JCI The Journal of Clinical Investigation

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

Tomoko Nakanishi, … , J. Brent Richards, Andrea Ganna

J Clin Invest. 2021. https://doi.org/10.1172/JCI152386.

Clinical Medicine In-Press Preview COVID-19 Genetics

BACKGROUND. There is considerable variability in COVID-19 outcomes amongst 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=7,185 hospitalized) from 17 cohorts in nine 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 meta-analyses 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 ≤60 years had higher odds of death or severe respiratory failure (OR 2.7, 95%CI 1.8-3.9) compared to those >60 years (OR 1.5, 95%CI 1.2-1.8, interaction-p=0.038). Amongst individuals ≤60 years who died or experienced severe respiratory failure, 32.3% were risk variant carriers, compared to 13.9% of those not experiencing these outcomes. The genetic risk improved the prediction of death or severe respiratory failure […]

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

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

[1]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.

[2]Department of Human Genetics, McGill University, Montréal, Québec, Canada.

[3]Centre for Clinical Epidemiology, Department of Medicine, Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Québec, Canada.

[4]Kyoto-McGill International Collaborative School in Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan.

[5]Research Fellow, 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]University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany.

[9]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada.

[10]Digestive Diseases Unit, Virgen del Rocio University Hospital, Institute of Biomedicine of Seville, University of Seville, Seville, Spain.

[11]Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBEREHD), Instituto de Salud Carlos III (ISCIII), Madrid, Spain.

[12]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.

[13]Centre de Génétique Humaine, Hôpital Erasme, Université Libre de Bruxelles (ULB); Brussels, Belgium.

[14]Liver Unit, Department of Internal Medicine, Hospital Universitari Vall d'Hebron, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain.

[15]Universitat Autònoma de Barcelona. Departament de Medicina.Bellatera, Barcelona, Spain.

[16]Vall d'Hebron Institut de Recerca (VHIR). Liver Diseases.

[17]Novo Nordisk Foundation Center for Protein Research, Disease Systems Biology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen.

[18]Division of Nephrology, Department of Medicine, Vagelos College of Physicians & Surgeons, Columbia University, New York, NY.

[19]GENYO. Centre for Genomics and Oncological Research: Pfizer / University of Granada / Andalusian Regional Government, Granada, Spain.

[20]Institute of Human Genetics, University Hospital Bonn, Medical Faculty University of Bonn, Bonn, Germany.

[21]Genetica Medica, Azienda Ospedaliero-Universitaria Senese, Italy.

[22]Medical Genetics, University of Siena, Italy.

[23]Med Biotech Hub and Competence Center, Department of Medical Biotechnologies, University of Siena, Italy.

[24]Technical University of Munich, School of Medicine, University Hospital rechts der Isar, Department of Internal Medicine II, Munich, Germany.

[25]Department of Transfusion Medicine and Hematology, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Università degli Studi di Milano, Milano, Italy.

[26]Mucosal Immunology Lab, Unit of Excellence Institute of Biomedicine and Molecular Genetics (IBGM), University of Valladolid-CSIC; Valladolid, Spain.

[27]Hospital Universitario Clinico San Cecilio, Granada, Spain.

[28]Instituto de Investigación Ibs.Granada. Granada, Spain.

[29]University of Liege, GIGA-Insitute, Liege, Belgium.

[30]Liege University Hospital (CHU of Liege), Liege, Belgium.

[31]Stroke Pharmacogenomics and Genetics Group, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain.

[32]Department of Microbiology, Oslo University Hospital, Oslo, Norway.

[33]Institute of Clinical Medicine, University of Oslo, Oslo, Norway.

[34]Department of Biochemistry and Genetics, School of Sciences, University of Navarra; Pamplona, Spain.

[35]Department of Surgical Sciences, Anaesthesiology and Intensive Care Medicine, Uppsala University, Uppsala, Sweden.

[36]Department of Biomedical Sciences, Humanitas University; Pieve Emanuele, Milan, Italy.

[37]IRCCS Humanitas Clinical and Research Hospital; Rozzano, Milan, Italy.

[38]Heart Institute (InCor)/Univ São Paulo Med Sch, São Paulo, Brazil.

[39]Osakidetza Basque Health Service, Donostialdea Integrated Health Organisation. Clinical Biochemistry Department, Sebastian, Spain.

[40]Norwegian PSC Research Center and Section of Gastroenterology, Dept Transplantation Medicine, Oslo University Hospital, Oslo, Norway.

[41]Research Institute of Internal Medicine, Oslo University Hospital, Oslo, Norway.

[42]Fonds de la Recherche Scientifique (FNRS), Brussels, Belgium.

[43]Institute for Biomedical Research of Barcelona (IIBB), National Spanish Research Council (CSIC), Barcelona, Spain.

[44]Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Barcelona, Spain.

[45]Institute for Environmental Medicine, Karolinska Institutet, 17167, Solna, Sweden.

[46]Institute of Virology, Technical University Munich/Helmholtz Zentrum München, Munich, Germany.

[47]Institute of Psychiatric Phenomics and Genomics, University Hospital, LMU Munich University, Munich, Germany.

[48]Department of Psychiatry, University Hospital, LMU Munich University, Munich, Germany.

[49]Department of Pathophysiology and Transplantation, Università degli Studi di Milano, Milano, Italy.

[50]Department of Transfusion Medicine and Hematology, Precision Medicine, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milano, Italy.

[51]Department of Neuroscience, Karolinska Institutet, Sweden.

[52]Max-Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

[53]Department of Twin Research, King's College London, London, United Kingdom.

[54]Massachusetts General Hospital, Harvard Medical School.

*Corresponding authors.

Andrea Ganna, J. Brent Richards

Andrea Ganna

FIMM-EMBL group leader

Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland

T: +358504729820

E: aganna@broadinstitute.org https://www.dsgelab.org/

J. Brent Richards, Professor of Medicine, McGill University

Senior Lecturer, King's College London (Honorary)

Pavilion H-413, Jewish General Hospital

3755 Côte-Ste-Catherine Montréal, Québec, Canada, H3T 1E2

T: +1 514 340 8222 x24362 F: +1 514 340 7529

E: brent.richards@mcgill.ca www.mcgill.ca/genepi

†These authors contributed equally to this study.

Conflict-of-interest statement:

JBR has served as an advisor to GlaxoSmithKline and Deerfield Capital. His institution has received investigator-initiated grant funding from Eli Lilly, GlaxoSmithKline and Biogen for projects unrelated to this paper. DP has served as an advisory board, and has received travel/research grants, speaking and teaching fees for Macopharma, Ortho Clinical Diagnostics, Grifols, Terumo, Immucor, Diamed, and Diatech Pharmacogenetics. THK has served an advisor to Novartis, Gilead, Intercept and Engitix. LV declares following; speaking fees: MSD, Gilead, AlfaSigma, AbbVie, Consulting: Gilead, Pfizer, Astra Zeneca, Novo Nordisk, Intercept pharmaceuticals, Diatech Pharmacogenetics, IONIS; Research grants: Gilead; CS reports grants,

personal fees, and non-financial support from AbbVie; grants, personal fees, and non-financial support from Apeiron; grants, personal fees from B. Braun Melsungen, grants from Cepheid, personal fees from Formycon, grants, personal fees, and non-financial support from Gilead Sciences; grants and personal fees from Eli Lilly; grants, personal fees, and non-financial support from Janssen-Cilag; personal fees from Molecular partners, grants, personal fees, and non-financial support from GSK/ViiV Healthcare; grants, personal fees, and non-financial support from MSD, outside the submitted work. JRH declares speaking fees and advisor role for Novartis, advisor role for Orkla Health and research support from Biogen. No other authors declare other relationships or activities that could appear to have influenced the submitted work.

