medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Genome-wide analysis in 756,646 individuals provides first genetic evidence that *ACE2* expression influences COVID-19 risk and yields genetic risk scores predictive of severe disease

J. E. Horowitz[1†], J. A. Kosmicki[1†], A. Damask[1†], D. Sharma[1], G. H. L. Roberts[2], A. E. Justice[3], N. Banerjee[1], M. V. Coignet[2], A. Yadav[1], J. B. Leader[3], A. Marcketta[1], D. S. Park[2], R. Lanche[1], E. Maxwell[1], S. C. Knight[2], X. Bai[1], H. Guturu[2], D. Sun[1], A. Baltzell[2], F. S. P. Kury[1], J. D. Backman[1], A. R. Girshick[2], C. O'Dushlaine[1], S. R. McCurdy[2], R. Partha[2], A. J. Mansfield[1], D. A. Turissini[2], A. H. Li[1], M. Zhang[2], J. Mbatchou[1], K. Watanabe[1], L. Gurski[1], S. E. McCarthy[1], H. M. Kang[1], L. Dobbyn[1], E. Stahl[1], A. Verma[4], G. Sirugo[4], Regeneron Genetics Center*, M. D. Ritchie[4], M. Jones[1], S. Balasubramanian[1], K. Siminovitch[1], W. J. Salerno[1], A. R. Shuldiner[1], D. J. Rader[4], T. Mirshahi[3], A. E. Locke[1], J. Marchini[1], J. D. Overton[1], D. J. Carey[3], L. Habegger[1], M. N. Cantor[1], K. A. Rand[2], E. L. Hong[2], J. G. Reid[1], C. A. Ball[2], A. Baras[1‡], G. R. Abecasis[1‡], M. A. Ferreira[1‡]

From:

[1]Regeneron Genetics Center, 777 Old Saw Mill River Rd., Tarrytown, NY 10591, USA

[2]AncestryDNA, 1300 West Traverse Parkway, Lehi, UT 84043, USA

[3]Geisinger, Danville, PA, 17822, USA

[4]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA

*A complete list of investigators is provided in the Supplementary Appendix.

[†] J. E. Horowitz, J. A. Kosmicki and A. Damask contributed equally to this manuscript.

[‡]A. Baras, G. R. Abecasis and M. A. Ferreira jointly supervised this work.

Correspondence to: manuel.ferreira@regeneron.com

This research has been conducted using the UK Biobank Resource (Project 26041)

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

ABSTRACT

SARS-CoV-2 enters host cells by binding angiotensin-converting enzyme 2 (ACE2). Through a genome-wide association study, we show that a rare variant (MAF = 0.3%, odds ratio 0.60, P=4.5x10^{-13}) that down-regulates *ACE2* expression reduces risk of COVID-19 disease, providing human genetics support for the hypothesis that ACE2 levels influence COVID-19 risk. Further, we show that common genetic variants define a risk score that predicts severe disease among COVID-19 cases.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

MAIN TEXT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causes coronavirus disease 2019 (COVID-19), which has lead to >3 million deaths worldwide since December 2019 [1]. Reported risk factors for severe COVID-19, defined here as death or hospitalization combined with respiratory failure [2], include male sex, older age, race, obesity, kidney, cardiovascular and respiratory diseases [3-5]. In this study, we used human genetics to identify genetic variants associated with severe COVID-19 and tested the utility of genetic risk scores to identify individuals at highest risk of severe disease.

We performed genome-wide association studies (GWAS) of COVID-19 outcomes across 52,630 individuals with COVID-19 and 704,016 individuals with no record of SARS-CoV-2 infection aggregated from four studies (Geisinger Health System [GHS], Penn Medicine BioBank [PMBB], UK Biobank [UKB] and AncestryDNA; **Supplementary Table 1**) and five continental ancestries. Of the COVID-19 cases, 6,911 (13.1%) were hospitalized and 2,184 (4.1%) had severe disease; hospitalized patients were more likely to be older, of non-European ancestry and to have pre-existing cardiovascular and lung disease (**Supplementary Table 2)**. Using these data, we defined two groups (risk and severity) of COVID-19 outcomes, ultimately resulting in five case-control comparisons related to risk of infection and two others related to disease severity among COVID-19 cases (**Table 1** and **Supplementary Table 3).** For each comparison, we performed ancestry-specific GWAS in each study and then combined results using a fixed-effects meta-analysis. Genomic inflation factors (λ_{GC}) for the meta-analyses were <1.05, suggesting no substantial impact of population structure or unmodeled relatedness (**Supplementary Table 4).**

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Our analysis provides independent support for several risk variants reported in previous GWAS of COVID-19 (**Supplementary Table 5**), including those recently reported by the COVID-19 Host Genetics Initiative (HGI) [6], to which we contributed an earlier version of these data (**Supplementary Table 6**). Details for these replicated loci follow below, but first we looked for novel genetic associations that might have been missed by the HGI. Across the seven risk and severity phenotypes, considering both common (MAF>0.5%, up to 13 million) and rare (MAF<0.5%, up to 76 million) variants, we observed one previously unreported association at a conservative $P<8\times10^{-11}$ (Bonferroni correction for seven phenotypes x 89 million variants). This association was between lower risk of infection and rs190509934:C on the X-chromosome (MAF=0.3%, OR=0.60, 95% CI 0.52-0.69, $P = 4.5\times10^{-13}$; **Figure 1A and 1B**). This rare variant is located on the X chromosome, 60 base pairs upstream of the angiotensin-converting enzyme 2 gene (*ACE2*), the primary cell entry receptor for SARS-CoV-2 [7].

Given the potential significance of these findings, we studied the association between the *ACE2* variant rs190509934 and COVID-19 outcomes in greater detail. We found that the variant was well imputed (imputation info score >0.6 for all studies), and that there was no evidence for differences in effect size (heterogeneity test $P>0.05$) across studies (**Figure 1C**) or ancestries (**Supplementary Table 7**). However, a significantly stronger association ($P=0.009$) was observed in males (OR=0.49, $P=7.0\times10^{-11}$; explaining 0.085% of the variance in disease liability [8], h^2) when compared to females (OR=0.72, $P=5\times10^{-4}$; $h^2=0.017\%$). There were no associations between rs190509934 and clinical risk factors for COVID-19 after correcting for multiple testing (**Supplementary Table 8**), suggesting that these were not likely confounders in the analysis. We then investigated the association between rs190509934 and severity among COVID-19 cases, and

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

found that carriers of rs190509934:C had numerically (but not significantly) lower risk of worse disease outcomes when compared to non-carriers (**Supplementary Table 9**). These results demonstrate that rs190509934 near *ACE2* confers protection against SAR-CoV-2 infection and potentially also modulates disease severity among infected individuals; since the variant is rare, a definite account of its role on disease severity will require larger numbers of severe cases.

We speculated that the protective rare variant near *ACE2* (rs190509934:C) might regulate *ACE2* expression. This variant was not characterized by GTEx [9] or other gene expression studies we queried (**Supplementary Table 10**). Thus, to test its association with *ACE2* expression, we analyzed RNA-seq data from liver tissue available in a subset of 2,035 individuals from the GHS study, including eight heterozygous and one hemizygous carrier for rs190509934:C. After adjusting for potential confounders (e.g. body mass index, liver disease), we found that rs190509934:C reduced *ACE2* expression by 0.87 standard deviation units (95% CI -1.18 to -0.57, P=2.7x10^{-8}; **Figure 1D**). When considering raw, pre-normalized *ACE2* expression levels, rs190509934:C was associated with a 39% reduction in expression relative to non-carriers. There was no association with the expression of 12 other nearby genes (within 500 kb) after accounting for multiple testing. These results are consistent with rs190509934:C lowering *ACE2* expression, which in turn confers protection from SARS-CoV-2 infection.

In addition to its role in viral infections, the normal physiologic role of ACE2 involves its hydrolysis and clearance of angiotensin II, a vasoconstrictive peptide that can lead to higher vascular tone or blood pressure [10]. Therefore, we investigated if rs190509934:C was associated with higher systolic blood pressure in the UKB study, but found no significant association (beta=0.009 SD-units, P=0.56; **Supplementary Table 11**). There was a trend for higher blood

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

pressure among carriers of ultra-rare coding variants in *ACE2* that are predicted to be full loss-of-function (beta=0.219, *P*=0.086; **Supplementary Table 11**), assayed through exome-sequencing as described previously [11]. These results need to be confirmed in larger datasets, but suggest that ACE2 loss-of-function may modestly increase blood pressure. This should be considered if ACE2 blockade is to be developed for COVID-19 treatment, although pharmacologic inhibition of ACE2 in such a setting would be expected to be short term and elevations in blood pressure could be managed with anti-hypertensives. Of note, *ACE2* expression in the airways was reported to be higher in smokers and patients with COPD [12] and to increase with age [13]. Collectively, these observations and our genetic findings are consistent with the hypothesis that ACE2 levels play a key role in determining COVID-19 risk.

As noted, our GWAS also identified associations at several loci reported in previous GWAS of COVID-19 outcomes. To explore previously reported signals in detail, we first attempted to replicate eight independent associations (r^2<0.05) with disease risk (**Supplementary Table 5)** reported in three recent GWAS that included >1,000 cases (**Supplementary Table 6**). After accounting for multiple testing, six variants had a significant (*P*<0.0012) and directionally consistent association in at least one of our five disease risk analyses (**Supplementary Table 12**), specifically those located in/near *LZTFL1*, *SLC6A20,* MHC, *ABO, DPP9* and *IFNAR2.* There was no evidence for heterogeneity in effect sizes across studies (**Supplementary Table 12**) or ancestries (**Supplementary Table 13**). We also explored the possibility that the association between these six variants and COVID-19 risk could have been confounded by disease status for relevant comorbidities. We found that only two of the six variants were modestly associated with a clinical risk factor (**Supplementary Table 8**), and we conclude that it is unlikely that the

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

126 association between the six variants and risk of COVID-19 is explained by these underlying
127 comorbidities.

128 To evaluate whether genetics could be used to predict severe disease, we first investigated which
129 replicated variants were associated with severity amongst COVID-19 cases. Among the six
130 replicated variants (in/near *LZTFL1*, *SLC6A20*, MHC, *ABO*, *DPP9* and *IFNAR2*), four were
131 significantly (P<0.05) associated with worse outcomes among infected individuals (in/near
132 *LZTFL1*, MHC, *DPP9* and *IFNAR2*), while those in *ABO* and near *SLC6A20* did not associate with
133 COVID-19 severity (**Supplementary Figure 1B** and **Supplementary Table 14**). Collectively,
134 these results highlight four variants associated with both COVID-19 risk and worse disease
135 outcomes, including respiratory failure and death. These variants may be used to identify
136 individuals at risk of severe COVID-19 and to guide the search for genes involved in the
137 pathophysiology of COVID-19.
138
139 We next evaluated whether variants identified by the COVID-19 HGI, a large worldwide effort to
140 identify genetic risk factors for COVID-19, could augment this set of four disease severity variants.
141 The latest HGI analyses[6] include data from 49,562 SARS-CoV-2 infected individuals and use >1.7
142 million individuals with no record of infection as controls (**Supplementary Table 15**). To identify
143 additional variants associated with severity, we started with variants associated with the phenotype
144 "reported infection" (infected vs. no record of infection) which, despite the sample overlap
145 between the HGI and our analyses, is statistically independent from severity among infected
146 individuals – since infection status (positive vs. negative or unknown) is uncorrelated with
147 hospitalization status once infected (hospitalized vs. not hospitalized). We found that two were
148 nominally associated with risks of hospitalization or severe disease among cases (rs11919389 near

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

149 *RPL24* and rs1886814 near *FOXP4*; **Supplementary Table 15**), suggesting that these loci also
150 modulate disease severity after infection with SARS-CoV-2.
151
152 Collectively, our association analyses highlighted six common variants identified in previous
153 GWAS or by the HGI - in/near *LZTFL1*, MHC, *DPP9*, *IFNAR2*, *RPL24* and *FOXP4* - that are
154 associated with risk of COVID-19 as well as disease severity among cases. To help identify genes
155 that might underlie the observed associations, we searched for functional protein-coding variants
156 (missense or predicted loss-of-function) in high LD (r^2>0.80) with each variant. We found eight
157 functional variants in five genes (**Supplementary Table 16**): *IFNAR2*, a cytokine receptor
158 component in the anti-viral type 1 interferon pathway, which is activated by SARS-CoV-2 and is
159 dysregulated in severe COVID-19 cases [14-16]; *CCHCR1*, a P-body protein associated with
160 cytoskeletal remodeling and mRNA turnover [17,18]; *TCF19*, a transcription factor associated with
161 hepatitis B [19]; and *C6orf15* and *PSORS1C1*, two functionally uncharacterized genes in the MHC.
162 These data indicate that the risk variants identified may have functional effects on these five genes.
163
164 Next, we asked if any of the six sentinel risk variants co-localized (*i.e.* were in high LD, r^2>0.80)
165 with published sentinel expression quantitative trait loci (eQTL) across 52 studies
166 (**Supplementary Table 10**), specifically focusing on 114 genes in *cis* (± 500 kb). We found co-
167 localization with sentinel eQTL for eight genes (**Supplementary Table 17)**: *SLC6A20* (eQTL
168 from lung), a proline transporter that binds the host SARS-CoV-2 receptor, ACE2 [20]; *NXPE3*
169 (esophagus), a gene of uknown function; *SENP7* (blood), a SUMO-specific protease that promotes
170 interferon signaling and that in mice is essential for innate defense against herpes simplex virus 1
171 infection[21]; *IFNAR2* and *TCF19* (multiple tissues), both discussed above; *LST1* (blood), an

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

immune modulatory protein that inhibits lymphocyte proliferation [22] and is upregulated in response to bacterial ligands [23]; *HLA-C* (adipose), a natural killer cell ligand, associated with viral infection [24] and autoimmunity [25]; and *IL10RB* (multiple tissues), a pleotropic cytokine receptor associated with persistent hepatitis B and autoimmunity [26,27].

Collectively, analysis of missense variation and eQTL catalogs suggests 12 potential effector genes in COVID-19 loci (*ACE2*, *C6orf15*, *CCHCR1*, *HLA*-C, *IFNAR2*, *IL10RB*, *LST1*, *NXPE3*, *PSORS1C1, SENP7*, *SLC6A20* and *TCL19*), though functional studies are required to confirm these predictions.

Next, we proceeded to evaluate if common genetic variants can help identify individuals at high risk of severe disease. Specifically, we focused on the six variants (in/near *LZTFL1*, MHC, *DPP9*, *IFNAR2, RPL24* and *FOXP4*) that associated with both risk of COVID-19 and disease severity among infected individuals. Using these variants, we created a weighted genetic risk score (GRS) for individuals with COVID-19 and then compared the risk of hospitalization and severe disease between those with a high GRS and all other cases, after adjusting for established risk factors. The weights used for each variant corresponded to the effect size (log of the odds ratio) reported in previous GWAS. When considering COVID-19 cases of European ancestry (N=44,958), we found that having a high GRS (top 10%) was associated with a 1.38-fold increased risk of hospitalization (95% CI 1.26 to 1.53, P=6x10^{-11}; **Figure 2A**) and 1.58-fold increased risk of severe disease (95% CI 1.36 to 1.82, P=7x10^{-10}; **Figure 2B**). In other ancestries, a high GRS also appeared to predict risk of hospitalization – including among individuals of African ancestry (N=2,598, 1.70-fold risk for high GRS, 95% CI 1.03 to 2.81, P=0.038), Admixed American ancestry (N=3,752, 1.56-fold risk, 95% CI 1.00 to 2.43, P = 0.05) and South Asian ancestry (N=760, 1.42-fold risk, 95% CI

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

194 0.72 to 2.82, P = 0.32, **Supplementary Table 18**). A similar pattern was observed for risk of
195 severe disease, though sample sizes were considerably smaller (**Supplementary Tables 19**).
196
197 We then compared the effect of the GRS between individuals with and without established risk
198 factors for severe COVID-19. In Europeans of both the AncestryDNA and UK Biobank studies,
199 we found that a high GRS (top 10%) was associated with risk of severe disease both among
200 individuals with and without established clinical risk factors for severe COVID-19 (**Figure 3**). In
201 the meta-analysis of the two studies, a high GRS was associated with a 1.65-fold (95% CI 1.39-
202 1.96, P=1x10^{-8}) and 1.75-fold (95% CI 1.28-2.40, P=4x10^{-4}) higher risk of severe disease,
203 respectively among individuals with and without established risk factors (**Supplementary Table**
204 **20**). There was no evidence for heterogeneity of GRS effect with clinical risk factor status
205 (P=0.30). Similar results were observed for (i) risk of hospitalization (**Supplementary Figure 2**
206 and **Supplementary Table 20**); (ii) when including in the GRS all 12 variants reported to associate
207 with risk of COVID-19 in previous GWAS (eight variants) and by the HGI (four novel variants
208 associated with reported infection; **Supplementary Figure 3**); and (iii) in individuals of Admixed
209 American ancestry (**Supplementary Figure 4**; stratitifed analysis not performed in other
210 ancestries due to small sample size). Lastly, we also found that expanding the GRS to include a
211 larger set of variants did not improve the observed associations (**Supplementary Figure 5**).
212 Overall, these results demonstrate that a GRS calculated using variants associated with disease risk
213 and severity can potentially be used to identify COVID-19 cases at high risk of developing poor
214 disease outcomes. We note that preeminent factor for severe COVID-19 outcomes remains age (as
215 illustrated in **Supplementary Figure 6**), but that common genetic variants appear to provide
216 complementary information that may be used to stratify risk among older individuals.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

The following caveats should be considered when interpreting results from this study. First, our study had greater power to identify associations with disease risk than with severity outcomes, given the relatively small sample size for the latter. Second, there was phenotypic heterogeneity among COVID-19 cases and controls and associated risk factors across our studies. One likely reason for this is that survey respondents from the AncestryDNA study were enriched for healthier individuals and milder COVID-19 cases, when compared to participants of the UKB, GHS and PMBB studies, who were ascertained in clinical settings and so were enriched for hospitalized and severe COVID-19 cases. Other sources of heterogeneity may include regional and temporal availability of COVID-19 testing and the inability to control for viral exposure among controls. While our meta-analysis collectively spans a broad phenotypic spectrum, these individual differences may account for variability in results across reported studies. Third, we used expression levels measured in liver to assess the impact of the ACE2 risk variant on gene expression. Liver is not the most relevant tissue to assess *ACE2* expression, but we note that cis eQTLs are often shared across tissues [9,28]. Fourth, the association between GRS and risk of severe disease was strongest in European individuals of the AncestryDNA (OR=1.72, $P=2\text{x}10^{-6}$) and UKB (OR=1.65, $P=6\text{x}10^{-6}$) studies when compared to the smaller GHS study (OR=1.03, $P=0.877$). The lower effect size in the latter may be due to differences in ascertainment of COVID-19 positive cases, as discussed above, or stochastic, given the smaller sample size. We also noted that the impact of the GRS on risk of hospitalization was attenuated in comparison to severe disease, which may be a reflection of the weighting schema for the variants comprising the score; the four largest GRS weights were derived from an analysis of critically ill individuals.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

In summary, we confirmed six common variant associations with risk of infection and further show that four of these variants modulate disease severity among cases. We also identified one novel association with disease risk which provides human genetic support for the hypothesis that ACE2 expression plays a key role SARS-CoV-2 infection and may constitute an attractive therapeutic target for prevention COVID-19 disease and its sequelae. Lastly, we demonstrate that a genetic risk score based on common variants validated in this study can be used to identify individuals at high risk of poor disease outcomes.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

ONLINE METHODS

Participating Studies

AncestryDNA COVID-19 Research Study. AncestryDNA customers over age 18, living in the United States, and who had consented to research, were invited to complete a survey assessing COVID-19 outcomes and other demographic information. These included SARS-CoV-2 swab and antibody test results, COVID-19 symptoms and severity, brief medical history, household and occupational exposure to SARS-CoV-2, and blood type. A total of 163,650 AncestryDNA survey respondents were selected for inclusion in this study [29]. Respondents selected for this study included all individuals with a positive COVID-19 test together with age and sex matched controls. DNA samples were genotyped as described previously[29]. Genotype data for variants not included in the array were then inferred using imputation to the Haplotype Reference Consortium (HRC) reference panel. Briefly, samples were imputed to HRC version 1.1, which consists of 27,165 total individuals and 36 million variants. The HRC reference panel does not include indels; consequently, indels are not present in the imputed data. We determined best-guess haplotypes with Eagle version 2.4.1 and performed imputation with Minimac4 version 1.0.1. We used 1,117,080 unique variants as input and 8,049,082 imputed variants were retained in the final data set. Variants with a Minimac4 R^2<0.30 were filtered from the analysis.

Geisinger Health System (GHS). The GHS MyCode Community Health Initiative is a health system-based cohort from central and eastern Pennsylvania (USA) with ongoing recruitment since 2006[30]. A subset of 144,182 MyCode participants sequenced as part of the GHS-Regeneron Genetics Center DiscovEHR partnership were included in this study. Information on COVID-19 outcomes were obtained through GHS's COVID-19 registry. Patients were identified as eligible

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

for the registry based on relevant lab results and ICD-10 diagnosis codes; patient charts were then reviewed to confirm COVID-19 diagnoses. The registry contains data on outcomes, comorbidities, medications, supplemental oxygen use, and ICU admissions. DNA from participants was genotyped on either the Illumina OmniExpress Exome (OMNI) or Global Screening Array (GSA) and imputed to the TOPMed reference panel (stratified by array) using the TOPMed Imputation Server. Prior to imputation, we retained variants that had a MAF >= 0.1%, missingness < 1% and HWE p-value > 10^{-15}. Following imputation, data from the OMNI and GSA datasets were merged for subsequent association analyses, which included an OMNI/GSA batch covariate, in addition to other covariates described below.

Penn Medicine BioBank (PMBB) study. PMBB contains ~70,000 study participants, all recruited through the University of Pennsylvania Health System (UPHS). Participants donate blood or tissue and allow access to EHR information [31]. The PMBB participants with COVID-19 infection were identified through the UPHS COVID-19 registry, which consists of qPCR results of all patients tested for SARS-CoV-2 infection within the health system. We then used electronic health records to classify COVID-19 patients into hospitalized and severe (ventilation or death) categories. DNA genotyping was performed with the Illumina Global Screening Array, and imputation performed using the TOPMed reference panel as described for GHS above.

UK Biobank (UKB) study. We studied the host genetics of SARS-CoV-2 infection in participants of the UK Biobank study, which took place between 2006 and 2010 and includes approximately 500,000 adults aged 40-69 at recruitment. In collaboration with UK health authorities, the UK Biobank has made available regular updates on COVID-19 status for all participants, including

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

293 results from four main data types: qPCR test for SARS-CoV-2, anonymized electronic health
294 records, primary care and death registry data. We report results based on phenotype data
295 downloaded on the 4th January 2021 and excluded from the analysis 28,547 individuals with a
296 death registry event prior to 2020. DNA samples were genotyped as described previously [32] using
297 the Applied Biosystems UK BiLEVE Axiom Array (N=49,950) or the closely related Applied
298 Biosystems UK Biobank Axiom Array (N=438,427). Genotype data for variants not included in
299 the arrays were inferred using the TOPMed reference panel, as described above.
300
301 **COVID-19 phenotypes used for genetic association analyses**
302 We grouped participants from each study into three broad COVID-19 disease categories
303 (**Supplementary Table 1**): (i) positive – those with a positive qPCR or serology test for SARS-
304 CoV-2, or with a COVID-19-related ICD10 code (U07), hospitalization or death; (ii) negative –
305 those with only negative qPCR or serology test results for SARS-CoV-2 and with no COVID-19-
306 related ICD10 code (U07), hospitalization or death; and (iii) unknown – those with no qPCR or
307 serology test results and no COVID-19-related ICD10 code (U07), hospitalization or death. We
308 then used these broad COVID-19 disease categories, in addition to hospitalization and disease
309 severity information, to create seven COVID-19-related phenotypes for genetic association
310 analyses, as detailed in **Supplementary Table 3**.
311 SARS-CoV-2 infection status (positive, negative or unknown) was determined based on a
312 qPCR test for SARS-CoV-2 in the UKB, GHS and PMBB studies; self-reported results for qPCR
313 or serology test for SARS-CoV-2 in the AncestryDNA study.
314 Hospitalization status (positive, negative or unknown) was determined based on COVID-
315 19-related ICD10 codes U071, U072, U073 in variable 'diag_icd10' (table 'hesin_diag') in the

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

316 UKB study; self-reported hospitalization due to COVID-19 in the AncestryDNA study; medical
317 records in the GHS and PMBB studies.

