Common genetic variants identify therapeutic targets for COVID-19 and individuals at high risk 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], A. Verma[4], G. Sirugo[4], Regeneron Genetics Center*, M. D. Ritchie[4], M. Jones[1], S. Balasubramanian[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. A. Kosmicki and J. E. Horowitz contributed equally to this manuscript.

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

Correspondence to: manuel.ferreira@regeneron.com and goncalo.abecasis@regeneron.com

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

ABSTRACT

The need to identify and effectively treat COVID-19 cases at highest risk for severe disease is critical. We identified seven common genetic variants (three novel) that modulate COVID-19 susceptibility and severity, implicating *IFNAR2, CCHCR1*, *TCF19*, *SLC6A20* and the hyaluronan pathway as potential therapeutic targets. A high genetic burden was strongly associated with increased risk of hospitalization and severe disease among COVID-19 cases, especially among individuals with few known risk factors.

34 **MAIN TEXT**

35 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[1] causes coronavirus disease
36 2019 (COVID-19)[2], which has lead to >1.5 million recorded deaths worldwide since December
37 2019 [3]. Reported risk factors for severe COVID-19, defined here as death or hospitalization
38 combined with respiratory failure [4], include male sex, older age, race, obesity, kidney,
39 cardiovascular and respiratory diseases [5-8]. Corticosteroids, repurposed anti-viral medicines,
40 and antibody therapies have been authorized to treat COVID-19 [9-12]. Presently, excess mortality
41 due to COVID-19 is high and the urgent need for therapies that treat or prevent severe disease
42 remains. Further, identifying individuals at highest risk of adverse outcomes may help prioritize
43 individuals for vaccines [13, 14] or monoclonal antibody treatments[15, 16], which are currently
44 in short supply. In this study, we used human genetics to identify potential therapeutic targets for
45 severe COVID-19 and developed genetic risk scores to identify individuals at highest risk of severe
46 disease with greater precision than can be attained with clinical and demographic variables alone.
47

48 We performed genetic association studies of COVID-19 outcomes across 11,356 individuals with
49 COVID-19 and 651,047 individuals with no record of SARS-CoV-2 infection aggregated from
50 four studies (**Supplementary Table 1**) and four ancestries (Admixed American, African,
51 European and South Asian). Of the COVID-19 cases, 2,175 (19%) were hospitalized and 847 (7%)
52 had severe disease; hospitalized patients were more likely to be older, of non-European ancestry
53 and to have pre-existing cardiovascular and lung disease (**Supplementary Table 2)**. Using these
54 data, we defined two groups of COVID-19 outcomes: five phenotypes related to disease risk and
55 two phenotypes related to disease severity among COVID-19 cases (**Supplementary Table 3).**
56 For each phenotype, when >30 cases were available, we performed ancestry-specific genetic

57 analyses in each study (**Supplementary Table 4)** and then combined results across studies and
58 ancestries using a fixed-effects meta-analysis.
59
60 To identify genetic variants that modulate disease severity following SARS-CoV-2 infection, we
61 first focused on replicating eight independent associations (r^2<0.05) with disease risk reported in
62 recent GWAS that included >1,000 cases [17-20] (**Supplementary Table 5**); we hypothesized
63 that some of these published risk variants could also modulate disease severity. After accounting
64 for multiple testing, six variants had a significant (*P*<0.0012) and directionally consistent
65 association with at least one of our five disease risk phenotypes (**Supplementary Table 6**):
66 rs73064425 in *LZTFL1* (strongest in contrast between COVID-19 hospitalized cases vs. COVID-
67 19 negative or unknown controls; MAF=7%, OR=1.42, P=7x10^{-11}); rs2531743 near *SLC6A20*
68 (COVID-19 positive vs. COVID-19 negative; MAF=42%, OR=1.06, P=9x10^{-4}); rs143334143 in
69 the MHC (COVID-19 severe vs. COVID-19 negative or unknown; MAF=7%, OR=1.36, P=6x10^{-}
70 ^{4}); rs9411378 in *ABO* (COVID-19 positive vs. COVID-19 negative or unknown; MAF=23%,
71 OR=1.12, P=6x10^{-10}); rs2109069 in *DPP9* (COVID-19 positive vs. COVID-19 negative or
72 unknown; MAF=31%, OR=1.06, P=10^{-4}); and rs2236757 in *IFNAR2* (COVID-19 hospitalized vs.
73 COVID-19 negative or unknown; MAF=29%, OR=1.13, P=2x10^{-4}). Effect sizes were comparable
74 across ancestries (**Supplementary Table 7**). Having established these variants as *bona fide* risk
75 factors for COVID-19, we then asked which (if any) were also associated with severity amongst
76 COVID-19 cases. We found that four of the six variants were significantly (*P*<0.05) associated
77 with worse outcomes among infected individuals **(Figure 1A)**. The exceptions were rs9411378 in
78 *ABO* and rs2531743 near *SLC6A20*, which did not associate with COVID-19 severity (**Figure 1B**).
79 Collectively, these results pinpoint four variants associated with worse disease outcomes,

80 including respiratory failure and death. These variants can be used to identify individuals at risk
81 of severe COVID-19 and suggest that nearby genes may represent targets for therapeutic
82 intervention.
83
84 We next looked for novel genetic associations with severe COVID-19. Across our two severity
85 phenotypes (hospitalized vs. non-hospitalized cases; severe vs. non-hospitalized cases), we found
86 no new associations at $P<5x10^{-8}$ (**Supplementary Figure 1**), indicating that these analyses were
87 underpowered for genome-wide discovery. To increase power, we combined results from our
88 COVID-19 hospitalization phenotype (2,175 cases vs. 651,047 controls) with those from two
89 published GWAS[19, 21], for a combined sample size of 5,461 hospitalized cases and 661,632
90 controls with no record of SARS-CoV-2 infection. In this larger analysis of disease risk, seven loci
91 reached genome-wide significance (**Figure 2**), including the four highlighted by the replication
92 analysis above (*LZTFL1*, MHC, *DPP9* and *IFNAR2*) and three novel associations (**Table 1)**:
93 rs79833209 near *CCNG1* (5q34; MAF=2%, OR=1.54, $P=2x10^{-8}$); rs4782327 in *ACSF3* (16q24.3;
94 MAF=22%, OR=1.17, $P=8x10^{-9}$); and rs12461764 in *FPR1* (19q13.41; MAF=35%, OR=1.18,
95 $P=10^{-8}$). Consistent associations were observed across studies and ancestries for the three novel
96 variants (**Supplementary Figure 2**). We then assessed the association between each novel variant
97 and disease severity and found that for all three, the risk allele was overrepresented ($P<0.05$)
98 amongst hospitalized and severe COVID-19 cases, relative to non-hospitalized cases (**Figure 1C**).
99 To further study the association between the three variants and worse disease outcomes, we
100 constructed a continuous score that combined nine different measures of severity self-reported by
101 4,448 COVID-19 cases from the AncestryDNA study. Lower severity scores were enriched for
102 asymptomatic cases and lower symptom severity, while higher scores were associated with

