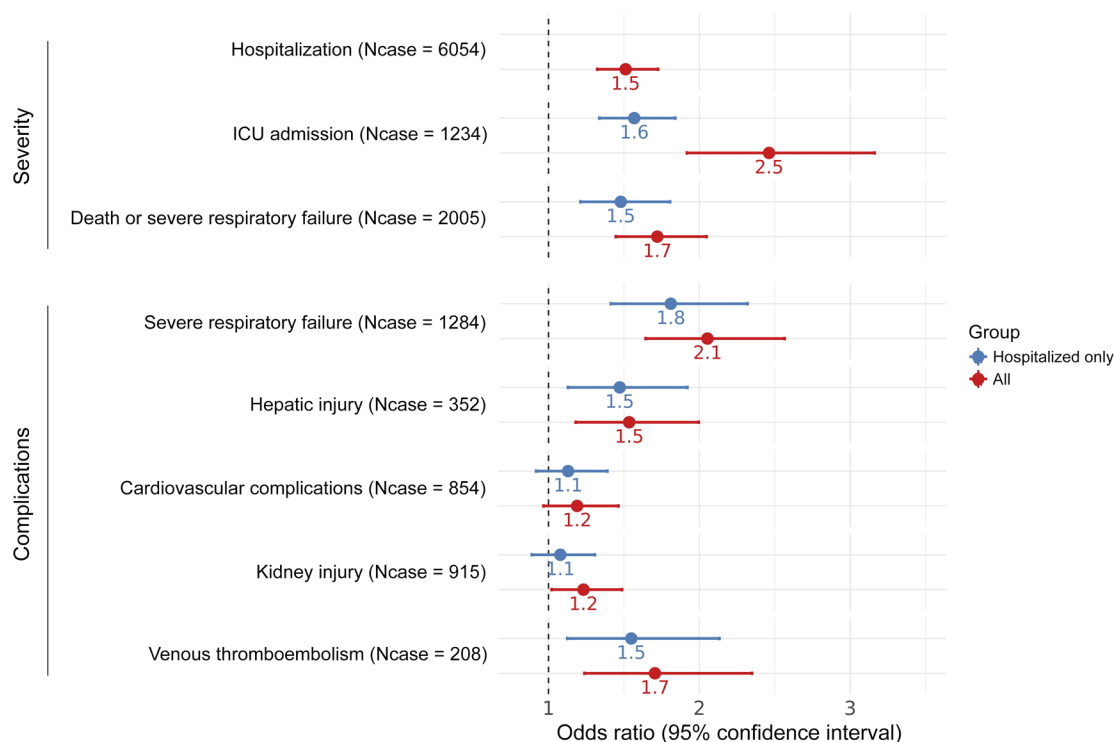
Figure 2. Associations between rs10490770 risk allele carrier status and COVID-19 severity and complications. The results described here were restricted to COVID-19 patients of European ancestry. Logistic regressions were fit to assess the associations of rs10490770 risk allele carrier status with COVID-19 severity and complications, adjusting for age, sex, and genetic PCs 1 to 5 as fixed effects, and a dummy variable representing the participating studies as random effects. Red: All participants (*n* = 12,091); blue: hospitalized participants only (*n* = 6054). The case counts demonstrated as Ncase are the case counts in the analyses of all participants. The full list of case and control counts in the analyses of all participants and those hospitalized only are described in Supplemental Table 5.

Associations with COVID-19 severity stratified by established clinical risk factors. We studied how the effects of rs10490770 risk allele carrier status on COVID-19 severity varied by other established clinical risk factors. Among individuals with no risk factors (BMI ≥ 30, smoking, cancer, chronic kidney disease, chronic obstructive pulmonary disease (COPD), heart failure, transplantation, and diabetes mellitus [DM]) prior to COVID-19, risk allele carriers had an OR of 1.8 for death or severe respiratory failure (95% CI, 1.0–3.4), whereas risk allele carrier status had more modest effects among those with 1 risk factor (OR, 1.6; 95% CI, 1.1–2.5) and more than 1 risk factor (OR, 1.4; 95% CI, 1.0–1.8) (*P* value for interaction = 0.091; Figure 3B and Supplemental Table 8).

Risk prediction compared with established clinical risk factors. We compared the risk discrimination conferred by the rs10490770 risk allele on COVID-19 severity with that observed for other established COVID-19 risk factors. To do so, we used multivariable regression in 7983 individuals of European ancestry, with complete ascertainment of clinical risk factors. rs10490770 risk allele carrier status was independent of other risk factors (Figure 4A and Supplemental Table 9) when examining the association with death or severe respiratory failure (OR, 2.0; 95% CI, 1.7–2.4, *P* = 1.7 × 10⁻¹³; frequency of risk allele carriers, 14.7%, cases/controls = 898/6454). The effect sizes were comparable, or larger, than those of other known risk factors such as DM (OR, 2.0; 95% CI, 1.7–2.4, *P* = 1.0 × 10⁻¹², frequency of DM, 12.5%). Stronger effects were observed among individuals 60 years or younger

(risk allele carrier status: OR, 3.5; 95% CI, 2.3–5.3, *P* = 1.4 × 10⁻⁹; frequency of risk allele carriers, 14.5%; cases/controls = 151/2348) relative to DM (OR, 2.7; 95% CI, 1.6–4.5, *P* = 4.4 × 10⁻⁴; frequency of DM, 5.7%; Figure 4A and Supplemental Table 9).

Consistent with the results from multivariable regression, adding the rs10490770 genotype to nongenetic risk factors modestly improved discrimination for death or severe respiratory failure among patients 60 years or younger (AUC: 0.82 vs. 0.84, *P* = 0.021, and net reclassification improvement [NRI], 0.41, *P* = 7.7 × 10⁻⁸; Table 3), and the performance of risk discrimination was similar to, or better than, that of most of established risk factors included in the study (Figure 4B and Supplemental Table 10).