318 Disease severity status (severe [ventilation or death] or not severe) was determined in the
319 UKB study based on (i) respiratory support ICD10 code Z998 in variable 'diag_icd10' (table
320 'hesin_diag'); (ii) the following respiratory support ICD10 codes in variable 'oper4' (table
321 'hesin_oper'): E85, E851, E852, E853, E854,E855, E856, E858, E859, E87, E871, E872, E873,
322 E874, E878, E879, E89, X56, X561, X562, X563, X568, X569, X58, X581, X588, X589; or (3)
323 COVID-19-related ICD10 codes U071, U072, U073 in cause of death (variable 'cause_icd10' in
324 table 'death_cause'). In the AncestryDNA study, disease severity was determined based on self-
325 reported ventilation or need for supplementary oxygen due to COVID-19. In the GHS and PMBB
326 study, it was determined based on ventilator or high-flow oxygen use.

327 For association analysis in the AncestryDNA study, we excluded from the COVID-19
328 unknown group individuals who had (i) a first-degree relative who was COVID-19 positive; or (ii)
329 flu-like symptoms.

330

331 **Genetic association analyses**

332 Association analyses in each study were performed using the genome-wide Firth logistic
333 regression test implemented in REGENIE [33]. In this implementation, Firth's approach is applied
334 when the p-value from standard logistic regression score test is below 0.05. We included in step 1
335 of REGENIE (*i.e.* prediction of individual trait values based on the genetic data) directly
336 genotyped variants with a minor allele frequency (MAF) >1%, <10% missingness, Hardy-
337 Weinberg equilibrium test *P*-value>10^{-15} and linkage-disequilibrium (LD) pruning (1000 variant
338 windows, 100 variant sliding windows and r^2<0.9). The association model used in step 2 of

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

REGENIE included as covariates age, age^2, sex, age-by-sex, and the first 10 ancestry-informative principal components (PCs) derived from the analysis of a stricter set of LD-pruned (50 variant windows, 5 variant sliding windows and r^2<0.5) common variants from the array (imputed for the GHS study) data.

Within each study, association analyses were performed separately for five different continental ancestries defined based on the array data: African (AFR), Admixed American (AMR), European (EUR) and South Asian (SAS). We determined continental ancestries by projecting each sample onto reference principle components calculated from the HapMap3 reference panel. Briefly, we merged our samples with HapMap3 samples and kept only SNPs in common between the two datasets. We further excluded SNPs with MAF<10%, genotype missingness >5% or Hardy-Weinberg Equilibrium test p-value < 10^{-5}. We calculated PCs for the HapMap3 samples and projected each of our samples onto those PCs. To assign a continental ancestry group to each non-HapMap3 sample, we trained a kernel density estimator (KDE) using the HapMap3 PCs and used the KDEs to calculate the likelihood of a given sample belonging to each of the five continental ancestry groups. When the likelihood for a given ancestry group was >0.3, the sample was assigned to that ancestry group. When two ancestry groups had a likelihood >0.3, we arbitrarily assigned AFR over EUR, AMR over EUR, AMR over EAS, SAS over EUR, and AMR over AFR. Samples were excluded from analysis if no ancestry likelihoods were >0.3, or if more than three ancestry likelihoods were > 0.3.

Results were subsequently meta-analyzed across studies and ancestries using an inverse variance-weighed fixed-effects meta-analysis.

Identification of putative targets of GWAS variants based on colocalization with eQTL

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

We identified as a likely target of a sentinel GWAS variant any gene for which a sentinel expression quantitative trait locus (eQTL) co-localized (*i.e.* had LD $r^2 > 0.80$) with the sentinel GWAS variant. That is, we only considered genes for which there was strong LD between a sentinel GWAS variant and a sentinel eQTL, which reduces the chance of spurious colocalization. Sentinel eQTL were defined across 174 published datasets (**Supplementary Table 10**), as described previously [34]. We did not use statistical approaches developed to distinguish colocalization from shared genetic effects because these have very limited resolution at high LD levels ($r^2 > 0.80$) [35].

Gene expression analysis in participants of the GHS study

For a subset of individuals from the GHS study (n=2,035, ascertained through the Geisinger Bariatric Surgery Clinic), RNA was extracted from liver biopsies conducted during bariatric surgery to evaluate liver disease. Individuals had class 3 obesity (BMI>40kg/m^2) or class 2 obesity (BMI 35-39 kg/m^2) with an obesity-related co-morbidity (e.g. type-2 diabetes, hypertension, sleep apnea, non-alcoholic fatty liver disease). RNA libraries were prepared using polyA-extraction and then sequenced with 75bp paired-end reads with two 10 bp index reads on the Illumina NovaSeq 6000 on S4 flow cells. RNA-seq data were then analyzed using the GTEx v8 workflow[36], using STAR [37] and RNASeqQC [38], except that GENCODE v32 was used in lieu of v26. Briefly: (i) raw expression counts were normalized with TMM (Trimmed Mean of M-values) as implemented in edgeR [38]; (ii) a rank-based inverse normal transformation was applied to the normalized expression values; (iii) principal components (PCs) analysis was performed on data from 25,078 genes with TPM >0.1 in >20% samples, to identify latent factors accounting for variation in gene expression; (iv) gene expression levels were adjusted for the top 100 PCs to improve power to identify cis-

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

385 regulatory effects. The association between adjusted *ACE2* expression and the imputed genotypes
386 of rs190509934 was then tested using REGENIE, with the following variables included as
387 covariates: age, age^2, four ancestry-informative principal components, steatosis status, fibrosis
388 status, diabetes status, and body mass index at the time of bariatric surgery.
389
390 **Genetic risk score (GRS) analysis of COVID-19 hospitalization and severity**
391 First, in each study (AncestryDNA, GHS, UKB and PMBB), we created a GRS for each COVID-
392 19 positive individual based on variants that were reported to associate with risk of COVID-19 in
393 previous GWAS and that we (i) independently replicated (except variants identified by the HGI);
394 and (ii) found to be associated with COVID-19 severity outcomes. We used as weights the effect
395 (beta) reported in previous GWAS (**Supplementary Table 5**). Second, we ranked COVID-19
396 individuals based on the GRS and created a new binary GRS predictor by assigning each individual
397 to a high (top 5%) or low (rest of the population) percentile group. Third, for studies with >100
398 hospitalized cases, we used logistic regression to test the association between the binary GRS
399 predictor and risk of hospitalization (hospitalized cases vs. all other cases), including as covariates
400 age, sex, age-by-sex interaction, and ten ancestry-informative PCs. In addition to age and sex, we
401 included as additional covariates established clinical risk factors for COVID-19 that are outlined
402 in the Emergency Use Authorisation treatment guidelines for casirivimab and imdevimb: BMI,
403 chronic kidney disease, diabetes, immunosuppressive disease, chronic obstructive pulmonary
404 disease or other chronic respiratory disease, cardiovascular disease and hypertension. We repeated
405 the association analysis (i) using different percentile cut-offs for the GRS (5%, 10%, 20%, 30%
406 and 40%); and (ii) to test the association with disease severity (severe cases vs. all other cases).
407 We then stratified COVID-19 cases by clinical risk (high versus lower) and evaluated the

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

association between the top 10% by GRS (i.e. high genetic risk) and risk of hospitalization or severe disease. The stratified analyses were performed with logistic regression, with sex and ancestry-informative PCs included as covariates. High clinical risk was defined as any one of the following: (i) age≥65; (ii) BMI≥35; (iii) chronic kidney disease, diabetes or immunosuppressive disease; (iv) age ≥55 and presence of chronic obstructive pulmonary disease/other chronic respiratory disease, cardiovascular disease, or hypertension.

Code availability

Upload Agent (v1.5.30) can be found at https://wiki.dnanexus.com/Downloads#Upload-Agent. bcl2fastq software (v2.19.0) can be found at https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html. BWA software (v0.7.17) for read alignment can be found at http://bio-bwa.sourceforge.net. Picard software (v1.141) can be found at https://broadinstitute.github.io/picard/. Samtools (v1.7) can be found at http://www.htslib.org. WeCall (v1.1.2) can be found at https://github.com/Genomicsplc/wecall. FastQC (v0.11.8) can be found at http://www.bioinformatics/babraham.ac.uk/projects/fastqc/. Bcftools, bgzip, and tabix (v1.7) can be found at http://www.htslib.org, bgzip/tabix v1.7. pigz (v2.3.4) can be found at https://zlib.net/pigz/. Eagle (v2.4.1) can be found at https://github.com/poruloh/Eagle. Minimac4 (v1.01) can be found at https://github.com/statgen/Minimac4. GLnexus (v0.4.0) can be found at https://github.com/dnanexus-rnd/GLnexus. PLINK (v1.90b6.21) can be found at https://www.cog-genomics.org/plink2/. PRIMUS can be found at https://primus.gs.washington.edu/primusweb/. REGENIE (v2.0.1) can be found at https://github.com/statgen/METAL.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Data availability

All genotype-phenotype association results reported in this study are available for browsing using the RGC's COVID-19 Results Browser (https://rgc-covid19.regeneron.com). Data access and use is limited to research purposes in accordance with the Terms of Use (https://rgc-covid19.regeneron.com/terms-of-use).

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Competing interests

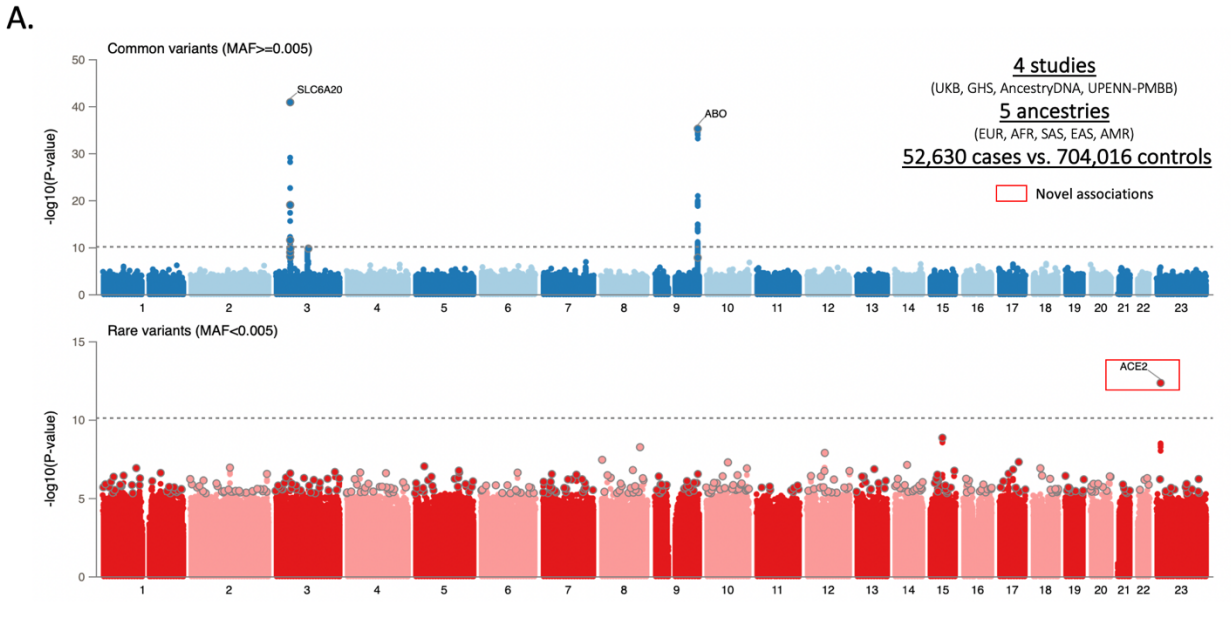
J.E.H., J.A.K., A.D., D.S., N.B, A.Y., A.M., R.L., E.M., X.B., D.S., F.S.P.K., J.D.B., C.O'D., A.J.M., D.A.T., A.H.L., J.M., K.W., L.G., S.E.M, H.M.K., L.D., E.S., M.J., S.B., K.S.M, W.J.S., A.R.S., A.E.L., J.M., J.O., L.H., M.N.C., J.G.R., A.B., G.R.A., and M.A.F. are current employees and/or stockholders of Regeneron Genetics Center or Regeneron Pharmaceuticals. G.H.L.R., M.V.C., D.S.P., S.C.K. A.Bal., A.R.G., S.R.M., R.P., M.Z., K.A.R., E.L.H., C.A.B. are current employees at AncestryDNA and may hold equity in AncestryDNA. The other authors declare no competing interests.

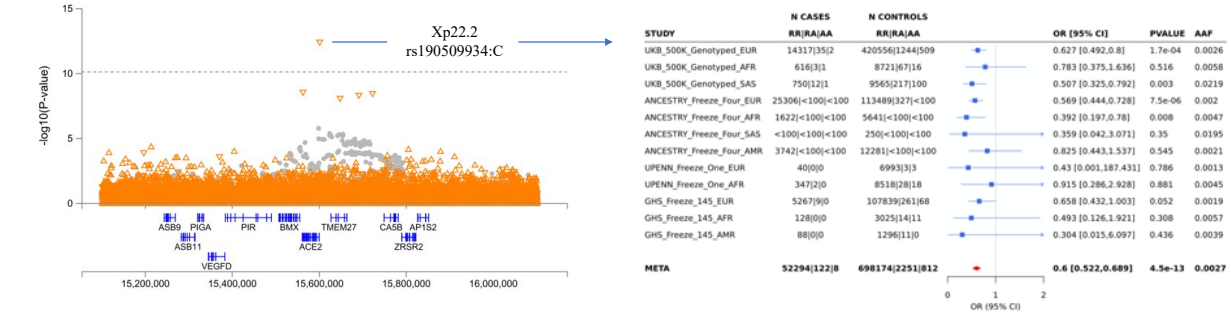
Acknowledgements

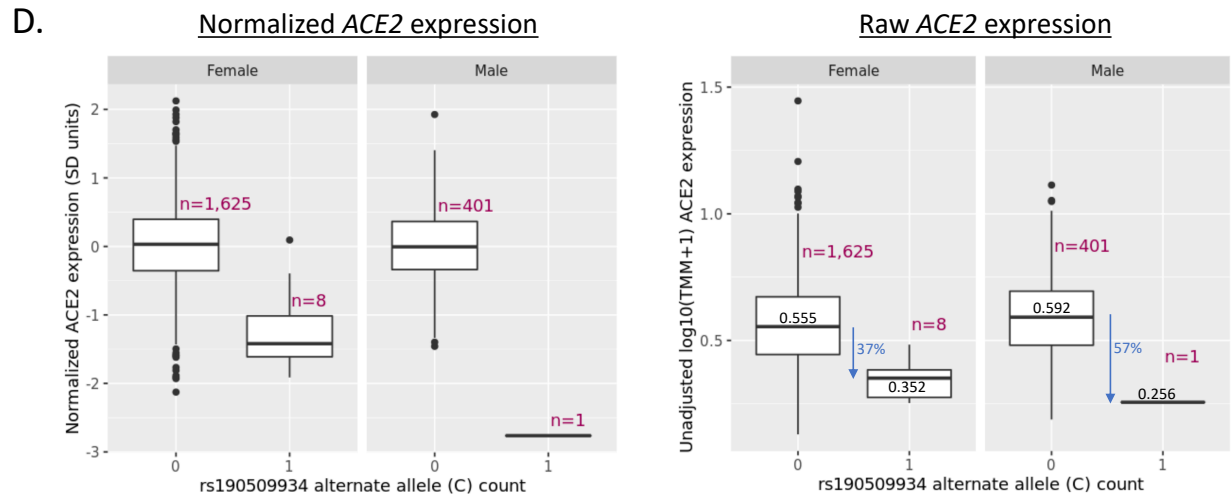
This research has been conducted using the UK Biobank Resource (Project 26041). The Penn Medicine BioBank is funded by a gift from the Smilow family, the National Center for Advancing Translational Sciences of the National Institutes of Health under CTSA Award Number UL1TR001878, and the Perelman School of Medicine at the University of Pennsylvania. We want to acknowledge the participants and investigators of the FinnGen study. We thank the AncestryDNA customers who voluntarily contributed information in the COVID-19 survey.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

FIGURES







medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Figure 1. GWAS of 52,630 COVID-19 positive cases vs. 704,016 COVID-19 negative or unknown controls identifies a novel association with a rare variant near *ACE2* that lowers gene expression and protects against COVID-19.

(A) Summary of association results for common (MAF>0.5%) and rare (MAF<0.5%) variants.

(B) Regional association plot centered around rs190509934 near ACE2.

(C) Breakdown of association results across studies included in the meta-analysis of rs190509934.

(D) Association between rs190509934:C and *ACE2* expression in liver measured in 2,035 individuals from the GHS study.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

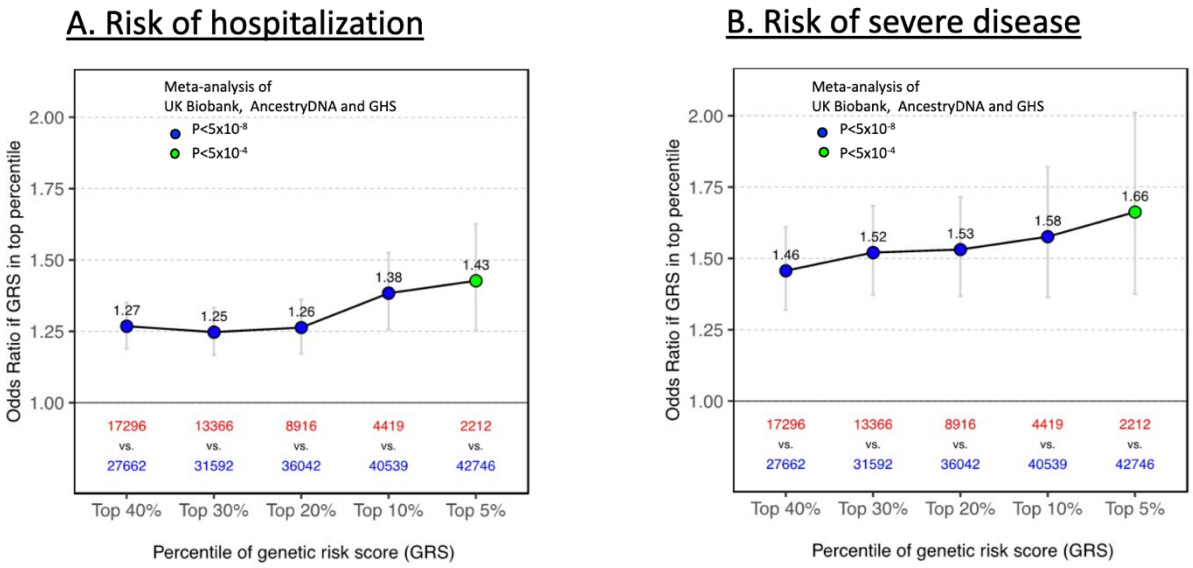


Figure 2. Association between a 6-SNP genetic risk score (GRS) and risk of hospitalization (A) and severe disease (B) among COVID-19 cases of European ancestry.

(A) Association between high genetic risk and hospitalization. Risk of hospitalization among cases is shown for individuals in the top GRS percentile, agnostic to the number of clinical risk factors present. The association was tested in three studies separately (AncestryDNA, UKB and GHS studies) using logistic regression, with established risk factors for COVID-19 included as covariates (see Methods for details). Results were then meta-analyzed across studies, for a combined sample size of 44,958 COVID-19 cases, including 6,138 hospitalized. N in red: number of COVID-19 cases in the top GRS percentile. Error bars represent 95% confidence intervals.

(B) Association between high genetic risk and severe disease. The association was tested as described above in three studies separately (AncestryDNA, UKB and GHS studies). Results were then meta-analyzed across studies, for a combined sample size of 44,958 COVID-19 cases, including 1,940 with severe disease.

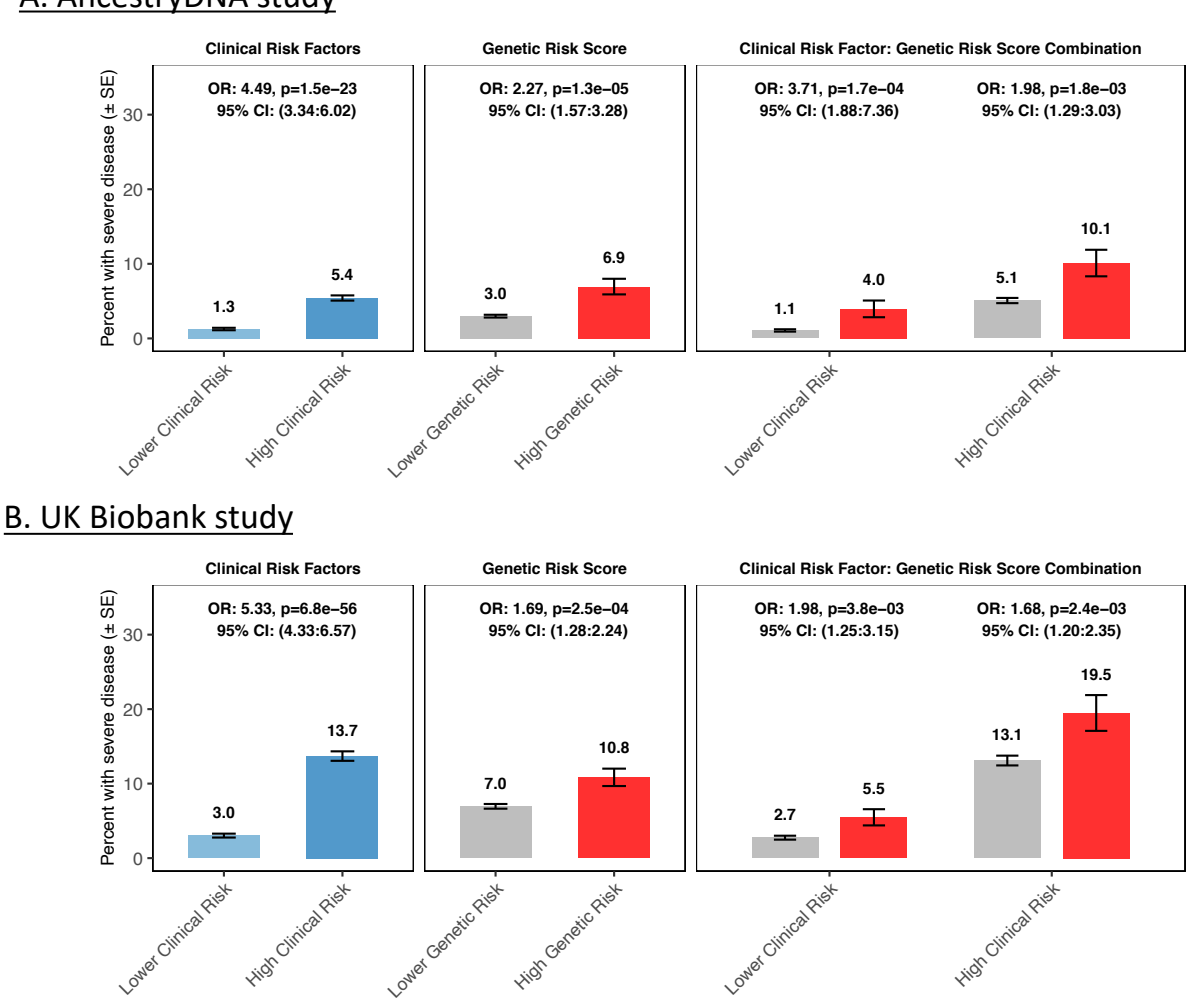
medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

482 N in red: number of COVID-19 cases in the top GRS percentile. N in grey: number of COVID-19

483 cases in the rest of population.