Role of funding source:

The funding sources had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Abstract:

Background. There is considerable variability in COVID-19 outcomes amongst 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=7,185 hospitalized) from 17 cohorts in nine 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 meta-analyses 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 ≤60 years had higher odds of death or severe respiratory failure (OR 2.7, 95%CI 1.8-3.9) compared to those >60 years (OR 1.5, 95%CI 1.2-1.8, interaction-p=0.038). Amongst individuals ≤60 years who died or experienced severe respiratory failure, 32.3% were risk variant carriers, compared to 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 amongst individuals ≤60 years. The effect was similar in magnitude and more common than most established clinical risk factors, suggesting potential implications for future clinical risk management.

Brief summary

The major common COVID-19 genetic risk factor on chromosome 3 was strongly associated with morbidities and mortality, with considerably larger effects in individuals ≤60 years.

Main Text:

Introduction

The COVID-19 pandemic has led to the death 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, whose impact 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 genome-wide association studies (GWAS) have identified multiple loci in the human genome associated with severity of COVID-19 (14–17). Amongst 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 (19) makes the identification of the causal allele, and the

gene(s) involved, challenging. The single nucleotide polymorphism (SNP) rs10490770 serves as a marker for this genetic risk factor (as well as other SNPs on the same haplotype (19)) and approximately 15% of individuals of European ancestry carry the C risk allele (20). 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 nine countries (Belgium, Brazil, Canada, Germany, Italy, Norway, Spain, Sweden, and UK) to assess the relationship between the chromosome 3 SNP rs10490770 with COVID-19 severity, complications and mortality, focusing on age-dependent effects. Last, 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, 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 5th, 2020 to February 7th, 2021. Table 1 illustrates the participants' demographic and clinical characteristics. By genetically inferring the ancestry using 1000G genetic superpopulations (21) as a reference, the majority of participants were of European descent (12,091; 87.1%). However, important numbers of non-European descent individuals were also included in meta-analyses: 389 (2.8%) were of South Asian

ancestry and 602 (4.3%) were of Admixed-American ancestry. 7,185 were hospitalized, amongst whom 1,695 (24.3%) were admitted to the ICU. 1,264 (10.0%) died following COVID-19 diagnosis and 1,704 (14.6%) met the criteria for severe respiratory failure (non-invasive ventilation, high flow oxygen therapy, or intubation), whose mean age was 62.9 and 31.2% of whom 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 the Supplemental material (Supplemental Figure 1, Supplemental Table 1).

Chromosome 3 genetic risk and a polygenic risk score

In order to tag the chromosome 3 locus, we selected the SNP rs10490770, which was most significantly associated with hospitalization in the COVID-19 genome-wide association study (GWAS) from The COVID-19 Host Genetics Initiative (HGI), since this is the largest genome-wide association study meta-analysis of COVID-19 severity (16) (cases / controls = 12,888 / 1,295,966). We then compared the predictive performance of rs10490770 and a polygenic risk score (PRS). By using the COVID-19 HGI GWAS release 6 (https://www.covid19hg.org/results/r6/), we first meta-analyzed GWAS results from cohorts which were not included in our study (Supplemental Table 2) and calculated PRSs using a pruning and thresholding method. A PRS with $p=5x10^{-4}$ 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 Table 3-4). 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 less 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 two 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 non-carriers. According to the population frequencies in gnomAD (20), we estimate that 14.4% of individuals of European descent carry at least one rs10490770 C allele, as well as 9.5% of Admixed-American, 2.4% of African, 47.1% of South Asians and 0.4% of East Asians. The carrier frequency was 16.2% amongst individuals of European descent in our cohort.

Association with mortality

We first estimated the hazard ratio (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 the Supplemental material. All analyses were based on mixed-effects model adjusted for age, sex and the first five genetic principal components (PCs) as fixed effects and 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 to non-carriers (HR 1.4, 95%CI 1.2–1.7, $p=4.5x10^{-5}$, dead / alive = 870 / 8,829) over a median follow-up duration of 43 days (interquartile range [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, $p=4.5x10^{-7}$, dead / alive = 750 / 8,829) (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 / 3,206, and HR for COVID-19 related mortality 1.3, 95% CI 1.1-1.6, $p=1.1x10^{-3}$, dead / alive = 750 / 3,206), 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, $p=1.2x10^{-9}$, cases / controls = 6,054 / 6,004). A stronger effect was observed for ICU admission (OR 2.5, 95%CI 1.9-3.2, $p=1.6x10^{-12}$, cases / controls = 1,234 / 6,004) and death or severe respiratory failure (OR 1.7, 95%CI 1.5-2.1, $p=9.0x10^{-10}$, cases / controls = 2,005 / 7,047) (Figure 2, Supplemental Table 5). Restricting analyses to hospitalized individuals, we observed consistent results, some of which were with diminished effect sizes (Figure 2, 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, $p=3.5x10^{-8}$, cases / controls = 1,234 / 4,820).

We next explored the association of the rs10490770 risk allele with laboratory values, which are known to be associated with the severity of COVID-19 (22–26). 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, $p=3.5x10^{-7}$, D-dimer: 0.14 SD increase, $p=2.1x10^{-3}$ and interleukin-6: 0.16 SD increase, $p=8.7x10^{-3}$; Supplemental Table 6**,** Supplemental Figure 2-3).

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 respiratory failure (OR 2.1, 95%CI 1.6-2.6, $p=2.3x10^{-10}$, Cases / Controls = 1,284 / 7,047), VTE (OR 1.7, 95%CI 1.2-2.4, $p=1.1x10^{-3}$, Cases / Controls = 208 / 8,936) and hepatic injury (OR 1.5, 95%CI 1.2-2.0, $p=1.4x10^{-3}$, Cases / Controls = 352 / 9,541). No significant effect was observed for cardiovascular complications (OR 1.2, 95%CI 1.0-1.5, p=0.10, Cases / Controls = 854 / 8,890), although this might be due to lack of statistical power to detect such effects. Similar results were observed when restricting to hospitalized patients (Figure 2**,** 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. Amongst 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 non-carriers ($p=1.6x10^{-4}$, N=2,005; Figure 3A, Supplemental Table 5). Stratifying by age, we found that amongst those who were ≤ 60 years, 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 amongst those >60 years with an OR of 1.5 (95%CI 1.2-1.9, p-value interaction=0.038, Figure 3B, Supplemental Table 5, 7). Amongst all participants ≤ 60 years 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 to 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 non-carriers. However, we did not detect a different effect in the association between rs10490770 risk allele carriers and these additional severity phenotypes amongst those who were ≤60 vs >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.