Metaanalyses. We next metaanalyzed the European ancestry results presented above with those of non-European ancestry participants and 2 external cohorts. We confirmed similar effects in the associations with mortality (Supplemental Figure 4), COVID-19 severity (Supplemental Figure 5), COVID-19 complications (Supplemental Figure 6), and age-dependent effects (Supplemental Figure 7). Given the small sample size of non-European participants, we lacked sufficient statistical power to investigate whether the association between rs10490770 risk allele carriers and COVID-19 outcomes was different when comparing individuals of non-European and European ancestry.

Sensitivity analysis. Finally, we performed several sensitivity analyses to evaluate the robustness of our results. First, we removed the study variables from the covariates (Supplemental
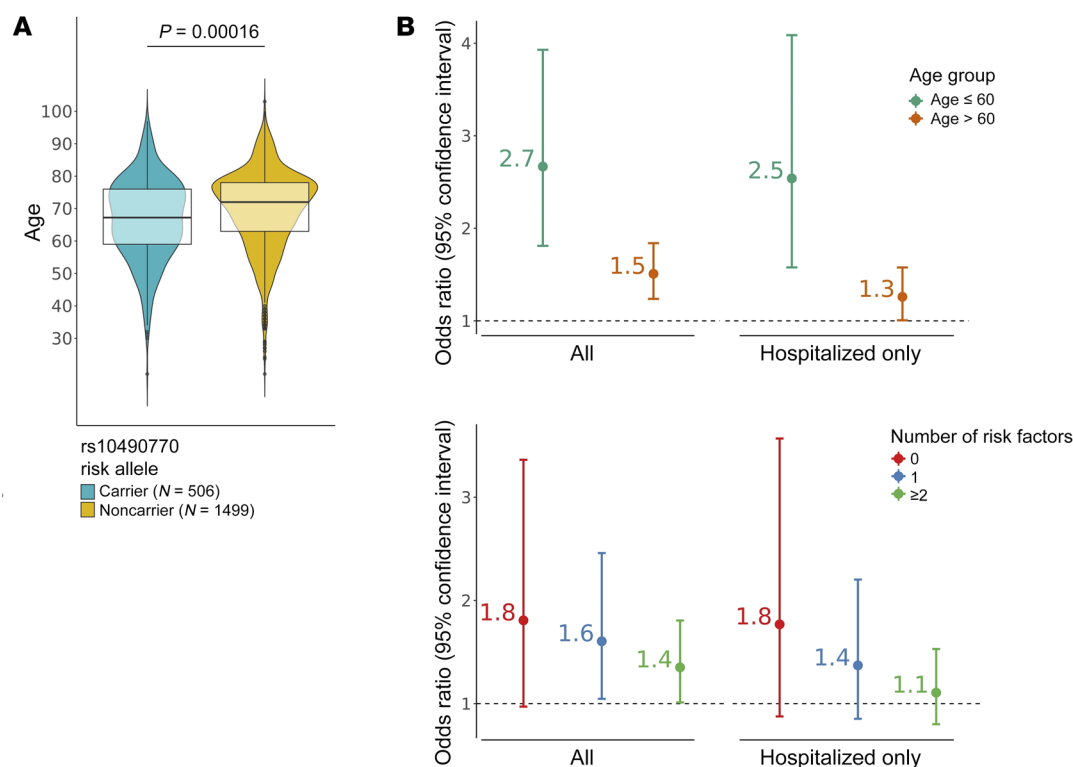
Figure 3. Influence of age and clinical risk factors for the effect of rs10490770 risk allele carrier status on death or severe respiratory failure. (**A**) Age distribution in COVID-19 patients of European ancestry who died or experienced severe respiratory failure (n = 2005). Median (IQR) age was 67.2 (range, 59–76) years in carriers (n = 506) and 72 (range, 63–78) years in noncarriers (n = 1499). (**B**) ORs of rs10490770 risk allele carrier status for death or severe respiratory failure. Regressions were performed within subgroups stratified by age (age ≤60 years and age >60 years) (cases/controls = 2005/7047) or by the number of established risk factors (0, 1, or ≥2); BMI ≥30, smoking, cancer, chronic kidney disease, COPD, chronic heart failure, transplantation, and DM (cases/controls = 898/6454). All analyses were adjusted for age, sex, genetic PCs 1 to 5 as fixed effects, and a dummy variable representing the participating studies as random effects.

Tables 11 and 12). Second, we included participating studies themselves either as fixed or random effects (Supplemental Tables 11 and 12). Third, we restricted the analyses to individuals of European descent from UK Biobank (UKB), a cohort that was not developed to study COVID-19 and thus is less prone to selection bias. These UKB analyses generated similar results (Supplemental Table 13). Fourth, we explored different cutoffs for age-stratified analyses (Supplemental Table 14). Finally, we excluded related individuals (Supplemental Table 15). All sensitivity analyses were consistent with the results from the main analyses.

Discussion

Combining individual-level clinical and genomic data from 13,888 individuals ascertained for COVID-19 outcomes from 17 cohorts in 9 countries, we found that the major genetic risk factor for severe COVID-19 on chromosome 3 was strongly associated with COVID-19–related mortality and clinical complications, such as respiratory failure and VTE.

The risk allele is common. We estimated that 14.4% of individuals of European ancestry are risk allele carriers at rs10490770. Further, 9.5% of admixed Americans, 2.4% of Africans, 47.1% of South Asians, and 0.4% of East Asians are risk allele carriers (20). Consequently, a large proportion of humans carry this risk factor.

The effect of carrying the risk allele on COVID-19 severity was stronger in younger individuals. First, among those 60 years or

younger, the odds of death or severe respiratory failure increased 2.7-fold for risk allele carriers. We found that 32% of individuals 60 years or younger who died or experienced severe respiratory failure were risk allele carriers compared with 14% of individuals not requiring supplemental oxygen. Second, among individuals who died or experienced severe respiratory failure, risk allele carriers were on average 2.3 years younger than noncarriers. Finally, the risk discrimination for death and severe respiratory COVID-19 provided by the risk allele was similar to, or larger than, established clinical risk factors in individuals 60 years or younger. Other common diseases have also demonstrated larger effects of genetic risk factors at a younger age (10, 11, 13). Genetic risk factors are often clinically valuable for risk stratification in younger age groups because the frequency of other established risk factors for COVID-19, such as DM, is often reduced, while the frequency of the genetic variant remains high. Moreover, this specific variant is not associated with any known COVID-19 risk factor (16) and therefore provides orthogonal information compared with existing risk assessment tools. Although vaccination development for SARS-CoV-2 has successfully reduced COVID-19 disease burden in many countries (26, 27), SARS-CoV-2 will likely become endemic in the human population, and it is still not known how long vaccine protection will last. Therefore, this genetic variant may aid in future public health strategies, including selecting individuals for early therapy and potentially for subsequent vaccination prioritization programs.

