Common genetic variants identify 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], 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)

ABSTRACT

The need to identify and effectively treat COVID-19 cases at highest risk for severe disease remains critical. We identified five common genetic loci (two novel) that modulate both COVID-19 susceptibility and severity, implicating *TMPRSS2*, *IFNAR2, CCHCR1*, *TCF19* and *SLC6A20* as potential 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 clinical 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 >2 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 [5-9]. Corticosteroids, repurposed biologics and anti-viral
40 medicines, as well as antibody therapies have been authorized to treat COVID-19 [10-14].
41 Presently, excess mortality due to COVID-19 is high and the urgent need for therapies that treat
42 or prevent severe disease remains. Further, identifying individuals at highest risk of adverse
43 outcomes may help prioritize individuals for vaccines [15, 16] or monoclonal antibody treatments
44 [17, 18], which are currently in short supply. In this study, we used human genetics to identify
45 potential targets for severe COVID-19 and developed genetic risk scores to identify individuals at
46 highest risk of severe disease with greater precision than can be attained with clinical and
47 demographic variables alone.

48

49 We performed genetic association studies of COVID-19 outcomes across 20,177 individuals with
50 COVID-19 and 653,646 individuals with no record of SARS-CoV-2 infection aggregated from
51 four studies (**Supplementary Table 1**) and five ancestries (Admixed American, African,
52 European, South Asian and East Asian). Of the COVID-19 cases, 3,056 (15%) were hospitalized
53 and 1,034 (5%) had severe disease; hospitalized patients were more likely to be older, of non-
54 European ancestry and to have pre-existing cardiovascular and lung disease (**Supplementary**
55 **Table 2)**. Using these data, we defined two groups (risk and severity) of COVID-19 outcomes:
56 five phenotypes related to disease risk and two phenotypes related to disease severity among

57 COVID-19 cases (**Supplementary Table 3).** For each phenotype, when >25 cases were available,
58 we performed ancestry-specific genetic analyses in each study and then combined results across
59 studies and ancestries using a fixed-effects meta-analysis. Genomic inflation factors (λ_{GC}) for the
60 meta-analyses were all less than 1.05, suggesting no substantial impact of population structure or
61 unmodeled relatedness (**Supplementary Table 4).**
62
63 To identify genetic variants that modulate disease severity following SARS-CoV-2 infection, we
64 focused on first replicating eight independent associations (r^2<0.05) with disease risk reported in
65 recent GWAS that included >1,000 cases [19-22] (**Supplementary Table 5**) and then explored
66 the impact of these variants on disease severity. After accounting for multiple testing, five variants
67 had a significant (P<0.0012) and directionally consistent association with at least one of our five
68 disease risk phenotypes (**Supplementary Table 6**): rs73064425:T in *LZTFL1* (strongest in
69 comparison between COVID-19 hospitalized cases vs. COVID-19 negative or unknown controls;
70 MAF=7%, OR=1.37, P=1x10^{-11}); rs2531743:G near *SLC6A20* (COVID-19 positive vs. COVID-
71 19 negative; MAF=42%, OR=0.94, P=2x10^{-6}); rs143334143:A in the MHC (COVID-19 severe
72 vs. COVID-19 negative or unknown; MAF=7%, OR=1.32, P=3x10^{-4}); rs9411378:A in *ABO*
73 (COVID-19 positive vs. COVID-19 negative or unknown; MAF=24%, OR=1.12, P=2x10^{-10}); and
74 rs2236757:A in *IFNAR2* (COVID-19 hospitalized vs. COVID-19 negative or unknown;
75 MAF=29%, OR=1.13, P=3x10^{-5}). Effect sizes were comparable across ancestries
76 (**Supplementary Table 7**). Of note, nominally significant and directionally consistent associations
77 were observed with the remaining three published risk variants (in/near *OAS3*, *DPP9* and
78 *RAVER1*; **Supplementary Table 6**), suggesting these associations likely represent true-positives.
79

80 Having established these five variants (in/near *LZTFL1*, *SLC6A20*, MHC, *ABO* and *IFNAR2*) as
81 *bona fide* risk factors for COVID-19, we then asked which (if any) of these variants were also
82 associated with severity amongst COVID-19 cases. Among these five variants, three were
83 significantly (P<0.05) associated with worse outcomes among infected individuals **(Figure 1A**
84 and **Supplementary Table 8),** while rs9411378 in *ABO* and rs2531743 near *SLC6A20* did not
85 associate with COVID-19 severity (**Figure 1B**). Collectively, these results highlight three variants
86 (in/near *LZTFL1*, *MHC* and *IFNAR2*) associated with both COVID-19 susceptibility and worse
87 disease outcomes, including respiratory failure and death. These variants may be used to identify
88 individuals at risk of severe COVID-19 and suggest nearby genes may represent targets for
89 therapeutic intervention.
90
91 We next looked for novel genetic associations with severe COVID-19. No new associations at
92 $P<5x10^{-8}$ were observed across our two severity phenotypes (hospitalized vs. non-hospitalized
93 cases; severe vs. non-hospitalized cases; **Supplementary Figure 1**), suggesting these analyses
94 were underpowered for genome-wide discovery. To increase power, we combined results from our
95 COVID-19 hospitalization phenotype (3,056 cases vs. 653,022 controls) with data from three
96 publicly available GWAS (FinnGen[19], Ellinghaus et al.[20] and GenOMICC[21]) for a
97 combined sample size of 6,993 hospitalized cases and 904,801 controls with no record of SARS-
98 CoV-2 infection. In this larger analysis of hospitalization risk, seven independent variants in six
99 loci reached genome-wide significance (**Figure 2**), including five variants that were reported in
100 previous GWAS (*LZTFL1* [two variants], MHC, *DPP9* and *IFNAR2*) and two novel associations
101 (**Table 1)**: rs9903642 near *KAT7* (5q34; MAF=4%, OR=1.32, $P=1.4x10^{-8}$), which is required for
102 the generation and maintainenace of peripheral T cells [23]; and rs2298661 in *TMPRSS2* (16q24.3;

103 MAF=22%, OR=0.87, P=1.9x10^{-8}), a host protease that promotes SARS-CoV-2 entry [24].
104 Consistent associations were observed across studies and ancestries for the *KAT7* and *TMPRSS2*
105 variants (**Supplementary Figure 2**). We then assessed the association between each of these two
106 novel variants and disease severity and found that the risk alleles for both were overrepresented
107 (P<0.05) amongst hospitalized cases (**Figure 1C and Supplementary Table 9**). Therefore, our
108 two novel susceptibility loci for COVID-19 risk are also associated with increased risk of worse
109 disease outcomes.
110
111 Collectively, our association analyses highlighted five common variants: three known (in/near
112 *LZTFL1*, MHC, *IFNAR2*) and two novel (in/near *KAT7*, *TMPRSS2*) associated with COVID-19
113 susceptibility (P<5x10^{-8}) as well as disease severity among cases (P<0.05). As all five variants are
114 either intronic or intergenic, it is not immediately clear how these variants influence gene function.
115 To help identify gene targets, and thus potential therapeutic targets for severe COVID-19, we
116 searched for functional protein-coding variants (missense or predicted loss-of-function) in high
117 LD (r^2>0.80) with each variant. We found nine functional variants in six genes (**Supplementary**
118 **Table 10**): *TMPRSS2*, discussed in greater detail below; *IFNAR2*, a cytokine receptor component
119 in the anti-viral type 1 interferon pathway, which is activated by SARS-CoV-2 and is dysregulated
120 in severe COVID-19 cases [25-27]; *CCHCR1*, a P-body protein associated with cytoskeletal
121 remodeling and mRNA turnover [28, 29], which can interact with human papillomavirus-16 [30,
122 31]; *TCF19*, a transcription factor associated with hepatitis B [32, 33]; and *C6orf15* and
123 *PSORS1C1*, two functionally uncharacterized genes in the MHC that are genetically associated
124 with autoimmunity [34]. These data indicate that the risk variants we identified may have
125 functional effects on these six genes.