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. Amongst individuals with no risk factors (BMI ≥ 30, smoking, cancer, chronic kidney disease, chronic obstructive pulmonary disease, heart failure, transplantation, and 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 amongst those with one risk factor (OR 1.6, 95%CI 1.1-2.5) and more than one risk

factors (OR 1.4, 95%CI 1.0-1.8) (p-value for interaction=0.091; Figure 3B, Supplemental Table 8).

Risk prediction compared to 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 7,983 individuals of European ancestry with complete ascertainment of clinical risk factors. rs10490770 risk allele carrier status was independent of other risk factors (Figure 4A, Supplemental Table 9) when examining the association with death or severe respiratory failure (OR 2.0, 95%CI 1.7-2.4, $p=1.7x10^{-13}$, frequency of risk allele carriers 14.7%, Cases / Controls = 898 / 6,454). 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, $p=1.0x10^{-12}$, frequency of DM 12.5%). Stronger effects were observed amongst individuals ≤60 years (risk allele carrier status: OR 3.5, 95%CI 2.3-5.3, $p=1.4x10^{-9}$, frequency of risk allele carriers 14.5%, Cases / Controls = 151 / 2,348) relative to DM (OR 2.7, 95%CI 1.6-4.5, $p=4.4x10^{-4}$, frequency of DM: 5.7%) (Figure 4A, Supplemental Table 9).

Consistent with the results from multivariable regression, adding rs10490770 genotype to non-genetic risk factors modestly improved discrimination for death or severe respiratory failure amongst ≤ 60 years (AUC: 0.82 vs 0.84, p=0.021 and NRI: 0.41, $p=7.7x10^{-8}$, 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, Supplemental Table 10).

Meta-analyses

We next meta-analyzed the European ancestry results presented above with those of non-European ancestry participants and two 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

Last, we performed several sensitivity analyses to evaluate the robustness of our results. First, we removed the study variables from the covariates (Supplemental Table 11-12). Second, we included participating studies themselves either as fixed or random effects (Supplemental Table 11-12). Third, we restricted to individuals of European descent from UKB, a cohort which 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 cut-offs for age-stratified analyses (Supplemental Table 14). Last, 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 nine 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 venous thromboembolism.

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-American, 2.4% of African, 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, amongst those ≤60 years, the odds of death or severe respiratory failure increased 2.7-fold for risk allele carriers. We found that 32% of individuals ≤60 years who died, or experienced severe respiratory failure, were risk allele carriers, compared to 14% of individuals not requiring supplemental oxygen. Second amongst individuals who died or experienced severe respiratory failure, risk allele carriers were on average 2.3 years younger than non-carriers. Last, 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. 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 diabetes mellitus, are 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 to existing risk assessment tools. Although vaccination development for SARS-CoV-2 has successfully reduced COVID-19 disease burden in many countries (27, 28), SARS-CoV-2 will likely become endemic in the human population, and it is still not known how long vaccines 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.

A polygenic risk score (PRS) for COVID-19 severity derived from release 6 of the COVID-19 HGI (https://www.covid19hg.org/results/r6/) had a stronger association with COVID-19 outcomes, compared to 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* (29, 30), *LZTFL1, FYCO1* (31), and the chemokine receptors; *CCR9* (30), *CXCR6* (32), *XCR1*. There are other chemokine receptors amongst flanking genes; *CCR1, CCR2* and *CCR3* (33–35), 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, 30) have been trying to pinpoint a 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 (36). Nevertheless, it may be less prone to selection bias than the COVID-19 cohorts. Selection bias is inherent to most COVID-19 observational studies (37) 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 (37, 38). To mitigate against these issues, we combined data from observational studies with different ascertainment strategies, including national healthcare systems, studies that were established prior to the COVID-19 pandemic and thus 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, on-going efforts should enable larger sample sizes in these ancestries 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 (39). Moreover, we deposited a subset of this harmonized data to European Genome-Phenome Archive (EGAS00001005304), for the 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 (7,185 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 one 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 was centrally collected at 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 the Supplemental material.

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 genome-wide association study (GWAS) from The COVID-19 Host Genetics Initiative, since this is the largest genome-wide association study meta-analysis of COVID-19 severity (16) (cases / controls = 12,888 / 1,295,966). Each participating study performed genotyping and imputation separately following a recommended quality control pipeline (https://docs.google.com/document/d/16ethjgi4MzlQeO0KAW_yDYyUHdB9kKbtfuGW4XYVKQg/edit). Detailed methods describing genotyping and imputation are available in the Supplemental material. Ancestry was inferred by performing projection onto the principal component analysis (PCA) space from the 1000G (21) Phase 3 population using HapMap3 SNPs (40) with minor allele frequency > 1% (detailed methods are in the Supplemental material). (Supplemental Table 16, Supplemental Figure 1).

Statistical analyses

To test the association between rs10490770 and all phenotypes, we applied a dominant model by grouping participants into two 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 non-carriers. We chose this model because it had the lowest Akaike Information Criterion (AIC), compared to additive and recessive models (see the Supplemental material for detail, Supplemental Table 17), in 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 the Supplemental materials. All analyses were based on mixed-effects model adjusted for age, sex and the first five genetic principal components (PCs) as fixed effects and study groups were also included as random effects to account for the 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 (detail is described in the Supplemental material). We further estimated the frequency of rs10490770 risk allele carrier status from the population frequencies reported in external database (the Genome Aggregation Database v 3.1 [gnomAD (20)]), assuming this variant follows Hardy-Weinberg equilibrium.

Association with mortality

The hazard ratio (HR) for all-cause mortality was estimated by Cox proportional hazard models using the "coxme v2.2-16" R package. Individuals entered the follow-up when diagnosed with COVID-19 or if a diagnosis date was missing, the date when they were hospitalized or when their symptoms started. They were considered as an event at the date of death and censored at the last date of follow-up (details are described in the Supplemental material). We additionally performed competing risk analyses to estimate the sub-distribution hazard ratio for COVID-19

related mortality using the "cmprsk v2.2-10" 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=9,699). Cause of death was defined by doctor-diagnoses, medical chart reviews or ICD-10 codes (details are described in the Supplemental material).

Association with COVID-19 severity and complications

To understand the clinical implications of the chromosome 3 locus, we fit mixed-effects regression models to assess the association of rs10490770 risk allele [C] carrier status with three types of COVID-19 outcomes: COVID-19 severity, COVID-19 complications and laboratory values. To do so, we defined three COVID-19 severity outcomes, with appropriate control definitions amongst SARS-CoV-2 positive individuals: 1) hospitalization; 2) intensive care unit (ICU) admission and 3) death or severe respiratory failure. Hospitalization cases were COVID-19 cases admitted to the hospital (corresponding to WHO clinical progression scale (41) ≥4, Supplemental Table 18), whereas controls were individuals who did not experience hospitalization (corresponding to WHO clinical progression scale (41) 1-3, 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 of the use of ventilator ("E851","E852"), corresponding to WHO clinical progression scale (41) ≥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 (41) 1-4 (Supplemental Table 18).