103 increased symptom severity and elevated hospitalization rates (**Supplementary Figure 3**). The
104 mean severity score was elevated for carriers of each of the three novel risk alleles
105 (**Supplementary Table 8**). Therefore, the three putative novel susceptibility loci we identified for
106 COVID-19 risk are also associated with increased risk of severe disease.
107
108 Collectively, our association analyses highlighted seven common variants (four known, three
109 novel) associated with COVID-19 susceptibility ($P<5x10^{-8}$), as well as disease severity among
110 cases ($P<0.05$). As all seven variants are either intronic or intergenic, it is not immediately clear
111 how these variants influence gene function. To help identify gene targets and thus potential
112 therapeutic targets for severe COVID-19, we searched for functional protein-coding variants
113 (missense or predicted loss-of-function) in high LD ($r^2>0.95$) with each variant. We found six
114 functional variants in three genes: two missense variants in *IFNAR2*, a component of the
115 heterodimeric type-1 interferon receptor; three missense and a stop-gain variant in *CCHCR1*, a P-
116 body protein associated with cytoskeletal remodeling and mRNA turnover [22, 23], which can
117 interact with human papillomavirus-16 [24, 25]; and a missense variant in *TCF19*, a transcription
118 factor associated with hepatitis B [26, 27] **(Supplementary Table 9**). These data indicate that the
119 risk variants identified in *IFNAR2*, *CCHCR1* and *TCF19* may have functional effects on these
120 genes. Next, we asked if the sentinel risk variants co-localized *(*$r^2>0.95$) with published sentinel
121 expression quantitative trait loci **(Supplementary Table 10**), specifically focusing on 168 genes
122 in *cis* (± 500 kb). We observed evidence for co-localization for *IFNAR2, TCF19* and *SLC6A20*
123 **(Supplementary Table 11)**. *SLC6A20* encodes a proline transporter that binds the host SARS-
124 CoV-2 receptor, angiotensin-converting enzyme 2 [28]. Our results suggest that higher *SLC6A20*
125 expression in the lung (and potentially increased viral uptake) is associated with higher risk of

126 severe disease (**Supplementary Table 11)**. In contrast, for *IFNAR2*, our analysis suggests that
127 severe COVID-19 is associated with lower expression in immune cells and lung tissue
128 (**Supplementary Table 11)**, consistent with the deficient interferon response observed in severe
129 COVID-19 patients [29]. Collectively, these analyses identified four specific putative effector
130 genes in COVID-19 risk loci *(IFNAR2, CCHCR1, TCF19* and *SLC6A20*), though functional
131 studies are required to confirm these predictions.
132
133 Co-localization analyses did not identify any likely target genes for the three new risk variants.
134 However, we note that two of these (rs12461764 and rs79833209) are near genes related to
135 hyaluronan (HA), a major component of the lung extracellular matrix (**Supplementary Figure 2**):
136 *HAS1* (16 Kb away) and *HMMR* (160 Kb away), which respectively encode HA synthase 1 and a
137 cell-surface receptor for HA. This observation is noteworthy because elevated HA (i) has been
138 reported in the lungs of COVID-19 patients [30, 31]; and (ii) is associated with inflammation and
139 other infections [31], including influenza [33]. Furthermore, *Hmmr*-deficient mice are protected
140 in acute fibrotic lung injury models [34]. If confirmed, these findings would support the use of
141 hyaluronidases in reducing accumulation of HA and preventing long-term lung damage in
142 COVID-19 patients [35].
143
144 The second aim of this study was to determine if genetics can help identify individuals at high risk
145 of severe disease, who may be prioritized for prophylactic or therapeutic interventions.
146 Specifically, we focused on the four variants (in/near *LZTFL1*, MHC, *DPP9* and *IFNAR2*) with
147 confirmed association with COVID-19 susceptibility and that also modulate COVID-19 severity.
148 Using these variants, we created a weighted genetic risk score (GRS) for individuals with COVID-

149 19 and then compared the risk of hospitalization and severe disease between those with a high
150 GRS and all other cases, after adjusting for established risk factors (*e.g.* age, sex, comorbidities).
151 When considering COVID-19 cases of European ancestry (N=9,277), we found that having a high
152 GRS (top 10%) was associated with a 2.0-fold increased risk of hospitalization (95% CI 1.66-2.55,
153 $P=3x10^{-11}$; **Supplementary Figure 4A**) and 1.8-fold increased risk of severe disease (95% CI
154 1.40-2.41, $P=10^{-5}$; **Figure 3A**). A consistent pattern was observed in other ancestries, though
155 sample sizes were considerably smaller (**Supplementary Tables 12** and **13**). Lastly, we compared
156 the effect of the GRS between individuals with and without established risk factors for severe
157 COVID-19. We found that a high GRS (top 10%) was strongly associated with risk of severe
158 disease among individuals with one (OR=3.26, 95% CI 2.03-5.23, $P=10^{-6}$) or two (OR=2.19, 95%
159 CI 1.29-3.72, $P=0.004$) established risk factors (**Figure 3B** and **3C; Supplementary Table 14**).
160 In contrast, there was a much weaker association between a high GRS and risk of severe disease
161 among individuals with many (three or more) risk factors (OR=1.26, 95% CI 0.75-2.11, $P=0.388$),
162 although evidence for heterogeneity of GRS effect with number of risk factors was not significant
163 ($P=0.16$). Similar results were observed for risk of hospitalization (**Supplementary Figure 4B**
164 and **4C; Supplementary Table 14**). Collectively, these results demonstrate that a GRS calculated
165 using variants associated with disease risk and severity can be used to identify COVID-19 cases
166 at high risk of developing poor disease outcomes. This is important as many of these individuals
167 might not be prioritized for prophylactic or therapeutic interventions according to current
168 guidelines[35].
169
170 In summary, we confirmed four common variant associations with COVID-19 susceptibility and
171 further show that they modulate disease severity and are likely to influence the function of

172 *IFNAR2*, *CCHCR1*, *TCF19* and *SLC6A20*. We also identified three novel associations with disease
173 severity, which potentially provide genetic evidence for a role of hyaluronan in severe COVID-
174 19. Lastly, we show that a genetic risk score based on the four common variants that we validated
175 can be used to identify individuals at high risk of poor disease outcomes. Collectively, our analyses
176 point to potential novel therapies and help identify patients at high risk of severe COVID-19.

177 **ONLINE METHODS**

178 **Participating Studies**

179 *AncestryDNA COVID-19 Research Study.* AncestryDNA customers over age 18, living in the
180 United States, and who had consented to research, were invited to complete a survey assessing
181 COVID-19 outcomes and other demographic information. These included SARS-CoV-2 swab and
182 antibody test results, COVID-19 symptoms and severity, brief medical history, household and
183 occupational exposure to SARS-CoV-2, and blood type. A total of 83,930 AncestryDNA survey
184 respondents were selected for inclusion in this study [36, 37]. Respondents selected for this study
185 included all individuals with a positive COVID-19 test together with age and sex matched controls.
186 DNA samples were genotyped as described previously[36]. Genotype data for variants not
187 included in the array were then inferred using imputation to the Haplotype Reference Consortium
188 (HRC) reference panel. Briefly, samples were imputed to HRC version 1.1, which consists of
189 27,165 total individuals and 36 million variants. The HRC reference panel does not include indels;
190 consequently, indels are not present in the imputed data. We determined best-guess haplotypes
191 with Eagle version 2.4.1 and performed imputation with Minimac4 version 1.0.1. We used
192 1,117,080 unique variants as input and 8,049,082 imputed variants were retained in the final data
193 set. Variants with a Minimac4 $R^2<0.30$ were filtered from the analysis. For this study, an additional
194 continuous severity phenotype was defined for a semi-overlapping set of in 4,448 survey-
195 respondents that reported a positive COVID-19 test. The continuous severity score was derived by
196 computing the first principal component across nine survey fields related to COVID-19 clinical
197 outcomes. Six of the nine questions were binary: hospitalization, ICU admittance with oxygen,
198 ICU admittance with ventilation, septic shock, respiratory failure, and organ failure due to
199 COVID-19. Binary responses were encoded as 0 for "No" and 1 for "Yes". Three symptom

