

1 A year of COVID-19 GWAS results from the GRASP portal reveals potential SARS-CoV-2  
2 modifiers

3

4 Florian Thibord<sup>1</sup>, Melissa V. Chan<sup>1</sup>, Ming-Huei Chen<sup>1</sup>, Andrew D. Johnson<sup>1,\*</sup>

5

6 <sup>1</sup>Division of Intramural Research, Population Sciences Branch, Framingham, MA 01702

7 \*corresponding author: johnsonad2@nhlbi.nih.gov

8

9

10 Abstract

11 Host genetic variants influence the susceptibility and severity of several infectious  
12 diseases, and the discovery of novel genetic associations with Covid-19 phenotypes could help  
13 developing new therapeutic strategies to reduce its burden.

14 Between May 2020 and February 2021, we used Covid-19 data released periodically by  
15 UK Biobank and performed over 400 Genome-Wide Association Studies (GWAS) of Covid-19  
16 susceptibility (N=15,738 cases), hospitalization (N=1,916), severe outcomes (N=935) and death  
17 (N=828), stratified by ancestry and sex.

18 In coherence with previous studies, we observed 2 independent signals at the  
19 chr3p21.31 locus (rs73062389-A, OR=1.22,  $P=7.64 \times 10^{-14}$  and rs13092887-A, OR=1.73,  
20  $P=2.38 \times 10^{-8}$ , in Europeans) modulating susceptibility and severity, respectively, and a signal  
21 influencing susceptibility at the *ABO* locus (rs9411378-A, OR=1.10,  $P=7.36 \times 10^{-10}$ , in Europeans),  
22 which was more significant in men than in women ( $P=0.01$ ). In addition, we detected 7 genome-

23 wide significant signals in the last data release analyzed (on February 24th 2021), of which 4  
24 were associated with susceptibility (*SCRT2*, *LRMDA*, chr15q24.2, *MIR3681HG*), 2 with  
25 hospitalization (*ANKS1A*, chr12p13.31) and 1 for severity (*ADGRE1*). Finally, we identified over  
26 300 associations which increased in significance over time, and reached at least  $P < 10^{-5}$  in the  
27 last data release analyzed. We replicated 2 of these signals in an independent dataset: a variant  
28 downstream of *CCL3* (rs2011959) associated with severity in men, and a variant located in an  
29 *ATP5PO* intron (rs12482569) associated with hospitalization.

30 These results, freely available on the GRASP portal, provide new insights on the host  
31 genetic architecture of Covid-19 phenotypes.

32

33 Introduction

34

35 The severe acute respiratory syndrome – coronavirus 2 (SARS-CoV-2) is responsible for  
36 the coronavirus disease 2019 (Covid-19) which affects individuals with variable severity, ranging  
37 from asymptomatic patients to mild respiratory symptoms, hypercytokinemia, pneumonia,  
38 thrombosis and even death.<sup>1,2</sup> Understanding the mechanisms leading to heterogeneous  
39 symptoms and susceptibility is essential in order to develop efficient treatments and improve  
40 patient care. Host genetic diversity has been shown to influence the effects of infection to  
41 several viruses,<sup>3</sup> such as variations in *CCR5* [MIM: 601373] leading to HIV resistance<sup>4</sup> [MIM:  
42 609423] or *IRF7* [MIM: 605047] deficiency affecting *Influenza* susceptibility [MIM: 614680].<sup>5</sup>

43 In order to discover human genetic determinants to Covid-19 susceptibility and severity,  
44 several biobanks and research groups worldwide collaborated to perform Genome-Wide

45 Association Studies (GWAS) and meta-analyses of the GWAS. In June 2020, a study involving  
46 1,980 Covid-19 infected patients with respiratory failure was the first to reveal genome-wide  
47 significant ( $P < 5 \times 10^{-8}$ ) associations at the 3p21.31 locus, encompassing *SLC6A20* [MIM:  
48 605616] and several chemokine receptors, and at the *ABO* [MIM: 110300] locus on  
49 chromosome 9.<sup>6</sup> These 2 signals were later validated in independent analyses for both Covid-19  
50 susceptibility and severity<sup>7,8</sup> while additional significant associations were observed at loci  
51 involved in immune response or inflammation, such as *IFNAR2* [MIM: 602376], *DPP9* [MIM:  
52 608258], *TYK2* [MIM: 176941], *CCHCR1* [MIM: 605310] and *OAS1* [MIM: 164350]. Notably,  
53 these findings implicate common and uncommon variants, while studies trying to identify  
54 associations of rare variants have been unsuccessful so far.<sup>9</sup> The largest effort is currently led by  
55 the Covid-19 host genetics initiative (Covid-19hgi),<sup>10</sup> which completed meta-analyses of results  
56 shared by 46 studies as of January 18<sup>th</sup> 2021, and plan to release new results as additional data  
57 is made available. A major contributor to this group is the UK Biobank (UKB)<sup>11</sup> which  
58 periodically releases the results of Covid-19 tests and related deaths, as well as health care data  
59 for its nearly 500,000 consented participants, to approved researchers.