We next defined five COVID-19 related complications, which were diagnosed at hospital. These included: 1) Severe respiratory failure, which was defined by the use of respiratory support or individuals with administrative codes for ARDS, respiratory failure or ventilatory support as described above, corresponding to WHO clinical progression scale (41) 6-9 (Supplemental Table 18); 2) Hepatic injury was defined as individuals with at least one of the following: doctor-diagnosed hepatic complications, highest alanine aminotransferase > 3 times upper limit of normal (ULN), or ICD-10 codes for acute hepatic failure (“K720”); 3) Cardiovascular complications were defined by at least one of the following: doctor-diagnosed acute myocardial infarction (AMI) or stroke, highest troponin T or troponin I > ULN, or ICD-10 codes for AMI or stroke (“I21*”,“I61”, “I62”, “I63”, “I64”, “I65” ,”I66*”).; 4) Kidney injury was defined by at least one of the following: doctor-diagnosed acute kidney injury (AKI), highest creatinine > 1.5 times ULN, or ICD-10 codes for AKI (“N17*”); 5) Venous thromboembolism (VTE) was defined by at least one 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 (41) 1-4 (Supplemental Table 18), whereas controls for other complications were defined as those who did not meet the corresponding case criteria and were alive.

Last we considered the laboratory values of complete blood count and biochemistry tests available at hospital (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 (22–26). 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 the Supplemental material (Supplemental Figure 2).

Age-dependent associations with COVID-19 severity

We evaluated the age-dependent effects of the risk allele carrier status on COVID-19 three severity phenotypes by performing two sets of analyses: 1) linear regressions between age at diagnosis and risk allele carrier status amongst severe cases, adjusting for the same covariates as the main analyses, and 2) 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 or >60 years and repeated the same logistic regressions, and we estimated the frequency of the risk allele carriers in the two 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)(44).

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 ≥30 kg/m^2 (a definition of obesity (44)), smoking (ever-smoker vs never-smoker), cancer, chronic kidney disease, chronic obstructive pulmonary disease (COPD), chronic heart failure, transplantation, and diabetes mellitus (DM), all of which were curated as established clinical risk factors for severe illness of COVID-19 according to the Centre for Disease Control website (44). All of the eight risk factors were defined by doctor-diagnoses, medical chart reviews or ICD-10 codes (details are described in the Supplemental material). We then tested the difference of the magnitude of the associations of the risk allele carrier status compared to the eight clinical risk factors. Clinical risk factors stratified analysis and prediction assessment (described below) were restricted to individuals with complete information for demographics, clinical risk factors and rs10490770 genotype information (N=7,983). The majority of this subset were from UK Biobank (N=7,461), and only 145 individuals were included from the first discovery GWAS (14).

Risk prediction compared to established clinical risk factors

To better understand the prediction improvement by adding of the chromosome 3 genetic risk in addition to the eight clinical risk factors, we performed multivariable regressions in individuals with complete information as described above (N=7,983). We evaluated whether the rs10490770 risk allele improved the risk prediction discrimination for severe COVID-19 outcomes by calculating the area under receiver operation curve (AUC) and the continuous net reclassification improvement (NRI) using “pROC v1.16.2” and “PredictABEL v1.2-4” R packages.

Meta-analyses

As secondary analyses, we meta-analyzed the results with non-European ancestries and two 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 meta-analyses were performed under a fixed effect and random effects models using the "meta v4.16-1" R package when the appropriate phenotypes were available and case counts, control counts, and the rs10490770 risk allele carrier counts were larger than ten 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 ratio. 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 two sets of sensitivity analyses where we included, 1) only five genetic PCs without including the study of origin as random or fixed effects, and 2) all participating studies either as fixed or random effects. Next, we performed the same analyses using UK Biobank (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 cut-offs for age-stratified analyses. In order to understand if results could have been influenced by related individuals within the samples, we selected one individual from a pair of

relatives with PI-HAT (proportion of identity by descent calculated by PLINK (45)) >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 two 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 non-carriers. All analyses were based on mixed-effects model adjusted for age, sex and the first five genetic principal components (PCs) as fixed effects and study groups were also included as random effects to account for the 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 less than 0.05 was considered significant since all the outcomes tested were related to COVID-19 severity and not independent of each other.

Study approval

All institutions contributing cohorts to the COVID-19 Host Genetics Initiative received ethics approval from their respective research ethics review boards. Genetic modifiers for COVID-19 related illness (BelCovid_1) was approved by the Erasme Ethics committee (protocol P2020_209). 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 Ethics Committee of the Medical Faculty of the University of Bonn. BQC19 received ethical approval from the JGH research ethics board (2020-2137). The BRACOVID study has been approved by the Hospital das Clinicas, Sao Paulo University Medical School and by Brazilian National IRB, CONEP. COMRI and the COVID-19 biobank of the Faculty of Medicine at Technical University Munich received ethical approval from the local research ethics board (TUM 217/20, TUM 221/20S, TUM 440/20S). San Sebastian Hospital and Basque Biobank (COVID19-Host(a)ge_1) was approved by the Euskadi Ethics Committee on April 6, 2020 (approval number PI2020064). The study in Hospital Universitario Valle Hebron and CIberehd del Instituto Carlos III. Barcelona (COVID19-Host(a)ge_2) was approved by Vall d'Hebron Ethical Committee. COVID GWAs, Premed COVID-19 (COVID19-Host(a)ge_3) was approved by COVID GWAs (ethics id: 0886-N-20) and Premed Covid (ethics id: 1954-N-20). Genetics against coronavirus (GENIUS), Humanitas University (COVID19-Host(a)ge_4) was approved by the ethic committee (approval number reference number 316/20). FoGS was approved by the ethics committee (approval number 342_2020). The GEN-COVID is a multicentre academic observational study was approved by the IRB of each participating centre. The INMUNGEN-CoV2 study was reviewed and approved by the Ethical Committee of the Hospital Clinic of Barcelona (CEIm number: Reg.HCB/2020/0357). NorCoV2 was approved by the Regional Committee for Medical and Health Research Ethics in South-Eastern Norway (project no. 132550). SPGRX was reviewed and approved by the Valladolid Ethics Committee (PI-201716) and the Granada Ethics Committee (no number given) on March 24th, 2020 and April 13th, 2020, respectively. SweCovid was approved by the National Ethical Review Agency (EPM; 2020-01623). UK Biobank 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 HUS coordinating Ethics committee. The Columbia University Biobank was approved by the Columbia University IRB. In all studies, informed consent was received from each participant.

Author contributions:

AG and JBR contributed equally to this study. Conception and design: TN, GBL, BNJ, FG, RF, MRG, KUL, MB, SR, MEAR, ECS, THK, LV, HZ, JBR and AG. Formal analysis: TN, SP, FD, MC, GBL, DMM, BNJ, YB, MN, DE, MMB, KUL, MEAR, LV, HZ, BR and AG. Data curation: TN, FD, MC, GBL, DMM, BNJ, YB, MN, DE, MMB, SR, SA, LR, FF, CDS, FG, IFC, JCH, RF, RA, ACP, LB, JRH, IM, AR, KUL, MB, ECS, JBR and AG. Interpretation of data: TN, GBL, BNJ, SR, RF, MRG, IM, KUL, MEAR, LV, HZ, BR, and AG. Funding acquisition: DMM, SA, FF, CDS, DP, DB, FG, GD, JCH, RF, SD, MRG, JRH, IM, AR, KUL, MB, SR, MEAR, ECS, THK, JBR, and AG. Investigation: TN, GBL, DMM, BNJ, YB, RF, IM, KUL, MEAR, BR and AG. Methodology: TN, GBL, MMB, MEAR, HZ, JBR and AG. Project administration: TN, FD, DMM, SR, CDS, DP, DB, FG, GD, JCH, JB, JRH, IM, KUL, SR, ECS, AF, THK, LV, JBR and AG. Resources: FG, GD, MRG, IM, SR, MEAR, JBR and AG. Supervision: DMM, BNJ, FG, MRG, IM, KUL, SR, MEAR, JBR and AG. Validation: TN, SP, FD, DE, AK, KK, and AG. Visualization: TN and AG. Writing—original draft: TN, JBR and AG. Writing—review and editing: TN, GBL, DMM, BNJ, AP, SR, IFC, JCH, RF, KK, SD, RA, LB, JRH, IM, AR, AMP, KUL, MEAR, THK, LV, HZ, JBR and AG. All authors were involved in further drafts of the manuscript and revised it critically for content. All authors gave final approval of the version to be published. The corresponding authors attest that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Acknowledgments:

We thank the patients who volunteered to contribute to all of the participating studies in such difficult times, and the research staffs in every cohort who recruited patients at personal risk.

See Supplemental Acknowledgments for details about The COVID-19 HGI (Supplemental Table 20).

Funding:

AG has received support by NordForsk Nordic Trial Alliance (NTA) grant, by Academy of Finland Fellow grant N. 323116 and the Academy of Finland for PREDICT consortium N. 340541.

The Richards research group is supported by the Canadian Institutes of Health Research (CIHR) (365825 and 409511), the Lady Davis Institute of the Jewish General Hospital, the Canadian Foundation for Innovation (CFI), the NIH Foundation, Cancer Research UK, Genome Québec, the Public Health Agency of Canada, the McGill Interdisciplinary Initiative in Infection and Immunity and the Fonds de Recherche Québec Santé (FRQS). TN is supported by a research fellowship of the Japan Society for the Promotion of Science for Young Scientists. GBL is supported by a CIHR scholarship and a joint FRQS and Québec Ministry of Health and Social Services scholarship. JBR is supported by an FRQS Clinical Research Scholarship. Support from Calcul Québec and Compute Canada is acknowledged. TwinsUK is funded by the Welcome Trust, the Medical Research Council, the European Union, the National Institute for Health Research-funded BioResource and the Clinical Research Facility and Biomedical Research

Centre based at Guy's and St. Thomas' NHS Foundation Trust in partnership with King's College London. The Biobanque Québec COVID19 is funded by FRQS, Genome Québec and the Public Health Agency of Canada, the McGill Interdisciplinary Initiative in Infection and Immunity and the Fonds de Recherche Québec Santé. These funding agencies had no role in the design, implementation or interpretation of this study.

The COVID19-Host(a)ge study received infrastructure support from the DFG Cluster of Excellence 2167 "Precision Medicine in Chronic Inflammation (PMI)" (DFG Grant: "EXC2167"). The COVID19-Host(a)ge study was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the Computational Life Sciences funding concept (CompLS grant 031L0165). Genotyping in COVID19-Host(a)ge was supported by a philantropic donation from Stein Erik Hagen.

The COVID GWAs, Premed COVID-19 study (COVID19-Host(a)ge_3) was supported by "Grupo de Trabajo en Medicina Personalizada contra el COVID-19 de Andalucia"and also by the Instituto de Salud Carlos III (CIBERehd and CIBERER). Funding comes from COVID-19-GWAS, COVID-PREMED initiatives. Both of them are supported by "Consejeria de Salud y Familias" of the Andalusian Government. DMM is currently funded by the the Andalussian government (Proyectos Estratégicos-Fondos Feder PE-0451-2018).

The Columbia University Biobank was supported by Columbia University and the National Center for Advancing Translational Sciences, NIH, through Grant Number UL1TR001873. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or Columbia University. AK was supported by K25(K25DK128563) from NIH/NIDDK and TL1(UL1TR001873) from NIH/NCATS.

The SPGRX study was supported by the Consejería de Economía, Conocimiento, Empresas y Universidad #CV20-10150.

The GEN-COVID study was funded by: the MIUR grant "Dipartimenti di Eccellenza 2018-2020" to the Department of Medical Biotechnologies University of Siena, Italy; the "Intesa San Paolo 2020 charity fund" dedicated to the project NB/2020/0119; and philanthropic donations to the Department of Medical Biotechnologies, University of Siena for the COVID-19 host genetics research project (D.L n.18 of March 17, 2020). Part of this research project is also funded by Tuscany Region "Bando Ricerca COVID-19 Toscana" grant to the Azienda Ospedaliero Universitaria Senese (CUP I49C20000280002). The Italian Ministry of University and Research for funding within the "Bando FISR 2020" in COVID-19 and the Istituto Buddista Italiano Soka Gakkai for funding the project "PAT-COVID: Host genetics and pathogenetic mechanisms of COVID-19" (ID n. 2020-2016_RIC_3). Authors are grateful to: the CINECA consortium for providing computational resources; the Network for Italian Genomes (NIG) (http://www.nig.cineca.it) for its support; the COVID-19 Host Genetics Initiative (https://www.covid19hg.org/); the Genetic Biobank of Siena, member of BBMRI-IT, Telethon Network of Genetic Biobanks (project no. GTB18001), EuroBioBank, and RD-Connect, for managing specimens.

Genetics against coronavirus (GENIUS), Humanitas University (COVID19-Host(a)ge_4) was supported by Ricerca Corrente (Italian Ministry of Health), intramural funding (Fondazione Humanitas per la Ricerca). The generous contribution of Banca Intesa San Paolo and of the Dolce&Gabbana Fashion Firm is gratefully acknowledged.

Data acquisition and sample processing was supported by COVID-19 Biobank, Fondazione IRCCS Cà Granda Milano; LV group was supported by MyFirst Grant AIRC n.16888, Ricerca

Finalizzata Ministero della Salute RF-2016-02364358, Ricerca corrente Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, the European Union (EU) Programme Horizon 2020 (under grant agreement No. 777377) for the project LITMUS- "Liver Investigation: Testing Marker Utility in Steatohepatitis", Programme "Photonics" under grant agreement "101016726" for the project "REVEAL: Neuronal microscopy for cell behavioural examination and manipulation", Fondazione Patrimonio Ca' Granda "Liver Bible" PR-0361. DP was supported by Ricerca corrente Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, CV PREVITAL "Strategie di prevenzione primaria nella popolazione Italiana" Ministero della Salute, and Associazione Italiana per la Prevenzione dell'Epatite Virale (COPEV). DB was supported by "Programa Estratégico Instituto de Biología y Genética Molecular (IBGM) Junta de Castilla y León" (CCVC8485), "Proyectos COVID-19 de la Junta de Castilla y León" (07.04.467B04.74011.0), "Consejo Superior de Investigaciones científicas" (CSIC-COV19-016/202020E155) and the European Union -NextGenerationEU-.

Genetic modifiers for COVID-19 related illness (BeLCovid_1) was supported by the "Fonds Erasme". The Host genetics and immune response in SARS-Cov-2 infection (BelCovid_2) study was supported by grants from Fondation Léon Fredericq, from the Walloon region and from Fonds de la Recherche Scientifique (FNRS).