Table 2. Age and risk allele carrier status by COVID-19 severity outcomes

| | Death or severe respiratory failure | COVID positive but no oxygen supplementation | |
		Hospitalized	All
All			
Carrier	25.2% [23.4; 27.2] (506)	16.2% [14.5; 18.1] (261)	13.8% [13; 14.6] (974)
Noncarrier	74.8% [72.8; 76.6] (1499)	83.8% [81.9; 85.5] (1346)	86.2% [85.4; 87] (6073)
Total	100% (2,005)	100% (1607)	100% (7047)
Age ≤60 years old			
Carrier	32.3% [28.3; 36.7] (151)	14.6% [11.3; 18.7] (52)	13.9% [12.6; 15.2] (366)
Noncarrier	67.7% [63.3; 71.7] (316)	85.4% [81.3; 88.7] (304)	86.1% [84.8; 87.4] (2274)
Total	100% (467)	100% (356)	100% (2640)
Age >60 years old			
Carrier	23.1% [21; 25.3] (355)	16.7% [14.7; 18.9] (209)	13.8% [12.8; 14.8] (608)
Noncarrier	76.9% [74.7; 79] (1183)	83.3% [81.1; 85.3] (1042)	86.2% [85.2; 87.2] (3799)
Total	100% (1538)	100% (1251)	100% (4407)

Frequency of rs10490770 risk variant carriers in individuals of European descent stratified by age and COVID-19 severe outcomes. Square brackets indicate 95% CI; parentheses show sample size.

A PRS for COVID-19 severity derived from release 6 of the COVID-19 HGI had a stronger association with COVID-19 outcomes compared with the rs10490770 risk allele alone. Nevertheless, the aim of this study is to explore the clinical implications of the major genetic risk factors of COVID-19, and future studies should investigate the role of PRSs in COVID-19 severity prediction.

The biology of how the chromosome 3 genetic risk has an effect on COVID-19 severity is still unknown. This locus on chromosome 3p21 includes the putative SARS-CoV-2 coreceptors *SCL6A20* (28, 29), *LZTFL1*, and *FYCO1* (30) and the chemokine receptors *CCR9* (29), *CXCR6* (31), and *XCR1*. There are other chemokine receptors among flanking genes, *CCR1*, *CCR2*, and *CCR3* (32–34), whose involvement in SARS-CoV-2 infection has been suggested and could explain the biology of the striking effect of this genetic risk. Many studies (15, 29) have been trying to pinpoint a single gene or a set of causal genes, but a robust biological consensus has not been built to date.

This study has important limitations. Each cohort has its own selection bias and ascertainment bias. Several studies were enriched for severe patients, whereas UKB is a non–COVID-19 cohort, with evidence of healthy volunteer bias (35). Nevertheless, it may be less prone to selection bias than the COVID-19 cohorts. Selection bias is inherent to most COVID-19 observational studies (36), and this influences the generalizability of the results outside the study populations. Indeed, the estimated protective effects of smoking for COVID-19 severity likely reflect the collider bias due to selection of study participants. Further, other COVID-19 epidemiological studies demonstrated similar effects (36, 37). To mitigate against these issues, we combined data from observational studies with different ascertainment strategies, including national health care systems, studies that were established prior to the COVID-19 pandemic so that recruitment was not dependent upon COVID-19 status, and hospital-based studies. This allowed for an increased representation of individuals with severe COVID-19 outcomes. We also provide analyses restricted to hospitalized patients, which is an ascertained, but clinically relevant, population. Although we were motivated to estimate whether homozygous individuals were at greater risk than heterozygous carriers, we could not draw any meaningful conclusions due to the low sample size (n = 135 homozygous carriers, of whom 92 were of European ancestry). While we included information from participants who were of non-European ancestry, ongoing efforts should enable larger sample sizes to better define the importance of the chromosome 3 risk locus in these ancestries. This further emphasizes the importance of developing genomics-enabled studies in individuals of non-European ancestry.

Since the beginning of the pandemic, we aimed to aggregate and harmonize individual-level clinical and genotype data from multiple cohorts from diverse countries. Due to the nature of the heterogeneity of health care systems, our data from multiple countries substantially increases the generalizability of our research findings (38). Moreover, we deposited a subset of this harmonized data to the European Genome-Phenome Archive (EGAS00001005304) for future use by all bona fide researchers to further improve our ability to understand the COVID-19 pandemic.

In summary, the major genetic COVID-19 risk locus is common and has moderate to large effects on COVID-19 outcomes, including mortality. These effects are age dependent, such that the magnitude of risk increases in younger individuals. These findings suggest potential implications of genetic information in clinical risk management.

Methods

Study participants. We gathered clinical and genomic data from 13,888 COVID-19 cases (7185 of whom were hospitalized) with genetic information available, harmonizing individual-level data from 17 studies. COVID-19 cases were defined as individuals having at least 1 confirmed SARS-CoV-2 viral nucleic acid amplification test from relevant biologic fluids or whose SARS-CoV-2 status was confirmed by ICD-10 codes, using codes U071 and/or U072. We combined data from hospital-based studies that recruited participants after COVID-19 outbreak and a population-based biobank in which recruitment was not dependent upon COVID-19 status. Data were centrally collected at the Institute for Molecular Medicine Finland and harmonized through a standardized data dictionary (https://docs.google.com/spreadsheets/d/1hwBeqckB3_qC8nnavT0kLLntOh3GrmWRJQHeO9zwG8w/edit#gid=665246845). Detailed information for data collection in each individual study is described in Supplemental Methods.