484

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .



485

486 **Figure 3. Association between a 6-SNP genetic risk score (GRS) and risk of severe disease**
487 **among COVID-19 cases of European ancestry after stratifying by the presence of clinical**
488 **risk factors.**

489 (A) Rate of severe disease in the AncestryDNA study (25,353 COVID-19 cases, including 667
490 with severe disease).

491 (B) Rate of severe disease in the UK Biobank study (14,320 COVID-19 cases, including 951 with
492 severe disease).

493 High genetic risk (red bars): top 10% of the GRS. Low genetic risk (grey bars): bottom 90% of the
494 GRS (*i.e.* all other COVID-19 cases). Error bars (black) represent 95% confidence intervals.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

TABLES

Table 1. Seven COVID-19 phenotypes analyzed in this study.

Broad phenotype category	Phenotype	Description	Group	Sample size with genetic data
Risk of infection	COVID-19 positive vs. COVID-19 negative or unknown	Risk of infection	Cases	52,630
			Controls	704,016
	COVID-19 positive vs. COVID-19 negative	Risk of infection among individuals tested for SARS-CoV-2	Cases	52,630
			Controls	109,605
	COVID-19 positive and not hospitalized vs. COVID-19 negative or unknown	Risk of infection that does not require hospitalization	Cases	45,641
			Controls	704,016
	COVID-19 positive and hospitalized vs. COVID-19 negative or unknown	Risk of infection that requires hospitalization	Cases	6,911
			Controls	689,620
	COVID-19 positive and severe vs. COVID-19 negative or unknown	Risk of infection with severe outcomes	Cases	2,184
			Controls	689,620
Risk of severe outcomes among infected individuals	COVID-19 positive and hospitalized vs. COVID-19 positive and not hospitalized	Risk of hospitalization among infected individuals	Cases	6,911
			Controls	45,185
	COVID-19 positive and severe vs. COVID-19 positive and not hospitalized	Risk of severe disease among infected individuals	Cases	2,184
			Controls	45,185

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

SUPPLEMENTARY FIGURES

Provided in a separate file.

Supplementary Figure 1. Comparison of effect sizes across COVID-19 risk and severity outcomes for six previously reported novel risk variants that validated in this study.

Six variants were reported to associatd with risk of COVID-19 in previous studies and replicated in our analysis. Of these, four variants also associated with disease severity among COVID-19 cases (in/near *LZTFL1*, *CCHCR1*, *DPP9* and *IFNAR2*), whereas two variants did not (in *ABO* and *SLC6A20*). Error bars represent 95% confidence intervals.

Supplementary Figure 2. Association between a 6-SNP genetic risk score (GRS) and risk of hospitalization among COVID-19 cases of European ancestry after stratifying by the presence of clinical risk factors.

(A) Rate of hospitalization in the AncestryDNA study (25,353 COVID-19 cases, including 1,484 hospitalized).

(B) Rate of hospitalization in the UK Biobank study (14,320 COVID-19 cases, including 3,878 hospitalized).

High genetic risk (red bars): top 10% of the GRS. Low genetic risk (grey bars): bottom 90% of the GRS (*i.e.* all other COVID-19 cases).

Supplementary Figure 3. Association between a 6- and 12-SNP genetic risk score (GRS) and risk of hospitalization and severe disease among COVID-19 cases of European ancestry. To evaluate if the association between the GRS and worse disease outcomes was dependent on the

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

list of variants selected for analysis, we compared results between GRS calculated using different sets of variants. We considered a GRS calculated using: (i) the six variants that were reported in previous GWAS of COVID-19 and that we validated the published association and further showed that they were associated with risk of hospitalization or severe disease among COVID-19 cases (in/near *LZTFL1*, MHC, *DPP9*, *IFNAR2*, *RPL24* and *FOXP4*; see Supplementary **Figure 1**); or (ii) all 12 variants reported in previous GWAS of COVID-19 (in/near *LZTFL1* [two variants], MHC, *ABO*, *OAS3*, *DPP9*, *RAVER1*, *IFNAR2*; and four novel risk variants discovered by the HGI in/near *RPL24, DNAH5, FOXP4* and *PLEKHA4*; see **Supplementary Tables 5 and 15**). Analyses were performed separately in the UK Biobank, AncestryDNA and GHS studies (risk of hospitalization only) after stratifying COVID-19 cases by the presence of clinical risk factors, considering individuals with lower clinical risk (blue circles), high clinical risk (green triangles) or all individuals (grey squares). Association results were then meta-analyzed across studies.

Supplementary Figure 4. Association between a six-SNP genetic risk score (GRS) and risk of hospitalization (A) and (B) severe disease among COVID-19 cases of Admixed American ancestry.

Supplementary Figure 5. Association between risk of severe disease among COVID-19 cases of European ancestry and genetic risk scores determined based on different criteria. We compared GRS based on variants (i) that were reported in the literature and validated in this study (Literature.HGI.1var: rs73064425 in LZTFL1; Literature.HGI.5var: variants from our 6-SNP model, with the exception of rs73064425 in LZTFL1; Literature.HGI.6var: all six variants from our 6-SNP model); (ii) obtained through pruning and thresholding applied to results from the risk

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

of infection phenotype reported by the HGI, using different association P-value and LD r^2 thresholds; (iii) the LDpred approach [39] applied to risk of infection reported by the HGI, considering different ρ parameters.

Supplementary Figure 6. Association between risk of severe disease among COVID-19 cases of European ancestry from the UK Biobank study and combination of clinical risk factors, genetic risk (6-SNP GRS model) and age.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

552 **SUPPLEMENTARY TABLES**

553 Provided in a separate file.

554

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

References

1. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine* **382**, 727-733 (2020).
2. Guan, W.J. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* **382**, 1708-1720 (2020).
3. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395**, 1054-1062 (2020).
4. Cummings, M.J. *et al.* Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. *Lancet* **395**, 1763-1770 (2020).
5. Lopez, L., 3rd, Hart, L.H., 3rd & Katz, M.H. Racial and Ethnic Health Disparities Related to COVID-19. *JAMA* (2021).
6. The COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19 by worldwide meta-analysis. in *medRxiv* (2021).
7. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).
8. Pawitan, Y., Seng, K.C. & Magnusson, P.K. How many genetic variants remain to be discovered? *PLoS One* **4**, e7969 (2009).
9. Consortium, G.T. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).
10. Keidar, S., Kaplan, M. & Gamliel-Lazarovich, A. ACE2 of the heart: From angiotensin I to angiotensin (1-7). *Cardiovasc Res* **73**, 463-9 (2007).
11. Van Hout, C.V. *et al.* Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* (2020).
12. Saheb Sharif-Askari, N. *et al.* Airways Expression of SARS-CoV-2 Receptor, ACE2, and TMPRSS2 Is Lower in Children Than Adults and Increases with Smoking and COPD. *Mol Ther Methods Clin Dev* **18**, 1-6 (2020).
13. Bunyavanich, S., Do, A. & Vicencio, A. Nasal Gene Expression of Angiotensin-Converting Enzyme 2 in Children and Adults. *JAMA* **323**, 2427-2429 (2020).
14. Hadjadj, J. *et al.* Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718-724 (2020).
15. van der Made, C.I. *et al.* Presence of Genetic Variants Among Young Men With Severe COVID-19. *JAMA* (2020).
16. Zhang, Q. *et al.* Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **370**(2020).
17. Ling, Y.H. *et al.* CCHCR1 interacts with EDC4, suggesting its localization in P-bodies. *Exp Cell Res* **327**, 12-23 (2014).
18. Tervaniemi, M.H. *et al.* Intracellular signalling pathways and cytoskeletal functions converge on the psoriasis candidate gene CCHCR1 expressed at P-bodies and centrosomes. *BMC Genomics* **19**, 432 (2018).
19. Kim, Y.J. *et al.* A genome-wide association study identified new variants associated with the risk of chronic hepatitis B. *Hum Mol Genet* **22**, 4233-8 (2013).

medRxiv preprint doi: https://doi.org/10.1101/2020.12.14.20248176; this version posted June 10, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

20. Vuille-dit-Bille, R.N. *et al.* Human intestine luminal ACE2 and amino acid transporter expression increased by ACE-inhibitors. *Amino Acids* **47**, 693-705 (2015).
21. Cui, Y. *et al.* SENP7 Potentiates cGAS Activation by Relieving SUMO-Mediated Inhibition of Cytosolic DNA Sensing. *PLoS Pathog* **13**, e1006156 (2017).
22. Rollinger-Holzinger, I. *et al.* LST1: a gene with extensive alternative splicing and immunomodulatory function. *J Immunol* **164**, 3169-76 (2000).
23. Mulcahy, H., O'Rourke, K.P., Adams, C., Molloy, M.G. & O'Gara, F. LST1 and NCR3 expression in autoimmune inflammation and in response to IFN-gamma, LPS and microbial infection. *Immunogenetics* **57**, 893-903 (2006).
24. Apps, R. *et al.* Influence of HLA-C expression level on HIV control. *Science* **340**, 87-91 (2013).
25. Kulkarni, S. *et al.* Genetic interplay between HLA-C and MIR148A in HIV control and Crohn disease. *Proc Natl Acad Sci U S A* **110**, 20705-10 (2013).
26. Begue, B. *et al.* Defective IL10 signaling defining a subgroup of patients with inflammatory bowel disease. *Am J Gastroenterol* **106**, 1544-55 (2011).
27. Frodsham, A.J. *et al.* Class II cytokine receptor gene cluster is a major locus for hepatitis B persistence. *Proc Natl Acad Sci U S A* **103**, 9148-53 (2006).
28. Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat Commun* **9**, 2282 (2018).
29. Roberts, G.H.L. AncestryDNA COVID-19 Host Genetic Study Identifies Three Novel Loci. *medRxiv* (2020).
30. Dewey, F.E. *et al.* Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* **354**(2016).
31. Park, J. *et al.* A genome-first approach to aggregating rare genetic variants in LMNA for association with electronic health record phenotypes. *Genet Med* **22**, 102-111 (2020).
32. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).
33. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics* (2021).
34. Ferreira, M.A. *et al.* Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat Genet* **49**, 1752-1757 (2017).
35. Chun, S. *et al.* Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat Genet* **49**, 600-605 (2017).
36. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
37. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
38. Robinson, M.D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25 (2010).
39. Vilhjalmsson, B.J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet* **97**, 576-92 (2015).

MAIN FIGURES

A.

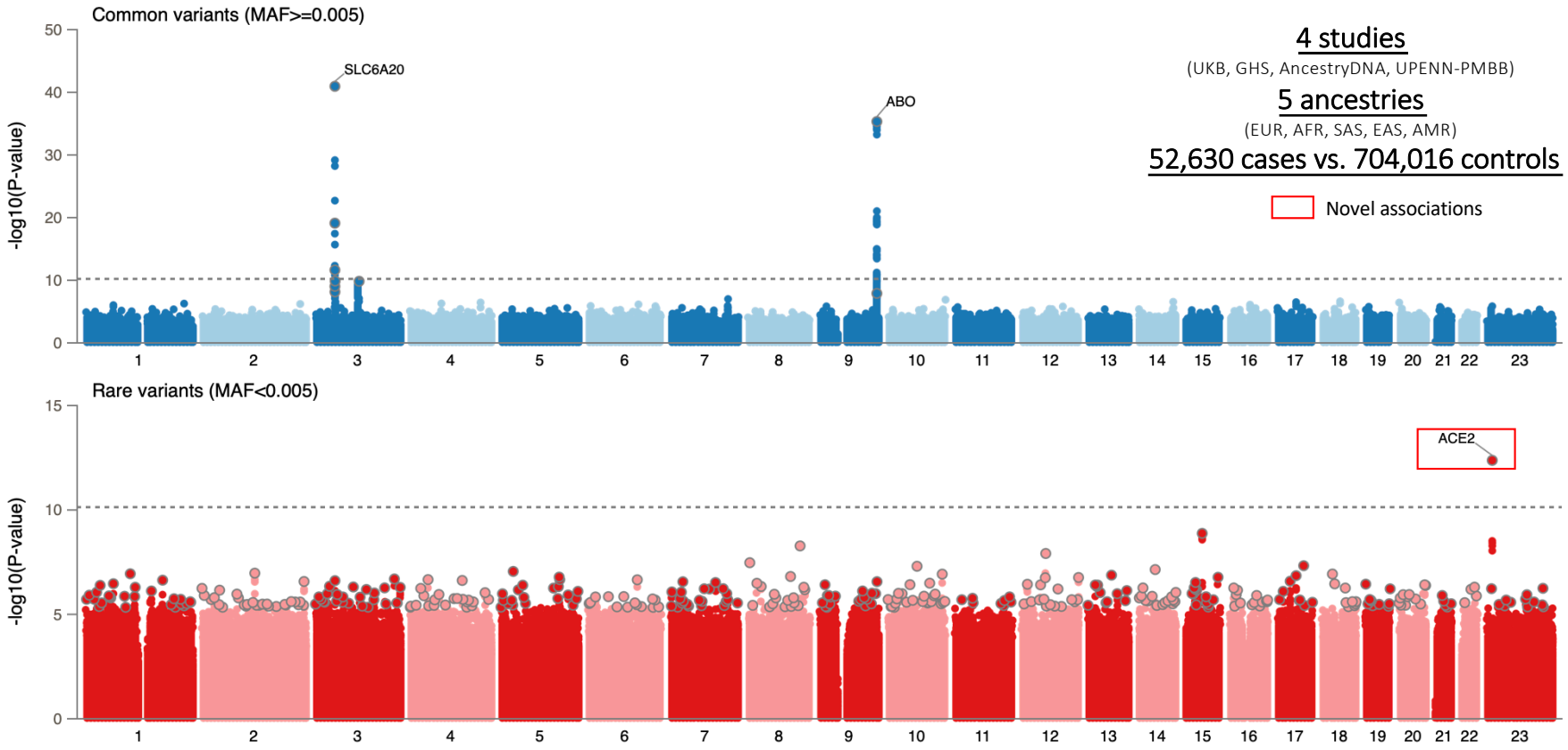
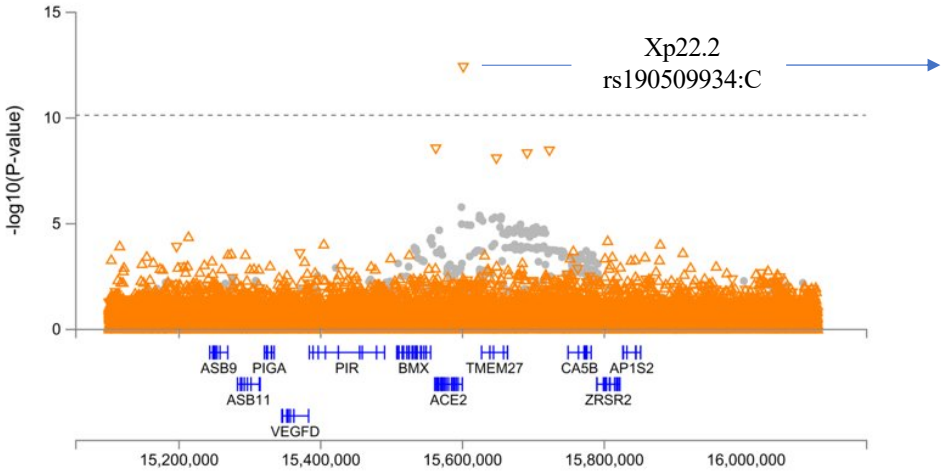


Figure 1. GWAS of 52,630 COVID-19 positive cases vs. 704,016 COVID-19 negative or unknown controls identifies a novel association with a rare variant near *ACE2* that lowers gene expression and protects against COVID-19.

(A) Summary of association results for common (MAF>0.5%) and rare (MAF<0.5%) variants.

B.



C.

STUDY	N CASES RR|RA|AA	N CONTROLS RR|RA|AA	OR [95% CI]	PVALUE	AAF
UKB_500K_Genotyped_EUR	14317|35|2	420556|1244|509	0.627 [0.492,0.8]	1.7e-04	0.0026
UKB_500K_Genotyped_AFR	616|3|1	8721|67|16	0.783 [0.375,1.636]	0.516	0.0058
UKB_500K_Genotyped_SAS	750|12|1	9565|217|100	0.507 [0.325,0.792]	0.003	0.0219
ANCESTRY_Freeze_Four_EUR	25306|<100|<100	113489|327|<100	0.569 [0.444,0.728]	7.5e-06	0.002
ANCESTRY_Freeze_Four_AFR	1622|<100|<100	5641|<100|<100	0.392 [0.197,0.78]	0.008	0.0047
ANCESTRY_Freeze_Four_SAS	<100|<100|<100	250|<100|<100	0.359 [0.042,3.071]	0.35	0.0195
ANCESTRY_Freeze_Four_AMR	3742|<100|<100	12281|<100|<100	0.825 [0.443,1.537]	0.545	0.0021
UPENN_Freeze_One_EUR	40|0|0	6993|3|3	0.43 [0.001,187.431]	0.786	0.0013
UPENN_Freeze_One_AFR	347|2|0	8518|28|18	0.915 [0.286,2.928]	0.881	0.0045
GHS_Freeze_145_EUR	5267|9|0	107839|261|68	0.658 [0.432,1.003]	0.052	0.0019
GHS_Freeze_145_AFR	128|0|0	3025|14|11	0.493 [0.126,1.921]	0.308	0.0057
GHS_Freeze_145_AMR	88|0|0	1296|11|0	0.304 [0.015,6.097]	0.436	0.0039
META	**52294|122|8**	**698174|2251|812**	**0.6 [0.522,0.689]**	**4.5e-13**	**0.0027**

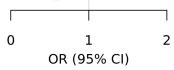


Figure 1 (cont). **GWAS of 52,630 COVID-19 positive cases vs. 704,016 COVID-19 negative or unknown controls identifies a rare variant near ACE2 that lowers gene expression and protects against COVID-19.**

(B) Regional association plot centered around rs190509934 near ACE2.
(C) Breakdown of association results across studies included in the meta-analysis of rs190509934.

D.

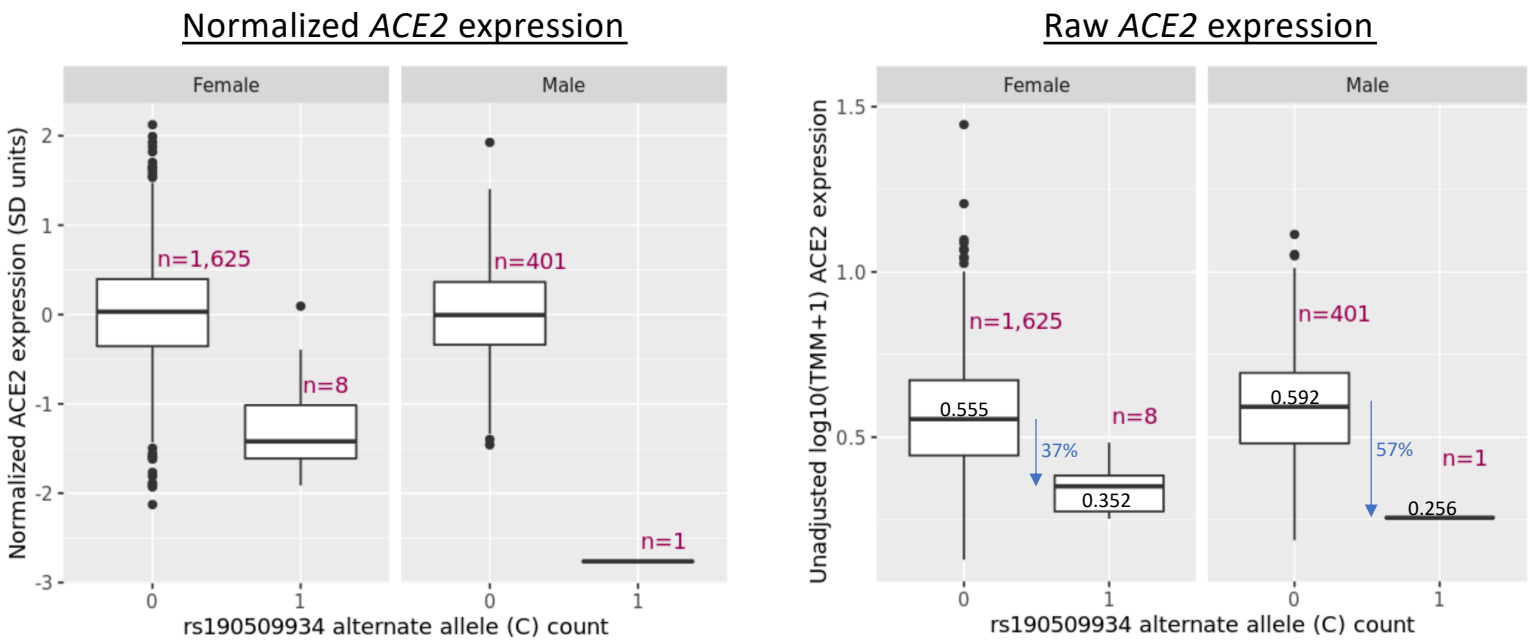


Figure 1 (cont). GWAS of 52,630 COVID-19 positive cases vs. 704,016 COVID-19 negative or unknown controls identifies a rare variant near ACE2 that lowers gene expression and protects against COVID-19.

(D) Association between rs190509934:C and ACE2 expression in liver measured in 2,035 individuals (80% female) from the GHS study.