200 questions related to shortness of breath, fever, and nausea/vomiting symptoms were encoded as a
201 unit-scaled variable based on the following mapping: 0="None", 0.2="Very mild", 0.4="Mild",
202 0.6="Moderate", 0.8="Severe", and 1.0="Very Severe". The three symptoms were chosen based
203 on prior literature indicating their positive association with COVID-19 hospitalization[37]. The
204 resulting score was standardized prior to association analysis.
205
206 *Geisinger Health System (GHS).* The GHS MyCode Community Health Initiative is a health
207 system-based cohort from central and eastern Pennsylvania (USA) with ongoing recruitment since
208 2006. A subset of 144,182 MyCode participants sequenced as part of the GHS-Regeneron Genetics
209 Center DiscovEHR partnership were included in this study. Information on COVID-19 outcomes
210 were obtained through GHS's COVID-19 registry. Patients were identified as eligible for the
211 registry based on relevant lab results and ICD-10 diagnosis codes; patient charts were then
212 reviewed to confirm COVID-19 diagnoses. The registry contains data on outcomes, comorbidities,
213 medications, supplemental oxygen use, and ICU admissions. DNA from participants was
214 genotyped on either the Illumina OmniExpress Exome (OMNI) or Global Screening Array (GSA)
215 and imputed to the TOPMed reference panel (stratified by array) using the TOPMed Imputation
216 Server. Prior to imputation, we retained variants that had a MAF >= 0.1%, missingness < 1% and
217 HWE p-value > 10^{-15}. Following imputation, data from the OMNI and GSA datasets were merged
218 for subsequent association analyses, which included an OMNI/GSA batch covariate, in addition
219 to other covariates described below.
220
221

222 *Penn Medicine BioBank (PMBB) study*. PMBB study participants are recruited through the
223 University of Pennsylvania Health System, which enrolls participants during hospital or clinic
224 visits. Participants donate blood or tissue and allow access to EHR information[38]. The PMBB
225 COVID-19 registry consists of patients who have positive qPCR testing for SARS-COV-2. We
226 then used electronic health records to classify COVID-19 patients into hospitalized and severe
227 (ventilation or death) categories. DNA genotyping was performed with the Illumina Global
228 Screening Array, and imputation performed using the TOPMed reference panel as described for
229 GHS above.
230
231 *UK Biobank (UKB) study*. We studied the host genetics of SARS-CoV-2 infection in participants
232 of the UK Biobank study, which took place between 2006 and 2010 and includes approximately
233 500,000 adults aged 40-69 at recruitment. In collaboration with UK health authorities, the UK
234 Biobank has made available regular updates on COVID-19 status for all participants, including
235 results from four main data types: qPCR test for SARS-CoV-2, anonymized electronic health
236 records, primary care and death registry data. We report results based on the 12 September 2020
237 data refresh and excluded from the analysis 28,547 individuals with a death registry event prior to
238 2020. DNA samples were genotyped as described previously [39] using the Applied Biosystems
239 UK BiLEVE Axiom Array (N=49,950) or the closely related Applied Biosystems UK Biobank
240 Axiom Array (N=438,427). Genotype data for variants not included in the arrays were inferred
241 using three reference panels (Haplotype Reference Consortium, UK10K and 1000 Genomes
242 Project phase 3) as described previously [39].
243
244 **COVID-19 phenotypes used for genetic association analyses**

245 We grouped participants from each study into three broad COVID-19 disease categories
246 (**Supplementary Table 1**): (i) positive – those with a positive qPCR or serology test for SARS-
247 CoV-2, or a COVID-19-related ICD10 code (U07), hospitalization or death; (ii) negative – those
248 with only negative qPCR or serology test results for SARS-CoV-2 and no COVID-19-related
249 ICD10 code (U07), hospitalization or death; and (iii) unknown – those with no qPCR or serology
250 test results and no COVID-19-related ICD10 code (U07), hospitalization or death. We then used
251 these broad COVID-19 disease categories, in addition to hospitalization and disease severity
252 information, to create seven COVID-19-related phenotypes for genetic association analyses, as
253 detailed in **Supplementary Table 3**. For association analysis in the AncestryDNA study, we
254 excluded from the COVID-19 unknown group individuals who had (i) a first-degree relative who
255 was COVID-19 positive; or (ii) flu-like symptoms.
256
257 **Genetic association analyses**
258 Association analyses in each study were performed using the genomewide Firth logistic regression
259 test implemented in REGENIE [40]. In this implementation, Firth's approach is applied when the
260 p-value from standard logistic regression score test is below 0.05. We included in step 1 of
261 REGENIE (*i.e.* prediction of individual trait values based on the genetic data) directly genotyped
262 variants with a minor allele frequency (MAF) >1%, <10% missingness, Hardy-Weinberg
263 equilibrium test P-value>10^{-15} and linkage-disequilibrium (LD) pruning (1000 variant windows,
264 100 variant sliding windows and r^2<0.9). The association model used in step 2 of REGENIE
265 included as covariates age, age^2, sex, age-by-sex, and the first 10 ancestry-informative principal
266 components (PCs) derived from the analysis of a stricter set of LD-pruned (50 variant windows, 5

267 variant sliding windows and r^2<0.5) common variants from the array (imputed for the GHS study)
268 data.
269
270 Within each study, association analyses were performed separately for four different ancestries
271 defined based on the array data (African [AFR], admixed American [AMR], European [EUR] and
272 South Asian [SAS]). We retained association results for variants with an imputation information
273 score ≥0.3 and MAC ≥5, and either (i) MAF>0.5% or (ii) a protein-altering consequence (*i.e.*
274 pLOF, missense or splice variants). Results were subsequently meta-analyzed across studies and
275 ancestries using an inverse variance-weighed fixed-effects meta-analysis.
276
277 **Identification of putative targets of GWAS variants based on colocalization with eQTL**
278 We identified as a likely target of a sentinel GWAS variant any gene for which a sentinel
279 expression quantitative trait locus (eQTL) co-localized (*i.e.* had LD r^2 > 0.95) with the sentinel
280 GWAS variant. That is, we only considered genes for which there was strong LD between a
281 sentinel GWAS variant and a sentinel eQTL, which reduces the chance of spurious colocalization.
282 Sentinel eQTL were defined across 174 published datasets (**Supplementary Table 10**), as
283 described previously[41]. We did not use statistical approaches developed to distinguish
284 colocalization from shared genetic effects because these have very limited resolution at high LD
285 levels (r^2 > 0.8)[42].
286
287 **Genetic risk score (GRS) analysis of COVID-19 hospitalization and severity**
288 First, in each study (AncestryDNA, GHS, UKB and PMBB), we created a GRS for each COVID-
289 19 positive individual based on variants that were reported to associate with COVID-19

290 susceptibility in previous GWAS and that we (i) independently replicated; and (ii) found to be
291 associated with COVID-19 severity. We used as weights the effect (beta) reported in previous
292 GWAS (**Supplementary Table 5**). Second, we ranked COVID-19 individuals based on the GRS
293 and created a new binary GRS predictor by assigning each individual to a high (top 5%) or low
294 (rest of the population) percentile group. Third, for studies with >100 hospitalized cases, the we
295 used logistic regression to test the association between the binary GRS predictor and risk of
296 hospitalization (hospitalized cases vs. all other cases), including as covariates age, sex and ten
297 ancestry-informative PCs. In addition to age and sex, we included as additional covariates
298 established clinical risk factors for COVID-19 that were significantly associated in each respective
299 study: BMI, smoking status, hypertension and chronic kidney disease for GHS; BMI, smoking
300 status, hypertension and chronic obstructive pulmonary disease in Europeans of the UKB;
301 hypertension, CKD, COPD, diabetes in AncestryDNA. We repeated the association analysis (i)
302 using different percentile cut-offs for the GRS (5%, 10%, 20%, 30% and 40%); and (ii) to test the
303 association with disease severity (severe cases vs. all other cases).
304
305 **Results availability**
306 All genotype-phenotype association results reported in this study are available for browsing using
307 the RGC's COVID-19 Results Browser (https://rgc-covid19.regeneron.com). Data access and use
308 is limited to research purposes in accordance with the Terms of Use (https://rgc-
309 covid19.regeneron.com/terms-of-use).
310
311