Genotyping and ancestry assignment. In order to tag the chromosome 3 locus, we selected the SNP rs10490770, which was most significantly associated with hospitalization in the COVID-19 GWAS from the COVID-19 HGI, since this is the largest GWAS metaanalysis of COVID-19 severity (ref. 16; cases/controls = 12,888/1,295,966). Each participating study used genotyping and imputation separately following a recommended quality control pipeline (https://docs.google.com/document/d/16ethjgi4MzlQeO0KAW_yDYyUHdB9k-KbtfuGW4XYVKQg/edit). Detailed methods describing genotyping
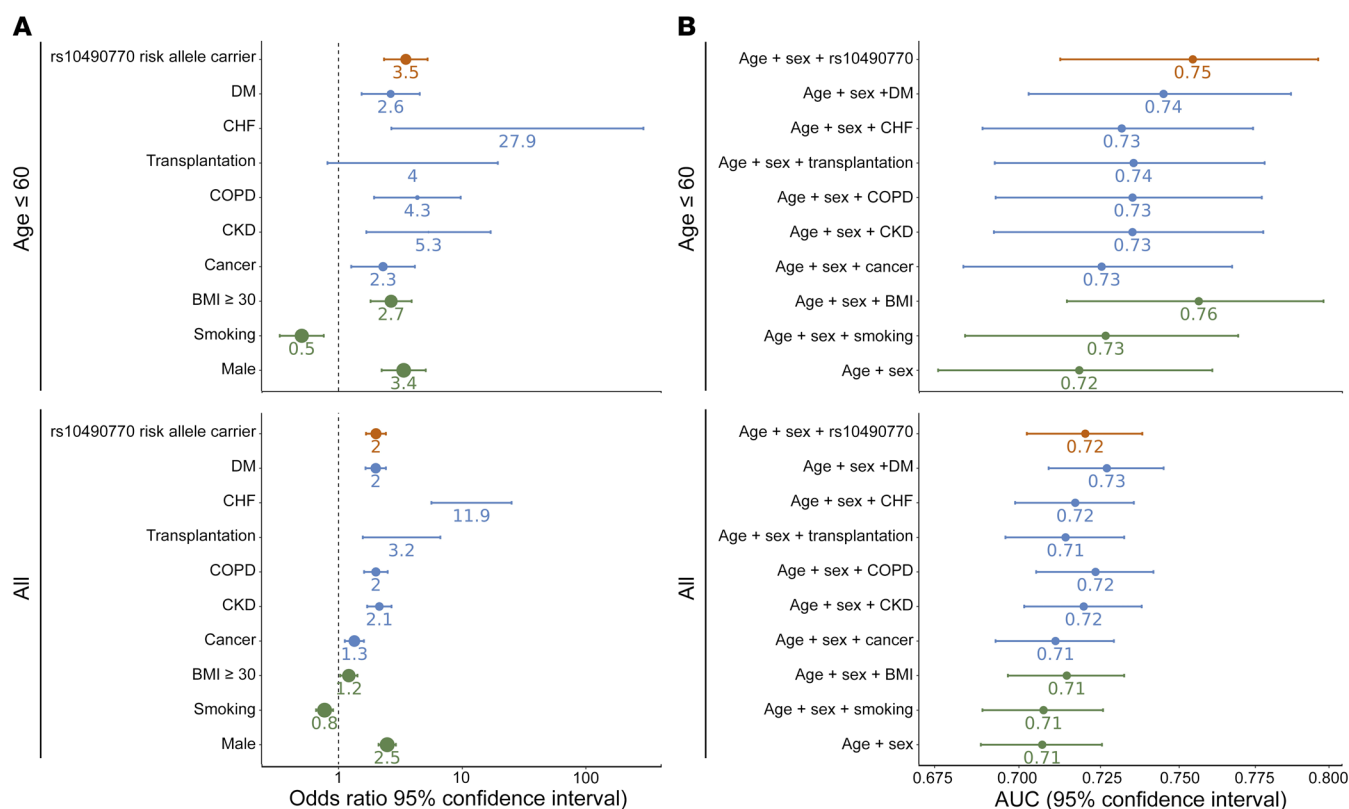
Figure 4. Multivariable regression models and risk prediction estimates for death or severe respiratory failure. Multivariable regression analyses for death or severe respiratory failure were restricted to European-ancestry individuals with complete information of demographic variables (green), comorbidities (blue), and rs10490770 risk allele status (red). *n* = 7352 for all and *n* = 2499 for age ≤60. CKD, chronic kidney disease; CHF, chronic heart failure. Error bars indicate 95% CIs. (**A**) Forest plots comparing ORs from multivariable regression models. The size of each dot represents the frequency of the risk factors. (**B**) Comparison of AUCs of predictions for COVID-19 outcomes. rs10490770 risk allele and nongenetic clinical risk factors were included separately in addition to age and sex in multivariable regression models.

and imputation are available in Supplemental Methods. Ancestry was inferred by performing projection onto the PC analysis (PCA) space from the 1000G (20) phase 3 population using HapMap3 SNPs (39) with minor allele frequency greater than 1% (detailed methods are in Supplemental Methods; Supplemental Table 16 and Supplemental Figure 1).

Statistical analyses. To test the association between rs10490770 and all phenotypes, we applied a dominant model by grouping participants into 2 groups according to their genotype at rs10490770. C is the allele associated with COVID-19 severity; those with TC genotype or CC genotype were labeled as carriers, and those with TT genotype were labeled as noncarriers. We chose this model because it had the lowest Akaike information criterion (AIC) compared with additive and recessive models (see the Supplemental Methods and Supplemental Table 17 for details) for a logistic regression for death or severe respiratory failure outcome (defined below). All analyses were performed separately for each ancestry group. Because the sample size in non-Europeans was limited, we reported the results from individuals of European descent as the main analyses, but the results from non-European ancestry individuals are in Supplemental Figures 4–7. All analyses were based on mixed-effects models adjusted for age and sex, and the first 5 genetic PCs as fixed effects and study groups were also included as random effects to account for study variability. Five study groups, mostly reflecting the country of origin of the study, were created by combining small participating studies with few cases and controls to

reduce the risk of collinearity (details are described in Supplemental Methods. We further estimated the frequency of rs10490770 risk allele carrier status from population frequencies reported in an external database (the Genome Aggregation Database, version 3.1 [gnomAD]; ref. 19), assuming this variant follows Hardy-Weinberg equilibrium.