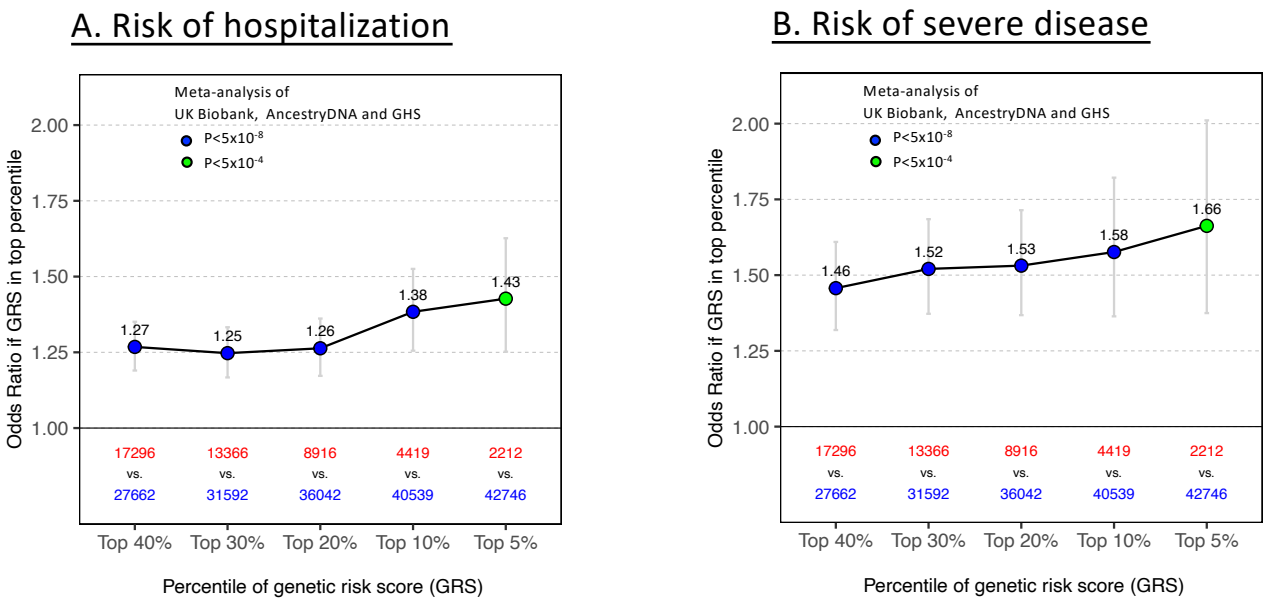
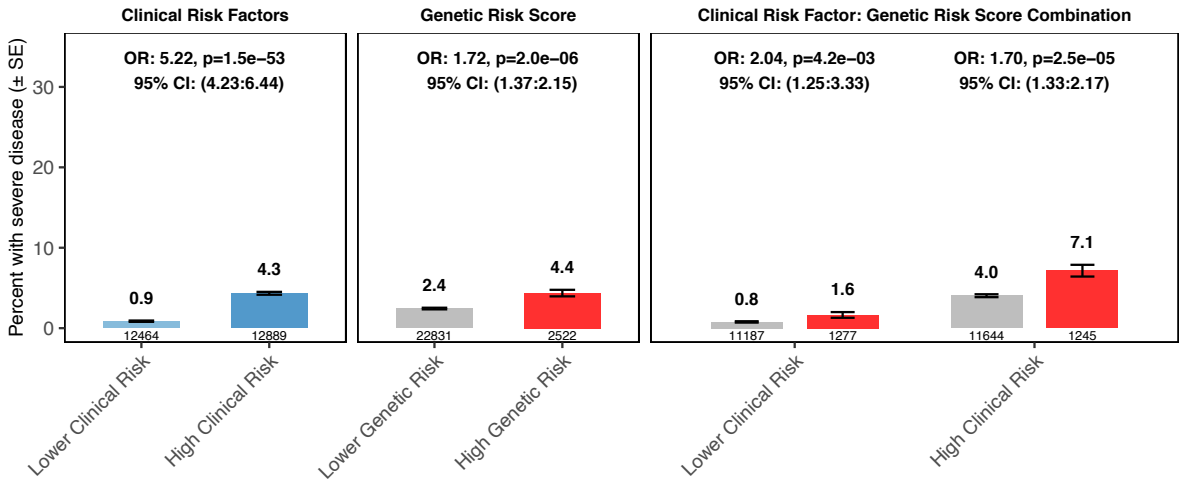
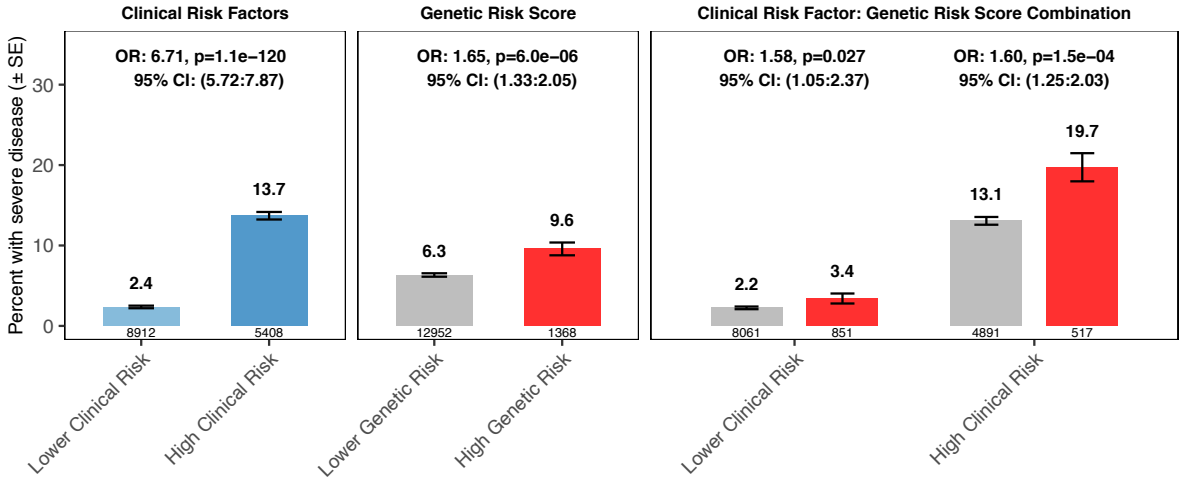


Figure 2. Association between a 6-SNP genetic risk score (GRS) and risk of hospitalization (A) and severe disease (B) among COVID-19 cases of European ancestry (N=44,958).

A. AncestryDNA study



B. UK Biobank study



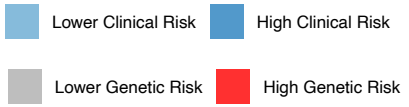
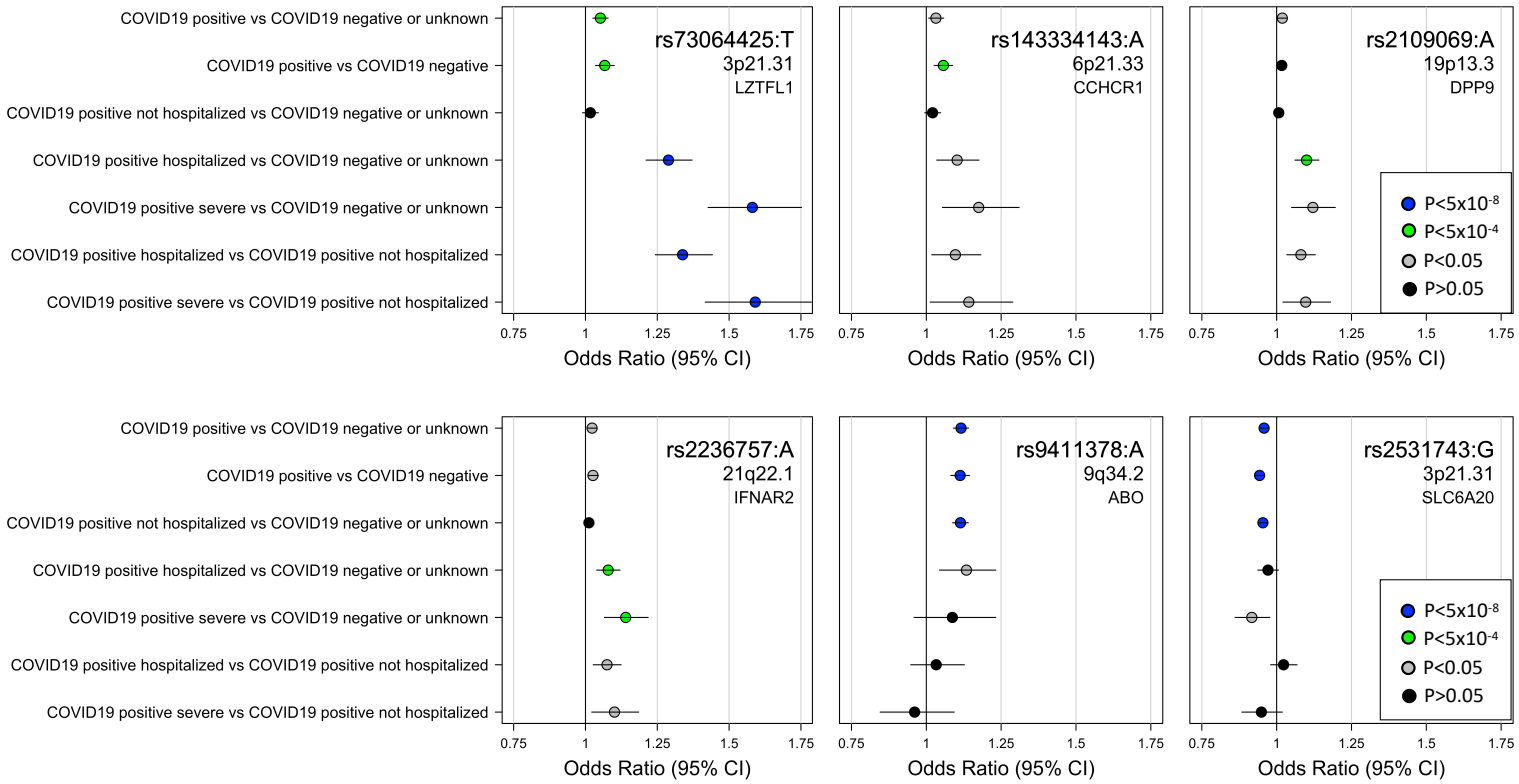


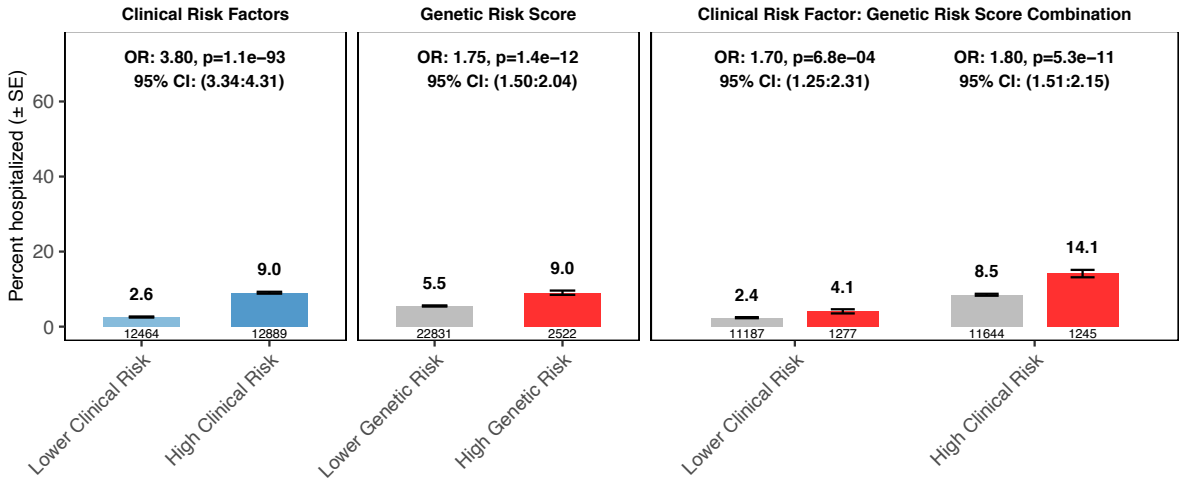
Figure 3. Association between a 6-SNP genetic risk score (GRS) and risk of severe disease among COVID-19 cases of European ancestry.

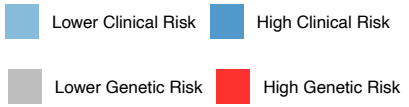
SUPPLEMENTARY FIGURES



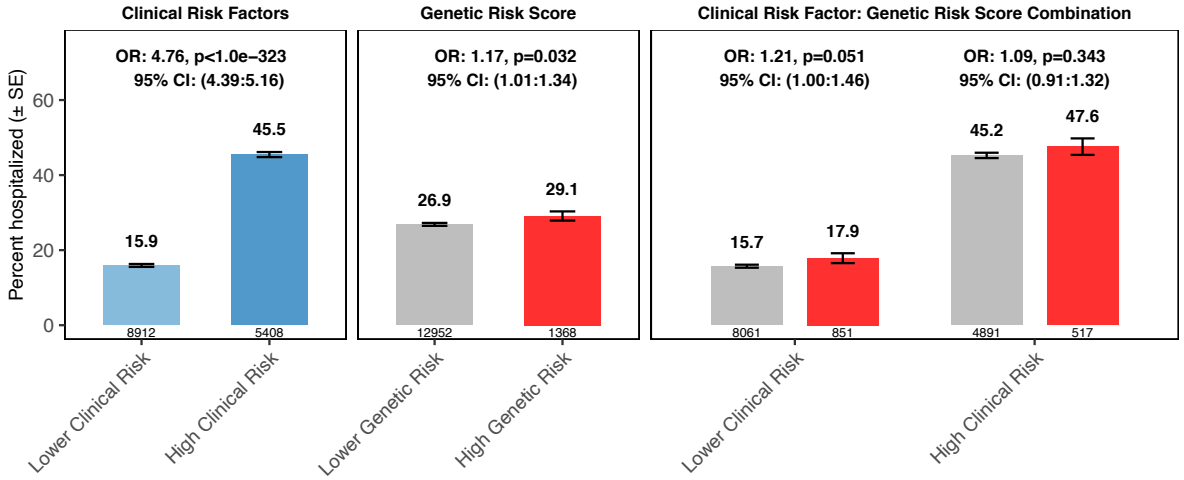
Supplementary Figure 1. Comparison of effect sizes across COVID-19 susceptibility and severity outcomes for six previously reported risk variants that validated in our study.

A. AncestryDNA study

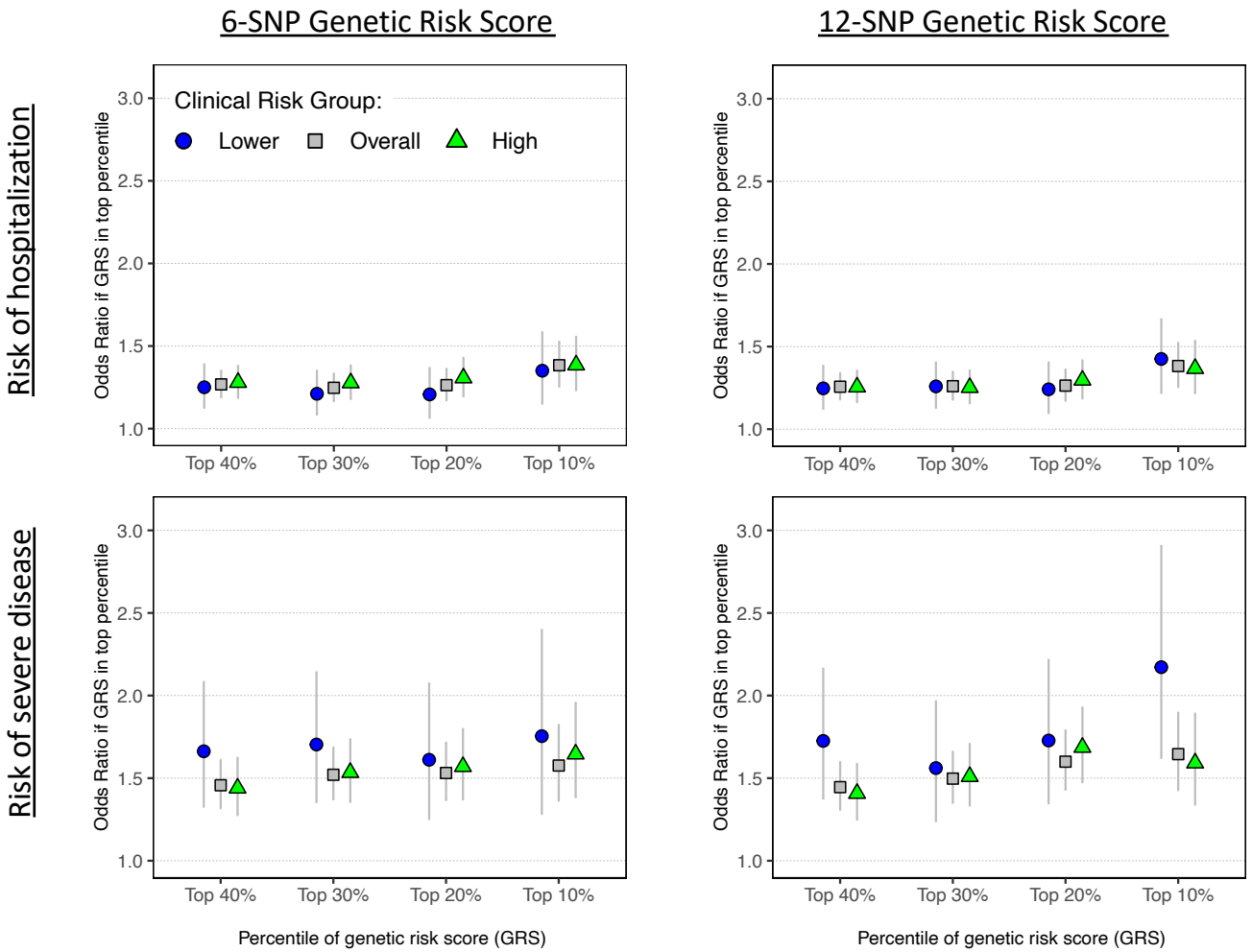




B. UK Biobank study

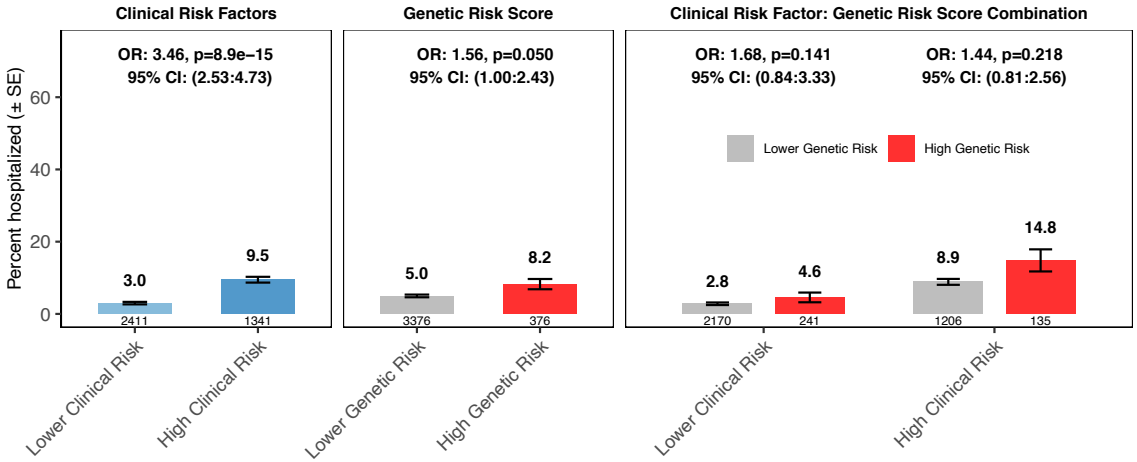


Supplementary Figure 2. **Association between a six-SNP genetic risk score (GRS) and risk of hospitalization among COVID-19 cases of European ancestry.**

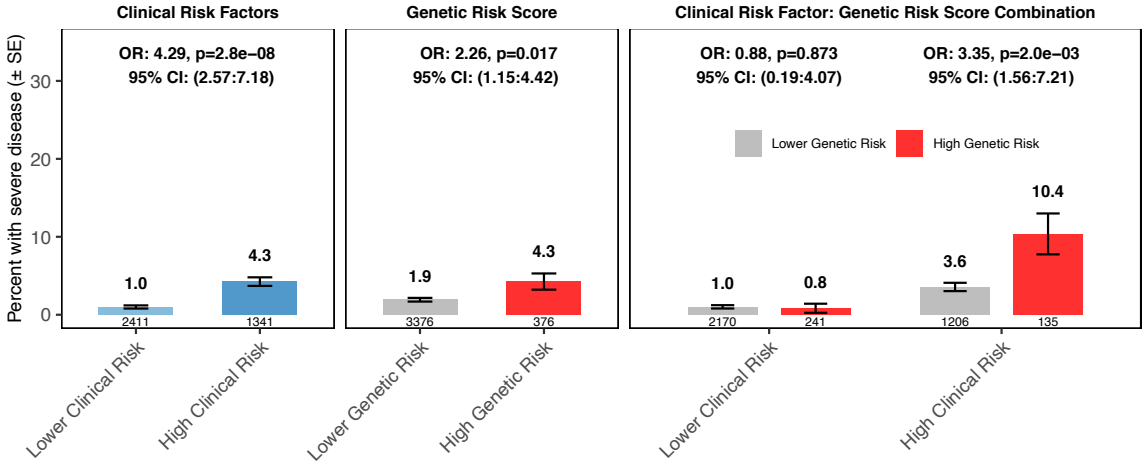


Supplementary Figure 3. Association between a 6-SNP and 12-SNP genetic risk score (GRS) and risk of hospitalization and severe disease among COVID-19 cases of European ancestry.