126

127 Next, we asked if the sentinel risk variants co-localized (*i.e.* were in high LD, r^2>0.80) with
128 published sentinel expression quantitative trait loci (eQTL) across 52 studies (**Supplementary**
129 **Table 11**), specifically focusing on 121 genes in *cis* (± 500 kb). We found co-localization with
130 sentinel eQTL for seven genes (**Supplementary Table 12)**: *CCR2* (eQTL from blood), a
131 monocyte-expressed chemokine receptor upregulated in COVID-19 [35-37] and ARDS [38];
132 *SLC6A20* (lung), a proline transporter that binds the host SARS-CoV-2 receptor, angiotensin-
133 converting enzyme 2 [39]; *IFNAR2* and *TCF19* (multiple tissues), both discussed above*; LST1*
134 (blood)*,* an immune modulatory protein that inhibits lymphocyte proliferation [40] and is
135 upregulated in response to bacterial ligands [41]; *HLA-C* (adipose), a natural killer cell ligand,
136 associated with viral infection [42] and autoimmunity [43]; and *IL10RB* (multiple tissues), a
137 pleotropic cytokine receptor associated with persistent hepatitis B and autoimmunity [44, 45].
138 Collectively across missense and regulatory variants, these analyses identified 11 specific putative
139 effector genes in COVID-19 risk loci, though functional studies are required to confirm these
140 predictions.

141 Of these 11 putative target genes, *TMPRSS2* stands out because of its demonstrated role in
142 promoting viral infections. *TMPRSS2* encodes a host protease that cleaves viral surface
143 glycoproteins to promote viral entry, including those from influenza hemagglutinin [46] and both
144 Middle East respiratory syndrome-related coronavirus [47] and SARS-CoV-1 spike proteins [48].
145 Further, *Tmprss2* deficient mice show reduced viral spread and immunopathology in the lungs in
146 response to coronavirus infection [49], and SARS-CoV-2 viral entry is blocked by protease
147 inhibitors that inhibit *TMPRSS2* activity [24]. These observations suggest that inhibition of

148 TMPRSS2 might be protective against COVID-19. Consistent with this hypothesis, we found that
149 the missense variant that was associated with protection from COVID-19 (rs12329760, p.V160M;
150 **Supplementary Table 11**) is predicted to be deleterious by SIFT [50], PolyPhen2-HVAR and
151 PolyPhen2-HDIV [51] and is also predicted to decrease protein stability in structural models [52,
152 53]. These results provide human genetic support for loss of *TMPRSS2* in SARS-CoV-2 infection
153 and suggest that TMPRSS2 blockade might protect against worse COVID-19 outcomes.
154
155 The second aim of this study was to determine if common genetic variants can help identify
156 individuals at high risk of severe disease, who may be prioritized for prophylactic or therapeutic
157 interventions. Specifically, we focused on the three variants (in/near *LZTFL1*, MHC and *IFNAR2*)
158 with confirmed association with COVID-19 susceptibility and that also modulate COVID-19
159 severity. Using these variants, we created a weighted genetic risk score (GRS) for individuals with
160 COVID-19 and then compared the risk of hospitalization and severe disease between those with a
161 high GRS and all other cases, after adjusting for established risk factors (*e.g.* age, sex,
162 comorbidities). The weights used for each variant corresponded to the effect size (log of the odds
163 ratio) reported in previous GWAS. When considering COVID-19 cases of European ancestry
164 (N=16,467), we found that having a high GRS (top 10%) was associated with a 1.45-fold increased
165 risk of hospitalization (95% CI 1.25-1.69, P=2x10^{-6}; **Figure 3A**) and 1.88-fold increased risk of
166 severe disease (95% CI 1.51-2.35, P=3x10^{-8}; **Figure 3B**). A consistent pattern was observed in
167 other ancestries, though sample sizes were considerably smaller (**Supplementary Tables 13** and
168 **14**).
169

170 We then compared the effect of the GRS between individuals with and without established risk
171 factors for severe COVID-19. In both the AncestryDNA and UK Biobank studies, we found that
172 a high GRS (top 10%) was associated with risk of severe disease both among individuals with and
173 without established clinical risk factors for severe COVID-19 (**Figure 4**). In the meta-analysis of
174 the two studies, a high GRS was associated with a 1.79-fold (95% CI 1.37-2.33, P=2x10^{-5}) and
175 2.42-fold (95% CI 1.65-3.55, P=6x10^{-6}) higher risk of severe disease, respectively among
176 individuals with and without established risk factors (**Supplementary Table 15**). There was no
177 evidence for heterogeneity of GRS effect with clinical risk factor status (P=0.21). Similar results
178 were observed for risk of hospitalization (**Supplementary Figure 3** and **Supplementary Table**
179 **15**) and when including in the GRS all eight variants reported to associate with COVID-19
180 susceptibility in previous GWAS (**Supplementary Figure 4**). Collectively, these results
181 demonstrate that a GRS calculated using variants associated with disease risk and severity can be
182 used to identify COVID-19 cases at high risk of developing poor disease outcomes. This is
183 important as many of these individuals might benefit from prioritization for prophylactic or
184 therapeutic interventions [54].
185
186 In summary, we confirmed five common variant associations with COVID-19 susceptibility and
187 further show that three of these variants modulate disease severity among cases. We also identified
188 two novel associations with both disease susceptibility and severity, one of which provides human
189 genetic support for TMPRSS2 blockade in protection from COVID-19. Lastly, we demonstrate
190 that a genetic risk score based on common variants validated in this study can be used to identify
191 individuals at high risk of poor disease outcomes. Collectively, our results point to potential novel
192 targets for COVID-19 and can be applied to predict severe outcomes in COVID-19 patients.

193 **ONLINE METHODS**

194 **Participating Studies**

195 *AncestryDNA COVID-19 Research Study.* AncestryDNA customers over age 18, living in the
196 United States, and who had consented to research, were invited to complete a survey assessing
197 COVID-19 outcomes and other demographic information. These included SARS-CoV-2 swab and
198 antibody test results, COVID-19 symptoms and severity, brief medical history, household and
199 occupational exposure to SARS-CoV-2, and blood type. A total of 83,930 AncestryDNA survey
200 respondents were selected for inclusion in this study [55]. Respondents selected for this study
201 included all individuals with a positive COVID-19 test together with age and sex matched controls.
202 DNA samples were genotyped as described previously[55]. Genotype data for variants not
203 included in the array were then inferred using imputation to the Haplotype Reference Consortium
204 (HRC) reference panel. Briefly, samples were imputed to HRC version 1.1, which consists of
205 27,165 total individuals and 36 million variants. The HRC reference panel does not include indels;
206 consequently, indels are not present in the imputed data. We determined best-guess haplotypes
207 with Eagle version 2.4.1 and performed imputation with Minimac4 version 1.0.1. We used
208 1,117,080 unique variants as input and 8,049,082 imputed variants were retained in the final data
209 set. Variants with a Minimac4 $R^2<0.30$ were filtered from the analysis.