312 **Competing interests**

313 J.E.H., J.A.K., A.D., D.S., N.B, A.Y., A.M., R.L., E.M., X.B., H.G., D.S., F.S.P.K., J.D.B., C.O.,
314 A.J.M., D.A.T., A.H.L., J.M., K.W., L.G., S.E.M, M.J., S.B., W.J.S., A.R.S., A.E.L., J.M., J.O.,
315 L.H., M.N.C., J.G.R., A.B., G.R.A., and M.A.F. are current employees and/or stockholders of
316 Regeneron Genetics Center or Regeneron Pharmaceuticals. G.H.L.R., M.V.C., D.S.P., S.C.K.
317 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
318 AncestryDNA and may hold equity in AncestryDNA. The other authors declare no competing
319 interests.
320

321 **FIGURES**

322

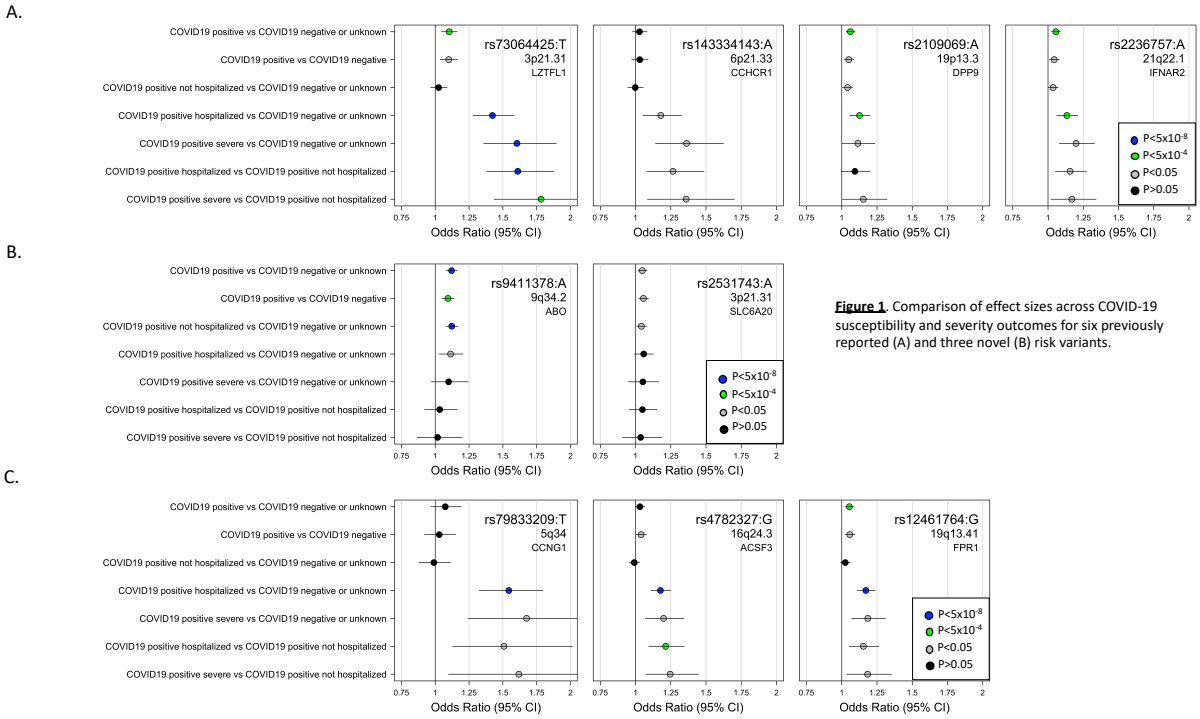


Figure 1. Comparison of effect sizes across COVID-19 susceptibility and severity outcomes for six previously reported (A) and three novel (B) risk variants.

323

324 **Figure 1. Comparison of effect sizes across COVID-19 susceptibility and severity outcomes**

325 **for six previously reported (A, B) and three novel (C) risk variants.**

326 **(A, B)** Six variants that were reported to associated with COVID-19 susceptibility in previous

327 studies and replicated in our analysis. Of these, four variants also associated with disease severity

328 among COVID-19 cases (in/near *LZTFL1*, *CCHCR1*, *DPP9* and *IFNAR2*; panel A), whereas two

329 variants did not (in *ABO* and *SLC6A20*, panel B).

330 **(C)** Three novel risk variants for COVID-19 discovered in the trans-ancestry meta-analysis of risk

331 of COVID-19 hospitalization across six studies (AncestryDNA, GHS, UKB, PMBB,

332 Ellinghaus[19] et al. and Pairo-Castineira et al.[18]).

333

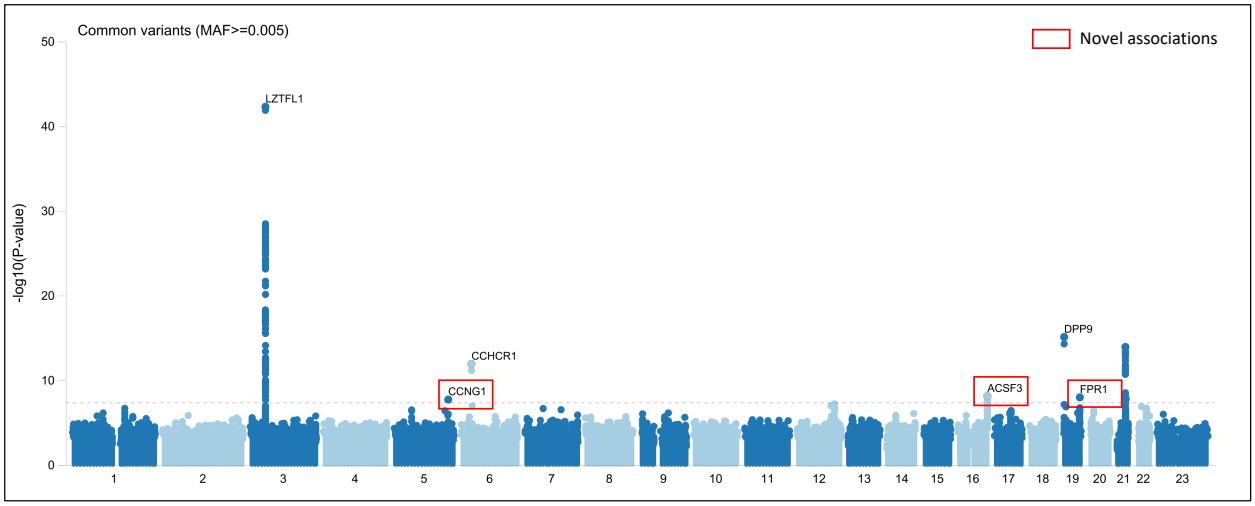


Figure 2. Summary of association results from the trans-ancestry meta-analysis of risk of COVID-19 hospitalization. The meta-analysis included results from six studies (AncestryDNA, GHS, UKB, PMBB, Ellinghaus et al. [19] and Pairo-Castineira et al.[18] and four ancestries (AFR, AMR, EUR and SAS), totaling 5,461 hospitalized cases and 661,632 controls with no record of SARS-CoV-2 infection.