A. Risk of hospitalization



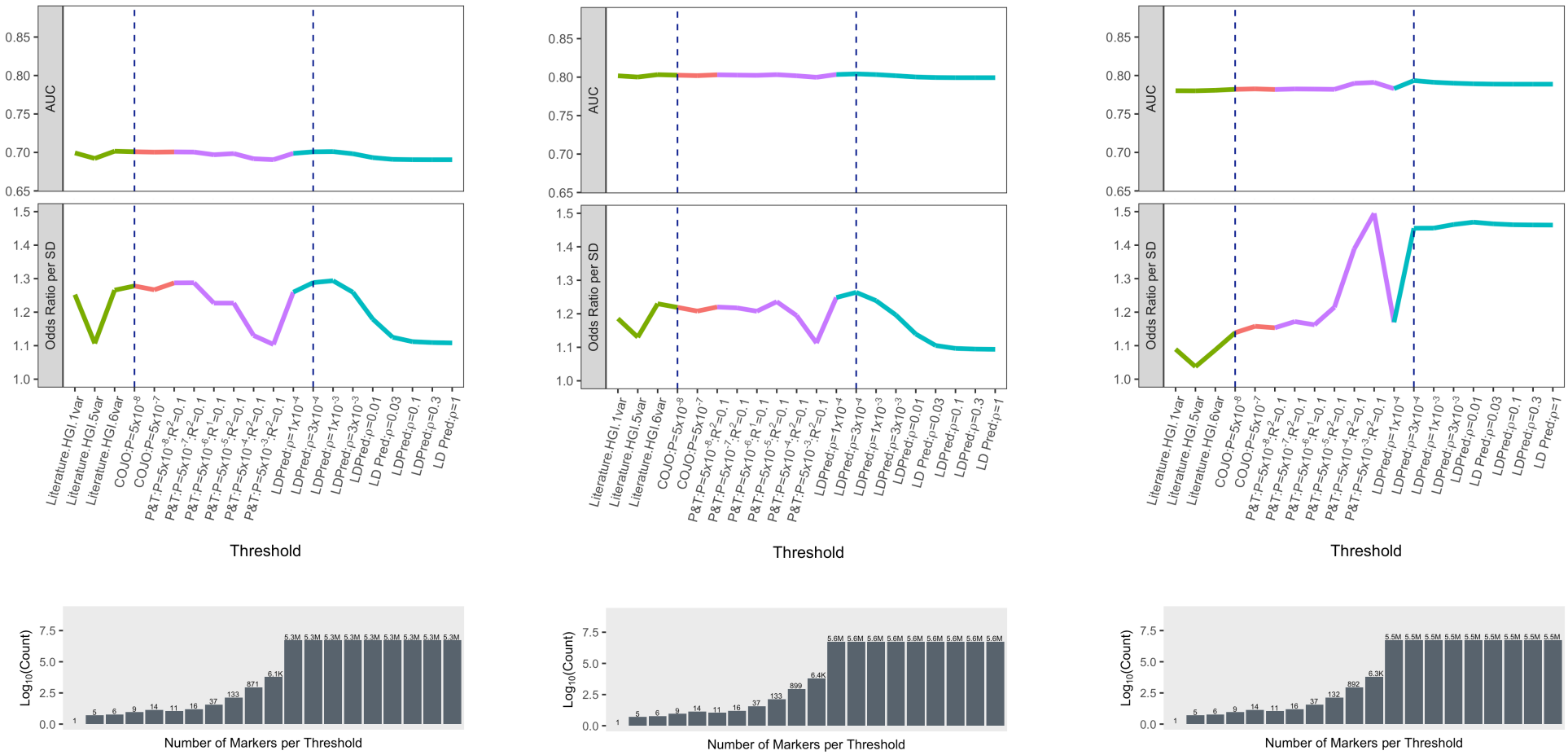
B. Risk of severe disease



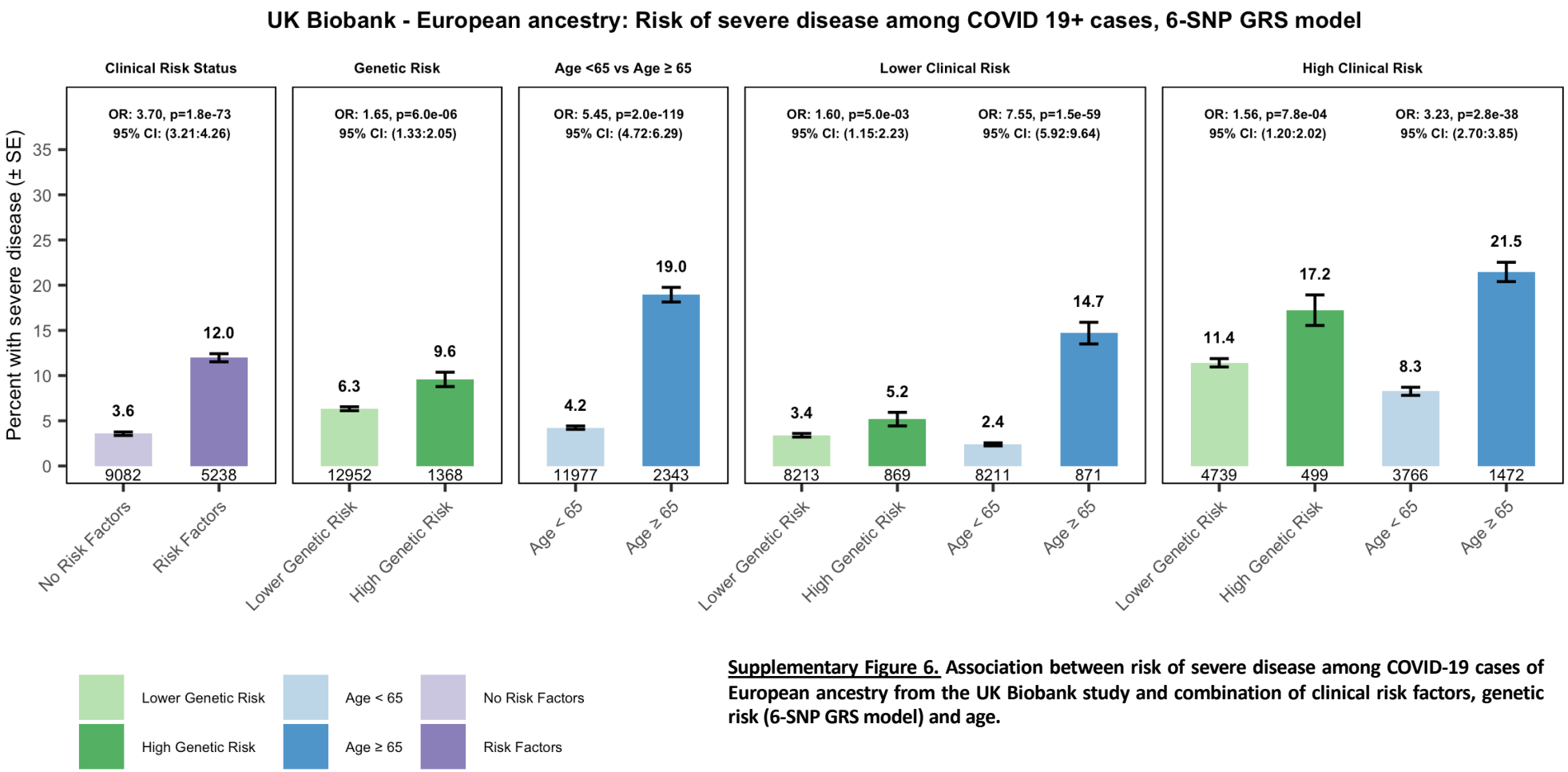
AncestryDNA study

Admixed American ancestry

Supplementary Figure 4. Association between a six-SNP genetic risk score (GRS) and risk of hospitalization (A) and (B) severe disease among COVID-19 cases of Admixed American ancestry.



<u>Supplementary Figure 5.</u> Association between risk of severe disease among COVID-19 cases of European ancestry and genetic risk scores determined based on (i) variants that were reported in the literature and validated in this study (Literature.HGI.1var: rs73064425 in LZTFL1; Literature.HGI.5var: variants from our 6-SNP model, with the exception of rs73064425 in LZTFL1; Literature.HGI.6var: all six variants from our 6-SNP model); (ii) pruning and thresholding applied to results from the risk of infection phenotype reported by the HGI, using different association P-value and LD r2 thresholds; (iii) LDpred applied to risk of infection reported by the HGI, considering different *ϱ* parameters.



Supplementary Figure 6. **Association between risk of severe disease among COVID-19 cases of European ancestry from the UK Biobank study and combination of clinical risk factors, genetic risk (6-SNP GRS model) and age.**

SUPPLEMENTARY TABLES

Supplementary Table 1. Breakdown of COVID-19 status across the four studies included in the analysis.

COVID-19 status	Positive qPCR or serology for SARS-CoV-2	ICD10 U07 diagnosis or hospitalization	Severe COVID-19 (ventilation or death)	Negative qPCR or serology test for SARS-CoV-2	Sample size				
					AncestryDNA	UK Biobank	GHS	PMBB	Total
Positive	Yes	Yes	Yes	Yes or No or NA	817	793	326	22	1,958
	Yes	Yes	No or NA	Yes or No or NA	1,016	2810	471	122	4,419
	Yes	No or NA	Yes	Yes or No or NA	0	0	0	0	0
	Yes	No or NA	No or NA	Yes or No or NA	29,093	11,613	4695	240	45,641
	No or NA	Yes	Yes	No or NA	0	187	0	0	187
	No or NA	Yes	No or NA	No or NA	0	164	0	0	164
	No or NA	No or NA	Yes	No or NA	0	0	0	0	0
	No or NA	Yes	Yes	Yes	0	54	0	0	54
	No or NA	Yes	No or NA	Yes	0	202	0	5	207
	No or NA	No or NA	Yes	Yes	0	0	0	0	0
					Total = 30,926	Total =15,823	Total = 5492	Total = 389	Total = 52,630
Negative	No or NA	No or NA	No or NA	Yes	36,107	47,205	23,737	2556	109,605
Unknown	NA	No or NA	No or NA	NA	96,617	395,999	88,788	13,007	594,411
Total					163,650	459,027	118,017	15,952	756,646

Supplementary Table 2. Demographics and clinical characteristics of study participants.

Demographics	COVID-19 Positive		Covid-19 Negative or Unknown
	Hospitalized	Not Hospitalized	
AncestryDNA			
Total N	1818	29,093	132,724
AFR ancestry, n (%)	135 (7.34%)	1494 (5.14%)	5688 (4.29%)
AMR ancestry, n (%)	199 (10.95%)	3553 (12.21%)	13323 (10.04%)
EAS ancestry, n (%)	0 (0%)	110 (0.38%)	570 (0.43%)
EUR ancestry, n (%)	1484 (81.63%)	23869 (81.04%)	113882 (85.80%)
SAS ancestry, n (%)	<100 (0%)	<100 (0.23%)	261 (0.20%)
Average Age, y (% >60y)	57.84 (863)	50.1 (7940)	59.43 (48547)
Female, n (%)	726 (60.4%)	9870 (66.1%)	42909 (67.7%)
Hypertension, n (%)	670 (36.6%)	5402 (18.6%)	28912 (21.8%)
Cardiovascular Disease, n (%)	204 (11.1%)	1048 (3.6%)	6573 (5.0%)
Type 2 Diabetes, n (%)	392 (21.4%)	2330 (8.0%)	12432 (9.4%)
Chronic kidney disease, n (%)	<100 (5.0%)	323 (1.1%)	2123 (1.6%)
Asthma, n (%)	382 (20.8%)	3861 (13.3%)	18256 (13.8%)
COPD, n (%)	125 (6.8%)	523 (1.8%)	3563 (2.7%)
UK Biobank			
Total N	4188	11,613	443,204
AFR ancestry, n (%)	148 (3.53%)	472 (4.06%)	8804 (1.99%)
EAS ancestry, n (%)	0 (0%)	64 (0.55%)	2209 (0.50%)
EUR ancestry, n (%)	3884 (92.74%)	10470 (90.16%)	422309 (95.29%)
SAS ancestry, n (%)	156 (3.72%)	607 (5.23%)	9882 (2.23%)
Average Age, y (% >60y)	61 (52.3%)	51 (20.2%)	57 (37.2%)
Female, n (%)	1838 (43.9%)	6410 (55.2%)	244345 (55.1%)
Hypertension, n (%)	1851 (44.2%)	2407 (20.7%)	102820 (23.2%)
Cardiovascular Disease, n (%)	553 (13.2%)	566 (4.8%)	25998 (5.9%)
Type 2 Diabetes, n (%)	757 (18.0%)	766 (6.6%)	26743 (6%)
Chronic kidney disease, n (%)	280 (6.7%)	190 (1.6%)	8546 (1.9%)
Asthma, n (%)	829 (19.8%)	1784 (15.4%)	62740 (14.1%)
COPD, n (%)	392 (9.3%)	250 (2.1%)	10762 (2.4%)
GHS			
Total N	773	4827	116734
AFR ancestry, n (%)	0 (0%)	112 (2.32%)	3050 (2.61%)

AMR ancestry, n (%)	0 (0%)	80 (1.65%)	1307 (1.11%)
EUR ancestry, n (%)	773 (100%)	4503 (93.28%)	108168 (92.66%)
Average Age, y (% >60y)	68.68 (76.06%)	52.20 (33.39%)	55.41 (43.43%)
Female, n (%)	393 (50.84%)	3129 (66.64%)	70456 (62.61%)
Hypertension, n (%)	603 (78.00%)	2191 (46.66%)	54881 (48.77%)
Cardiovascular Disease, n (%)	328 (42.43%)	653 (13.90%)	17395 (15.45%)
Type 2 Diabetes, n (%)	348 (45.01%)	1049 (22.34%)	24152 (21.46%)
Chronic kidney disease, n (%)	298 (38.55%)	519 (11.05%)	14534 (12.91%)
Asthma, n (%)	74 (9.57%)	381 (8.11%)	7146 (6.35%)
COPD, n (%)	192 (24.83%)	376 (8.00%)	11028 (9.80%)
PMBB			
Total N	132	240	15563
AFR ancestry, n (%)	132 (100%)	217 (90.42%)	8564 (55.03%)
EUR ancestry, n (%)	0 (0%)	23 (9.58%)	6999 (44.97%)
Average Age, y (% >60y)	61.77 (58.33%)	50.45 (27.50%)	63.26 (61.95%)
Female, n (%)	69 (52.27%)	174 (72.50%)	8121 (52.18%)
Hypertension, n (%)	113 (85.61%)	141 (58.75%)	9466 (60.82%)
Cardiovascular Disease, n (%)	40 (30.30%)	34 (14.17%)	4842 (31.11%)
Type 2 Diabetes, n (%)	90 (68.18%)	70 (29.17%)	4546 (29.21%)
Chronic kidney disease, n (%)	74 (56.06%)	26 (10.83%)	3507 (22.53%)
Asthma, n (%)	29 (21.97%)	52 (21.67%)	2189 (14.07%)
COPD, n (%)	26 (19.70%)	15 (6.25%)	1706 (10.96%)

Supplementary Table 3. Definitions used for the seven COVID-19 phenotypes analyzed.

Broad phenotype category	Phenotype	Case/control group	Definition			Sample size with genetic data				
			COVID-19 status	Hospitalized	Severe disease	AncestryDNA	UK Biobank	GHS	PMBB	Total
Risk of infection	COVID-19 positive vs. COVID-19 negative or unknown	Cases	Positive	Yes, No or NA	Yes, No or NA	30,926	15,823	5,492	389	52,630
		Controls	Negative or unknown	No or NA	No or NA	132,724	443,204	112,525	15,563	704,016
	COVID-19 positive vs. COVID-19 negative	Cases	Positive	Yes, No or NA	Yes, No or NA	30,926	15,823	5,492	389	52,630
		Controls	Negative	No or NA	No or NA	36,107	47,205	23,737	2,556	109,605
	COVID-19 positive and not hospitalized vs. COVID-19 negative or unknown	Cases	Positive	No	No	29,093	11,613	4,695	240	45,641
		Controls	Negative or unknown	No or NA	No or NA	132,724	443,204	112,525	15,563	704,016
	COVID-19 positive and hospitalized vs. COVID-19 negative or unknown	Cases	Positive	Yes (or death)	Yes, No or NA	1,818	4,188	773	132	6,911
		Controls	Negative or unknown	No or NA	No or NA	131,893	440,995	108,168	8,564	689,620
	COVID-19 positive and severe vs. COVID-19 negative or unknown	Cases	Positive	Yes, No or NA	Yes	810	1,028	321	25	2,184
		Controls	Negative or unknown	No or NA	No or NA	131,893	440,995	108,168	8,564	689,620
Risk of adverse outcomes amongst infected individuals	COVID-19 positive and hospitalized vs. COVID-19 positive and not hospitalized	Cases	Positive	Yes (or death)	Yes, No or NA	1,818	4,188	773	132	6,911
		Controls	Positive	No	No	28,916	11,549	4,503	217	45,185
	COVID-19 positive and severe vs. COVID-19 positive and not hospitalized	Cases	Positive	Yes, No or NA	Yes	810	1,028	321	25	2,184
		Controls	Positive	No	No	28,916	11,549	4,503	217	45,185

Supplementary Table 4. Genomic inflation factor (lambda) across the seven COVID-19 phenotypes analyzed.

Index	Phenotype	Study	N cases	N controls	Total N	Common variants (MAF>0.5%)		Rare variants (MAF<0.5%)	
						N variants	Lambda	N variants	Lambda
1	COVID19 positive vs COVID19 negative or unknown	Meta-analysis	52630	704016	756646	12659604	1.054	76347805	1.011
		ANCESTRY Freeze Four EUR	25353	113882	139235	9114485	1.032	17505860	1.021
		UKB 500K EUR	14354	422309	436663	10518951	1.034	41557338	1.038
		GHS Freeze 145 EUR	5276	108168	113444	10517411	1.013	31213671	1.078
		ANCESTRY Freeze Four AMR	3752	12323	16075	10360579	1.126	16342242	1.142
		ANCESTRY Freeze Four AFR	1629	5688	7317	16651249	1.083	7924824	1.111
		UKB 500K SAS	763	9882	10645	11223467	1.047	16126339	1.147
		UKB 500K AFR	620	8804	9424	19365480	1.028	23403726	1.126
		UPENN-PMBB Freeze One TOPMED AFR	349	8564	8913	19201924	0.993	24334265	0.877
		GHS Freeze 145 AFR	128	3050	3178	19261006	0.793	13222396	0.775
		ANCESTRY Freeze Four EAS	120	570	690	8287783	0.955	388659	1.276
		GHS Freeze 145 AMR	88	1307	1395	13559291	0.784	5975072	0.804
		UKB 500K EAS	86	2209	2295	9472040	0.994	5274937	0.808
		ANCESTRY Freeze Four SAS	<100	261	333	8772668	0.772	86	0.772
		UPENN-PMBB Freeze One TOPMED EUR	40	6999	7039	10569481	0.914	11542106	0.189
2	COVID19 positive vs COVID19 negative	Meta-analysis	52630	109605	162235	15917507	1.018	48430743	1.047
		ANCESTRY Freeze Four EUR	25353	30667	56020	9116923	1.069	17300691	1.065
		UKB 500K EUR	14354	44903	59257	10521738	1.031	31741985	1.05
		GHS Freeze 145 EUR	5276	22848	28124	10518718	1.021	26548178	1.082
		ANCESTRY Freeze Four AMR	3752	3403	7155	10409037	1.056	12259804	1.071
		ANCESTRY Freeze Four AFR	1629	1783	3412	16657740	1.037	5856801	1.063
		UKB 500K SAS	763	1104	1867	11217149	0.986	4212996	1.03
		UKB 500K AFR	620	1029	1649	19377829	0.988	6949159	1.032
		UPENN-PMBB Freeze One TOPMED AFR	349	1804	2153	19210279	1.001	9238302	1.159
		GHS Freeze 145 AFR	128	583	711	19220711	0.684	2145450	1.005
		ANCESTRY Freeze Four EAS	120	174	294	7368136	0.816	NA	NA
		GHS Freeze 145 AMR	88	306	394	12615699	0.625	NA	NA
		UKB 500K EAS	86	169	255	8096105	0.767	NA	NA
		ANCESTRY Freeze Four SAS	<100	<100	152	7499011	0.477	NA	NA
		UPENN-PMBB Freeze One TOPMED EUR	40	752	792	10649166	0.902	1084967	0.686
3	COVID19 positive not hospitalized vs COVID19 negative or unknown	Meta-analysis	45641	704016	749657	12664016	1.049	76053657	0.971
		ANCESTRY Freeze Four EUR	23869	113882	137751	9114352	1.033	17505215	1.02
		UKB 500K EUR	10470	422309	432779	10518900	1.027	41512665	1.004
		GHS Freeze 145 EUR	4503	108168	112671	10517274	1.015	31211655	1.112
		ANCESTRY Freeze Four AMR	3553	12323	15876	10359235	1.126	16286078	1.142
		ANCESTRY Freeze Four AFR	1494	5688	7182	16650531	1.080	7876530	1.119
		UKB 500K SAS	607	9882	10489	11224048	1.036	16007835	1.085
		UKB 500K AFR	472	8804	9276	19365078	1.037	23196596	1.048
		UPENN-PMBB Freeze One TOPMED AFR	217	8564	8781	19201513	0.981	24110114	0.655
		GHS Freeze 145 AFR	112	3050	3162	19259910	0.798	13175822	0.706
		ANCESTRY Freeze Four EAS	110	570	680	8277196	0.937	370881	1.238
		GHS Freeze 145 AMR	80	1307	1387	13555149	0.805	5935171	0.752
		ANCESTRY Freeze Four SAS	<100	261	328	8746716	0.734	59	0.737
		UKB 500K EAS	64	2209	2273	9469792	1.011	5218675	0.653
		UPENN-PMBB Freeze One TOPMED EUR	23	6999	7022	10568880	0.842	11520379	0.103
		Meta-analysis	6911	689620	696531	12311356	1.019	72100390	0.774
		UKB 500K EUR	3884	422309	426193	10518717	1.029	41438313	0.869

4	COVID19 positive hospitalized vs COVID19 negative or unknown	ANCESTRY_Freeze_Four_EUR	1484	113882	115366	9113935	1.020	17492079	1.095
		GHS_Freeze_145_EUR	773	108168	108941	10517626	0.989	31200896	0.679
		ANCESTRY_Freeze_Four_AMR	199	12323	12522	10355552	0.954	15145563	0.66
		UKB_500K_SAS	156	9882	10038	11226298	0.956	15643828	0.594
		UKB_500K_AFR	148	8804	8952	19364991	0.959	22729568	0.524
		ANCESTRY_Freeze_Four_AFR	135	5688	5823	16650183	0.845	7301363	0.704
		UPENN-PMBB_Freeze_One_TOPMED_AFR	132	8564	8696	19203151	0.954	23970759	0.463
5	COVID19 positive severe vs COVID19 negative or unknown	Meta-analysis	2184	689620	691804	12317578	0.996	71887891	0.436
		UKB_500K_EUR	953	422309	423262	10518693	1.017	41399045	0.518
		ANCESTRY_Freeze_Four_EUR	667	113882	114549	9113896	1.010	17491621	0.797
		GHS_Freeze_145_EUR	321	108168	108489	10517572	0.944	31199381	0.337
		ANCESTRY_Freeze_Four_AMR	<100	12323	12404	10356352	0.929	15099237	0.371
		ANCESTRY_Freeze_Four_AFR	<100	5688	5750	16648952	0.925	7269674	0.481
		UKB_500K_AFR	43	8804	8847	19365845	0.985	22580462	0.17
		UKB_500K_SAS	32	9882	9914	11226767	0.918	15544218	0.168
		UPENN-PMBB_Freeze_One_TOPMED_AFR	25	8564	8589	19201950	0.910	23788706	0.093
6	COVID19 positive hospitalized vs COVID19 positive not hospitalized	Meta-analysis	6911	45185	52096	17736142	0.980	24498210	1.056
		UKB_500K_EUR	3884	10470	14354	10524169	1.040	16832015	1.071
		ANCESTRY_Freeze_Four_EUR	1484	23869	25353	9117639	1.011	15595929	1.148
		GHS_Freeze_145_EUR	773	4503	5276	10521676	1.008	8477328	1.132
		ANCESTRY_Freeze_Four_AMR	199	3553	3752	10392795	1.006	8906919	1.083
		UKB_500K_SAS	156	607	763	11125523	0.881	900209	1.018
		UKB_500K_AFR	148	472	620	19346132	0.963	1173739	1.121
		ANCESTRY_Freeze_Four_AFR	135	1494	1629	16632746	0.879	3878224	1.088
		UPENN-PMBB_Freeze_One_TOPMED_AFR	132	217	349	17386629	0.878	NA	NA
7	COVID19 positive severe vs COVID19 positive not hospitalized	Meta-analysis	2184	45185	47369	17565432	0.977	22301026	0.981
		UKB_500K_EUR	953	10470	11423	10524284	1.046	14265668	1.086
		ANCESTRY_Freeze_Four_EUR	667	23869	24536	9117378	1.028	15460239	0.885
		GHS_Freeze_145_EUR	321	4503	4824	10521257	0.902	7888063	0.971
		ANCESTRY_Freeze_Four_AMR	<100	3553	3634	10395661	0.993	8745342	0.685
		ANCESTRY_Freeze_Four_AFR	<100	1494	1556	16630281	0.978	3750324	0.792
		UKB_500K_AFR	43	472	515	19296565	0.892	266926	0.693
		UKB_500K_SAS	32	607	639	11034836	0.814	467149	0.548
		UPENN-PMBB_Freeze_One_TOPMED_AFR	25	217	242	15603727	0.998	NA	NA

Supplementary Table 5. Eight variants associated with COVID-19 susceptibility in previous GWAS.