210

211 *Geisinger Health System (GHS).* The GHS MyCode Community Health Initiative is a health
212 system-based cohort from central and eastern Pennsylvania (USA) with ongoing recruitment since
213 2006[56]. A subset of 144,182 MyCode participants sequenced as part of the GHS-Regeneron
214 Genetics Center DiscovEHR partnership were included in this study. Information on COVID-19
215 outcomes were obtained through GHS's COVID-19 registry. Patients were identified as eligible

216 for the registry based on relevant lab results and ICD-10 diagnosis codes; patient charts were then
217 reviewed to confirm COVID-19 diagnoses. The registry contains data on outcomes, comorbidities,
218 medications, supplemental oxygen use, and ICU admissions. DNA from participants was
219 genotyped on either the Illumina OmniExpress Exome (OMNI) or Global Screening Array (GSA)
220 and imputed to the TOPMed reference panel (stratified by array) using the TOPMed Imputation
221 Server. Prior to imputation, we retained variants that had a MAF >= 0.1%, missingness < 1% and
222 HWE p-value > 10^{-15}. Following imputation, data from the OMNI and GSA datasets were merged
223 for subsequent association analyses, which included an OMNI/GSA batch covariate, in addition
224 to other covariates described below.
225
226 *Penn Medicine BioBank (PMBB) study*. PMBB study participants are recruited through the
227 University of Pennsylvania Health System (UPHS), which traditionally enrolled participants
228 during hospital or clinic visits. Due to COVID-19, recruitment pivoted to electronic consenting.
229 Participants donate blood or tissue and allow access to EHR information [57]. Patients were
230 identified through the UPHS COVID-19 registry, and subsequently invited to participate in
231 PMBB. The PMBB COVID-19 registry consists of patients who have positive qPCR testing for
232 SARS-COV-2 and also enrolled in PMBB. We then used electronic health records to classify
233 COVID-19 patients into hospitalized and severe (ventilation or death) categories. DNA genotyping
234 was performed with the Illumina Global Screening Array, and imputation performed using the
235 TOPMed reference panel as described for GHS above.
236
237 *UK Biobank (UKB) study*. We studied the host genetics of SARS-CoV-2 infection in participants
238 of the UK Biobank study, which took place between 2006 and 2010 and includes approximately

239 500,000 adults aged 40-69 at recruitment. In collaboration with UK health authorities, the UK
240 Biobank has made available regular updates on COVID-19 status for all participants, including
241 results from four main data types: qPCR test for SARS-CoV-2, anonymized electronic health
242 records, primary care and death registry data. We report results based on phenotype data
243 downloaded on the 4th January 2021 and excluded from the analysis 28,547 individuals with a
244 death registry event prior to 2020. DNA samples were genotyped as described previously [58]
245 using the Applied Biosystems UK BiLEVE Axiom Array (N=49,950) or the closely related
246 Applied Biosystems UK Biobank Axiom Array (N=438,427). Genotype data for variants not
247 included in the arrays were inferred using the TOPMed reference panel, as described above.
248

249 **COVID-19 phenotypes used for genetic association analyses**

250 We grouped participants from each study into three broad COVID-19 disease categories
251 (**Supplementary Table 1**): (i) positive – those with a positive qPCR or serology test for SARS-
252 CoV-2, or a COVID-19-related ICD10 code (U07), hospitalization or death; (ii) negative – those
253 with only negative qPCR or serology test results for SARS-CoV-2 and no COVID-19-related
254 ICD10 code (U07), hospitalization or death; and (iii) unknown – those with no qPCR or serology
255 test results and no COVID-19-related ICD10 code (U07), hospitalization or death. We then used
256 these broad COVID-19 disease categories, in addition to hospitalization and disease severity
257 information, to create seven COVID-19-related phenotypes for genetic association analyses, as
258 detailed in **Supplementary Table 3**. For association analysis in the AncestryDNA study, we
259 excluded from the COVID-19 unknown group individuals who had (i) a first-degree relative who
260 was COVID-19 positive; or (ii) flu-like symptoms.
261

262 **Genetic association analyses**

263 Association analyses in each study were performed using the genome-wide Firth logistic
264 regression test implemented in REGENIE [59]. In this implementation, Firth's approach is applied
265 when the p-value from standard logistic regression score test is below 0.05. We included in step 1
266 of REGENIE (*i.e.* prediction of individual trait values based on the genetic data) directly
267 genotyped variants with a minor allele frequency (MAF) >1%, <10% missingness, Hardy-
268 Weinberg equilibrium test *P*-value>10^{-15} and linkage-disequilibrium (LD) pruning (1000 variant
269 windows, 100 variant sliding windows and r^2<0.9). The association model used in step 2 of
270 REGENIE included as covariates age, age^2, sex, age-by-sex, and the first 10 ancestry-informative
271 principal components (PCs) derived from the analysis of a stricter set of LD-pruned (50 variant
272 windows, 5 variant sliding windows and r^2<0.5) common variants from the array (imputed for the
273 GHS study) data.

274 Within each study, association analyses were performed separately for five different
275 continental ancestries defined based on the array data: African (AFR), Admixed American (AMR),
276 European (EUR) and South Asian (SAS). We determined continental ancestries by projecting each
277 sample onto reference principle components calculated from the HapMap3 reference panel.
278 Briefly, we merged our samples with HapMap3 samples and kept only SNPs in common between
279 the two datasets. We further excluded SNPs with MAF<10%, genotype missingness >5% or
280 Hardy-Weinberg Equilibrium test p-value < 10^{-5}. We calculated PCs for the HapMap3 samples
281 and projected each of our samples onto those PCs. To assign a continental ancestry group to each
282 non-HapMap3 sample, we trained a kernel density estimator (KDE) using the HapMap3 PCs and
283 used the KDEs to calculate the likelihood of a given sample belonging to each of the five
284 continental ancestry groups. When the likelihood for a given ancestry group was >0.3, the sample

285 was assigned to that ancestry group. When two ancestry groups had a likelihood >0.3, we
286 arbitrarily assigned AFR over EUR, AMR over EUR, AMR over EAS, SAS over EUR, and AMR
287 over AFR. Samples were excluded from analysis if no ancestry likelihoods were >0.3, or if more
288 than three ancestry likelihoods were > 0.3.

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

291

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

293 We identified as a likely target of a sentinel GWAS variant any gene for which a sentinel
294 expression quantitative trait locus (eQTL) co-localized (*i.e.* had LD $r^2 > 0.80$) with the sentinel
295 GWAS variant. That is, we only considered genes for which there was strong LD between a
296 sentinel GWAS variant and a sentinel eQTL, which reduces the chance of spurious colocalization.
297 Sentinel eQTL were defined across 174 published datasets (**Supplementary Table 10**), as
298 described previously [60]. We did not use statistical approaches developed to distinguish
299 colocalization from shared genetic effects because these have very limited resolution at high LD
300 levels ($r^2 > 0.80$) [61].