The INMUNGEN-CoV2 study was funded by the Spanish National Research Council (CSIC, grant n. 202020E086), and received samples from the COVIDBANK of the HCB-IDIBAPS Biobank funded in part by Fundació Glòria Soler.

KUL is supported by the German Research Foundation (LU 1944/3-1).

SweCovid is funded by the SciLifeLab/KAW national COVID-19 research program project grant to Michael Hultström (KAW 2020.0182) and the Swedish Research Council (2014-02569

and 2014-07606) as well as The Swedish Kidney Foundation (F2020-0054) to RF. HZ is supported by Jeansson Stiftelser, Magnus Bergvalls Stiftelse.

The COMRI cohort is funded by Technical University of Munich, Munich, Germany. Genotyping for the COMRI cohort was performed and funded by the Genotyping Laboratory of Institute for Molecular Medicine Finland FIMM Technology Centre, University of Helsinki, Helsinki, Finland.

NorCoV2 study supported by grants from Research Council of Norway grant no 312780, and a philanthropic donation from Vivaldi Invest A/S owned by Jon Stephenson von Tetzchner.

These funding agencies had no role in the design, implementation or interpretation of this study.

Data and materials availability:

All code for data management and analysis is archived online at https://github.com/tomoconaka/COVID19-chr3 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, SPGRX), and Sweden (SweCovid) was deposited at the European Genome-phenome Archive (EGA) under EGAS00001005304.

Regarding the data from genetic modifiers for COVID-19 related illness (BelCovid_1), individual level data were acquired and shared with FIMM during the sanitary crisis under an emergency consent and an ethical approval which were specific to this particular project and do not cover deposition to public repositories. Upon contact with Françoise Wilkin (Françoise.Wilkin@erasme.ulb.ac.be), Isabelle Migeotte (Isabelle.Migeotte@erasme.ulb.ac.be),

or Guillaume Smits (Guillaume.Smits@erasme.ulb.ac.be), 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 2-4 months from initial contact.

Regarding the BoSCO study, individual-level genotype and clinical data for 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) or Markus M. Nöthen (markus.noethen@uni-bonn.de), an institutional data transfer agreement can be established and data shared if the aims of data use is covered by ethical approvals 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 COMRI cohort, data protection legislation does not allow for deposition of individual level data in public repositories. Upon direct contact with Prof Ulrike Protzer (protzer@tum.de, genetic data) and Dr 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 approvals 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 2-3 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 a specific consent. Participant written consent also does not cover public sharing of data for use for unknown purposes. Upon contact with professor 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 approvals and patient consent. The procedure will involve the request for an amendment to the ethical approvals, as well as review by legal departments at both institutions and the process will typically take 1-2 months from initial contact.

Regarding Norwegian data (NorCoV2), 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 professor Tom H Karlsen (t.h.karlsen@medisin.uio.no) or professor 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 is covered by ethical approvals and patient consent. The procedure will involve an update to the ethical approvals, as well as review by legal departments at both institutions and the process will typically take 1-2 months from initial contact.

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

References

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 [Internet]. *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 [Internet]. *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 [Internet]. *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 [Internet]. *Science (80-.).* 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 [Internet]. *Hum. Genomics 2021 151* 2021;15(1):1–13.

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

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

8. Chaudhry R, Dranitsaris G, Mubashir T, Bartoszko J, Riazi S. A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes [Internet]. *EClinicalMedicine* 2020;25:100464.

9. Baric RS. Emergence of a Highly Fit SARS-CoV-2 Variant [Internet]. *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 [Internet]. *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 [Internet]. *Eur. Heart J.* 2013;34(45):3478–3490.

12. Feng Y-CA et al. Findings and insights from the genetic investigation of age of first reported occurrence for complex disorders in the UK Biobank and FinnGen [Internet]. *medRxiv* 2020;2020.11.20.20234302.

13. Olarte L et al. Apolipoprotein E ε4 and age at onset of sporadic and familial Alzheimer disease in Caribbean Hispanics [Internet]. *Arch. Neurol.* 2006;63(11):1586–1590.

14. The Severe Covid-19 GWAS Group. Genomewide Association Study of Severe Covid-19 with Respiratory Failure [Internet]. *N. Engl. J. Med.* 2020;NEJMoa2020283.

15. Pairo-Castineira E et al. Genetic mechanisms of critical illness in Covid-19 [Internet]. *Nature* 2020;1–1.

16. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19 [Internet]. *Nature* 2021;1–8.

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

18. Bastard P et al. Autoantibodies against type I IFNs in patients with life-threatening COVID-19 [Internet]. *Science (80-.).* 2020;370(6515). doi:10.1126/SCIENCE.ABD4585

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

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

21. Auton A et al. A global reference for human genetic variation [Internet]. *Nature* 2015;526(7571):68–74.

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

23. Merrill JT, Erkan D, Winakur J, James JA. Emerging evidence of a COVID-19 thrombotic syndrome has treatment implications [Internet]. *Nat. Rev. Rheumatol.* 2020;16(10):581–589.

24. 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.* [published online ahead of print: January 12, 2021]; doi:10.1016/j.livres.2021.01.001

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

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

27. Polack FP et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine [Internet]. *N. Engl. J. Med.* 2020;383(27):2603–2615.

28. 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.

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

30. 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. [Internet]. *Signal Transduct. Target. Ther.* 2021;6(1):85.

31. Smieszek SP, Polymeropoulos MH. Role of FYVE and Coiled-Coil Domain Autophagy Adaptor 1 in severity of COVID-19 1 infection 2 [Internet]. *medRxiv* 2021;2021.01.22.21250070.

32. Payne DJ et al. The CXCR6/CXCL16 axis links inflamm-aging to disease severity in COVID-19 patients [Internet]. *bioRxiv* 2021;2021.01.25.428125.

33. Khalil BA, Elemam NM, Maghazachi AA. Chemokines and chemokine receptors during COVID-19 infection. *Comput. Struct. Biotechnol. J.* 2021;19:976–988.

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

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

36. Fry A et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population [Internet]. *Am. J. Epidemiol.* 2017;186(9):1026–1034.

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

38. Miyara M et al. Low incidence of daily active tobacco smoking in patients with symptomatic COVID-19 [Internet]. *Qeios* [published online ahead of print: April 21, 2020]; doi:10.32388/WPP19W.3

39. Di Maria E, Latini A, Borgiani P, Novelli G. Genetic variants of the human host influencing the coronavirus-associated phenotypes (SARS, MERS and COVID-19): Rapid systematic review and field synopsis [Internet]. *Hum. Genomics* 2020;14(1):30.