Locus index	Independent signal index	Reference (PMID)	First author	Phenotype	rs ID	hg38	Nearest gene	Effect allele	Locus	Odds Ratio	LCI	UCI	P-value	N cases	N controls	Top independent association in locus
1	1	33307546	Pairo-Castineira	covid19_severe_vs_negative_or_unknown	rs73064425	3:45859597:C:T	*LZTFL1*	T	3p21.31	2.14	1.88	2.45	4.77E-30	1676	8380	1
1	2	33888907	Shelton	covid_test_positive_vs_negative	rs2531743	3:45796808:G:A	*SLC6A20*	G	3p21.31	0.917	0.89	0.94	7.60E-10	12972	101268	1
2	1	33307546	Pairo-Castineira	covid19_severe_vs_negative_or_unknown	rs143334143	6:31153649:G:A	*CCHCR1*	A	6p21.33	1.85	1.61	2.13	8.82E-18	1676	8380	1
3	1	33888907	Shelton	covid_test_positive_vs_negative	rs9411378	9:133270015:A:C	*ABO*	A	9q34.2	1.17	1.12	1.20	5.30E-20	12972	101268	1
4	1	33307546	Pairo-Castineira	covid19_severe_vs_negative_or_unknown	rs10735079	12:112942203:G:A	*OAS3*	A	12q24.13	1.29	1.18	1.42	1.65E-08	1676	8380	1
5	1	33307546	Pairo-Castineira	covid19_severe_vs_negative_or_unknown	rs2109069	19:4719431:G:A	*DPP9*	A	19p13.3	1.36	1.25	1.48	3.98E-12	1676	8380	1
6	1	33307546	Pairo-Castineira	covid19_severe_vs_negative_or_unknown	rs74956615	19:10317045:T:A	*RAVER1*	A	19p13.2	1.59	1.35	1.87	2.31E-08	1676	8380	1
7	1	33307546	Pairo-Castineira	covid19_severe_vs_negative_or_unknown	rs2236757	21:33252612:A:G	*IFNAR2*	A	21q22.1	1.28	1.17	1.41	5.00E-08	1676	8380	1

Supplementary Table 6. Sample overlap between our study and those queried to identify variants associated with COVID-19 susceptibility in previous GWAS.

First author	Reference (PMID)	N cases	N controls	Sample overlap with our study
Three studies with minimal or no sample overlap with our study *(eight variants selected for replication with five susceptibility phenotypes: in/near LZTFL1, SLC6A20, MHC, ABO, OAS3, DPP9, RAVER1 and IFNAR2)*				
Ellinghaus	32558485	1,610	2,205	No known overlap
Pairo-Castineira	33307546	1,676	8,380	~1% of samples from our analysis of UKB were included as controls in this study
Shelton	33888907	15,434	1,035,598	No known overlap
One study with substantial sample overlap with our study *(Four variants associated with reported infection [phenotype "C2"] selected for look-up in our independent analysis of severity among cases: in/near RPL24, DNAH5, FOXP4 and PLEKHA4)*				
HGI (freeze 5)	NA	49,562	1.7 million	~100% of samples from UKB, GHS and UPENN, and ~60% from AncestryDNA contributed to HGI, but using an earlier phenotype freeze.

Supplementary Table 7. Comparison of effect sizes between ancestries for the novel ACE2 risk variant.

Locus, nearest gene	rs ID	hg38	Effect allele	Phenotype	Ancestry	Odds Ratio (95% CI)	P-value	N cases with 0|1|2 copies of the effect allele	N controls with 0|1|2 copies of the effect allele	Effect allele frequency	Cross-ancestry heterogeneity test P-value
Xp22.2 ACE2	rs190509934	23:15602217:T:C	C	COVID19 positive vs COVID19 negative or unknown	AFR	0.578[0.373,0.894]	0.014	2713|12|1	25905|146|55	0.005	0.676
					AMR	0.792[0.431,1.456]	0.45	3830|10|0	13577|44|9	0.002	
					EUR	0.606[0.516,0.711]	8.80E-10	44930|87|6	648877|1835|646	0.002	
					SAS	0.500[0.323,0.773]	0.0018	821|13|1	9815|226|102	0.020	

Supplementary Table 8. Association results in the UK Biobank study between six established clinical risk factors for COVID-19 and (i) the novel risk variant in ACE2; (ii) the six published risk variants for COVID-19 that validated in this study; and (iii) the two novel risk variants identified in the HGI analysis of C2 and that we found to also associate with disease severity.

Nearest gene	Variant	Effect allele	Disease	Odds Ratio	P-value	Effect allele frequency	N cases	N controls
(i) Novel risk variant in ACE2								
ACE2	23:15602217:T:C	C	Coronary artery disease	0.917[0.798,1.054]	0.2245	0.0027	28605	254598
			Kidney disease	0.927[0.808,1.064]	0.2800	0.0030	35472	211418
			Hypertension	0.899[0.821,0.984]	0.0210	0.0030	105117	269090
			COPD	0.904[0.739,1.105]	0.3200	0.0030	12839	372779
			Asthma	1.016[0.896,1.151]	0.8100	0.0030	39372	343873
			Type-2 diabetes	0.957[0.831,1.103]	0.5500	0.0030	25722	433026
(ii) Six published risk variants that validated in this study								
SLC6A20	3:45796808:G:A	G	Coronary artery disease	0.997[0.981,1.014]	0.7500	0.4070	38269	416904
			Kidney disease	1.000[0.983,1.018]	0.9700	0.4160	35472	211418
			Hypertension	0.995[0.984,1.007]	0.4100	0.4160	105117	269090
			COPD	1.011[0.984,1.038]	0.4300	0.4160	12839	372779
			Asthma	0.993[0.978,1.009]	0.3800	0.4160	39372	343873
			Type-2 diabetes	0.999[0.980,1.018]	0.9000	0.4160	25722	433026
LZTFL1	3:45859597:C:T	T	Coronary artery disease	0.986[0.957,1.017]	0.3700	0.0700	38269	416904
			Kidney disease	1.018[0.985,1.052]	0.2900	0.0690	35472	211418
			Hypertension	1.003[0.982,1.025]	0.7700	0.0690	105117	269090
			COPD	1.053[1.002,1.107]	0.0410	0.0690	12839	372779
			Asthma	1.026[0.996,1.056]	0.0940	0.0690	39372	343873
			Type-2 diabetes	1.002[0.966,1.039]	0.9200	0.0690	25722	433026
CCHCR1	6:31153649:G:A	A	Coronary artery disease	1.048[1.016,1.081]	0.0030	0.0650	38269	416904
			Kidney disease	1.016[0.982,1.051]	0.3500	0.0650	35472	211418
			Hypertension	0.989[0.968,1.011]	0.3400	0.0650	105117	269090
			COPD	0.999[0.949,1.052]	0.9800	0.0650	12839	372779
			Asthma	0.913[0.885,0.941]	6.80E-09	0.0650	39372	343873
			Type-2 diabetes	1.083[1.045,1.123]	1.50E-05	0.0660	25722	433026
ABO	9:133271182:T:C	T	Coronary artery disease	1.022[1.004,1.040]	0.0180	0.2550	38269	416904
			Kidney disease	0.963[0.945,0.982]	0.0001	0.2590	35472	211418
			Hypertension	0.990[0.978,1.003]	0.1400	0.2570	105117	269090
			COPD	1.009[0.980,1.039]	0.5500	0.2570	12839	372779
			Asthma	1.012[0.995,1.030]	0.1800	0.2570	39372	343873
			Type-2 diabetes	1.042[1.021,1.064]	0.0001	0.2570	25722	433026
			Coronary artery disease	1.009[0.992,1.025]	0.3100	0.3230	38269	416904
			Kidney disease	0.982[0.964,1.000]	0.0460	0.3200	35472	211418

DPP9	19:4719431:G:A	A	Hypertension	0.998[0.987,1.010]	0.7700	0.3200	105117	269090
			COPD	0.992[0.965,1.019]	0.5600	0.3200	12839	372779
			Asthma	1.000[0.984,1.016]	0.9900	0.3200	39372	343873
			Type-2 diabetes	1.004[0.984,1.024]	0.7200	0.3200	25722	433026
IFNAR2	21:33252612:A:G	A	Coronary artery disease	1.012[0.994,1.029]	0.1900	0.2810	38269	416904
			Kidney disease	1.006[0.987,1.025]	0.5400	0.2860	35472	211418
			Hypertension	1.008[0.996,1.021]	0.1800	0.2870	105117	269090
			COPD	0.992[0.965,1.021]	0.6000	0.2870	12839	372779
			Asthma	1.011[0.995,1.028]	0.1900	0.2870	39372	343873
			Type-2 diabetes	1.013[0.992,1.033]	0.2300	0.2870	25722	433026
(iii) Two novel risk variants identified in the HGI analysis of C2 and that we found to also associate with disease severity								
RPL24	3:101705614:T:C	C	Coronary artery disease	1.006[0.990,1.022]	0.4900	0.3550	38269	416904
			Kidney disease	1.004[0.987,1.022]	0.6300	0.3550	35472	211418
			Hypertension	1.012[1.001,1.024]	0.0400	0.3540	105117	269090
			COPD	1.004[0.978,1.031]	0.7400	0.3550	12839	372779
			Asthma	0.980[0.965,0.996]	0.0120	0.3550	39372	343873
			Type-2 diabetes	0.991[0.973,1.010]	0.3700	0.3550	25722	433026
FOXP4	6:41534945:A:C	C	Coronary artery disease	0.977[0.924,1.033]	0.4100	0.0210	38269	416904
			Kidney disease	0.989[0.933,1.050]	0.7300	0.0210	35472	211418
			Hypertension	0.996[0.958,1.035]	0.8200	0.0210	105117	269090
			COPD	0.982[0.897,1.076]	0.7000	0.0210	12839	372779
			Asthma	0.976[0.926,1.030]	0.3800	0.0210	39372	343873
			Type-2 diabetes	1.009[0.946,1.077]	0.7800	0.0210	25722	433026

Supplementary Table 9. Association between two severity phenotypes and a novel risk variant for COVID-19 near ACE2 identified in this study.

Locus, nearest gene	rs ID	hg38	Phenotype	Effect allele	Odds Ratio* [95% CI]	P-value	N cases with 0|1|2 copies of the effect allele	N controls with 0|1|2 copies of the effect allele	Effect allele frequency	N cases	N controls	Cross-study heterogeneity test P-value
Xp22.2 ACE2	rs190509934	23:15602217:T:C	COVID19 positive hospitalized vs COVID19 positive not hospitalized	C	0.694[0.419,1.150]	1.57E-01	6766|13|0	44854|106|8	0.001	6779	44968	0.222
			COVID19 positive severe vs COVID19 positive not hospitalized	C	0.544[0.210,1.408]	2.10E-01	2157|2|0	44854|106|8	0.001	2159	44968	0.916

* Estimated based on trans-ancestry inverse-variance meta-analysis across UKB, AncestryDNA, GHS and UPENN-PMBB.

Supplementary Table 10. Published gene expression datasets used to identify sentinel expression quantitative trait loci (eQTL) that co-localized (LD r2>0.8) with sentinel GWAS variants.

First author	PMID	Tissue/cell type	Sample size
Alasoo	29379200	MACROPHAGES-IFNg	86
		MACROPHAGES-IFNg-salmonella	86
		MACROPHAGES-naive	86
		MACROPHAGES-salmonella	86
Andiappan	26259071	NEUTROPHILS	114
Barreiro	22233810	DENDRITIC-NotInfected	65
		DENDRITIC-TBInfected	65
Battle	24092820	WHOLE-BLOOD	922
		WHOLE-BLOOD-ase	922
		WHOLE-BLOOD-splice	922
Brown	29058714	LCLS	506
		SKIN	471
		WHOLE-BLOOD	242
Brumpton	27155841	LCLS	356
Caliskan	25874939	PBMCS-baseline	98
		PBMCS-rhinovirus	98
		PBMCS-rhinovirus-reQTL	98
Chen	27863251	CD4TCELLS	Up to 197
		MONOCYTES	Up to 197
		NEUTROPHILS	Up to 197
Chiang	28369037	LUNG	Up to 147
		SKIN	Up to 147
		WHOLEBLOOD	Up to 147
Davenport 2016	26917434	LEUCOCYTES	265
Davenport 2018	30340504	WHOLEBLOOD	157
		WHOLEBLOOD-IFN-interaction	157
Dimas	19644074	LCLS	75
		TCELLS	75
Di Narzo	27336838	WHOLE-BLOOD	149
Ding	21129726	LESIONAL-SKIN	57
		NORMAL-SKIN	53
		UNINVOLVED-SKIN	53
Dixon	17873877	LCLS	400
Fairfax 2012	22446964	BCELLS	283
		MONOCYTES	283
Fairfax 2014	24604202	MONOCYTES-IFN	367
		MONOCYTES-LPS2	261
		MONOCYTES-LPS24	322
		MONOCYTES-NAIVE	414
Fehrmann	21829388	WHOLE-BLOOD	1469
Ferraro	24610777	Tconv	65
		Tregs	65
Franco	23878721	WHOLE-BLOOD-Influenza	247
Lappalainen	24037378	LCLS	373
Grundberg	22941192	LCLS	856
		SKIN	856
GTEx v8	32912332	49 individual tissues	73-706
Hao	23209423	LUNG	1111
Huang	25951796	LCLs	368
		PBMCs	240
		SKIN	110
Jansen	28165122	WHOLE-BLOOD	4896
Joahnes	28122634	WHOLE-BLOOD	5257
Kasela	28248954	CD4TCELLS	293
		CD8TCELLS	293
KimHellmuth	28814792	MONOCYTES-LPS6H	134
		MONOCYTES-LPS90	134
		MONOCYTES-MDP6H	134
		MONOCYTES-MDP90	134
		MONOCYTES-RNA6H	134

Kim	25327457	MONOCYTES-Baseline	137
		MONOCYTES-Differential	137
		MONOCYTES-LPS	137
Kukurba	27197214	WHOLE-BLOOD	922
Lee	24604203	DENDRITIC-Baseline	528
		DENDRITIC-Flu	342
		DENDRITIC-Flu-delta	342
		DENDRITIC-IFNb	284
		DENDRITIC-IFNb-delta	284
		DENDRITIC-LPS	356
		DENDRITIC-LPS-delta	356
LloydJones	28065468	WHOLE-BLOOD	2765
Luo	26102239	SMALL-AIRWAYS	105
Momozawa	29930244	BCELLS	323
		CD4TCELLS	323
		CD8TCELLS	323
		GRANULOCYTES	323
		MONOCYTES	323
Murphy	20833654	CD4-TCELLS	200
Naranbhai	26151758	NEUTROPHILS	101
Nedelec	27768889	MACROPHAGES-baseline	95
		MACROPHAGES-baseline-asQTL	95
		MACROPHAGES-listeria	95
		MACROPHAGES-listeria-asQTL	95
		MACROPHAGES-listeria-reQTL	95
		MACROPHAGES-salmonella	95
		MACROPHAGES-salmonella-asQTL	95
		MACROPHAGES-salmonella-reQTL	95
Pala	28394350	LEUCOCYTES	624
Peters	27015630	BCELLS	80
		CD4-TCELLS	121
		CD8-TCELLS	108
		MONOCYTES	124
		NEUTROPHILS	121
Qiu 2011	21949713	SPUTUM	131
Qiu 2014	24770851	LCLS-steroids	117
Quach	27768888	MONOCYTES-baseline	100
		MONOCYTES-IAV	100
		MONOCYTES-LPS	100
		MONOCYTES-Pam3CSK4	100
		MONOCYTES-R848	100
Raj	24786080	CD4-TCELLS	407
		MONOCYTES	401
vanderWijst	29610479	BCELLS	45
		CD4TCELLS	45
		CD8TCELLS	45
		DENDRITIC	45
		MONOCYTES	45
		MONOCYTES-classical	45
		MONOCYTES-nonclassical	45
		NKCELLS	45
		PBMCS	45
Vosa	BioRxiv	WHOLE-BLOOD+PBMCS	31684
Walsh	27140173	WHOLE-BLOOD	377
Westra	24013639	WHOLE-BLOOD	5311
Yao	28285768	WHOLEBLOOD	5257
Ye	25214635	CD4-TCELLS-48h328	348
		CD4-TCELLS-48hTh17	348
		CD4-TCELLS-4h328	348
		CD4-TCELLS-4hIFNb	348
		CD4-TCELLS-UNST	348
Zeller	20502693	MONOCYTES	1490
Zhernakova	27918533	WHOLEBLOOD-exon-primary	2116
		WHOLEBLOOD-exonratio-primary	2116
		WHOLEBLOOD-gene-primary	2116
		WHOLEBLOOD-polyAratio-primary	2116

Supplementary Table 11. Association between systolic and diastolic blood pressure in Europeans of the UK Biobank study and (i) the rare non-coding variant near ACE2 that was found to be associated with COVID-19 risk; and (ii) a burden of ultra-rare coding variants in ACE2 that are predicted to be loss-of-function.

Variant	Effect allele	Phenotype	Effect[95% CI]	P-value	N individuals with 0|1|2 copies of the effect allele	Effect allele frequency
23:15602217:T:C	C	Systolic_blood_pressure	0.009[-0.022,0.041]	5.60E-01	406539|1179|493	3.00E-03
		Diastolic_blood_pressure	0.017[-0.017,0.051]	3.30E-01	406543|1179|493	3.00E-03
Burden of predicted loss-of-function (pLOF) variants	Minor allele for any pLOF variant	Systolic_blood_pressure	0.219[-0.031,0.469]	8.60E-02	408174|33|4	5.00E-05
		Diastolic_blood_pressure	0.195[-0.070,0.461]	1.50E-01	408178|33|4	5.00E-05

Supplementary Table 12. Association between eight published risk variants for COVID-19 and five disease risk phenotypes. P-values in red were significant after correcting for the 40 tests performed (P<0.00125).

Locus, nearest gene	rs ID	hg38	Phenotype	Effect allele	Odds Ratio* [95% CI]	P-value	N cases with 0|1|2 copies of the effect allele	N controls with 0|1|2 copies of the effect allele	Effect allele frequency	N cases	N controls	Heterozygote Odds Ratio** [95% CI]	Homozygote Odds Ratio** [95% CI]	Cross-study heterogeneity test P-value
3p21.31 SLC6A20	rs2531743	3:45796808:G:A	COVID19_positive_hospitalized_vs_COVID19_negative_or_unknown	G	0.971[0.937,1.006]	1.07E-01	2398|3345|1168	234336|336160|119124	0.416	6911	689620	0.971[0.921,1.023]	0.954[0.889,1.024]	0.1686
			COVID19_positive_not_hospitalized_vs_COVID19_negative_or_unknown	G	0.954[0.940,0.969]	8.21E-10	16157|22128|7356	240080|342626|121310	0.415	45641	704016	0.966[0.946,0.988]	0.913[0.886,0.94]	0.6444
			COVID19_positive_severe_vs_COVID19_negative_or_unknown	G	0.917[0.861,0.977]	7.19E-03	776|1072|336	234336|336160|119124	0.416	2184	689620	0.957[0.872,1.05]	0.841[0.739,0.956]	0.5430
			COVID19_positive_vs_COVID19_negative	G	0.943[0.927,0.958]	2.95E-12	18592|25509|8529	36660|53558|19387	0.416	52630	109605	0.951[0.929,0.974]	0.89[0.861,0.919]	0.6468
			COVID19_positive_vs_COVID19_negative_or_unknown	G	0.958[0.945,0.971]	1.19E-09	18592|25509|8529	240080|342626|121310	0.415	52630	704016	0.968[0.948,0.987]	0.919[0.894,0.945]	0.6063
3p21.31 LZTFL1	rs73064425	3:45859597:C:T	COVID19_positive_hospitalized_vs_COVID19_negative_or_unknown	T	1.289[1.211,1.371]	9.71E-16	5713|1127|71	592971|92456|4193	0.073	6911	689620	1.285[1.204,1.371]	1.699[1.336,2.16]	0.0306
			COVID19_positive_not_hospitalized_vs_COVID19_negative_or_unknown	T	1.017[0.990,1.045]	2.25E-01	39073|6308|260	605892|93863|4261	0.073	45641	704016	1.03[1.001,1.06]	0.899[0.79,1.024]	0.5419
			COVID19_positive_severe_vs_COVID19_negative_or_unknown	T	1.581[1.427,1.752]	2.16E-18	1732|421|31	592971|92456|4193	0.073	2184	689620	1.585[1.423,1.766]	2.591[1.808,3.712]	0.5770
			COVID19_positive_vs_COVID19_negative	T	1.067[1.035,1.100]	2.67E-05	44855|7443|332	94427|14495|683	0.074	52630	109605	1.076[1.043,1.11]	0.958[0.836,1.098]	0.8732
			COVID19_positive_vs_COVID19_negative_or_unknown	T	1.052[1.026,1.079]	7.32E-05	44855|7443|332	605892|93863|4261	0.073	52630	704016	1.063[1.035,1.091]	1.001[0.891,1.124]	0.6960
6p21.33 CCHCR1	rs143334143	6:31153649:G:A	COVID19_positive_hospitalized_vs_COVID19_negative_or_unknown	A	1.103[1.035,1.176]	2.53E-03	5839|1017|55	592380|93262|3978	0.073	6911	689620	1.103[1.031,1.179]	1.379[1.055,1.803]	0.4370
			COVID19_positive_not_hospitalized_vs_COVID19_negative_or_unknown	A	1.021[0.995,1.048]	1.07E-01	38299|7008|334	604383|95543|4090	0.074	45641	704016	1.039[1.011,1.068]	1.021[0.909,1.147]	0.7970
			COVID19_positive_severe_vs_COVID19_negative_or_unknown	A	1.175[1.054,1.310]	3.63E-03	1814|347|23	592380|93262|3978	0.073	2184	689620	1.166[1.039,1.308]	1.727[1.143,2.611]	0.8028
			COVID19_positive_vs_COVID19_negative	A	1.057[1.026,1.088]	2.00E-04	44198|8040|392	93495|15427|683	0.079	52630	109605	1.065[1.033,1.098]	1.104[0.969,1.256]	0.4270
			COVID19_positive_vs_COVID19_negative_or_unknown	A	1.032[1.008,1.058]	9.47E-03	44198|8040|392	604383|95543|4090	0.074	52630	704016	1.047[1.021,1.074]	1.069[0.96,1.191]	0.5346
9q34.2 ABO	rs879055593 (proxy for rs9411378)	9:133271182:T:C	COVID19_positive_hospitalized_vs_COVID19_negative_or_unknown	T	1.096[1.054,1.139]	4.06E-06	3743|2658|510	384940|259919|44761	0.253	6911	689620	1.077[1.024,1.132]	1.221[1.112,1.34]	0.7870
			COVID19_positive_not_hospitalized_vs_COVID19_negative_or_unknown	T	1.098[1.081,1.116]	7.88E-30	24427|18060|3154	393656|264837|45523	0.254	45641	704016	1.114[1.091,1.137]	1.156[1.111,1.203]	0.3500
			COVID19_positive_severe_vs_COVID19_negative_or_unknown	T	1.096[1.022,1.175]	1.01E-02	1174|852|158	384940|259919|44761	0.253	2184	689620	1.09[0.998,1.191]	1.19[1.007,1.407]	0.4745
			COVID19_positive_vs_COVID19_negative	T	1.097[1.078,1.117]	9.96E-24	28215|20747|3668	61047|41288|7270	0.259	52630	109605	1.114[1.089,1.14]	1.136[1.088,1.187]	0.6010
			COVID19_positive_vs_COVID19_negative_or_unknown	T	1.098[1.081,1.114]	7.07E-34	28215|20747|3668	393656|264837|45523	0.254	52630	704016	1.109[1.088,1.13]	1.165[1.123,1.209]	0.1825
12q24.13 OAS3	rs10735079	12:112942203:G:A	COVID19_positive_hospitalized_vs_COVID19_negative_or_unknown	G	0.968[0.934,1.003]	7.67E-02	2988|3067|856	283922|316030|89668	0.359	6911	689620	0.945[0.898,0.994]	0.949[0.879,1.025]	0.6392
			COVID19_positive_not_hospitalized_vs_COVID19_negative_or_unknown	G	0.997[0.983,1.012]	7.18E-01	19607|20340|5694	290868|322016|91132	0.358	45641	704016	0.976[0.956,0.997]	0.992[0.961,1.024]	0.4075
			COVID19_positive_severe_vs_COVID19_negative_or_unknown	G	0.975[0.915,1.039]	4.31E-01	942|973|269	283922|316030|89668	0.359	2184	689620	0.953[0.87,1.043]	0.957[0.835,1.097]	0.9016
			COVID19_positive_vs_COVID19_negative	G	0.996[0.979,1.012]	6.09E-01	22636|23437|6557	45628|49969|14008	0.353	52630	109605	0.97[0.948,0.993]	0.987[0.953,1.022]	0.7406
			COVID19_positive_vs_COVID19_negative_or_unknown	G	0.993[0.979,1.007]	3.11E-01	22636|23437|6557	290868|322016|91132	0.357	52630	704016	0.972[0.953,0.991]	0.986[0.958,1.016]	0.3908
19p13.3 DPP9	rs2109069	19:4719431:G:A	COVID19_positive_hospitalized_vs_COVID19_negative_or_unknown	A	1.100[1.061,1.141]	3.44E-07	3180|2954|777	327458|294675|67487	0.312	6911	689620	1.055[1.003,1.11]	1.227[1.134,1.328]	0.0945
			COVID19_positive_not_hospitalized_vs_COVID19_negative_or_unknown	A	1.007[0.992,1.023]	3.59E-01	22268|19042|4331	335629|299944|68443	0.310	45641	704016	0.994[0.973,1.015]	1.027[0.992,1.064]	0.1766
			COVID19_positive_severe_vs_COVID19_negative_or_unknown	A	1.121[1.050,1.196]	5.93E-04	994|936|254	327458|294675|67487	0.312	2184	689620	1.073[0.981,1.173]	1.308[1.139,1.502]	0.8598
			COVID19_positive_vs_COVID19_negative	A	1.017[0.999,1.034]	6.21E-02	25489|22028|5113	53011|46169|10425	0.306	52630	109605	1.012[0.989,1.035]	1.051[1.012,1.092]	0.5871
			COVID19_positive_vs_COVID19_negative_or_unknown	A	1.019[1.004,1.034]	1.02E-02	25489|22028|5113	335629|299944|68443	0.310	52630	704016	1.003[0.984,1.022]	1.054[1.021,1.089]	0.5211
19p13.2 RAVER1	rs74956615	19:10317045:T:A	COVID19_positive_hospitalized_vs_COVID19_negative_or_unknown	A	1.092[1.011,1.180]	2.59E-02	6201|681|29	620702|67104|1814	0.051	6911	689620	1.052[0.971,1.139]	1.668[1.155,2.41]	0.0060
			COVID19_positive_not_hospitalized_vs_COVID19_negative_or_unknown	A	1.012[0.978,1.046]	4.97E-01	41354|4180|107	634281|67914|1821	0.051	45641	704016	1.007[0.973,1.042]	0.959[0.784,1.174]	0.3301
			COVID19_positive_severe_vs_COVID19_negative_or_unknown	A	1.244[1.087,1.424]	1.58E-03	1937|238|9	620702|67104|1814	0.051	2184	689620	1.188[1.037,1.36]	1.672[0.867,3.225]	0.1839
			COVID19_positive_vs_COVID19_negative	A	1.022[0.984,1.060]	2.60E-01	47425|4863|136	98596|10384|282	0.050	52424	109262	1.008[0.972,1.046]	1.068[0.865,1.319]	0.9762
			COVID19_positive_vs_COVID19_negative_or_unknown	A	1.022[0.991,1.055]	1.64E-01	47630|4864|136	634281|67914|1821	0.051	52630	704016	1.012[0.98,1.044]	1.061[0.886,1.27]	0.3694
21q22.1 IFNAR2	rs2236757	21:33252612:A:G	COVID19_positive_hospitalized_vs_COVID19_negative_or_unknown	A	1.079[1.039,1.120]	7.34E-05	3369|2898|644	351679|280661|57280	0.287	6911	689620	1.079[1.026,1.134]	1.165[1.07,1.269]	0.4116
			COVID19_positive_not_hospitalized_vs_COVID19_negative_or_unknown	A	1.012[0.996,1.028]	1.51E-01	22463|18988|4190	358003|286693|59320	0.289	45641	704016	1.021[1,1.042]	1.031[0.995,1.069]	0.2857
			COVID19_positive_severe_vs_COVID19_negative_or_unknown	A	1.140[1.066,1.218]	1.10E-04	1032|930|222	351679|280661|57280	0.287	2184	689620	1.125[1.029,1.23]	1.295[1.118,1.5]	0.3644
			COVID19_positive_vs_COVID19_negative	A	1.026[1.009,1.045]	3.50E-03	25861|21920|4849	55434|44836|9335	0.293	52630	109605	1.029[1.005,1.052]	1.054[1.013,1.096]	0.8180
			COVID19_positive_vs_COVID19_negative_or_unknown	A	1.023[1.008,1.038]	2.51E-03	25861|21920|4849	358003|286693|59320	0.289	52630	704016	1.028[1.008,1.048]	1.048[1.014,1.083]	0.6665