301

302 **Genetic risk score (GRS) analysis of COVID-19 hospitalization and severity**

303 First, in each study (AncestryDNA, GHS, UKB and PMBB), we created a GRS for each COVID-
304 19 positive individual based on variants that were reported to associate with COVID-19
305 susceptibility in previous GWAS and that we (i) independently replicated; and (ii) found to be
306 associated with COVID-19 severity. We used as weights the effect (beta) reported in previous
307 GWAS (**Supplementary Table 5**). Second, we ranked COVID-19 individuals based on the GRS

308 and created a new binary GRS predictor by assigning each individual to a high (top 5%) or low
309 (rest of the population) percentile group. Third, for studies with >100 hospitalized cases, we used
310 logistic regression to test the association between the binary GRS predictor and risk of
311 hospitalization (hospitalized cases vs. all other cases), including as covariates age, sex, age-by-sex
312 interaction, and ten ancestry-informative PCs. In addition to age and sex, we included as additional
313 covariates established clinical risk factors for COVID-19 that are outlined in the Emergency Use
314 Authorisation treatment guidelines for casirivimab and imdevimb: BMI, chronic kidney disease,
315 diabetes, immunosuppressive disease, chronic obstructive pulmonary disease or other chronic
316 respiratory disease, cardiovascular disease and hypertension. We repeated the association analysis
317 (i) using different percentile cut-offs for the GRS (5%, 10%, 20%, 30% and 40%); and (ii) to test
318 the association with disease severity (severe cases vs. all other cases). We then stratified COVID-
319 19 cases by clinical risk (high versus lower) and evaluated the association between the top 10%
320 by GRS (i.e. high genetic risk) and risk of hospitalization or severe disease. The stratified analyses
321 were performed with logistic regression, with sex and ancestry-informative PCs included as
322 covariates. High clinical risk was defined as any one of the following: (i) age≥65; (ii) BMI≥35;
323 (iii) chronic kidney disease, diabetes or immunosuppressive disease; (iv) age ≥55 and presence of
324 chronic obstructive pulmonary disease/other chronic respiratory disease, cardiovascular disease,
325 or hypertension.
326

327 **Results availability**

328 All genotype-phenotype association results reported in this study are available for browsing using
329 the RGC's COVID-19 Results Browser (https://rgc-covid19.regeneron.com). Data access and use

330 is limited to research purposes in accordance with the Terms of Use (https://rgc-
331 covid19.regeneron.com/terms-of-use).

332 **Competing interests**

333 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.,
334 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.,
335 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
336 and/or stockholders of Regeneron Genetics Center or Regeneron Pharmaceuticals. G.H.L.R.,
337 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
338 employees at AncestryDNA and may hold equity in AncestryDNA. The other authors declare no
339 competing interests.

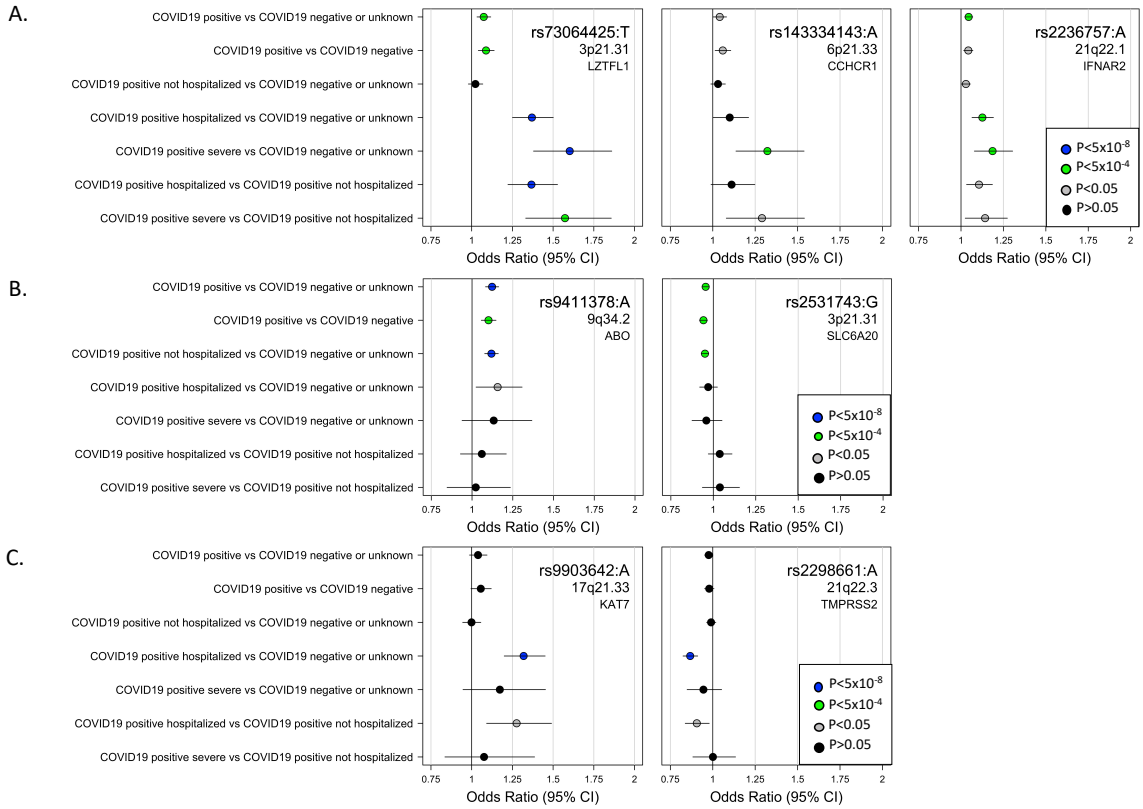
340

341 **Acknowledgements**

342 This research has been conducted using the UK Biobank Resource (Project 26041). The Penn
343 Medicine BioBank is funded by a gift from the Smilow family, the National Center for Advancing
344 Translational Sciences of the National Institutes of Health under CTSA Award Number
345 UL1TR001878, and the Perelman School of Medicine at the University of Pennsylvania. We want
346 to acknowledge the participants and investigators of the FinnGen study.

347 **FIGURES**

348



349

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

351 **for five previously reported (A, B) and two novel (C) risk variants.**

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

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

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

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

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

357 of COVID-19 hospitalization across seven studies (AncestryDNA, GHS, UKB, PMBB, FinnGen,