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

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

42. Données COVID-19 par âge et sexe au Québec | INSPQ [Internet]https://www.inspq.qc.ca/covid-19/donnees/age-sexe. cited February 14, 2021

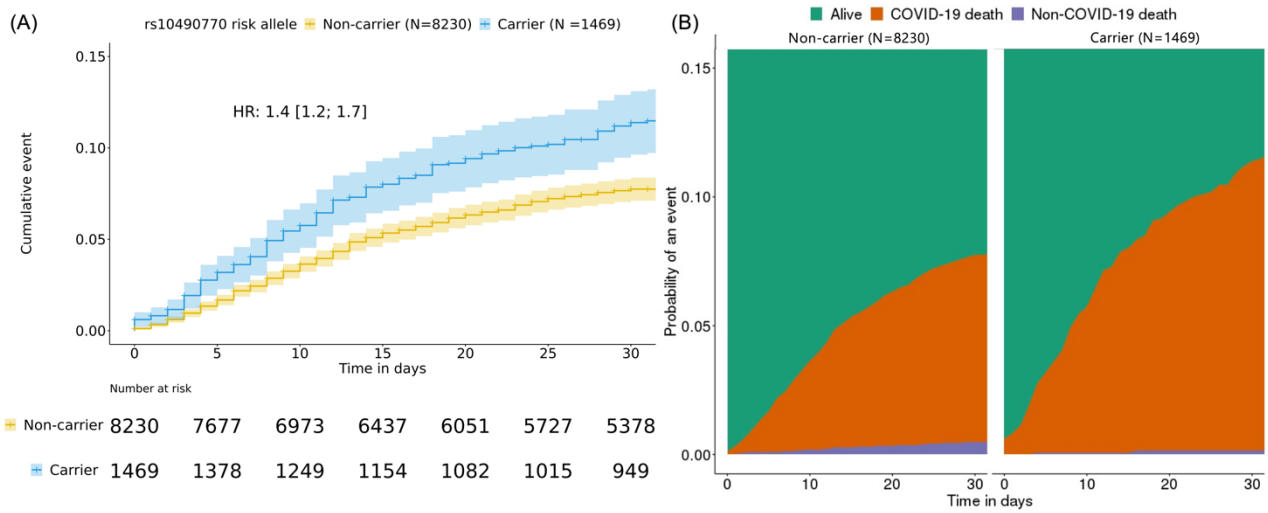
43. Signorelli C, Odone A. Age-specific COVID-19 case-fatality rate: no evidence of changes over time [Internet]. *Int. J. Public Health* 2020;65(8):1435–1436.

44. Certain Medical Conditions and Risk for Severe COVID-19 Illness | CDC [Internet]https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html. cited January 29, 2021

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

Figure Legends

Figure 1. Associations with mortality.

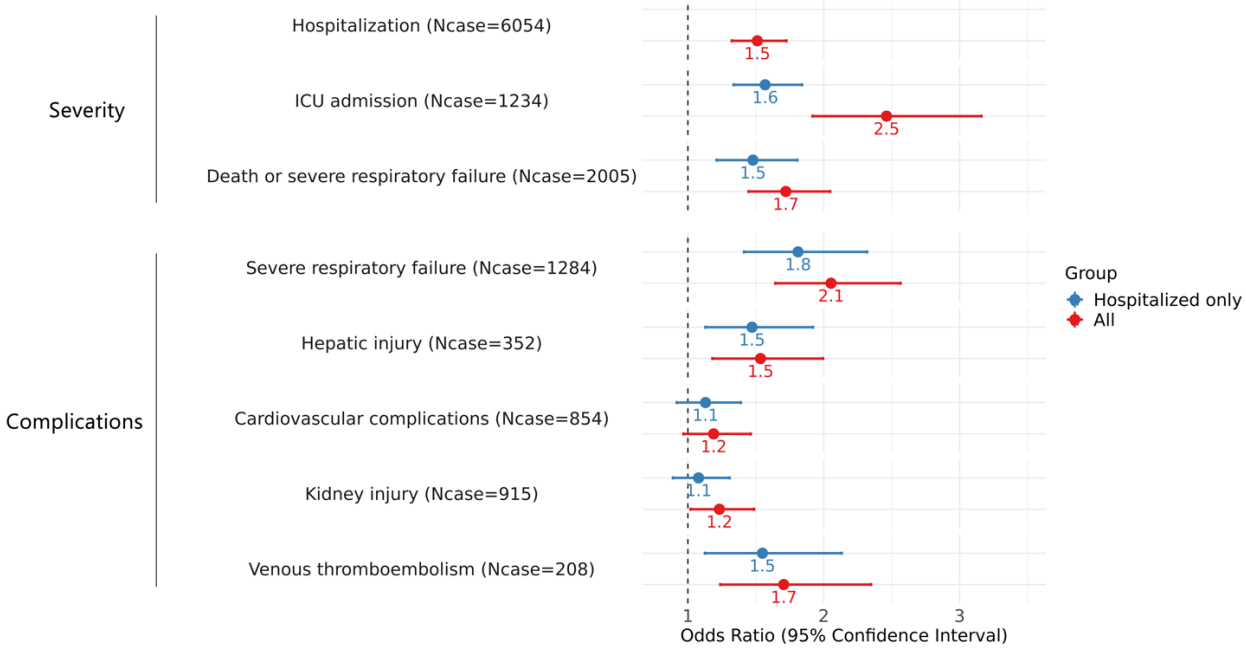


The results described here were restricted to 9,699 COVID-19 patients of European ancestry with available follow-up and cause of death information.

(A) Survival analysis using Cox-proportional hazard model. Kaplan-Meier curves stratified by rs10490770 risk allele carrier status. (Carriers: N=1,469 vs non-carriers: N=8,230). Hazard ratios (HR) were calculated by adjusting for age, sex, genetic PCs 1 to 5 as fixed effects, and groups indicating participating studies as random effects.

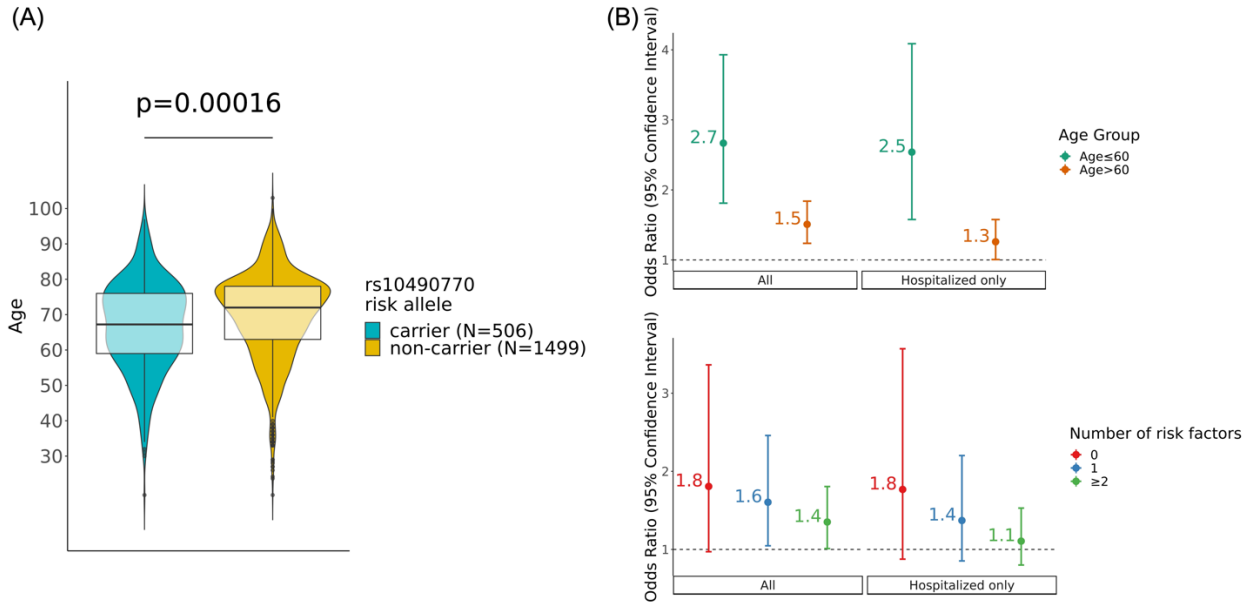
(B) Cumulative incidence curves for COVID-19 related death and COVID-19 unrelated death amongst the same individuals as described in (A).