* Estimated based on trans-ancestry inverse-variance meta-analysis across UKB, AncestryDNA, GHS and UPENN-PMBB.

** Estimated based on ancestry-specific genotype counts in UKB, AncestryDNA, GHS and UPENN-PMBB, using the Mantel-Haenszel procedure.

Supplementary Table 13. Comparison of effect sizes between ancestries for the six published risk variants that were validated in this study.

Locus, nearest gene	rs ID	hg38	Effect allele	Phenotype	Ancestry	Odds Ratio (95% CI)	P-value	N cases with 0|1|2 copies of the effect allele	N controls with 0|1|2 copies of the effect allele	Effect allele frequency	Cross-ancestry heterogeneity test P-value
3p21.31 SLC6A20	rs2531743	3:45796808:G:A	G	COVID19 positive vs COVID19 negative	AFR	0.9 (0.84, 0.98)	0.0091	716|1353|657	1200|2672|1327	0.504	0.647
					AMR	0.91 (0.85, 0.98)	0.0155	1606|1778|456	1457|1727|525	0.362	
					EAS	0.96 (0.6, 1.54)	0.8615	142|60|4	228|101|14	0.179	
					EUR	0.95 (0.93, 0.96)	8.41E-10	15747|21938|7338	33271|48507|17392	0.416	
					SAS	0.94 (0.79, 1.12)	0.498	381|380|74	504|551|129	0.331	
3p21.31 LZTFL1	rs73064425	3:45859597:C:T	T	COVID19 positive severe vs COVID19 negative or unknown	AFR	1.11 (0.42, 2.95)	0.8284	126|3|1	22413|641|2	0.014	0.577
					AMR	1.63 (0.81, 3.27)	0.1731	69|11|1	11052|1236|35	0.053	
					EUR	1.6 (1.44, 1.78)	4.32E-18	1522|391|28	553956|86912|3491	0.073	
					SAS	1.26 (0.68, 2.31)	0.4591	15|16|1	5550|3667|665	0.253	
6p21.33 CCHCR1	rs143334143	6:31153649:G:A	A	COVID19 positive vs COVID19 negative	AFR	1.0 (0.86, 1.17)	0.9888	2412|307|7	4640|542|17	0.057	0.427
					AMR	1.02 (0.92, 1.14)	0.6816	2937|846|57	2904|747|58	0.121	
					EAS	0.8 (0.44, 1.44)	0.4582	166|36|4	285|55|3	0.096	
					EUR	1.06 (1.03, 1.1)	0.0001	38029|6686|308	84722|13856|592	0.077	
					SAS	1.09 (0.87, 1.35)	0.4527	654|165|16	944|227|13	0.111	
9q34.2 ABO	rs879055593 (proxy for rs9411378)	9:133271182:T:C	T	COVID19 positive vs COVID19 negative or unknown	AFR	1.1 (1.01, 1.2)	0.0224	1781|858|87	18064|7277|765	0.171	0.183
					AMR	1.16 (1.08, 1.24)	2.42E-05	2391|1272|177	8706|4386|538	0.203	
					EAS	1.23 (0.93, 1.64)	0.1535	123|75|8	1815|851|113	0.196	
					EUR	1.1 (1.08, 1.11)	7.89E-30	23347|18306|3370	358276|249302|43780	0.260	
					SAS	0.95 (0.83, 1.1)	0.491	573|236|26	6795|3021|327	0.180	
19p13.3 DPP9	rs2109069	19:4719431:G:A	A	COVID19 positive hospitalized vs COVID19 negative or unknown	AFR	1.32 (1.11, 1.56)	0.0014	229|155|31	14025|7882|1149	0.221	0.094
					AMR	0.94 (0.74, 1.2)	0.6154	116|72|11	6722|4760|841	0.261	
					EUR	1.09 (1.05, 1.14)	4.87E-06	2743|2672|726	300516|278789|65054	0.317	
					SAS	1.15 (0.85, 1.56)	0.3533	92|55|9	6195|3244|443	0.209	
21q22.1 IFNAR2	rs2236757	21:33252612:A:G	A	COVID19 positive hospitalized vs COVID19 negative or unknown	AFR	1.05 (0.88, 1.25)	0.5825	245|145|25	13867|7992|1197	0.225	0.412
					AMR	1.25 (1.02, 1.53)	0.0355	55|111|33	4813|5759|1751	0.377	
					EUR	1.08 (1.03, 1.12)	0.0003	3011|2578|552	329458|262225|52676	0.285	
					SAS	1.03 (0.81, 1.32)	0.7863	58|64|34	3541|4685|1656	0.405	

Supplementary Table 14. Association between two severity phenotypes and six published risk variants for COVID-19 that validated in this study.

Locus, nearest gene	rs ID	hg38	Phenotype	Effect allele	Odds Ratio* [95% CI]	P-value	N cases with 0|1|2 copies of the effect allele	N controls with 0|1|2 copies of the effect allele	Effect allele frequency	N cases	N controls	Heterozygote Odds Ratio** [95% CI]	Homozygote Odds Ratio** [95% CI]	Cross-study heterogeneity test P-value
3p21.31 SLC6A20	rs2531743	3:45796808:G:A	COVID19_positive_hospitalized_vs_COVID19_positive_not_hospitalized_combined	G	1.023[0.980,1.068]	2.99E-01	2398|3345|1168	15913|21963|7309	0.406	6911	45185	1.016[0.958,1.078]	1.038[0.959,1.124]	0.1999
			COVID19_positive_severe_vs_COVID19_positive_not_hospitalized_combined	G	0.949[0.884,1.019]	1.47E-01	776|1072|336	15913|21963|7309	0.405	2184	45185	1.000[0.909,1.1]	0.92[0.806,1.051]	0.5371
3p21.31 LZTFL1	rs73064425	3:45859597:C:T	COVID19_positive_hospitalized_vs_COVID19_positive_not_hospitalized_combined	T	1.338[1.243,1.442]	1.50E-14	5713|1127|71	38671|6256|258	0.077	6911	45185	1.269[1.18,1.365]	1.835[1.393,2.418]	0.0522
			COVID19_positive_severe_vs_COVID19_positive_not_hospitalized_combined	T	1.591[1.417,1.786]	4.28E-15	1732|421|31	38671|6256|258	0.077	2184	45185	1.545[1.381,1.728]	2.81[1.916,4.121]	0.5010
6p21.33 CCHCR1	rs143334143	6:31153649:G:A	COVID19_positive_hospitalized_vs_COVID19_positive_not_hospitalized_combined	A	1.097[1.018,1.182]	1.54E-02	5839|1017|55	37929|6929|327	0.084	6911	45185	1.055[0.98,1.137]	1.363[1.011,1.836]	0.2440
			COVID19_positive_severe_vs_COVID19_positive_not_hospitalized_combined	A	1.142[1.013,1.289]	3.06E-02	1814|347|23	37929|6929|327	0.084	2184	45185	1.115[0.99,1.255]	1.71[1.111,2.631]	0.6952
9q34.2 ABO	rs879055593 (proxy for rs9411378)	9:133271182:T:C	COVID19_positive_hospitalized_vs_COVID19_positive_not_hospitalized_combined	T	1.022[0.976,1.071]	3.56E-01	3743|2658|510	24131|17915|3139	0.268	6911	45185	0.972[0.919,1.028]	1.049[0.945,1.165]	0.9457
			COVID19_positive_severe_vs_COVID19_positive_not_hospitalized_combined	T	1.000[0.926,1.080]	9.98E-01	1174|852|158	24131|17915|3139	0.268	2184	45185	0.979[0.894,1.073]	1.023[0.861,1.216]	0.4760
19p13.3 DPP9	rs2109069	19:4719431:G:A	COVID19_positive_hospitalized_vs_COVID19_positive_not_hospitalized_combined	A	1.081[1.034,1.129]	5.66E-04	3180|2954|777	21966|18903|4316	0.308	6911	45185	1.067[1.008,1.128]	1.223[1.119,1.337]	0.0077
			COVID19_positive_severe_vs_COVID19_positive_not_hospitalized_combined	A	1.097[1.021,1.180]	1.20E-02	994|936|254	21966|18903|4316	0.306	2184	45185	1.082[0.986,1.186]	1.281[1.111,1.479]	0.4921
21q22.1 IFNAR2	rs2236757	21:33252612:A:G	COVID19_positive_hospitalized_vs_COVID19_positive_not_hospitalized_combined	A	1.075[1.027,1.124]	1.76E-03	3369|2898|644	22305|18785|4095	0.299	6911	45185	1.064[1.006,1.125]	1.137[1.033,1.251]	0.6332
			COVID19_positive_severe_vs_COVID19_positive_not_hospitalized_combined	A	1.101[1.022,1.185]	1.09E-02	1032|930|222	22305|18785|4095	0.299	2184	45185	1.102[1.006,1.208]	1.257[1.081,1.463]	0.3129

* Estimated based on trans-ancestry inverse-variance meta-analysis across UKB, AncestryDNA, GHS and UPENN-PMBB.

** Estimated based on ancestry-specific genotype counts in UKB, AncestryDNA, GHS and UPENN-PMBB, using the Mantel-Haenszel procedure.

Supplementary Table 15. Fifteen variants associated with COVID-19 in the latest meta-analyses performed by the COVID-19 Host Genetics Initiative*.

Locus index	rs ID	hg38	Nearest gene	HGI phenotype	Effect allele	Odds Ratio	P-value	Locus reported in previous GWAS	HGI phenotype independent from disease severity phenotypes among cases	Association with risk of hospitalization among cases in our study						Association with risk of severe disease among cases in our study					
										Odds Ratio (95% CI)	P-value	N cases RR|RA|AA	N controls RR|RA|AA	AAF	Cross-study heterogeneity test P-value	Odds Ratio (95% CI)	P-value	N cases RR|RA|AA	N controls RR|RA|AA	AAF	Cross-study heterogeneity test P-value
1	rs10490770	3:45823240:T:C	*LZTFL1*	*C2 - Reported infection*	C	1.16	9.72E-30	Yes	Yes	Variants in four loci that were reported in previous GWAS. Association between these loci and disease severity is included in Supplementary Table 10.											
2	rs912805253	9:133274084:C:T	*ABO*		T	0.9	1.45E-39	Yes	Yes												
3	rs10774671	12:112919388:G:A	*OAS1*		A	1.06	1.61E-11	Yes	Yes												
4	rs2109069	19:4719431:G:A	*DPP9*		A	1.05	4.08E-09	Yes	Yes												
5	rs74956615	19:10317045:T:A	*RAVER1*	*B2 - Risk of hospitalization*	T	1.27	5.05E-10	Yes	No												
6	rs13050728	21:33242905:T:C	*IFNAR2*		C	0.86	2.72E-20	Yes	No												
7	rs67579710	1:155203736:G:A	*THBS3*	*B2 - Risk of hospitalization*	A	0.87	3.38E-08	No	No	Variants in five new loci discovered in the HGI analysis of phenotypes (A2 and B2) that are not independent of the two disease severity phenotypes tested in our study. As such, we did not test if these variants were associated with disease severity among casess.											
8	rs1381109	2:166061783:G:T	*SCN1A*		T	0.91	4.21E-08	No	No												
9	rs72711165	8:124324323:T:C	*TMEM65*		C	1.37	2.13E-09	No	No												
10	rs1819040	17:46142465:T:A	*KANSL1*		A	0.88	1.83E-10	No	No												
11	rs77534576	17:49863303:C:T	*TAC4*	*A2 - rusk of severe disease*	T	1.45	4.37E-09	No	No												
12	rs11919389	3:101705614:T:C	*RPL24*	*C2 - Reported infection*	C	0.94	3.40E-15	No	Yes	0.962[0.921,1.005]	0.084	3142|3025|744	20405|19847|4933	0.328	0.751	0.923[0.859,0.992]	0.029	1015|961|208	20405|19847|4933	0.328	0.767
13	rs10070196	5:13939721:A:C	*DNAH5*		C	1.05	2.30E-08	No	Yes	0.989[0.947,1.033]	0.622	793|3038|3080	5100|19947|20138	0.666	0.451	0.956[0.890,1.027]	0.218	242|1017|925	5100|19947|20138	0.666	0.739
14	rs1886814	6:41534945:A:C	*FOXP4*		C	1.11	2.40E-08	No	Yes	1.159[1.031,1.302]	0.013	6435|458|18	41747|3252|186	0.040	0.126	1.255[1.039,1.517]	0.018	2024|152|8	41747|3252|186	0.040	0.445
15	rs4801778	19:48867352:G:T	*PLEKHA4*		T	0.95	1.20E-08	No	Yes	0.999[0.946,1.055]	0.966	4670|2028|213	30955|12887|1343	0.173	0.165	0.925[0.845,1.012]	0.090	1513|609|62	30955|12887|1343	0.172	0.919

* Reported in Table 1 of https://www.medrxiv.org/content/10.1101/2021.03.10.21252820v1. If a variant was associated with multiple phenotypes, we show results for C2 over B2 over A2.

Supplementary Table 16. Missense or predicted loss-of-function variants in high linkage disequilibrium (LD, r2>0.80) with sentinel GWAS variants.

Locus	Sentinel GWAS variant			Missense or pLOF in high LD with lead variant				
	rs ID	hg38	Nearest gene	Gene	rs ID	hg38	r2 with sentinel GWAS variant*	Variant effect
21q22.1	rs2236757	21:33252612:A:G	*IFNAR2*	*IFNAR2*	rs1051393	21:33241950:T:G	0.88	Missense:Phe10Val
					rs1131668	21:33262573:G:A	0.86	Missense:Ala285Thr
6p21.33	rs143334143	6:31153649:G:A	*CCHCR1*	*CCHCR1*	rs130072	6:31144707:C:T	1.00	Missense:Arg716Gln
					rs11540822	6:31151121:A:T	1.00	Missense:Leu268Gln
					rs72856718	6:31157480:C:A	1.00	Stop_gained:Glu41X
				TCF19	rs2073724	6:31161930:C:T	1.00	Missense:Pro241Leu
				C6orf15	rs2233976	6:31112217:C:T	0.87	Missense:Gly48Arg
				PSORS1C1	rs1265097	6:31138682:C:A	0.84	Missense:Pro24Thr
* LD estimated based on individuals of European ancestry from the 1000 Genomes project								

Supplementary Table 17. Expression quantitative trait loci (eQTL) that co-localized (LD r2>0.80) with sentinel GWAS variants.

Sentinel GWAS variant				Sentinel eQTL									
Locus	rs ID	hg38	Disease predisposing allele	rs ID	hg38	Gene	eQTL dataset	P-value	Effect allele	Reference allele	Effect	r2 with sentinel GWAS variant*	Allele in phase with disease predisposing allele
3p21.31	rs73064425	3:45859597:C:T	T	rs17763537	3:45801823:C:T	SLC6A20	Hao__Lung	1.00E-16	C	T	-9.32	0.961	T
3p21.31	rs11919389	3:101705614:T:C	T	rs36002990	3:101719018:G:A	NXPE3	GTeXv8__Esophagus_Muscularis	3.27E-40	A	G	-0.753	0.99	G
	rs11919389	3:101705614:T:C	T	rs13100262	3:101695258:T:C	NXPE3	Pala__Leucocytes__isoQTL	6.25E-17	T	C	0.535	0.962	T
	rs11919389	3:101705614:T:C	T	rs13100262	3:101695258:T:C	NXPE3	GTeXv8__Esophagus_Gastroesophageal_Junction	1.88E-16	C	T	-0.452	0.962	T
	rs11919389	3:101705614:T:C	T	rs11923273	3:101619019:T:C	SENP7	Zhernakova__Whole_Blood__gene	6.53E-40	C	T	-13.222	0.81	T
	rs11919389	3:101705614:T:C	T	rs11923273	3:101619019:T:C	SENP7	Pala__Leucocytes	4.25E-14	T	C	0.372	0.81	T
6p21.33	rs143334143	6:31153649:G:A	A	rs130072	6:31144707:C:T	TCF19	Yao__Whole_Blood	2.14E-96	C	T	-0.092	0.999	T
	rs143334143	6:31153649:G:A	A	rs130072	6:31144707:C:T	TCF19	GTeXv8__Lung	8.91E-11	T	C	0.416	0.999	T
	rs143334143	6:31153649:G:A	A	rs114881571	6:31160828:C:T	TCF19	Jansen__Whole_Blood	1.20E-35	T	NA	0.416	0.998	T
	rs143334143	6:31153649:G:A	A	rs112279190	6:31155370:C:T	TCF19	GTeXv8__Artery_Aorta	2.14E-22	T	C	1.019	0.998	T
	rs143334143	6:31153649:G:A	A	rs150765547	6:31191804:G:A	LST1	Vosa__Whole_Blood	2.01E-27	A	G	-10.849	0.99	A
	rs143334143	6:31153649:G:A	A	rs11540822	6:31151121:A:T	HLA-C	GTeXv8__Adipose_Subcutaneous	1.28E-10	T	A	0.407	0.989	T
	rs143334143	6:31153649:G:A	A	rs72863830	6:31221764:G:A	TCF19	GTeXv8__Whole_Blood	9.62E-19	A	G	0.324	0.98	A
	rs143334143	6:31153649:G:A	A	rs17197087	6:31158455:C:T	TCF19	Zhernakova__Whole_Blood__gene	4.90E-156	T	C	26.612	0.976	T
	rs143334143	6:31153649:G:A	A	rs17197087	6:31158455:C:T	TCF19	Joehanes__Whole_Blood	1.61E-98	C	T	-0.091	0.976	T
	rs143334143	6:31153649:G:A	A	rs17197087	6:31158455:C:T	TCF19	Ratnapryia__Retina	4.25E-30	T	C	1.371	0.976	T
	rs143334143	6:31153649:G:A	A	rs17197087	6:31158455:C:T	TCF19	GTeXv8__Brain_Cortex	3.28E-22	T	C	1.142	0.976	T
	rs143334143	6:31153649:G:A	A	rs17197087	6:31158455:C:T	TCF19	GTeXv8__Pituitary	4.58E-22	T	C	1.108	0.976	T
	rs143334143	6:31153649:G:A	A	rs17197087	6:31158455:C:T	TCF19	Chen__CD4Tcells	1.35E-18	T	C	1.82	0.976	T
	rs143334143	6:31153649:G:A	A	rs17197087	6:31158455:C:T	TCF19	GTeXv8__Brain_Cerebellar_Hemisphere	1.95E-17	T	C	1.215	0.976	T
	rs143334143	6:31153649:G:A	A	rs17197087	6:31158455:C:T	TCF19	GTeXv8__Brain_Hypothalamus	8.32E-17	T	C	1.383	0.976	T
	rs143334143	6:31153649:G:A	A	rs17197087	6:31158455:C:T	TCF19	GTeXv8__Muscle_Skeletal	6.93E-14	T	C	0.474	0.976	T
	rs143334143	6:31153649:G:A	A	rs17197087	6:31158455:C:T	TCF19	GTeXv8__Brain_Caudate_basal_ganglia	4.73E-10	T	C	1.02	0.976	T
	rs143334143	6:31153649:G:A	A	rs17197087	6:31158455:C:T	TCF19	GTeXv8__Brain_Frontal_Cortex_BA9	6.53E-10	T	C	0.961	0.976	T
	rs143334143	6:31153649:G:A	A	rs17197087	6:31158455:C:T	TCF19	GTeXv8__Brain_Amygdala	1.14E-09	T	C	1.241	0.976	T
	rs143334143	6:31153649:G:A	A	rs17197101	6:31158632:G:T	TCF19	Battle__Whole_Blood	3.43E-99	NA	NA	0	0.975	T
	rs143334143	6:31153649:G:A	A	rs17197101	6:31158632:G:T	TCF19	Pala__Leucocytes	2.95E-49	G	T	-1.648	0.975	T
	rs143334143	6:31153649:G:A	A	rs17197101	6:31158632:G:T	TCF19	Davenport__Whole_Blood	1.50E-36	T	G	0.438	0.975	T
	rs143334143	6:31153649:G:A	A	rs17197101	6:31158632:G:T	TCF19	Walsh__Whole_Blood	2.54E-15	NA	NA	0	0.975	T
	rs143334143	6:31153649:G:A	A	rs2233976	6:31112217:C:T	TCF19	Raj__CD4Tcells	2.02E-27	NA	NA	0	0.872	T
21q22.1	rs2236757	21:33252612:A:G	A	rs3153	21:33237200:G:A	IFNAR2	Westra__Whole_Blood	1.97E-71	A	G	17.871	0.997	A
	rs2236757	21:33252612:A:G	A	rs574424283	21:33232654:T:TAA	IL10RB	Brown__LCLs	5.24E-11	NA	NA	0	0.997	TAA
	rs2236757	21:33252612:A:G	A	rs12053666	21:33246134:G:A	IL10RB	Kasela__CD4Tcells	4.07E-13	A	G	7.253	0.995	A
	rs2236757	21:33252612:A:G	A	rs2834163	21:33248146:G:A	IFNAR2	GTeXv8__Muscle_Skeletal	1.01E-15	A	G	-0.239	0.883	G
	rs2236757	21:33252612:A:G	A	rs2252639	21:33245424:A:G	IFNAR2	LloydJones__Whole_Blood	4.60E-44	A	G	0.411	0.88	A
	rs2236757	21:33252612:A:G	A	rs17860115	21:33230000:C:A	IFNAR2	Huang__LCLs	1.85E-24	C	NA	0.483	0.88	A
	rs2236757	21:33252612:A:G	A	rs17860115	21:33230000:C:A	IFNAR2	Fairfax__Monocytes__naive	1.84E-20	NA	NA	0	0.88	A
	rs2236757	21:33252612:A:G	A	rs17860115	21:33230000:C:A	IFNAR2	Huang__PBMCs	5.92E-20	C	NA	0.678	0.88	A
	rs2236757	21:33252612:A:G	A	rs17860115	21:33230000:C:A	IFNAR2	Hao__Lung	1.00E-16	C	A	11	0.88	A
	rs2236757	21:33252612:A:G	A	rs2834158	21:33244908:T:C	IFNAR2	Fairfax__Monocytes__IFN	2.22E-13	NA	NA	0	0.88	T
	rs2236757	21:33252612:A:G	A	rs17860115	21:33230000:C:A	IFNAR2	Momozawa__Monocytes	4.09E-11	A	C	-0.118	0.88	A
	rs2236757	21:33252612:A:G	A	rs12482556	21:33230629:T:C	IFNAR2	Momozawa__CD4Tcells	1.54E-16	C	T	-0.147	0.879	C
	rs2236757	21:33252612:A:G	A	rs2236758	21:33253108:A:G	IFNAR2	Battle__Whole_Blood__splice	1.92E-147	NA	NA	0	0.878	A
	rs2236757	21:33252612:A:G	A	rs2236758	21:33253108:A:G	IFNAR2	Fehrmann__Whole_Blood	3.80E-19	A	G	8.94	0.878	A
	rs2236757	21:33252612:A:G	A	rs2250226	21:33260011:T:C	IL10RB	Westra__Whole_Blood	2.51E-13	C	T	7.319	0.878	C
	rs2236757	21:33252612:A:G	A	rs17860241	21:33258828:A:G	IFNAR2	Brown__LCLs	1.17E-30	NA	NA	0	0.876	G
	rs2236757	21:33252612:A:G	A	rs17860241	21:33258828:A:G	IFNAR2	GTeXv8__Adipose_Subcutaneous	3.71E-18	G	A	-0.287	0.876	G