358 Ellinghaus et al. and Pairo-Castineira et al).

359

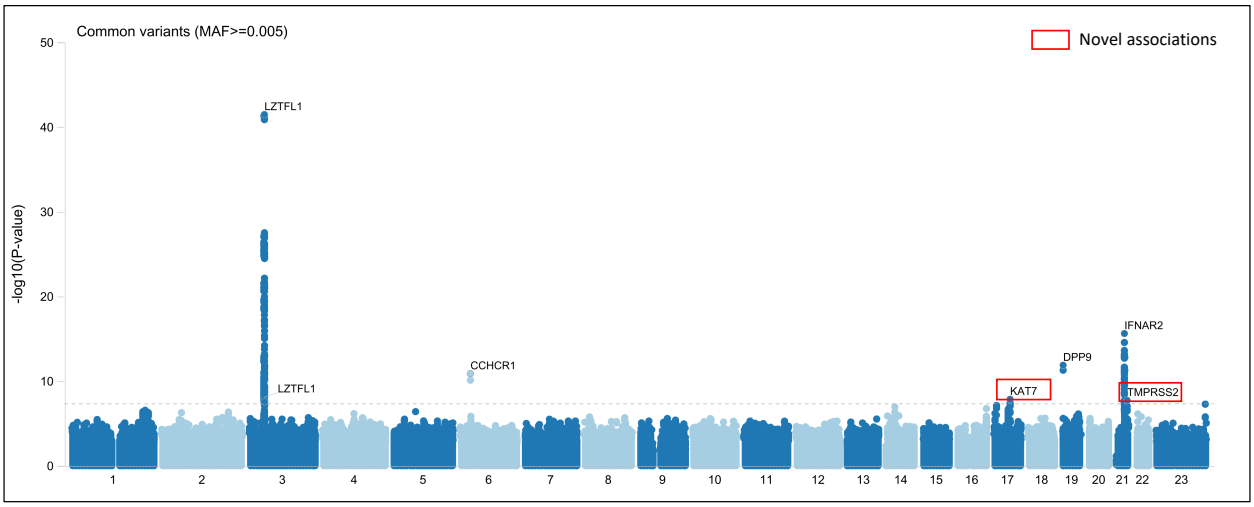
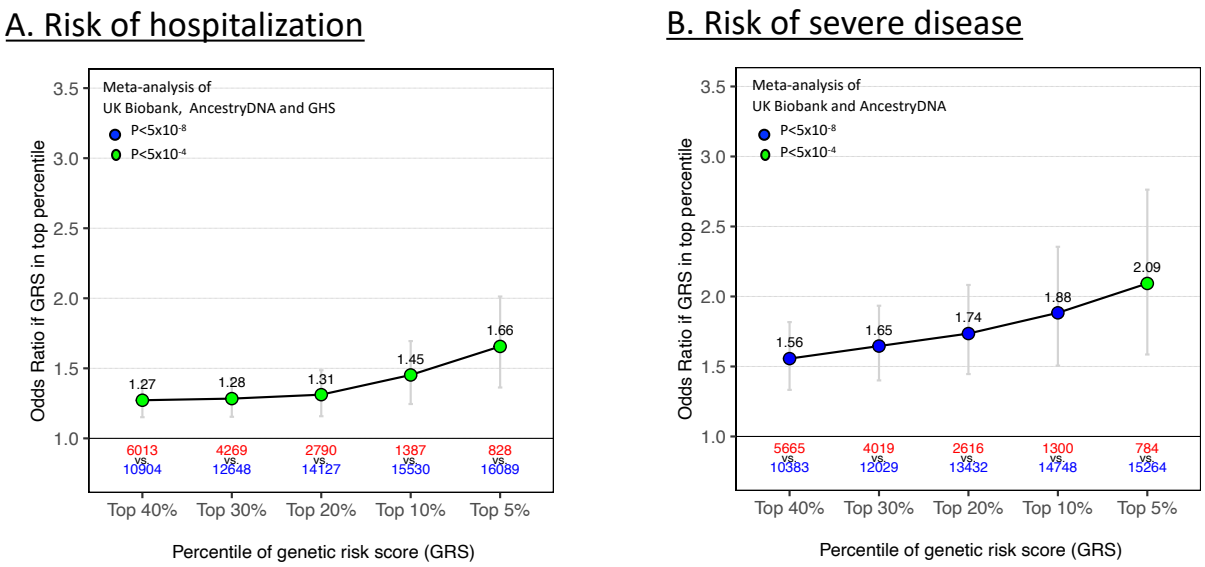


Figure 2. Summary of association results from the trans-ancestry meta-analysis of risk of COVID-19 hospitalization. The meta-analysis included results from seven studies (AncestryDNA, GHS, UKB, PMBB, FinnGen, Ellinghaus et al. and Pairo-Castineira et al. and five ancestries (AFR, AMR, EUR, SAS and EAS), totaling 6,993 hospitalized cases and 904,801 controls with no record of SARS-CoV-2 infection.

368



369

370 **Figure 3. Association between a three-SNP genetic risk score (GRS) and risk of**
371 **hospitalization (A) and severe disease (B) among COVID-19 cases of European ancestry.**
372 **(A)** Association between high genetic risk and hospitalization. Risk of hospitalization among cases
373 is shown for individuals in the top GRS percentile, agnostic to the number of clinical risk factors
374 present. The association was tested in three studies separately (AncestryDNA, UKB and GHS
375 studies) using logistic regression, with established risk factors for COVID-19 included as
376 covariates (see Methods for details). Results were then meta-analyzed across studies, for a
377 combined sample size of 16,917 COVID-19 cases, including 2,640 hospitalized. N in red: number
378 of COVID-19 cases in the top GRS percentile.
379 **(B)** Association between high genetic risk and severe disease. The association was tested as
380 described above in two studies separately (AncestryDNA and UKB studies). Results were then
381 meta-analyzed across studies, for a combined sample size of 16,048 COVID-19 cases, including
382 816 with severe disease.

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

384 cases in the rest of population.