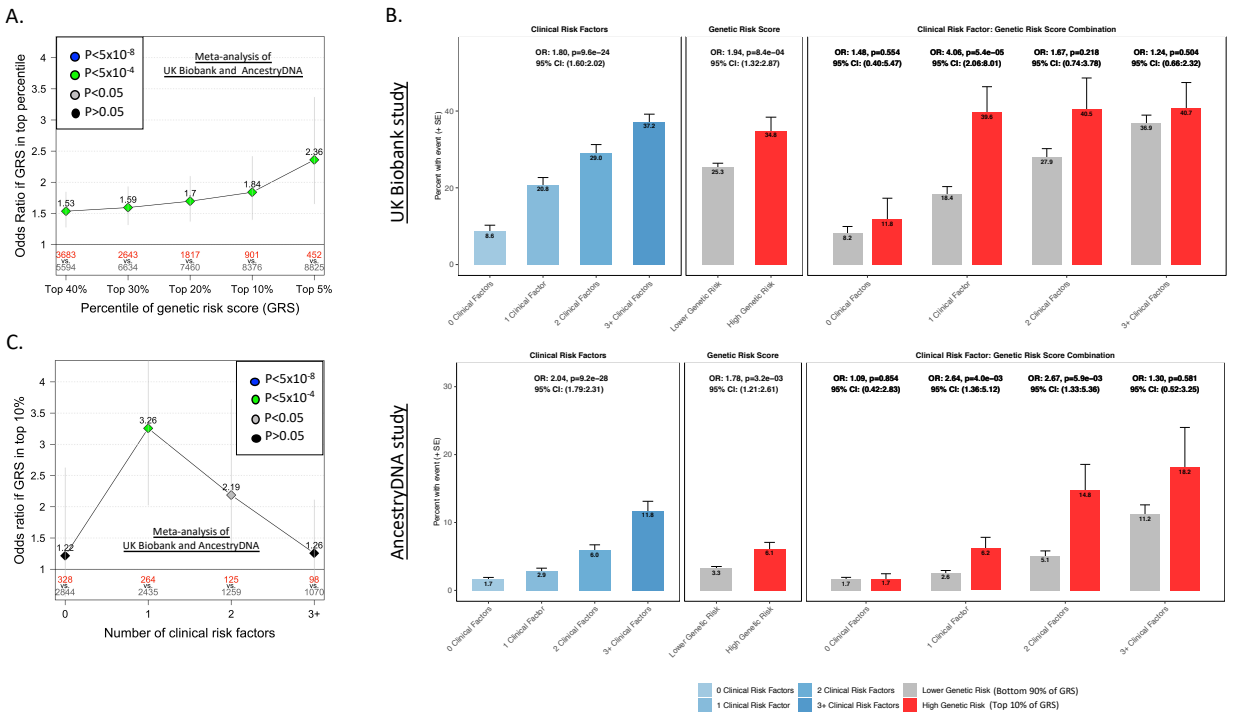


Figure 3. Association between a four-SNP genetic risk score (GRS) and risk of severe disease among COVID-19 cases of European ancestry.

(A) Association between high genetic risk and severe disease. Overall risk of severe disease is shown for individuals in the top GRS percentile, agnostic to the number of clinical risk factors present. The association was tested in two studies separately (AncestryDNA and UKB 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 8,423 COVID-19 cases, including 744 with severe disease. N in red: number of COVID-19 cases in the top GRS percentile. N in grey: number of COVID-19 cases in the rest of population. **(B)** Rate of severe disease in UK Biobank and AncestryDNA study, after stratifying by the number of clinical risk factors and genetic risk score. High genetic risk: top 10% of the GRS. Low genetic risk: bottom 90% of the GRS (*i.e.* all other COVID-19 cases).

354 **(C)** Association between high genetic risk (top 10% of GRS vs. all other COVID-19 cases) and
355 severe disease, after stratifying by the number of pre-existing clinical risk factors. The association
356 was tested in the UK Biobank and AncestryDNA studies, with results combined using inverse-
357 variance meta-analysis.
358

359 **TABLES**

360

361 **Table 1. Top associations identified in the trans-ancestry meta-analysis of risk of COVID-19**

362 **hospitalization.** The meta-analysis included results from six studies (AncestryDNA, GHS, UKB,

363 PMBB, Pairo-Castineira et al.[18] and Ellinghaus et al. [19]) and four ancestries (AFR, AMR,

364 EUR and SAS), totaling 5,461 hospitalized cases and 661,632 controls with no record of SARS-

365 CoV-2 infection.

rsID	Position hg38	Effect allele	Odds ratio (LCI,UCI)	P-value	Effect allele frequency	N cases (N with 0|1|2 copies of the effect allele)	N controls (N with 0|1|2 copies of the effect allele)	Nearest gene
Previously reported associations								
rs73064425	3:45859597	T	1.716 (1.589, 1.854)	5.42E-43	0.074	5461 (4205|1165|87)	661632 (568579|88970|4081)	*LZTFL1*
rs143334143	6:31153649	A	1.335 (1.233, 1.447)	1.29E-12	0.072	5461 (4525|881|52)	661632 (569685|88265|3678)	*CCHCR1*
rs2277732	19:4723658	A	1.225 (1.166, 1.287)	8.40E-16	0.300	5461 (2371|2411|677)	661632 (325425|276058|60146)	*DPP9*
rs13050728	21:33242905	T	1.210 (1.153, 1.270)	1.24E-14	0.315	5461 (2237|2491|729)	661632 (311728|283741|66160)	*IFNAR2*
Novel associations								
rs79833209	5:163300447	T	1.544 (1.326, 1.796)	2.02E-08	0.022	5387 (5072|308|4)	652327 (624117|27910|296)	*CCNG1*
rs4782327	16:89117727	G	1.171 (1.110, 1.236)	8.15E-09	0.216	5461 (3109|2002|346)	661632 (407670|222489|31470)	*ACSF3*
rs12461764	19:51739497	G	1.181 (1.116, 1.250)	1.11E-08	0.346	3785 (1431|1770|582)	653252 (279857|295218|78176)	*FPR1*

366
367

368 **SUPPLEMENTARY FIGURES**

369 Provided in a separate file.

370

371 **Supplementary Figure 1. Summary of association results from the meta-analysis of two**
372 **phenotypes related to disease severity among COVID-19 cases**. **(A)** Phenotype COVID-19
373 positive hospitalized (N=2,052) vs COVID-19 positive and not hospitalized (N=9,098). **(B)**
374 Phenotype COVID-19 positive severe (N=815) vs COVID-19 positive and not hospitalized
375 (N=8,999).

376

377 **Supplementary Figure 2. Three novel loci identified in the trans-ancestry meta-analysis of**
378 **risk of COVID-19 hospitalization. (A)** rs79833209 (5:163300447:C:T, effect allele C) in 5q34,
379 near *CCNG1*. **(B)** rs4782327 (16:89117727:G:C, effect allele G) in 16q24.3 in *ACSF3*. **(C)**
380 rs12461764 (19:51739497:G:T, effect allele G) in 19q13.41 near *FPR1*. RR: number of individuals
381 homozygous for reference allele. RA: number of heterozygous individuals. RA: number of
382 individuals homozygous for alternative (effect) allele. AAF: alternative allele frequency. OR: odds
383 ratio.

384

385 **Supplementary Figure 3. Association between the continuous severity score and key**
386 **individual measures of severity in the AncestryDNA study.** The continuous severity score was
387 derived from the first principal component across nine survey fields related to COVID-19 clinical
388 outcomes, including three symptoms, hospitalization, ICU admittance, and other severe
389 complications due to COVID-19 illness (Methods). Plots reflect mean symptom severity (top three
390 panels) or incident prevalence (bottom three panels) for several fields as a function of ascending