Association with mortality. The HR for all-cause mortality was estimated by Cox's proportional hazard models using the coxme version 2.2-16 R package (https://cran.r-project.org/web/packages/coxme/). Individuals entered follow-up when diagnosed with COVID-19 or, if a diagnosis date was missing, when hospitalization occurred or when symptoms started. Date of death was considered an event, and data were censored at the last date of follow-up (details are described in Supplemental Methods). We additionally performed competing risk analyses to estimate the subdistribution HR for COVID-19–related mortality using the cmprsk version 2.2-10 R package, which accounts for the competing risk of non-COVID-19–related death: i.e., individuals who did not die of COVID-19 but died due to other causes (e.g., cancer). In the competing risk model, study groups were considered as fixed effects. Survival analyses were restricted to study participants with available follow-up and cause of death information (*n* = 9699). Cause of death was defined by doctor diagnoses, medical chart reviews, or ICD-10 codes (details are described in Supplemental Methods).

Association with COVID-19 severity and complications. To understand the clinical implications of the chromosome 3 locus, we

Table 3. Risk prediction performance for death or severe respiratory failure

Age range	Model	AUC[A]	AUC P value[B]	NRI[A]	NRI P value[B]
All Cases = 898 Controls = 6454	Baseline	0.76 [0.75; 0.78]	–	–	–
	Baseline and rs10490770	0.77 [0.76; 0.79]	1.4×10^{-4}	0.19 [0.13; 0.25]	4.4×10^{-11}
Age ≤60 Cases = 151 Controls = 2,348	Baseline	0.82 [0.79; 0.86]	–	–	–
	Baseline and rs10490770	0.84 [0.81; 0.88]	2.1×10^{-2}	0.41 [0.26; 0.56]	7.7×10^{-8}

Only individuals with complete information regarding clinical risk factors and genotype were included. Baseline model includes age, sex, BMI, smoking status (ever smoker versus never smoker), cancer, chronic kidney disease, COPD, chronic heart failure, transplantation, and DM. [A]Square brackets show 95% CI. [B]P values were calculated by comparing baseline model and baseline and rs10490770 model.

fit mixed-effects regression models to assess the association of rs10490770 risk allele (C) carrier status with 3 types of COVID-19–related measurements: COVID-19 severity, COVID-19 complications, and laboratory values. To do so, we defined 3 COVID-19 severity outcomes, with appropriate control definitions among SARS-CoV-2–positive individuals: (a) hospitalization; (b) ICU admission, and (c) death or severe respiratory failure. Hospitalization cases were COVID-19 cases admitted to the hospital (corresponding to WHO clinical progression scale [ref. 40] ≥ 4; Supplemental Table 18), whereas controls were individuals who did not experience hospitalization (corresponding to WHO clinical progression scale [ref. 40] 1 to 3; Supplemental Table 18). ICU cases were those COVID-19 cases admitted to the ICU, and controls were individuals who did not experience hospitalization. To assess potential selection bias, we also repeated the analyses using only individuals who were hospitalized. In these analyses, controls were defined as those who were hospitalized, but not admitted to the ICU. Death or severe respiratory failure cases were defined as individuals who died or required respiratory support (intubation, continuous positive airway pressure, bilevel positive airway pressure, or continuous external negative pressure, high-flow positive end expiratory pressure oxygen), had ICD-10 codes for acute respiratory distress syndrome (ARDS) or acute respiratory failure (J80, J9600, J9609, Z991), or OPCS codes for the use of a ventilator (E851, E852), corresponding to WHO clinical progression scale (40) ≥6 (Supplemental Table 18). Controls for the death or severe respiratory failure cases were defined as those requiring no oxygen therapy and who were alive, corresponding to WHO clinical progression scale (40) 1–4 (Supplemental Table 18).

We next defined 5 COVID-19–related complications, which were diagnosed in the hospital. These included the following: (a) severe respiratory failure, which was defined as individuals who used respiratory support or had administrative codes for ARDS, respiratory failure, or ventilatory support, as described above, corresponding to WHO clinical progression scale (40) 6 to 9 (Supplemental Table 18); (b) hepatic injury, which was defined as individuals with at least 1 of the following: doctor-diagnosed hepatic complications, highest alanine aminotransferase over 3 times the upper limit of normal (ULN), or ICD-10 codes for acute hepatic failure (K720); (c) cardiovascular complications, which were defined by at least 1 of the following: doctor-diagnosed acute myocardial infarction (AMI) or stroke, highest troponin T or troponin I greater than ULN, or ICD-10 codes for AMI or stroke (I21*, I61, I62, I63, I64, I65, I66*); (d) kidney injury, defined

by at least 1 of the following: doctor-diagnosed acute kidney injury (AKI), highest creatinine greater than 1.5 times ULN, or ICD-10 codes for AKI (N17*); and (e) VTE, defined by at least 1 of the following: doctor-diagnosed pulmonary embolism (PE) or deep venous thrombosis (DVT) or ICD-10 codes for PE or DVT (I26*, I81, I82*). Controls for severe respiratory failure were defined as those requiring no oxygen therapy and who were alive, corresponding to WHO clinical progression scale (40) 1 to 4 (Supplemental Table 18), whereas controls for other complications were defined as those who did not meet the corresponding case criteria and were alive.

Finally, we considered the laboratory values of complete blood count and biochemistry tests available at hospitals (Supplemental Table 6). To test the association with the chromosome 3 locus, we used the lowest value for lymphocyte counts and otherwise the highest value recorded per individual (21–25). This is because we were interested in using these laboratory values as a proxy for COVID-19 severity. Definitions and quality control of laboratory values and specific codes are described in Supplemental Methods and Supplemental Figure 2.