385

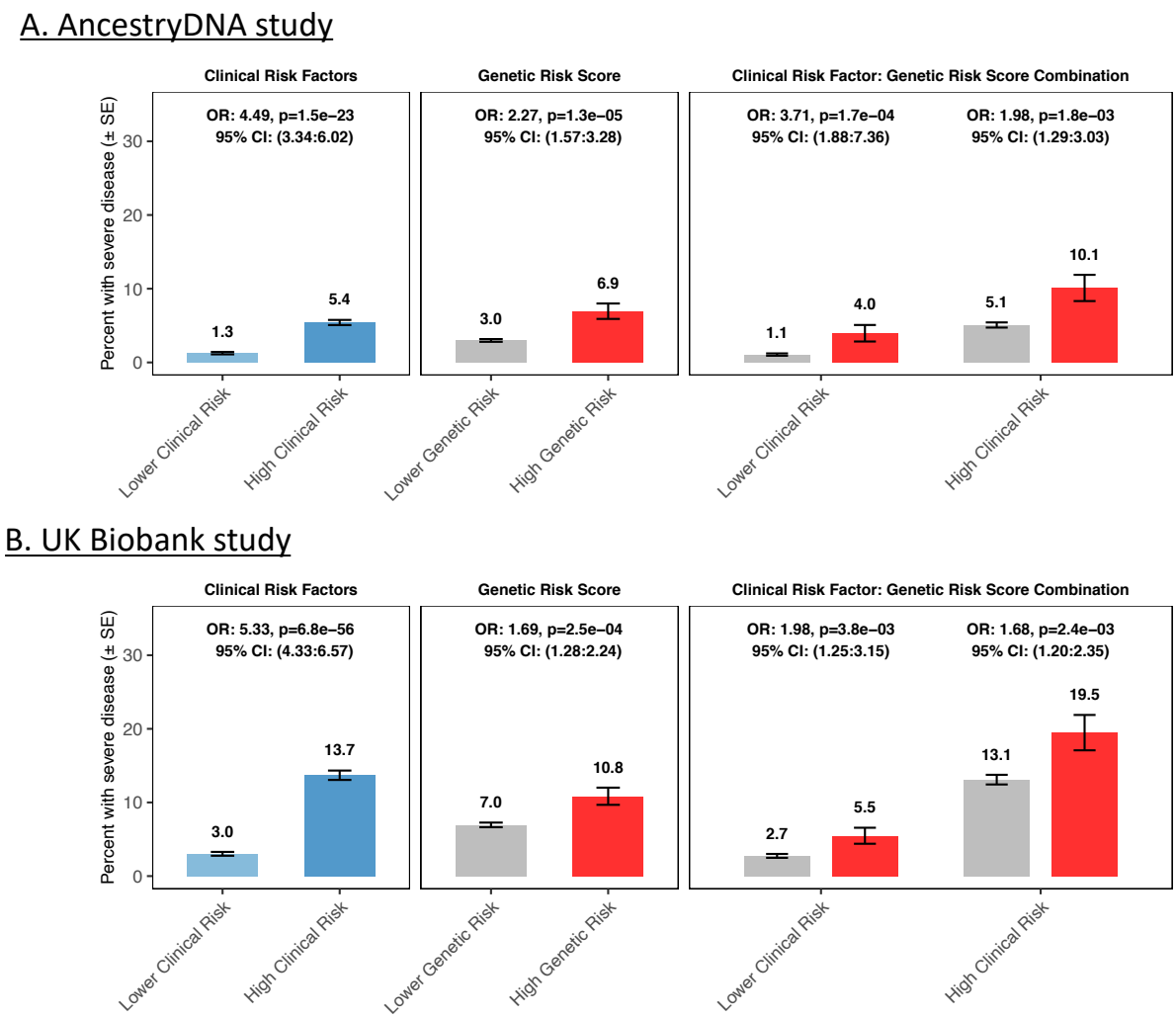


Figure 4. Association between a three-SNP genetic risk score (GRS) and risk of severe disease among COVID-19 cases of European ancestry after stratifying by the presence of clinical risk factors.

(A) Rate of severe disease in the AncestryDNA study (8,835 COVID-19 cases, including 287 with severe disease).

(B) Rate of severe disease in the UK Biobank study (7,213 COVID-19 cases, including 529 with severe disease).

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).

396 **TABLES**

397

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

399 **hospitalization.** The meta-analysis included results from seven studies (AncestryDNA, GHS,

400 UKB, PMBB, FinnGen [19], Pairo-Castineira et al. [20] and Ellinghaus et al. [21]) and five

401 ancestries (AFR, AMR, EUR, SAS and EAS), totaling 6,993 hospitalized cases and 904,801

402 controls with no record of SARS-CoV-2 infection.

Nearest gene	rsID	Position hg38	Effect allele	Odds ratio* (95% CI)	P-value	N cases (N with 0|1|2 copies of the effect allele)	N controls (N with 0|1|2 copies of the effect allele)	Effect allele frequency
Independent associations in previously reported risk loci								
LZTFL1	rs9814578	3:45821775:T:G	G	0.891 (0.855, 0.928)	2.68E-08	6993 (2160|3357|1467)	904801 (239479|446973|218337)	0.488
LZTFL1	rs67959919	3:45830416:G:A	A	1.602 (1.497, 1.714)	3.36E-42	6662 (5191|1363|102)	903146 (776403|121415|5321)	0.073
CCHCR1	rs111837807	6:31153455:T:C	C	1.272 (1.187, 1.364)	1.33E-11	6844 (5704|1075|59)	904056 (780913|118381|4754)	0.071
DPP9	rs2277732	19:4723658:C:A	A	1.171 (1.121, 1.223)	1.34E-12	6993 (3140|3028|819)	904801 (444323|378278|82190)	0.300
IFNAR2	rs2252639	21:33245424:A:G	A	1.192 (1.142,1.242)	2.46E-16	6993 (2871|3175|938)	904801 (423900|389958|90935)	0.316
Independent associations in putative novel risk loci								
KAT7	rs9903642	17:49791559:G:A	A	1.316 (1.197, 1.447)	1.40E-08	6756 (6043|687|19)	903616 (838183|63992|1434)	0.037
TMPRSS2	rs2298661	21:41473715:C:A	A	0.869 (0.827, 0.912)	1.92E-08	6993 (4550|2186|247)	904801 (548976|311126|44691)	0.221

403 * Estimated based on trans-ancestry inverse-variance meta-analysis across UKB, AncestryDNA, GHS, UPENN-PMBB and three publicly available GWAS (FinnGen, Ellinghaus et al., Pairo-Castineira et al.

404

405 **SUPPLEMENTARY FIGURES**

406 Provided in a separate file.

407

408 **Supplementary Figure 1. Summary of association results from the meta-analysis of two**
409 **phenotypes related to disease severity among COVID-19 cases**. **(A)** Phenotype COVID-19
410 positive hospitalized (N=3,056) vs COVID-19 positive and not hospitalized (N=17,018). **(B)**
411 Phenotype COVID-19 positive severe (N=1,007) vs COVID-19 positive and not hospitalized
412 (N=16,919).

413

414 **Supplementary Figure 2. Two putative novel loci identified in the trans-ancestry meta-**
415 **analysis of risk of COVID-19 hospitalization. (A)** rs9903642 (17:49791559:G:A, effect allele
416 A) at 17q21.33, near *KAT7*. **(B)** rs2298661 (21:41473715:C:A, effect allele A) at 21q22.3 in
417 *TMPRSS2*. RR: number of individuals homozygous for reference allele. RA: number of
418 heterozygous individuals. RA: number of individuals homozygous for alternative (effect) allele.
419 AAF: alternative allele frequency. OR: odds ratio.

420

421 **Supplementary Figure 3. Association between a three-SNP genetic risk score (GRS) and risk**
422 **of hospitalization among COVID-19 cases of European ancestry after stratifying by the**
423 **presence of clinical risk factors.**

424 (A) Rate of hospitalization in the AncestryDNA study (8,835 COVID-19 cases, including 636
425 hospitalized).

426 (B) Rate of hospitalization in the UK Biobank study (7,213 COVID-19 cases, including 1,824
427 hospitalized).

428 High genetic risk (red bars): top 10% of the GRS. Low genetic risk (grey bars): bottom 90% of the
429 GRS (*i.e.* all other COVID-19 cases).
430
431 **Supplementary Figure 4. Association between a 3- and 8-SNP genetic risk score (GRS) and**
432 **risk of hospitalization and severe disease among COVID-19 cases of European ancestry.** To
433 evaluate if the association between the GRS and worse disease outcomes was dependent on the
434 list of variants selected for analysis, we compared results between GRS calculated using different
435 sets of variants. We considered a GRS calculated using: (i) the three variants that were reported in
436 previous GWAS of COVID-19 susceptibility and that we validated the published association and
437 further showed that they were associated with risk of hospitalization or severe disease among
438 COVID-19 cases (in/near *LZTFL1*, MHC, *IFNAR2*; see **Figure 1A**); or (ii) all eight variants
439 reported in previous GWAS of COVID-19 susceptibility (in/near *LZTFL1* [two variants], MHC,
440 *ABO*, *OAS3*, *DPP9*, *RAVER1* and *IFNAR2*; see **Supplementary Table 5**). Analyses were
441 performed separately in the UK Biobank, AncestryDNA and GHS studies (risk of hospitalization
442 only) after stratifying COVID-19 cases by the presence of clinical risk factors, considering
443 individuals with lower clinical risk (blue circles), high clinical risk (green triangles) or all
444 individuals (grey squares). Association results were then meta-analyzed across studies.