391 severity decile. Symptom information was encoded as follows: 0=None, 0.2=Very Mild, 0.4=Mild,
392 0.6=Moderate, 0.8=Severe, and 1.0=Very Severe. A paucisymptomatic case corresponds to
393 reporting symptoms of mild intensity or less for any of the symptom severity questions.
394
395 **Supplementary Figure 4. Association between a four-SNP genetic risk score (GRS) and risk**
396 **of hospitalization among COVID-19 cases of European ancestry.**
397 **(A)** Association between high genetic risk and hospitalization. Overall risk of hospitalization is
398 shown for individuals in the top GRS percentile, agnostic to the number of clinical risk factors
399 present. The association was tested in three studies separately (AncestryDNA, GHS and UKB
400 studies) using logistic regression, with established risk factors for COVID-19 included as
401 covariates (see Methods for details). Results were then meta-analyzed across studies, for a
402 combined sample size of 9,277 COVID-19 cases, including 1,832 who were hospitalized. N in red:
403 number of COVID-19 cases in the top GRS percentile. N in grey: number of COVID-19 cases in
404 the rest of population.
405 **(B)** Rate of hospitalization in UK Biobank and AncestryDNA study, after stratifying by the number
406 of clinical risk factors and genetic risk score. High genetic risk: top 10% of the GRS. Low genetic
407 risk: bottom 90% of the GRS (*i.e.* all other COVID-19 cases).
408 **(C)** Association between high genetic risk (top 10% of GRS vs. all other COVID-19 cases) and
409 hospitalization, after stratifying by the number of pre-existing clinical risk factors. The association
410 was tested in the UK Biobank and AncestryDNA studies, with results combined using inverse-
411 variance meta-analysis.
412

413 **SUPPLEMENTARY TABLES**

414 Provided in a separate file.

415

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

417 **the analysis.**

418

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

420

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

422

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

424 **phenotypes analyzed.**

425

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

427 **GWAS.**

428

429 **Supplementary Table 6. Association between eight published risk variants for COVID-19**

430 **and five disease susceptibility phenotypes in this study.** P-values shown in red were significant

431 after correcting for the 40 tests performed (P<0.00125).

432

433 **Supplementary Table 7. Comparison of effect sizes between ancestries for the six published**

434 **risk variants that were validated in this study.**

435

436 **Supplementary Table 8. Association between a COVID-19 severity score and seven variants**
437 **(four known, three novel) associated with both disease risk and severity.**
438
439 **Supplementary Table 9. Missense or predicted loss-of-function variants in high linkage**
440 **disequilibrium (LD, r^2>0.95) with sentinel GWAS variants.**
441
442 **Supplementary Table 10. Published gene expression datasets used to identify sentinel**
443 **expression quantitative trait loci (eQTL) that co-localized (LD r^2>0.95) with sentinel GWAS**
444 **variants.**
445
446 **Supplementary Table 11. Expression quantitative trait loci (eQTL) that co-localized (LD**
447 **r^2>0.95) with sentinel GWAS variants.**
448
449 **Supplementary Table 12. Association between a four-SNP genetic risk score (GRS) and risk**
450 **of hospitalization among COVID-19 cases.**
451
452 **Supplementary Table 13. Association between a four-SNP genetic risk score (GRS) and risk**
453 **of severe disease among COVID-19 cases.**
454
455 **Supplementary Table 14. Association between a four-SNP genetic risk score (GRS) and risk**
456 **of hospitalization and severe disease, after stratifying COVID-19 cases by the number of pre-**
457 **existing clinical risk factors for severe COVID-19.**

458 **References**

459

460 1. Zhu, N., et al., *A Novel Coronavirus from Patients with Pneumonia in China, 2019.* New
461 England Journal of Medicine, 2020. **382**(8): p. 727-733.

462 2. Coronaviridae Study Group of the International Committee on Taxonomy of, V., *The*
463 *species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV*
464 *and naming it SARS-CoV-2.* Nat Microbiol, 2020. **5**(4): p. 536-544.

465 3. *World Health Organization Coronavirus Disease Weekly Epidemiological Update*. 2020;
466 Available from: https://covid19.who.int/.

467 4. Guan, W.J., et al., *Clinical Characteristics of Coronavirus Disease 2019 in China.* N
468 Engl J Med, 2020. **382**(18): p. 1708-1720.

469 5. Richardson, S., et al., *Presenting Characteristics, Comorbidities, and Outcomes Among*
470 *5700 Patients Hospitalized With COVID-19 in the New York City Area.* Jama, 2020.
471 **323**(20): p. 2052-2059.

472 6. Atkins, J.L., et al., *PREEXISTING COMORBIDITIES PREDICTING SEVERE COVID-*
473 *19 IN OLDER ADULTS IN THE UK BIOBANK COMMUNITY COHORT.* medRxiv,
474 2020: p. 2020.05.06.20092700.

475 7. Zhou, F., et al., *Clinical course and risk factors for mortality of adult inpatients with*
476 *COVID-19 in Wuhan, China: a retrospective cohort study.* Lancet, 2020. **395**(10229): p.
477 1054-1062.

478 8. Cummings, M.J., et al., *Epidemiology, clinical course, and outcomes of critically ill*
479 *adults with COVID-19 in New York City: a prospective cohort study.* Lancet, 2020.
480 **395**(10239): p. 1763-1770.

481 9. Rochwerg, B., *A living WHO guideline on drugs for covid-19.* BMJ 2020;370:m3379
482 2020.
483 10. Group, R.C., et al., *Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary*
484 *Report.* N Engl J Med, 2020.
485 11. FDA, *Investigational COVID-19 Convalescent Plasma.* 2020.
486 12. *WHO Solidarity Trial Consortium. Repurposed antiviral drugs for COVID-19 –interim*
487 *WHO SOLIDARITY trial results*. medRxiv 2020; Available from:
488 https://www.medrxiv.org/content/10.1101/2020.10.15.20209817v1?ijkey=6ff7378035b5
489 c9c24346a6b90cadb5e866a63f71&keytype2=tf_ipsecsha.
490 13. Jackson, L.A., et al., *An mRNA Vaccine against SARS-CoV-2 - Preliminary Report.* N
491 Engl J Med, 2020. **383**(20): p. 1920-1931.
492 14. Walsh, E.E., et al., *Safety and Immunogenicity of Two RNA-Based Covid-19 Vaccine*
493 *Candidates.* N Engl J Med, 2020.
494 15. Hansen, J., et al., *Studies in humanized mice and convalescent humans yield a SARS-*
495 *CoV-2 antibody cocktail.* Science, 2020. **369**(6506): p. 1010-1014.
496 16. Baum, A., et al., *Antibody cocktail to SARS-CoV-2 spike protein prevents rapid*
497 *mutational escape seen with individual antibodies.* Science, 2020. **369**(6506): p. 1014-
498 1018.
499 17. Shelton, J., F., Anjali J. Shastri, Chelsea Ye, Catherine H. Weldon, Teresa Filshtein-
500 Somnez, Daniella Coker, Antony Symons, Jorge Esparza-Gordillo, The 23andMe
501 COVID-19 Team, Stella Aslibekyan, Adam Auton, *Trans-ethnic analysis reveals genetic*
502 *and non-genetic associations with COVID-19 susceptibility and severity.* 2020.

503 18. Pairo-Castineira, E., et al, … The GenOMICC Investigators, The ISARIC4C
504 Investigators, The Covid-19 Human Genetics Initiative, Xia Shen, Kathy Rowan, Angie
505 Fawkes, Lee Murphy, Chris P Ponting, Albert Tenesa, Mark Caulfield, Richard
506 Scott, Peter JM Openshaw, Malcolm G Semple, Veronique Vitart, James F Wilson,
507 J Kenneth Baillie, *Genetic mechanisms of critical illness in Covid-19.* medRxiv, 2020: p.
508 2020.09.24.20200048.
509 19. Ellinghaus, D., et al., *Genomewide Association Study of Severe Covid-19 with*
510 *Respiratory Failure.* New England Journal of Medicine, 2020.
511 20. *The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host*
512 *genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic.* Eur J
513 Hum Genet, 2020. **28**(6): p. 715-718.
514 21. Tervaniemi, M.H., et al., *Intracellular signalling pathways and cytoskeletal functions*
515 *converge on the psoriasis candidate gene CCHCR1 expressed at P-bodies and*
516 *centrosomes.* BMC Genomics, 2018. **19**(1): p. 432.
517 22. Ling, Y.H., et al., *CCHCR1 interacts with EDC4, suggesting its localization in P-bodies.*
518 Exp Cell Res, 2014. **327**(1): p. 12-23.
519 23. Olejnik-Schmidt, A.K., et al., *Search for cellular partners of human papillomavirus type*
520 *16 E2 protein.* Arch Virol, 2008. **153**(5): p. 983-90.
521 24. Muller, M. and C. Demeret, *CCHCR1 interacts specifically with the E2 protein of human*
522 *papillomavirus type 16 on a surface overlapping BRD4 binding.* PLoS One, 2014. **9**(3):
523 p. e92581.
524 25. Matsuura, K., M. Isogawa, and Y. Tanaka, *Host genetic variants influencing the clinical*
525 *course of hepatitis B virus infection.* J Med Virol, 2016. **88**(3): p. 371-9.