Age-dependent associations with COVID-19 severity. We evaluated the age-dependent effects of the risk allele carrier status on the 3 COVID-19 severity phenotypes we defined above by performing 2 sets of analyses: (a) linear regressions between age at diagnosis and risk allele carrier status among severe cases, adjusting for the same covariates as the main analyses, and (b) adding a carrier status by age interaction term in the main regression models. Age was not dichotomized in these analyses. We also stratified participants by age 60 years or less or more than 60 years and repeated the same logistic regressions, and we estimated the frequency of the risk allele carriers in the 2 age groups. We used 60 years as a cut-point for age-stratified analyses because COVID-19 case fatality rates increase markedly after this age (https://www.inspq.qc.ca/covid-19/donnees/age-sexe) (41).

Associations with COVID-19 severity stratified by established clinical risk factors. In order to compare the association of rs10490770 risk allele carrier status with other risk factors, we similarly stratified participants by BMI of 30 kg/m² or more (a definition of obesity; ref. 42), smoking (ever smoker vs. never smoker), cancer, chronic kidney disease, COPD, chronic heart failure, transplantation, and DM, all of which were curated as established clinical risk factors for severe illness of COVID-19 according to the CDC website (42). All of the 8 risk factors were defined by doctor diagnoses, medical chart reviews, or ICD-10 codes (details are described in Supplemental Methods and Supplemental Table 19). We then tested the difference of the magnitude of the associations of the risk allele carrier status compared with the 8 clinical risk factors. Clinical risk factor–stratified analysis and prediction assessment (described below) were restricted to individuals with complete information for demographics, clinical risk factors, and rs10490770 genotype information (n = 7983). The majority of this subset were from UKB (n = 7461), and only 145 individuals were included from the first discovery GWAS (14).

Risk prediction compared with established clinical risk factors. To better understand the prediction improvement by addition of the chromosome 3 genetic risk in addition to the 8 clinical risk factors,

we performed multivariable regressions in individuals with complete information as described above (n = 7983). We evaluated whether the rs10490770 risk allele improved the risk prediction discrimination for severe COVID-19 outcomes by calculating the AUC and the continuous net reclassification improvement (NRI) using pROC, version 1.16.2 (https://cran.r-project.org/web/packages/pROC/index.html), and PredictABEL, version 1.2-4 R packages (https://cran.r-project.org/web/packages/PredictABEL/index.html).

Metaanalyses. As secondary analyses, we metaanalyzed the results for non-European ancestries and 2 external cohorts for which we did not have access to individual-level data: FinnGen and Columbia University COVID-19 Biobank (CUB). This resulted in a total study population of 15,064 individuals with COVID-19. Inverse-variance weighted metaanalyses were performed under a fixed effect and random effects model using the meta version 4.16-1 R package when the appropriate phenotypes were available and case counts, control counts, and the rs10490770 risk allele carrier counts were larger than 10 in each cohort.

Sensitivity analysis. Adjusting for participating studies may lead to reduced statistical power, given that some studies had only severe cases or had disproportional case-control ratios. To alleviate the collinearity issue, we grouped some small studies to account for study variability. This may not fully account for between-study variability. Thus, we performed 2 sets of sensitivity analyses where we included (a) only 5 genetic PCs without including the study of origin as random or fixed effects and (b) all participating studies either as fixed or random effects. Next, we performed the same analyses using UKB to provide estimates that are more representative of the general population, since this is not a COVID-19–specific cohort. We also tried binning by different cutoffs for age-stratified analyses. In order to understand whether results could have been influenced by related individuals within the samples, we selected 1 individual from a pair of relatives with PI-HAT (proportion of identity by descent calculated by PLINK; ref. 43) greater than 0.1875 (meaning between second and third-degree relatives) and repeated the main analyses.

Statistics. To test the association between rs10490770 and all phenotypes, we applied a dominant model by grouping participants into 2 groups according to their genotypes at rs10490770. C is the allele associated with COVID-19 severity; those with TC genotype or CC genotype were labeled as carriers, and those with TT genotype were labeled as noncarriers. All analyses were based on mixed-effects models adjusted for age, sex, and the first 5 genetic PCs as fixed effects. Study groups were also included as random effects to account for study variability. Five study groups, mostly reflecting the country of origin of the study, were created by combining small participating studies with few cases and controls to reduce the risk of collinearity. We did not apply a multiple-testing correction, and a P value of less than 0.05 was considered significant, since all the outcomes tested were related to COVID-19 severity and not independent of each other.

Data and materials availability. All code for data management and analysis is archived online at https://github.com/tomoconaka/COVID19-chr3 (commit 183ddb7) for review and reuse. The harmonized individual-level data of some participating cohorts from Belgium (BeLCovid_2), Brazil (BRACOVID), Italy (COVID19-Host(a)ge_4, GEN-COVID), Spain (COVID19-Host(a)ge_1,2,3, INMUNGEN-CoV2, Determining the Molecular Pathways and Genetic Predisposition of the Acute Inflammatory Process Caused by SARS-CoV-2

[SPGRX]), and Sweden (SweCovid) were deposited at the European Genome-Phenome Archive (EGA EGAS00001005304). Regarding the SweCovid study, an institutional data transfer agreement can be established and data may be shared if the aims of data use are covered by ethical approval and patient consent. Regarding the data from genetic modifiers for COVID-19–related illness (BelCovid_1), individual-level data were acquired and shared with FIMM during the early stages of the pandemic Upon contact with Isabelle Migeotte (Isabelle.Migeotte@erasme.ulb.ac.be), an institutional data transfer agreement can be established and data can be shared if the aims of data use are covered by ethical approval and patient consent. The procedure will involve an update to the ethical approval as well as review by legal departments at both institutions, and the process will typically take 2 to 4 months from initial contact.

Regarding the BoSCO study, individual-level genotype and clinical data for the purpose of this study were shared with FIMM under a legal, bilateral agreement and were specific to this particular project. Current participant consents and privacy regulations prohibit deposition of individual level data to public repositories. Upon contact with Kerstin Ludwig (kerstin.ludwig@uni-bonn.de), an institutional data transfer agreement can be established and data shared if the aims of data use are covered by ethical approval and patient consent. The procedure will involve review by legal departments at both institutions, and the process will typically take about 2 months from initial contact.