445 **SUPPLEMENTARY TABLES**

446 Provided in a separate file.

447

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

449 **the analysis.**

450

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

452

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

454

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

456 **phenotypes analyzed.**

457

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

459 **GWAS.**

460

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

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

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

464

465 **Supplementary Table 7. Comparison of effect sizes between ancestries for the five published**

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

467

468 **Supplementary Table 8. Association between two severity phenotypes and five published**
469 **risk variants for COVID-19 that validated in this study.**
470
471 **Supplementary Table 9. Association between two severity phenotypes and two novel risk**
472 **variants for COVID-19 identified in this study.**
473
474 **Supplementary Table 10. Missense or predicted loss-of-function variants in high linkage**
475 **disequilibrium (LD, r^2>0.80) with sentinel GWAS variants.**
476
477 **Supplementary Table 11. Published gene expression datasets used to identify sentinel**
478 **expression quantitative trait loci (eQTL) in high LD (r^2>0.80) with sentinel GWAS variants.**
479
480 **Supplementary Table 12. Expression quantitative trait loci (eQTL) in high LD (r^2>0.80) with**
481 **sentinel GWAS variants.**
482
483 **Supplementary Table 13. Association between a three-SNP genetic risk score (GRS) and risk**
484 **of hospitalization among COVID-19 cases.**
485
486 **Supplementary Table 14. Association between a three-SNP genetic risk score (GRS) and risk**
487 **of severe disease among COVID-19 cases.**
488

489 **Supplementary Table 15. Association between a three-SNP genetic risk score (GRS) and**
490 **risk of hospitalization and severe disease among COVID-19 cases, after stratifying COVID-**
491 **19 cases by pre-existing clinical risk factor status for severe COVID-19.**
492
493

494 **References**

495 1. Zhu, N., et al., *A Novel Coronavirus from Patients with Pneumonia in China, 2019.* New
496 England Journal of Medicine, 2020. **382**(8): p. 727-733.
497 2. Coronaviridae Study Group of the International Committee on Taxonomy of, V., *The*
498 *species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV*
499 *and naming it SARS-CoV-2.* Nat Microbiol, 2020. **5**(4): p. 536-544.
500 3. *World Health Organization Coronavirus Disease Weekly Epidemiological Update*. 2020;
501 Available from: https://covid19.who.int/.
502 4. Guan, W.J., et al., *Clinical Characteristics of Coronavirus Disease 2019 in China.* N Engl J
503 Med, 2020. **382**(18): p. 1708-1720.
504 5. Richardson, S., et al., *Presenting Characteristics, Comorbidities, and Outcomes Among*
505 *5700 Patients Hospitalized With COVID-19 in the New York City Area.* Jama, 2020.
506 **323**(20): p. 2052-2059.
507 6. Atkins, J.L., et al., *PREEXISTING COMORBIDITIES PREDICTING SEVERE COVID-19 IN OLDER*
508 *ADULTS IN THE UK BIOBANK COMMUNITY COHORT.* medRxiv, 2020: p.
509 2020.05.06.20092700.
510 7. Zhou, F., et al., *Clinical course and risk factors for mortality of adult inpatients with*
511 *COVID-19 in Wuhan, China: a retrospective cohort study.* Lancet, 2020. **395**(10229): p.
512 1054-1062.
513 8. Cummings, M.J., et al., *Epidemiology, clinical course, and outcomes of critically ill adults*
514 *with COVID-19 in New York City: a prospective cohort study.* Lancet, 2020. **395**(10239):
515 p. 1763-1770.
516 9. Lopez, L., 3rd, L.H. Hart, 3rd, and M.H. Katz, *Racial and Ethnic Health Disparities Related*
517 *to COVID-19.* JAMA, 2021.
518 10. Rochwerg, B., *A living WHO guideline on drugs for covid-19.* BMJ 2020;370:m3379 2020.
519 11. Group, R.C., et al., *Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary*
520 *Report.* N Engl J Med, 2020.
521 12. FDA, *Investigational COVID-19 Convalescent Plasma.* 2020.
522 13. *WHO Solidarity Trial Consortium. Repurposed antiviral drugs for COVID-19 –interim*
523 *WHO SOLIDARITY trial results*. medRxiv 2020; Available from:
524 https://www.medrxiv.org/content/10.1101/2020.10.15.20209817v1?ijkey=6ff7378035b
525 5c9c24346a6b90cadb5e866a63f71&keytype2=tf_ipsecsha.
526 14. Investigators, T.R.-C., *Interleukin-6 Receptor Antagonists in Critically Ill Patients with*
527 *Covid-19 – Preliminary report.* bioRxiv, 2021.
528 15. Jackson, L.A., et al., *An mRNA Vaccine against SARS-CoV-2 - Preliminary Report.* N Engl J
529 Med, 2020. **383**(20): p. 1920-1931.
530 16. Walsh, E.E., et al., *Safety and Immunogenicity of Two RNA-Based Covid-19 Vaccine*
531 *Candidates.* N Engl J Med, 2020.
532 17. Hansen, J., et al., *Studies in humanized mice and convalescent humans yield a SARS-CoV-*
533 *2 antibody cocktail.* Science, 2020. **369**(6506): p. 1010-1014.
534 18. Baum, A., et al., *Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational*
535 *escape seen with individual antibodies.* Science, 2020. **369**(6506): p. 1014-1018.