526 26. Kim, Y.J., et al., *A genome-wide association study identified new variants associated*
527 *with the risk of chronic hepatitis B.* Hum Mol Genet, 2013. **22**(20): p. 4233-8.
528 27. Vuille-dit-Bille, R.N., et al., *Human intestine luminal ACE2 and amino acid transporter*
529 *expression increased by ACE-inhibitors.* Amino Acids, 2015. **47**(4): p. 693-705.
530 28. Hadjadj, J., et al., *Impaired type I interferon activity and inflammatory responses in*
531 *severe COVID-19 patients.* Science, 2020. **369**(6504): p. 718-724.
532 29. Hellman, U., et al., *Presence of hyaluronan in lung alveoli in severe Covid-19: An*
533 *opening for new treatment options?* J Biol Chem, 2020. **295**(45): p. 15418-15422.
534 30. Kaber, G., et al., *Hyaluronan is abundant in COVID-19 respiratory secretions.* medRxiv,
535 2020.
536 31. Siiskonen, H., et al., *Hyaluronan synthase 1: a mysterious enzyme with unexpected*
537 *functions.* Front Immunol, 2015. **6**: p. 43.
538 32. Bell, T.J., et al., *Defective lung function following influenza virus is due to prolonged,*
539 *reversible hyaluronan synthesis.* Matrix Biol, 2019. **80**: p. 14-28.
540 33. Cui, Z., et al., *The Receptor for Hyaluronan-Mediated Motility (CD168) promotes*
541 *inflammation and fibrosis after acute lung injury.* Matrix Biol, 2019. **78-79**: p. 255-271.
542 34. Shi, Y., et al., *COVID-19 infection: the perspectives on immune responses.* Cell Death
543 Differ, 2020. **27**(5): p. 1451-1454.
544 35. *How CDC Is Making COVID-19 Vaccine Recommendations*. 2020 [cited 2020
545 December 2, 2020]; Available from: https://www.cdc.gov/coronavirus/2019-
546 ncov/vaccines/recommendations-process.html.
547 36. Roberts, G.H.L., *AncestryDNA COVID-19 Host Genetic Study Identifies Three Novel*
548 *Loci.* medRxiv, 2020.

549 37. Knight, S.C., *COVID-19 Susceptibilityand Severity Risks in a Survey of Over 500,000*
550 *People.* medRxiv, 2020.
551 38. Park, J., et al., *A genome-first approach to aggregating rare genetic variants in LMNA*
552 *for association with electronic health record phenotypes.* Genet Med, 2020. **22**(1): p.
553 102-111.
554 39. Bycroft, C., et al., *The UK Biobank resource with deep phenotyping and genomic data.*
555 Nature, 2018. **562**(7726): p. 203-209.
556 40. Mbatchou, J., et al., *Computationally efficient whole genome regression for quantitative*
557 *and binary traits.* bioRxiv, 2020: p. 2020.06.19.162354.
558 41. Ferreira, M.A., et al., *Shared genetic origin of asthma, hay fever and eczema elucidates*
559 *allergic disease biology.* Nat Genet, 2017. **49**(12): p. 1752-1757.
560 42. Chun, S., et al., *Limited statistical evidence for shared genetic effects of eQTLs and*
561 *autoimmune-disease-associated loci in three major immune-cell types.* Nat Genet, 2017.
562 **49**(4): p. 600-605.
563

MAIN FIGURES

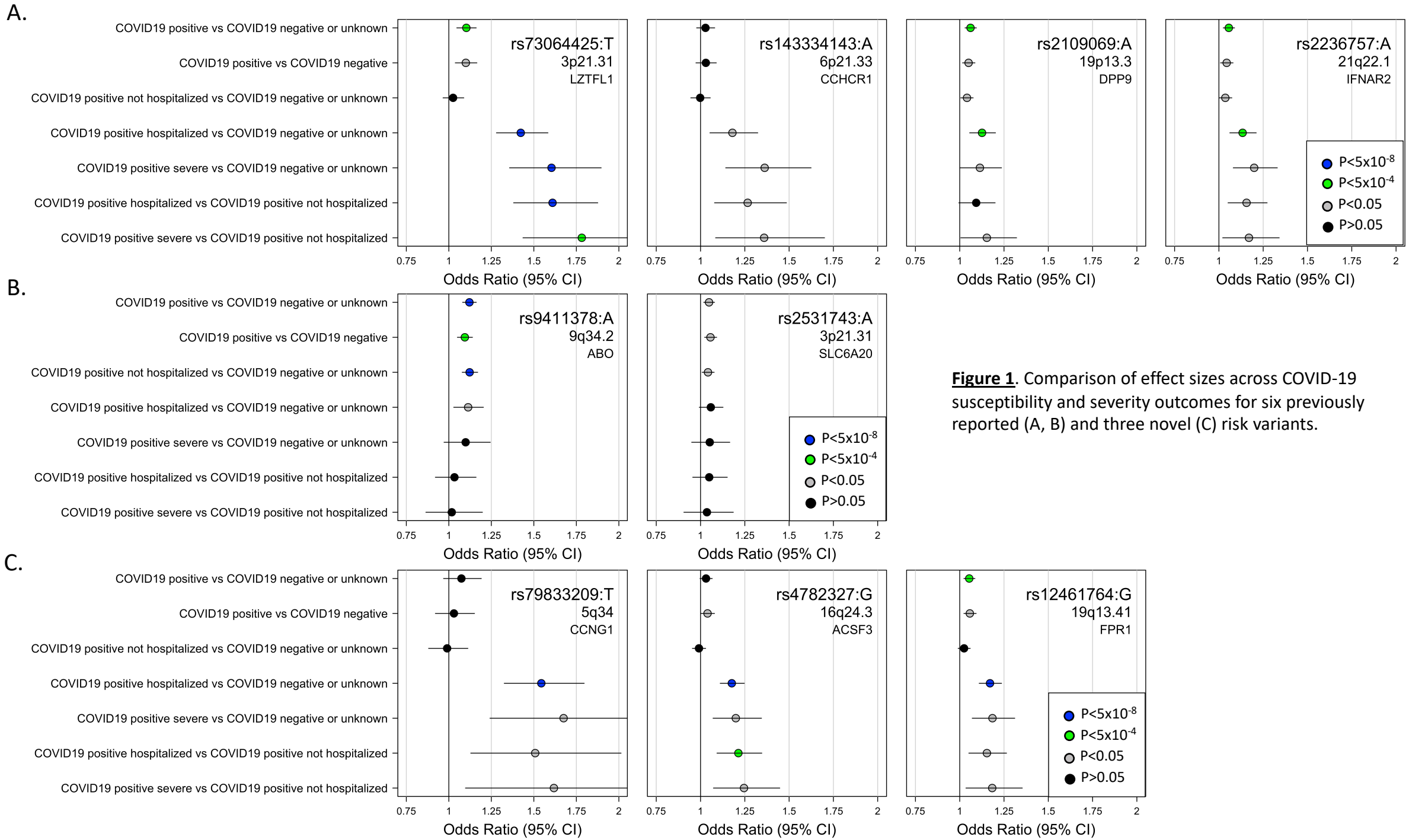
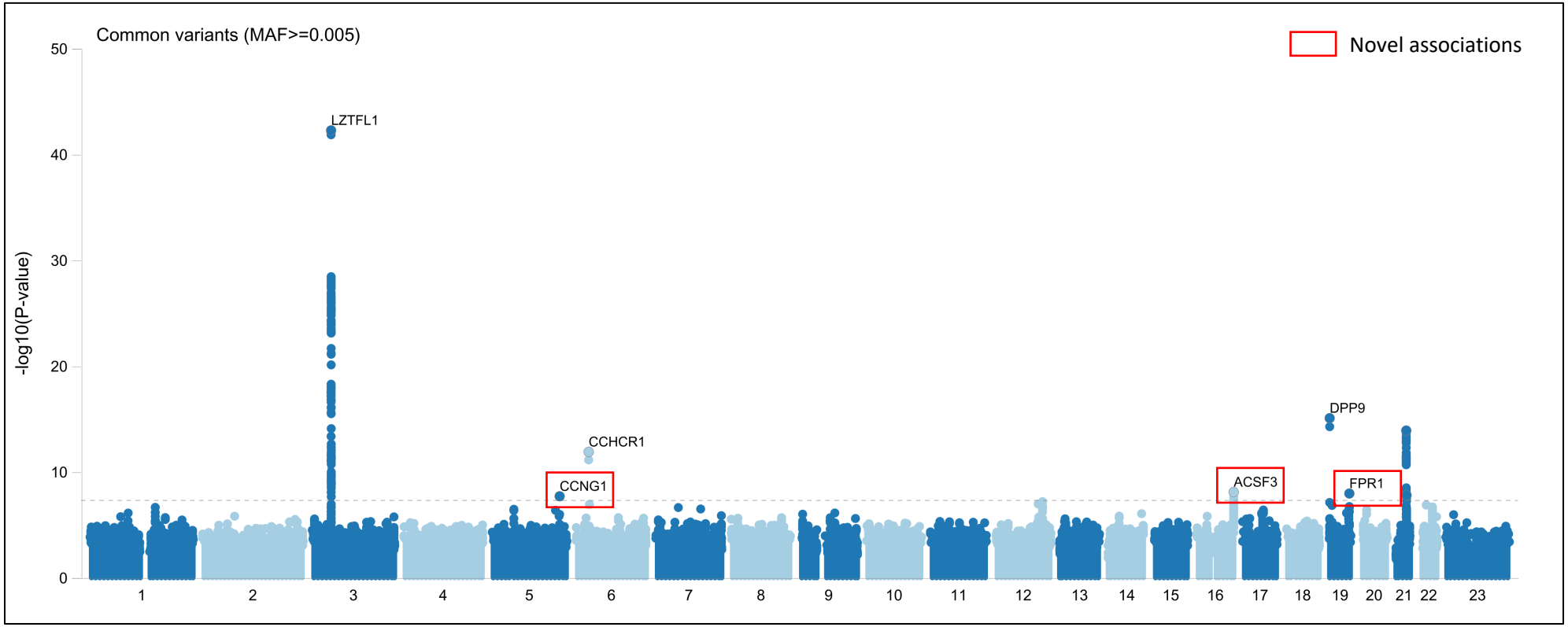