The BQC19 is an open science biobank. Instructions on how to access data for individuals from the BQC19 at the Jewish General Hospital site are available here: https://www.mcgill.ca/genepi/mcg-covid-19-biobank. Instructions on how to access data from other sites of the BQC19 are available here: https://www.bqc19.ca/en/access-data-samples.

For the COVID-19 Kohortenstudie am Klinikum München Rechts der Isar (COMRI) cohort, data protection legislation does not allow for deposition of individual level data in public repositories. Upon direct contact with Christoph Spinner (christoph.spinner@tum.de), an institutional data transfer agreement can be established and data will be shared if the aims of data use are covered by ethical approval and patient consent. The procedure will involve an update to the ethical approval as well as review by legal departments at both institutions, and the process will typically take 2 to 3 months from initial contact.

Regarding the Fondazione IRCCS Milan data (FOGS study), institutional data privacy regulations prohibit deposition of individual level data to public repositories without specific consent. Participant written consent also does not cover public sharing of data for use for unknown purposes. Upon contact with Luca Valenti (luca.valenti@unimi.it), an institutional data transfer agreement can be established and data shared if the aims of data use are covered by ethical approval and patient consent. The procedure will involve the request for an amendment to the ethical approval as well as review by legal departments at both institutions, and the process will typically take 1 to 2 months from initial contact.

Regarding Norwegian data (the Norwegian SARS-CoV-2 study), institutional data privacy regulations prohibit deposition of individual level data to public repositories. Participant written consent also does not cover public sharing of data for use for unknown purposes. Upon contact with Tom H. Karlsen (t.h.karlsen@medisin.uio.no) or Johannes R. Hov (j.e.r.hov@medisin.uio.no), an institutional data transfer agreement can be established and data shared if the aims of

data use are covered by ethical approval and patient consent. The procedure will involve an update to the ethical approval as well as review by legal departments at both institutions, and the process will typically take 1 to 2 months from initial contact.

The genetic and phenotype data sets from UKB are available via the UKB data access process (see http://www.ukbiobank.ac.uk/register-apply/).

Study approval. All institutions contributing cohorts to the COVID-19 HGI received ethics approval from their respective research ethics review boards. Study of genetic modifiers for COVID-19–related illness (BelCovid_1) was approved by the Erasme Ethics committee (protocol P2020_209). Study of host genetics and immune response in SARS-CoV-2 infection (BelCovid_2) was approved by the ethics committee of Liege University Hospital (approval number 2020-242). The BoSCO study was approved by the ethics committee of the Medical Faculty of the University of Bonn. Use of BQC19 received ethical approval from the Jewish General Hospital research ethics board (2020-2137). The BRACOVID study has been approved by the Hospital das Clinicas, São Paulo Medical School, and by the Brazilian National IRB, Comissão Nacional de Ética em Pesquisa (CONEP). COMRI and the COVID-19 biobank of the Faculty of Medicine at Technical University of Munich received ethical approval from the local research ethics board (TUM 217/20, TUM 221/20S, TUM 440/20S). The study at San Sebastian Hospital and use of the Basque Biobank (COVID19-Host(a)ge_1) were approved by the Euskadi Ethics Committee on April 6, 2020 (approval number PI2020064). The studies at Hospital Universitario Vall d'Hebron and CIBERehd del ISCIII (COVID19-Host(a)ge_2) were approved by the Vall d'Hebron Ethical Committee. COVID GWAS and Premed COVID-19 studies (COVID19-Host(a)ge_3) were approved by COVID GWAS (ethics id: 0886-N-20) and Premed COVID (ethics ID, 1954-N-20). The Genetics Against Coronavirus (GENIUS), Humanitas University (COVID19-Host(a)ge_4), study was approved by the ethics committee at Humanitas University (approval number reference number 316/20). FoGS was approved by the ethics committee at Milan area 2, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, via Francesco Sforza 28 20122 Milan Italy (approval number 342_2020). GEN-COVID is a multicenter academic observational study that was approved by the IRB of each participating center. The INMUNGEN-CoV2 study was reviewed and approved by the ethical committee of the Hospital Clinic of Barcelona (CEIm number: Reg.HCB/2020/0357). The Norwegian SARS-CoV-2 study was approved by the Regional Committee for Medical and Health Research Ethics in South-Eastern Norway (project no. 132550). The SPGRX study was reviewed and approved by the Valladolid Ethics Committee (PI-201716) and the Granada Ethics Committee on March 24, 2020, and April 13, 2020, respectively. SweCovid was approved by the National Ethical Review Agency (EPM; 2020-01623). UKB was approved by the Northwest Multi-Centre Research Ethics Committee, and informed consent was obtained from all participants prior to participation. This study was conducted under project ID 27449. FinnGen was approved by the HUS coordinating ethics committee. The Columbia University Biobank study was approved by the Columbia University IRB. For all studies, informed consent was received from each participant.

Note added in proof. Using chromosome conformation capture and gene-expression analysis, a recent study identified the gain-of-function SNP for LZTFL1, rs17713054G>A, as a probable causative variant conferring increased risk of respiratory failure with COVID-19 (44).

Address correspondence to: Andrea Ganna, Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. Phone: 358.504729820; Email: aganna@broadinstitute.org. Or to: J. Brent Richards, McGill University, Senior Lecturer, King's College London (Honorary), Pavilion H-413, Jewish General Hospital, 3755 Côte-Ste-Catherine Montréal, Quebec H3T 1E2, Canada. Phone: 1.514.340.8222 ext. 24362; Email: brent.richards@mcgill.ca.

1. McKee M, Stuckler D. If the world fails to protect the economy, COVID-19 will damage health not just now but also in the future. *Nat Med*. 2020;26(5):640–642.

2. Buitrago-Garcia D, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARSCoV-2 infections: A living systematic review and meta-analysis. *PLoS Med*. 2020;17(9):e1003346.