536 19. *The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host*
537 *genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic.* Eur J
538 Hum Genet, 2020. **28**(6): p. 715-718.
539 20. Ellinghaus, D., et al., *Genomewide Association Study of Severe Covid-19 with Respiratory*
540 *Failure.* New England Journal of Medicine, 2020.
541 21. Pairo-Castineira, E., et al., *Genetic mechanisms of critical illness in Covid-19.* Nature,
542 2020.
543 22. Shelton, J., F., Anjali J. Shastri, Chelsea Ye, Catherine H. Weldon, Teresa Filshtein-
544 Somnez, Daniella Coker, Antony Symons, Jorge Esparza-Gordillo, The 23andMe COVID-
545 19 Team, Stella Aslibekyan, Adam Auton, *Trans-ethnic analysis reveals genetic and non-*
546 *genetic associations with COVID-19 susceptibility and severity.* 2020.
547 23. Newman, D.M., et al., *Essential role for the histone acetyltransferase KAT7 in T cell*
548 *development, fitness, and survival.* J Leukoc Biol, 2017. **101**(4): p. 887-892.
549 24. Hoffmann, M., et al., *SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is*
550 *Blocked by a Clinically Proven Protease Inhibitor.* Cell, 2020. **181**(2): p. 271-280 e8.
551 25. Hadjadj, J., et al., *Impaired type I interferon activity and inflammatory responses in*
552 *severe COVID-19 patients.* Science, 2020. **369**(6504): p. 718-724.
553 26. van der Made, C.I., et al., *Presence of Genetic Variants Among Young Men With Severe*
554 *COVID-19.* JAMA, 2020.
555 27. Zhang, Q., et al., *Inborn errors of type I IFN immunity in patients with life-threatening*
556 *COVID-19.* Science, 2020. **370**(6515).
557 28. Tervaniemi, M.H., et al., *Intracellular signalling pathways and cytoskeletal functions*
558 *converge on the psoriasis candidate gene CCHCR1 expressed at P-bodies and*
559 *centrosomes.* BMC Genomics, 2018. **19**(1): p. 432.
560 29. Ling, Y.H., et al., *CCHCR1 interacts with EDC4, suggesting its localization in P-bodies.* Exp
561 Cell Res, 2014. **327**(1): p. 12-23.
562 30. Olejnik-Schmidt, A.K., et al., *Search for cellular partners of human papillomavirus type 16*
563 *E2 protein.* Arch Virol, 2008. **153**(5): p. 983-90.
564 31. Muller, M. and C. Demeret, *CCHCR1 interacts specifically with the E2 protein of human*
565 *papillomavirus type 16 on a surface overlapping BRD4 binding.* PLoS One, 2014. **9**(3): p.
566 e92581.
567 32. Matsuura, K., M. Isogawa, and Y. Tanaka, *Host genetic variants influencing the clinical*
568 *course of hepatitis B virus infection.* J Med Virol, 2016. **88**(3): p. 371-9.
569 33. Kim, Y.J., et al., *A genome-wide association study identified new variants associated with*
570 *the risk of chronic hepatitis B.* Hum Mol Genet, 2013. **22**(20): p. 4233-8.
571 34. Lee, K.Y., et al., *Discovering Genetic Factors for psoriasis through exhaustively searching*
572 *for significant second order SNP-SNP interactions.* Sci Rep, 2018. **8**(1): p. 15186.
573 35. Szabo, P.A., et al., *Analysis of respiratory and systemic immune responses in COVID-19*
574 *reveals mechanisms of disease pathogenesis.* medRxiv, 2020.
575 36. Zhou, Z., et al., *Heightened Innate Immune Responses in the Respiratory Tract of COVID-*
576 *19 Patients.* Cell Host Microbe, 2020. **27**(6): p. 883-890 e2.
577 37. Merad, M. and J.C. Martin, *Pathological inflammation in patients with COVID-19: a key*
578 *role for monocytes and macrophages.* Nat Rev Immunol, 2020. **20**(6): p. 355-362.

579 38. Aggarwal, N.R., L.S. King, and F.R. D'Alessio, *Diverse macrophage populations mediate*
580 *acute lung inflammation and resolution.* Am J Physiol Lung Cell Mol Physiol, 2014.
581 **306**(8): p. L709-25.
582 39. Vuille-dit-Bille, R.N., et al., *Human intestine luminal ACE2 and amino acid transporter*
583 *expression increased by ACE-inhibitors.* Amino Acids, 2015. **47**(4): p. 693-705.
584 40. Rollinger-Holzinger, I., et al., *LST1: a gene with extensive alternative splicing and*
585 *immunomodulatory function.* J Immunol, 2000. **164**(6): p. 3169-76.
586 41. Mulcahy, H., et al., *LST1 and NCR3 expression in autoimmune inflammation and in*
587 *response to IFN-gamma, LPS and microbial infection.* Immunogenetics, 2006. **57**(12): p.
588 893-903.
589 42. Apps, R., et al., *Influence of HLA-C expression level on HIV control.* Science, 2013.
590 **340**(6128): p. 87-91.
591 43. Kulkarni, S., et al., *Genetic interplay between HLA-C and MIR148A in HIV control and*
592 *Crohn disease.* Proc Natl Acad Sci U S A, 2013. **110**(51): p. 20705-10.
593 44. Begue, B., et al., *Defective IL10 signaling defining a subgroup of patients with*
594 *inflammatory bowel disease.* Am J Gastroenterol, 2011. **106**(8): p. 1544-55.
595 45. Frodsham, A.J., et al., *Class II cytokine receptor gene cluster is a major locus for hepatitis*
596 *B persistence.* Proc Natl Acad Sci U S A, 2006. **103**(24): p. 9148-53.
597 46. Bottcher, E., et al., *Proteolytic activation of influenza viruses by serine proteases*
598 *TMPRSS2 and HAT from human airway epithelium.* J Virol, 2006. **80**(19): p. 9896-8.
599 47. Kleine-Weber, H., et al., *Functional analysis of potential cleavage sites in the MERS-*
600 *coronavirus spike protein.* Sci Rep, 2018. **8**(1): p. 16597.
601 48. Matsuyama, S., et al., *Efficient activation of the severe acute respiratory syndrome*
602 *coronavirus spike protein by the transmembrane protease TMPRSS2.* J Virol, 2010.
603 **84**(24): p. 12658-64.
604 49. Iwata-Yoshikawa, N., et al., *TMPRSS2 Contributes to Virus Spread and Immunopathology*
605 *in the Airways of Murine Models after Coronavirus Infection.* J Virol, 2019. **93**(6).
606 50. Vaser, R., et al., *SIFT missense predictions for genomes.* Nat Protoc, 2016. **11**(1): p. 1-9.
607 51. Adzhubei, I., D.M. Jordan, and S.R. Sunyaev, *Predicting functional effect of human*
608 *missense mutations using PolyPhen-2.* Curr Protoc Hum Genet, 2013. **Chapter 7**: p.
609 Unit7.20.
610 52. Vishnubhotla R., V., N., et al. , *Genetic variants in TMPRSS2 and Structure of SARS-CoV-2*
611 *spike glycoprotein and TMPRSS2 complex.* bioRxiv, 2020.
612 53. David, A., Khanna, T., Beykou, M., Hanna, G., Sternberg, M. J. E., *Structure, function and*
613 *variants analysis of the androgen-regulated TMPRSS2, a drug target candidate for*
614 *COVID-19 infection.* bioRxiv, 2020.
615 54. *How CDC Is Making COVID-19 Vaccine Recommendations*. 2020 [cited 2020 December
616 2, 2020]; Available from: https://www.cdc.gov/coronavirus/2019-
617 ncov/vaccines/recommendations-process.html.
618 55. Roberts, G.H.L., *AncestryDNA COVID-19 Host Genetic Study Identifies Three Novel Loci.*
619 medRxiv, 2020.
620 56. Dewey, F.E., et al., *Distribution and clinical impact of functional variants in 50,726*
621 *whole-exome sequences from the DiscovEHR study.* Science, 2016. **354**(6319).

622 57. Park, J., et al., *A genome-first approach to aggregating rare genetic variants in LMNA for*
623 *association with electronic health record phenotypes.* Genet Med, 2020. **22**(1): p. 102-
624 111.
625 58. Bycroft, C., et al., *The UK Biobank resource with deep phenotyping and genomic data.*
626 Nature, 2018. **562**(7726): p. 203-209.
627 59. Mbatchou, J., et al., *Computationally efficient whole genome regression for quantitative*
628 *and binary traits.* bioRxiv, 2020: p. 2020.06.19.162354.
629 60. Ferreira, M.A., et al., *Shared genetic origin of asthma, hay fever and eczema elucidates*
630 *allergic disease biology.* Nat Genet, 2017. **49**(12): p. 1752-1757.
631 61. Chun, S., et al., *Limited statistical evidence for shared genetic effects of eQTLs and*
632 *autoimmune-disease-associated loci in three major immune-cell types.* Nat Genet, 2017.
633 **49**(4): p. 600-605.
634