	rs2236757	21:33252612:A:G	A	rs17860241	21:33258828:A:G	IFNAR2	GTeXv8__Esophagus_Mucosa	1.49E-16	G	A	-0.292	0.876	G
	rs2236757	21:33252612:A:G	A	rs11911133	21:33256870:A:G	IFNAR2	Vosa__Whole_Blood	2.09E-280	G	A	-35.782	0.875	G
	rs2236757	21:33252612:A:G	A	rs11911133	21:33256870:A:G	IFNAR2	Davenport__Whole_Blood	3.37E-16	G	A	-0.114	0.875	G
	rs2236757	21:33252612:A:G	A	rs9975538	21:33254549:C:T	IL10RB	GTeXv8__Muscle_Skeletal	3.91E-13	T	C	-0.253	0.875	C
	rs2236757	21:33252612:A:G	A	rs6517156	21:33261740:C:G	IL10RB	Vosa__Whole_Blood	1.16E-117	G	C	23.06	0.862	G
	rs2236757	21:33252612:A:G	A	rs6517156	21:33261740:C:G	IL10RB	Joehanes__Whole_Blood	5.21E-41	C	G	-0.038	0.862	G
	rs2236757	21:33252612:A:G	A	rs6517156	21:33261740:C:G	IFNAR2	Lappalainen__LCLs	7.65E-36	NA	NA	0	0.862	G
	rs2236757	21:33252612:A:G	A	rs6517156	21:33261740:C:G	IL10RB	GTeXv8__Cells_Cultured_fibroblasts	1.82E-29	G	C	-0.264	0.862	G
	rs2236757	21:33252612:A:G	A	rs6517156	21:33261740:C:G	IFNAR2	Brown__Adipose	2.87E-25	NA	NA	0	0.862	G
	rs2236757	21:33252612:A:G	A	rs6517156	21:33261740:C:G	IL10RB	GTeXv8__Brain_Cerebellum	4.75E-25	G	C	0.687	0.862	G
	rs2236757	21:33252612:A:G	A	rs6517156	21:33261740:C:G	IL10RB	GTeXv8__Brain_Cerebellar_Hemisphere	4.35E-17	G	C	0.586	0.862	G
	rs2236757	21:33252612:A:G	A	rs6517156	21:33261740:C:G	IL10RB	GTeXv8__Brain_Cortex	2.64E-12	G	C	0.427	0.862	G
	rs2236757	21:33252612:A:G	A	rs1131668	21:33262573:G:A	IFNAR2	Walsh__Whole_Blood	1.60E-30	NA	NA	0	0.86	A
	rs2236757	21:33252612:A:G	A	rs1131668	21:33262573:G:A	IFNAR2	Alasoo__Macrophages__naive	1.14E-12	NA	NA	0	0.86	A
	rs2236757	21:33252612:A:G	A	rs1131668	21:33262573:G:A	IFNAR2	Alasoo__Macrophages__IFNg	3.25E-10	NA	NA	0	0.86	A
	rs2236757	21:33252612:A:G	A	rs3216172	21:33262748:G:GT	IL10RB	GTeXv8__Nerve_Tibial	1.62E-14	GT	G	0.344	0.857	GT
* LD estimated based on individuals of European ancestry from the 1000 Genomes project													

Supplementary Table 18. Association between a 6-SNP genetic risk score (GRS) and risk of hospitalization among COVID-19 cases.

Population, total N COVID-19 cases	GRS Group	Risk of hospitalization among COVID-19 cases (N positive and hospitalized / N positive) x 100		Unadjusted Risk Difference	Adjusted Odds Ratio*				
		Highest GRS group	Rest of Population		Estimate	SE	LCI	UCI	P-value
European (EUR) ancestry									
AncestryDNA EUR N=25,353	Top 40% vs Rest of population	7.33 (729/9947)	4.90 (755/15406)	2.43	1.54	0.06	1.38	1.72	6.7E-15
	Top 30% vs Rest of population	7.40 (555/7501)	5.20 (929/17852)	2.20	1.46	0.06	1.31	1.64	4.2E-11
	Top 20% vs Rest of population	8.09 (407/5029)	5.30 (1077/20324)	2.79	1.59	0.06	1.40	1.80	2.5E-13
	Top 10% vs Rest of population	9.04 (228/2522)	5.50 (1256/22831)	3.54	1.75	0.08	1.50	2.04	1.4E-12
	Top 5% vs Rest of population	9.73 (120/1233)	5.66 (1364/24120)	4.08	1.75	0.11	1.43	2.16	1.1E-07
GHS EUR N=5,285	Top 40% vs Rest of population	16.23 (343/2114)	13.65 (433/3171)	2.57	1.23	0.09	1.03	1.48	0.024
	Top 30% vs Rest of population	16.77 (266/1586)	13.79 (510/3699)	2.98	1.25	0.10	1.03	1.51	0.024
	Top 20% vs Rest of population	18.07 (191/1057)	13.84 (585/4228)	4.23	1.32	0.11	1.07	1.64	0.011
	Top 10% vs Rest of population	18.34 (97/529)	14.28 (679/4756)	4.06	1.28	0.15	0.96	1.70	0.091
	Top 5% vs Rest of population	18.87 (50/265)	14.46 (726/5020)	4.41	1.27	0.20	0.86	1.88	0.230
UKB EUR N=14,320	Top 40% vs Rest of population	28.25 (1479/5235)	26.41 (2399/9085)	1.85	1.13	0.04	1.03	1.23	6.1E-03
	Top 30% vs Rest of population	28.37 (1214/4279)	26.53 (2664/10041)	1.84	1.13	0.05	1.03	1.23	0.011
	Top 20% vs Rest of population	28.09 (795/2830)	26.83 (3083/11490)	1.26	1.06	0.05	0.96	1.18	0.255
	Top 10% vs Rest of population	29.09 (398/1368)	26.87 (3480/12952)	2.23	1.17	0.07	1.01	1.34	0.032
	Top 5% vs Rest of population	30.53 (218/714)	26.90 (3660/13606)	3.63	1.24	0.10	1.03	1.49	0.024
Meta-analysis EUR N=44,958	Top 40% vs Rest of population				1.27	0.03	1.19	1.35	2.3E-13
	Top 30% vs Rest of population				1.25	0.03	1.17	1.33	7.0E-11
	Top 20% vs Rest of population				1.26	0.04	1.17	1.36	9.2E-10
	Top 10% vs Rest of population				1.38	0.05	1.26	1.53	6.5E-11
	Top 5% vs Rest of population				1.43	0.07	1.25	1.63	1.0E-07
African (AFR) ancestry									
AncestryDNA AFR N=1,629	Top 40% vs Rest of population	8.9	7.88	1.01	1.33	0.23	0.84	2.09	0.220
	Top 30% vs Rest of population	8.59	8.16	0.43	1.18	0.25	0.73	1.91	0.499
	Top 20% vs Rest of population	9.35	8.03	1.32	1.49	0.28	0.86	2.55	0.152
	Top 10% vs Rest of population	6.88	8.44	-1.57	1.00	0.41	0.45	2.21	0.993
	Top 5% vs Rest of population	6.1	8.4	-2.31	0.73	0.58	0.23	2.28	0.587
UKB AFR N=620	Top 40% vs Rest of population	28.63 (71/248)	20.70 (77/372)	7.93	1.62	0.24	1.00	2.60	0.048
	Top 30% vs Rest of population	27.57 (51/185)	22.30 (97/435)	5.27	1.37	0.26	0.83	2.27	0.220
	Top 20% vs Rest of population	28.23 (35/124)	22.78 (113/496)	5.44	1.43	0.29	0.80	2.52	0.224
	Top 10% vs Rest of population	37.70 (23/61)	22.36 (125/559)	15.34	2.61	0.37	1.25	5.42	0.010
	Top 5% vs Rest of population	41.94 (13/31)	22.92 (135/589)	19.02	3.12	0.52	1.13	8.58	0.028
PMBB AFR N=349	Top 40% vs Rest of population	45.00 (63/140)	33.01 (69/209)	11.99	1.92	0.41	1.20	5.88	0.016
	Top 30% vs Rest of population	43.81 (46/105)	35.25 (86/244)	8.56	2.43	0.45	1.38	7.97	7.2E-03
	Top 20% vs Rest of population	44.29 (31/70)	36.20 (101/279)	8.08	1.91	0.50	0.97	6.80	0.058
	Top 10% vs Rest of population	42.86 (15/35)	37.26 (117/314)	5.60	1.73	0.72	0.57	9.58	0.241
	Top 5% vs Rest of population	50.00 (9/18)	37.16 (123/331)	12.84	2.30	1.01	0.48	25.16	0.219
Meta-analysis AFR N=2,598	Top 40% vs Rest of population				1.52	0.15	1.12	2.05	0.007
	Top 30% vs Rest of population				1.39	0.17	1.00	1.93	5.1E-02
	Top 20% vs Rest of population				1.52	0.19	1.05	2.19	0.026

N=2,598	Top 10% vs Rest of population				1.70	0.26	1.03	2.81	0.038
	Top 5% vs Rest of population				1.71	0.36	0.84	3.47	0.140
Admixed American (AMR) and South Asian (SAS) ancestries									
AncestryDNA AMR N=3,752	Top 40% vs Rest of population	5.89	4.91	0.98	1.16	0.16	0.85	1.60	0.343
	Top 30% vs Rest of population	6.9	4.63	2.27	1.41	0.16	1.02	1.95	0.037
	Top 20% vs Rest of population	7.81	4.69	3.12	1.62	0.18	1.14	2.30	6.8E-03
	Top 10% vs Rest of population	8.24	4.98	3.27	1.56	0.23	1.00	2.43	0.050
	Top 5% vs Rest of population	7.98	5.16	2.82	1.22	0.32	0.65	2.28	0.530
UKB SAS N=760	Top 40% vs Rest of population	22.37 (68/304)	18.86 (86/456)	3.51	1.37	0.23	0.86	2.16	0.183
	Top 30% vs Rest of population	21.93 (50/228)	19.55 (104/532)	2.38	1.34	0.25	0.82	2.18	0.243
	Top 20% vs Rest of population	23.84 (36/151)	19.38 (118/609)	4.47	1.37	0.28	0.80	2.35	0.256
	Top 10% vs Rest of population	25.00 (19/76)	19.74 (135/684)	5.26	1.42	0.35	0.72	2.82	0.316
	Top 5% vs Rest of population	36.84 (14/38)	19.39 (140/722)	17.45	3.18	0.46	1.29	7.79	0.012

* Logistic regression model adjusted for age, sex, age x sex, BMI, ancestry-informative PCs, REGENIE genome-wide predictor (to account for family structure) and clinical risk factors associated with risk of hospitalization and severe disease: chronic kidney disease, diabetes, immunosuppressive disease, chronic obstructive pulmonary disease or other chronic respiratory disease, cardiovascular disease, hyptertension (see methods for details)

Supplementary Table 19. Association between a 6-SNP genetic risk score (GRS) and risk of severe disease among COVID-19 cases.

Population, total N COVID-19 cases	GRS Group	Risk of severe disease among COVID-19 cases (N positive and severe / N positive) x 100		Unadjusted Risk Difference	Adjusted Odds Ratio*				
		Highest GRS group	Rest of Population		Estimate	SE	LCI	UCI	P-value
European (EUR) ancestry									
AncestryDNA EUR N=25,353	Top 40% vs Rest of population	3.33 (331/9947)	2.18 (336/15406)	1.15	1.58	0.08	1.35	1.87	3.0E-08
	Top 30% vs Rest of population	3.60 (270/7501)	2.22 (397/17852)	1.38	1.69	0.08	1.43	1.99	7.9E-10
	Top 20% vs Rest of population	3.96 (199/5029)	2.30 (468/20324)	1.65	1.74	0.09	1.45	2.08	1.5E-09
	Top 10% vs Rest of population	4.36 (110/2522)	2.44 (557/22831)	1.92	1.72	0.11	1.37	2.15	2.0E-06
	Top 5% vs Rest of population	4.79	2.52	2.26	1.64	0.15	1.22	2.20	9.9E-04
GHS EUR N=5,285	Top 40% vs Rest of population	6.91 (146/2114)	5.55 (176/3171)	1.36	1.26	0.13	0.98	1.61	0.071
	Top 30% vs Rest of population	7.19 (114/1586)	5.62 (208/3699)	1.56	1.27	0.13	0.98	1.64	0.074
	Top 20% vs Rest of population	7.85 (83/1057)	5.65 (239/4228)	2.20	1.33	0.15	1.00	1.78	0.050
	Top 10% vs Rest of population	6.81 (36/529)	6.01 (286/4756)	0.79	1.03	0.20	0.69	1.53	0.877
	Top 5% vs Rest of population	6.04 (16/265)	6.10 (306/5020)	-0.06	0.76	0.30	0.42	1.36	0.355
UKB EUR N=14,320	Top 40% vs Rest of population	7.98 (418/5235)	5.87 (533/9085)	2.12	1.43	0.07	1.24	1.66	1.4E-06
	Top 30% vs Rest of population	8.39 (359/4279)	5.90 (592/10041)	2.49	1.49	0.08	1.28	1.73	2.5E-07
	Top 20% vs Rest of population	8.73 (247/2830)	6.13 (704/11490)	2.60	1.44	0.09	1.21	1.70	2.5E-05
	Top 10% vs Rest of population	9.58 (131/1368)	6.33 (820/12952)	3.24	1.65	0.11	1.33	2.05	6.0E-06
	Top 5% vs Rest of population	11.06 (79/714)	6.41 (872/13606)	4.66	2.00	0.14	1.52	2.64	7.5E-07
Meta-analysis EUR N=44,958	Top 40% vs Rest of population	NA			1.46	0.05	1.32	1.61	1.3E-13
	Top 30% vs Rest of population				1.52	0.05	1.37	1.68	1.2E-15
	Top 20% vs Rest of population				1.53	0.06	1.37	1.71	1.4E-13
	Top 10% vs Rest of population				1.58	0.07	1.36	1.82	7.1E-10
	Top 5% vs Rest of population				1.66	0.10	1.37	2.01	1.6E-07
African (AFR) ancestry									
AncestryDNA AFR N=1,629	Top 40% vs Rest of population	4.45	3.38	1.07	Number of severe cases too small for analysis (N<100)				
	Top 30% vs Rest of population	4.09	3.68	0.41					
	Top 20% vs Rest of population	4.36	3.67	0.69					
	Top 10% vs Rest of population	2.50	3.95	-1.45					
	Top 5% vs Rest of population	2.44	3.88	-1.44					
UKB AFR N=620	Top 40% vs Rest of population	5.65 (14/248)	7.80 (29/372)	-2.15	Number of severe cases too small for analysis (N=43)				
	Top 30% vs Rest of population	5.95 (11/185)	7.36 (32/435)	-1.41					
	Top 20% vs Rest of population	7.26 (9/124)	6.85 (34/496)	0.40					
	Top 10% vs Rest of population	11.48 (7/61)	6.44 (36/559)	5.04					
	Top 5% vs Rest of population	9.68 (3/31)	6.79 (40/589)	2.89					
PMBB AFR N=349	Top 40% vs Rest of population	6.43 (9/140)	7.66 (16/209)	-1.23	Number of severe cases too small for analysis (N=25)				
	Top 30% vs Rest of population	6.67 (7/105)	7.38 (18/244)	-0.71					
	Top 20% vs Rest of population	10.00 (7/70)	6.45 (18/279)	3.55					
	Top 10% vs Rest of population	8.57 (3/35)	7.01 (22/314)	1.57					
	Top 5% vs Rest of population	5.56 (1/18)	7.25 (24/331)	-1.70					
Admixed American (AMR) and South Asian (SAS) ancestries									
	Top 40% vs Rest of population	2.61	1.86	0.75					

AncestryDNA AMR N=3,752	Top 30% vs Rest of population	3.23	1.71	1.52	Number of severe cases too small for analysis (N<100)
	Top 20% vs Rest of population	3.90	1.73	2.17	
	Top 10% vs Rest of population	4.26	1.93	2.33	
	Top 5% vs Rest of population	4.79	2.02	2.77	
UKB SAS N=760	Top 40% vs Rest of population	4.61 (14/304)	3.95 (18/456)	0.66	Number of severe cases too small for analysis (N=32)
	Top 30% vs Rest of population	4.82 (11/228)	3.95 (21/532)	0.88	
	Top 20% vs Rest of population	5.30 (8/151)	3.94 (24/609)	1.36	
	Top 10% vs Rest of population	6.58 (5/76)	3.95 (27/684)	2.63	
	Top 5% vs Rest of population	7.89 (3/38)	4.02 (29/722)	3.88	

* Logistic regression model adjusted for age, sex, age x sex, BMI, ancestry-informative PCs, REGENIE genome-wide predictor (to account for family structure) and clinical risk factors associated with risk of hospitalization and severe disease: chronic kidney disease, diabetes, immunosuppressive disease, chronic obstructive pulmonary disease or other chronic respiratory disease, cardiovascular disease, hyptertension (see methods for details)

Supplementary Table 20. Association between a 6-SNP genetic risk score (GRS) and risk of hospitalization and severe disease in individuals of European ancestry, after stratifying COVID-19 cases by pre-existing clinical risk factor status for severe COVID-19.

Population, total N COVID-19 cases	Clinical Risk Level	N COVID-19 cases	Risk among COVID-19 cases (N positive and [hospitalized or severe] / N positive) x 100		Unadjusted Risk Difference	Adjusted Odds Ratio**				
			Top 10% GRS	Rest of Population		Estimate	SE	LCI	UCI	P-value
Risk of hospitalization										
AncestryDNA EUR	Lower	12,464	4.07	2.39	1.69	1.70	0.16	1.25	2.31	6.8E-04
N=25,353	High	12,889	14.14	8.49	5.64	1.80	0.09	1.51	2.15	5.3E-11
UKB EUR	Lower	8,912	17.86 (152/851)	15.72 (1267/8061)	2.14	1.21	0.10	1.00	1.46	0.051
N=14,320	High	5,408	47.58 (246/517)	45.25 (2213/4891)	2.34	1.09	0.10	0.91	1.32	0.343
GHS EUR	Lower	1,537	5.48 (8/146)	2.95 (41/1391)	2.53	1.98	0.42	0.87	4.53	0.105
N=5,285	High	3,748	23.24 (89/383)	18.96 (638/3365)	4.28	1.21	0.14	0.92	1.59	0.171
Meta-analysis EUR	Lower	22,913	NA			1.35	0.08	1.15	1.58	2.1E-04
N=44,958	High	22,045				1.39	0.06	1.23	1.56	3.7E-08
Risk of severe disease										
AncestryDNA EUR	Lower	12,464	1.64	0.78	0.87	2.04	0.25	1.25	3.33	4.2E-03
N=25,353	High	12,889	7.15	4.04	3.11	1.70	0.13	1.33	2.17	2.5E-05
UKB EUR	Lower	8,912	3.41 (29/851)	2.25 (181/8061)	1.16	1.58	0.21	1.05	2.37	0.027
N=14,320	High	5,408	19.73 (102/517)	13.06 (639/4891)	6.66	1.60	0.12	1.25	2.03	1.5E-04
Meta-analysis EUR	Lower	21,376	NA			1.75	0.16	1.28	2.40	4.2E-04
N=39,673	High	18,297				1.65	0.09	1.39	1.96	1.5E-08

* High clinical risk included indiviiduals with any of the following criteria: age≥65, BMI≥35, chronic kidney disease, diabetes, immunosuppressive disease, or age ≥55 and presence of chronic obstructive pulmonary disease, cardiovascular disease, or hypertension (see methods for details).

** Logistic regression model adjusted for sex, ancestry-informative PCs and REGENIE genome-wide predictor to account for family structure.