3. Dooling K, et al. The Advisory Committee on immunization practices' updated interim recommendation for allocation of COVID-19 vaccine - United States, December 2020. *MMWR Morb Mortal Wkly Rep*. 2021;69(5152):1657–1660.

4. Bubar KM, et al. Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. *Science*. 2021;371(6532):916–921.

5. Novelli G, et al. COVID-19 one year into the pandemic: from genetics and genomics to therapy, vaccination, and policy. *Hum Genomics*. 2021;15(1):1–13.

6. O'Driscoll M, et al. Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature*. 2020;590(7844):140–145.

7. Williamson EJ, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature*. 2020;584(7821):430–436.

8. Chaudhry R, et al. A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes. *EClinicalMedicine*. 2020;25:100464.

9. Baric RS. Emergence of a highly fit SARS-CoV-2 variant. *N Engl J Med*. 2020;383(27):2684–2686.

10. Van Der Kolk DM, et al. Penetrance of breast cancer, ovarian cancer and contralateral breast cancer in BRCA1 and BRCA2 families: High cancer incidence at older age. *Breast Cancer Res Treat*. 2010;124(3):643–651.

11. Nordestgaard BG, et al. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: Consensus Statement of the European Atherosclerosis Society. *Eur Heart J*. 2013;34(45):3478–3490.

12. Feng YCA, et al. Findings and insights from the genetic investigation of age of first reported occurrence for complex disorders in the UK Biobank and FinnGen [preprint]. https://doi.org/10.1101/2020.11.20.20234302. Posted on medRxiv November 25, 2020.

13. Olarte L, et al. Apolipoprotein E epsilon4 and age at onset of sporadic and familial Alzheimer disease in Caribbean Hispanics. *Arch Neurol*. 2006;63(11):1586–1590.

14. The Severe Covid-19 GWAS Group. Genome wide association study of severe Covid-19 with respiratory failure. *N Engl J Med*. 2020;383:1522–1534.

15. Pairo-Castineira E, et al. Genetic mechanisms of critical illness in COVID-19. *Nature*. 2020;591(7848):92–98.

16. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19 [published online July 8, 2021]. *Nature*. https://doi.org/10.1038/s41586-021-03767-x.

17. Kosmicki JA, et al. Pan-ancestry exome-wide association analyses of COVID-19 outcomes in 586,157 individuals. *Am J Hum Genet*. 2021;108(7):1350–1355.

18. Zeberg H, Pääbo S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature*. 2020;587(7835):610–612.

19. Karczewski KJ, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434–443.

20. Auton A, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74.

21. Del Valle DM, et al. An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nat Med*. 2020;(3):1–8.

22. Merrill JT, et al. Emerging evidence of a COVID-19 thrombotic syndrome has treatment implications. *Nat Rev Rheumatol*. 2020;16(10):581–589.

23. Higuera-de la Tijera F, et al. Impact of liver enzymes on SARS-CoV-2 infection and the severity of clinical course of COVID-19. *Liver Res*. 2021;5(1):21–27.

24. Vafadar Moradi E, et al. Increased age, neutrophil-to-lymphocyte ratio (NLR) and white blood cells count are associated with higher COVID-19 mortality. *Am J Emerg Med*. 2021;40:11–14.

25. Yan L, et al. An interpretable mortality prediction model for COVID-19 patients. *Nat Mach Intell*. 2020;2(5):283–288.

26. Polack FP, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med*. 2020;383(27):2603–2615.

27. Poustchi H, et al. SARS-CoV-2 antibody seroprevalence in the general population and high-risk occupational groups across 18 cities in Iran: a population-based cross-sectional study. *Lancet Infect Dis*. 2021;21(4):473–481.

28. Vuille-Dit-Bille RN, et al. Human intestine luminal ACE2 and amino acid transporter expression increased by ACE-inhibitors. *Amino Acids*. 2015;47(4):693–705.

29. Yao Y, et al. Genome and epigenome editing identify CCR9 and SLC6A20 as target genes at the 3p21.31 locus associated with severe COVID-19. *Signal Transduct Target Ther*. 2021;6(1):85.

30. Smieszek SP, Polymeropoulos MH. Role of FYVE and coiled-coil domain autophagy adaptor 1 in severity of COVID-19 1 infection 2 [preprint]. https://doi.org/10.1101/2021.01.22.21250070. Posted on medRxiv January 27, 2021.

31. Payne DJ, et al. The CXCR6/CXCL16 axis links inflamm-aging to disease severity in COVID-19 patients [preprint]. https://doi.org/10.1101/2021.01.25.428125. Posted on bioRxiv January 25, 2021.

32. Khalil BA, et al. Chemokines and chemokine receptors during COVID-19 infection. *Comput Struct Biotechnol J*. 2021;19:976–988.

33. Chua RL, et al. COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat Biotechnol*. 2020;38(8):970–979.

34. Liao M, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med*. 2020;26(6):842–844.

35. Fry A, et al. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am J Epidemiol*. 2017;186(9):1026–1034.

36. Griffith GJ, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun*. 2020;11(1):1–12.

37. Miyara M, et al. Low incidence of daily active tobacco smoking in patients with symptomatic COVID-19 [preprint]. https://doi.org/10.32388/WPP19W.3. Posted on Qeios April 21, 2020.

38. Di Maria E, et al. Genetic variants of the human host influencing the coronavirus-associated phenotypes (SARS, MERS and COVID-19): rapid systematic review and field synopsis. *Hum Genomics*. 2020;14(1):30.

39. Altshuler DM, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467(7311):52–58.

40. Marshall JC, et al. A minimal common outcome measure set for COVID-19 clinical research. *Lancet Infect Dis*. 2020;20(8):e192–e197.

41. Signorelli C, Odone A. Age-specific COVID-19 case-fatality rate: no evidence of changes over time. *Int J Public Health*. 2020;65(8):1435–1436.

42. Centers for Disease Control and Prevention. People With Certain Medical Conditions. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html. Updated August 20, 2021. Accessed January 29, 2021.

43. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–575.

44. Downes DJ, et al. Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus. *Nat Genet*. 2021;53(11):1606–1615.