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, genetic PCs 1 to 5 as fixed effects, and groups indicating participating studies as random effects. Red: All participants (N=12,091) Blue: Hospitalized participants only (N=6,054). 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 hospitalized-only were described in the Supplemental Table 5.

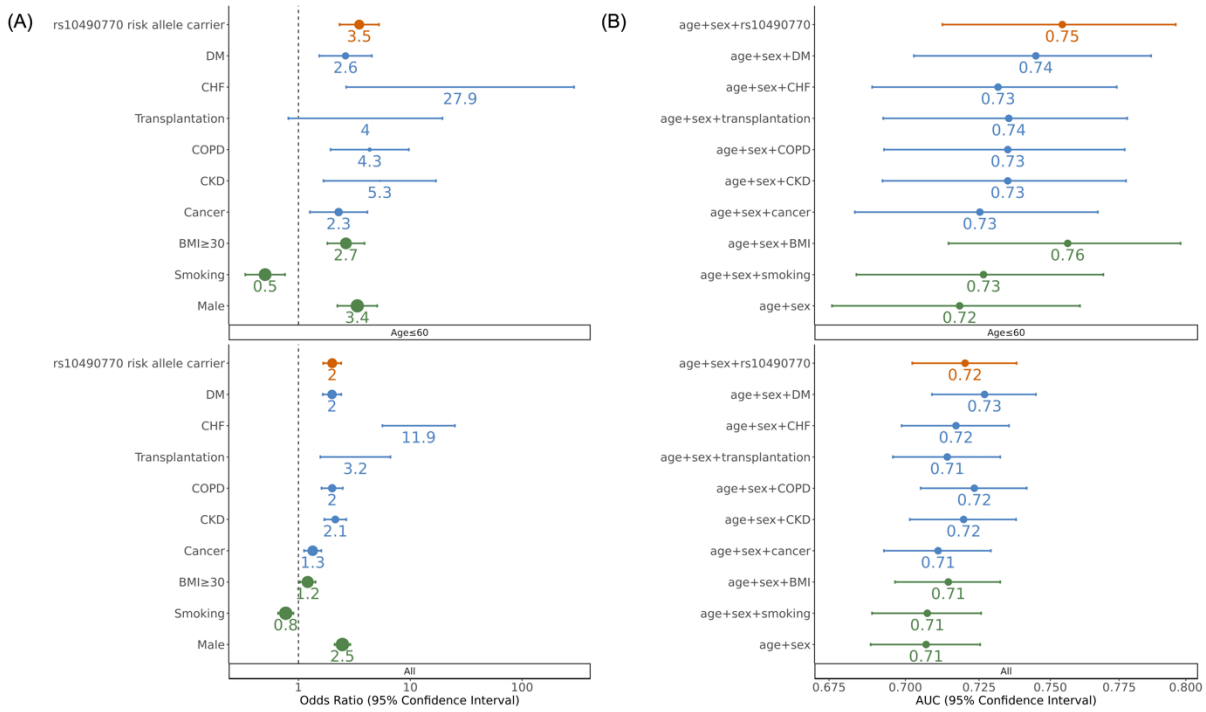
Figure 3. Influence of age and clinical risk factors for the effect of rs10490770 risk allele carrier status on death or severe respiratory failure.



(A) Age distribution in COVID-19 patients of European ancestry who died or experienced severe respiratory failure (N=2,005). Median (IQR) age was 67.2 (59-76) years in carriers (N=506) and 72 (63-78) years in non-carriers (N=1,499).

(B) Odds ratios 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 = 2,005 / 7,047) or by the number of established risk factors (0, 1, or ≥2); BMI≥30, smoking, cancer, chronic kidney disease, chronic obstructive pulmonary disease (COPD), chronic heart failure, transplantation, and diabetes mellitus (Cases / Controls = 898 / 6,454). All analyses were adjusted for age, sex, genetic PCs 1 to 5 as fixed effects, and groups indicating participating studies as random effects.

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=7,352 for all and N = 2,499 for Age ≤ 60), CKD: chronic kidney disease, COPD: chronic obstructive pulmonary disease, CHF: chronic heart failure, DM: diabetes mellitus. Error bars indicate 95% confidence intervals.

(A) Forest plots comparing odds ratios 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 non-genetic clinical risk factors were included separately in addition to age and sex in multivariable regression models. Error bars indicate 95% confidence intervals.

Table 1. Patients' characteristics.

	Hospitalized (N=7,185)	**Total (N=13,888)**
Female	2,866 (39.9%)	6,549 (47.2%)
Age (years)*	64.8 (14.7)	63.7 (12.8)
Ancestry		
European	6,054 (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	1,695 (24.3%)	1,695 (12.5%)
Death Status		
Survived	4,887 (79.3%)	11,369 (90.0%)
Deceased	1,264 (20.5%)	1,264 (10.0%)
Respiratory failure		
Severe respiratory failure	1,704 (30.2%)	1,704 (14.6%)
Oxygen supplementation	2,051 (36.4%)	2,051 (17.6%)
Hepatic injury	532 (10.8%)	536 (4.7%)
Cardiovascular complications	1,017 (19.6%)	1,040 (9.3%)
Kidney injury	1,172 (21.8%)	1,182 (10.0%)
Venous thromboembolism	288 (6.9%)	289 (2.7%)

Age*: Mean (SD), % was calculated amongst those with complete information. The missing rates per each study are listed in Supplemental Table 1. Others in ancestry were the rest of individuals who were not assigned as either of European, South Asian, African, Eat Asian or Admixed American descent.

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)
non-carrier	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% (1,607)	100% (7,047)
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)
non-carrier	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)
non-carrier	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. [95%CI] (Sample size)

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

Age range	Model	AUC †	AUC p-value*	NRI †	NRI p-value*
All Cases = 898 Controls = 6,454	Baseline	0.76 [0.75; 0.78]	-	-	-
	Baseline and rs10490770	0.77 [0.76; 0.79]	$1.4x10^{-4}$	0.19 [0.13; 0.25]	$4.4x10^{-11}$
Age≤60 Cases = 151 Controls = 2,348	Baseline	0.82 [0.79; 0.86]	-	-	-
	Baseline and rs10490770	0.84 [0.81; 0.88]	$2.1x10^{-2}$	0.41 [0.26; 0.56]	$7.7x10^{-8}$

Only individuals with complete information of clinical risk factors and genotype were included. Baseline model includes age, sex, BMI, smoking status (ever-smoker vs never-smoker), cancer, chronic kidney disease, chronic obstructive pulmonary disease (COPD), chronic heart failure, transplantation, and diabetes mellitus. *p-values were calculated by comparing baseline model and baseline and rs10490770 model. †: [95%CI]