Figure 1. Comparison of effect sizes across COVID-19 susceptibility and severity outcomes for six previously reported (A, B) and three novel (C) risk variants.



<u>Figure 2.</u> Summary of association results from the trans-ancestry meta-analysis of risk of COVID-19 hospitalization. The meta-analysis included results from six studies (AncestryDNA, GHS, UKB, PMBB, Ellinghaus et al. [19] and Pairo-Castineira et al. [18]) and four ancestries (AFR, AMR, EUR and SAS), totaling 5,461 hospitalized cases and 661,632 controls with no record of SARS-CoV-2 infection.

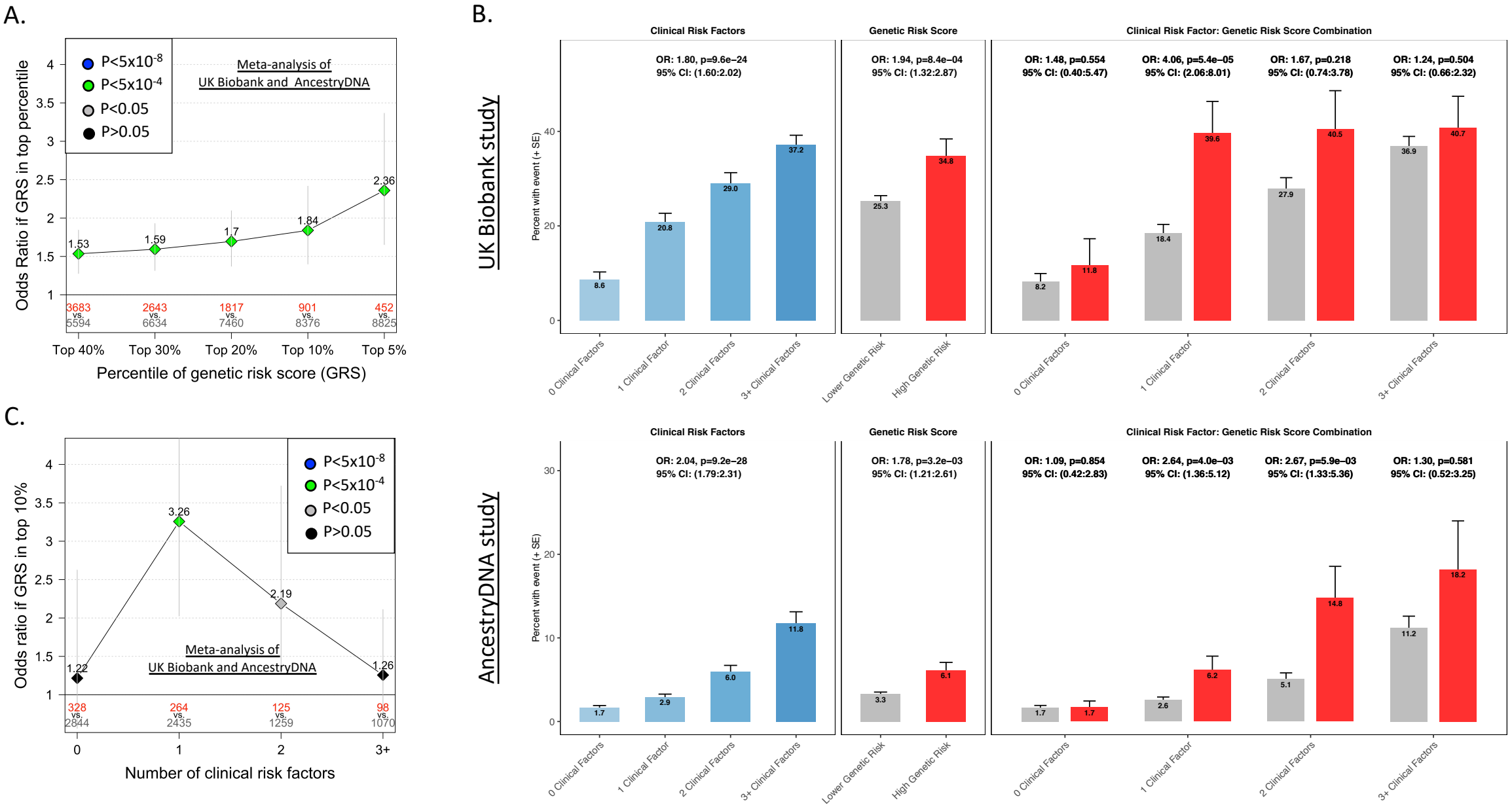


Figure 3. Association between a four-SNP genetic risk score (GRS) and risk of severe disease among COVID-19 cases of European ancestry.