60 We created a public Covid-19 GWAS results portal  
61 (<https://grasp.nih.gov/Covid19GWASResults.aspx>) in order to provide rapid deep  
62 annotation for emerging genetics results and facilitate open access to the scientific community.  
63 We contribute to this resource by performing GWAS on each Covid-19 data release from the UK  
64 Biobank, including sex-specific, ancestry-specific, and trans-ethnic Covid-19 related GWAS,  
65 along with a deep set of annotations for top variants (with  $P < 1 \times 10^{-5}$ ). For each release, up to  
66 65 GWAS have been generated including Covid-19 susceptibility, Covid-19 hospitalization,

67 severe Covid-19 with respiratory failure, and Covid-19 death. Here we describe the results of  
68 these GWAS, 434 in total as of February 24<sup>th</sup> 2021, and report the evolution of signals  
69 associated with these Covid-19 phenotypes over the consecutive datasets released by UKB  
70 since May 2020. The latter approach, tracking the evolution of genetic signatures iteratively in  
71 UKB, suggests a valuable new analytic approach in genetic biobank studies where emerging  
72 true signals may be identified before they reach genome-wide significance based on their  
73 trajectories of significance.

74

75

## 76 Materials and Methods

77

### 78 UK Biobank data

79 Analyses are based on v.3 of the UKB imputed dataset,<sup>12</sup> which provide genomic data  
80 for 487,320 participants from multiple ethnicities, including 459,250 of European ancestry  
81 (EUR), 7,644 of African ancestry (AFR), 9,417 of South Asian ancestry (SAS) and 11,009 of other  
82 ancestries (OTHERS). Participants were enrolled at ages ranging from 37 to 73 and are 51.16%  
83 female. UKB started to release Covid-19 test results of its participants on March 15<sup>th</sup> 2020, and  
84 periodically update this resource as new cases are reported. Furthermore, information about  
85 Covid-19 related death was made available from June 2020, while inpatient data and primary  
86 care data was first added during the summer of 2020 and are periodically updated. Details  
87 regarding the definition and selection of cases with Covid-19 susceptibility, Covid-19  
88 hospitalization, Covid-19 severity, and Covid-19 death are available in **Table S1**.

89

90 Phenotype definition

91            Depending on the Covid-19 phenotype analyzed (susceptibility, hospitalization, severity,  
92 or death), up to 3 different subsets of participants were used as controls. For Covid-19  
93 susceptibility, cases with positive test results were analyzed against either participants tested  
94 with negative results (labelled Tested), or against all participants without a positive test  
95 (labelled Population). For analyses of Covid-19 hospitalization, patients hospitalized due to  
96 Covid-19 were tested against non-hospitalized participants with a positive test (Positive), or  
97 non-hospitalized participants with a test (Tested), or against all non-hospitalized participants  
98 (Population). For analyses of Covid-19 severity, patients requiring invasive respiratory support  
99 or patients who died from complications were tested against non-severe participants with a  
100 positive test (Positive), or non-severe participants with a test (Tested) or all non-severe  
101 participants (Population). For analyses of Covid-19 death, patients with Covid-19 death were  
102 tested against participants with a positive test (Positive), or participants with a test (Tested) or  
103 against all participants (Population).

104

105 Analyses

106            Each GWAS was conducted with SAIGE v0.38,<sup>13</sup> which controls for population  
107 stratification, relatedness and case-control imbalance, and adjusted for baseline age (at  
108 enrollment), sex and 10 genetic principal components. For the results uploaded to the GRASP  
109 portal, variants were filtered on imputation quality ( $r^2 > 0.3$ ), minor allele count (MAC > 2), and  
110 minor allele frequency (MAF > 0.0001). However, for the results presented in this manuscript,

111 we applied a more stringent filter, and considered only well-imputed variants ( $r^2 > 0.8$ ) and  
112 common variants (MAF > 0.01). After applying this filter, the lambda (genomic control factor)  
113 ranged from 0.991 to 1.027 in all 65 analyses of the 02.24.21 data release (**Table S2**), indicating  
114 no systematic inflation. For analyses prior to the June 18<sup>th</sup> 2020 release, we conducted analyses  
115 on participants of European ancestry only, and started adding new analyses stratified by sex  
116 and ancestry from June 18<sup>th</sup> 2020 onward. GWAS were stratified for EUR, AFR, SAS, and OTHERS  
117 ancestries, and an additional trans-ancestry GWAS combining all participants (labelled ALL) as  
118 well as GWAS combining non-European (nEUR) participants were performed.

119 Associations are either reported as odd ratios (OR) and 95% confidence intervals or as  
120 beta coefficients ( $\beta$ ) and associated standard errors (SE). Linkage disequilibrium (LD) was  
121 estimated by squared correlation ( $r^2$ ) using UKB EUR imputed data. To test 2 observed effects  
122 are equal, we used the Z statistic:  $Z = \frac{b_1 - b_2}{\sqrt{SEb_1^2 + SEb_2^2}}$ , with  $b_1$  and  $b_2$  corresponding to the  
123 observed effects and  $SEb_1$  and  $SEb_2$  the associated standard errors.

124

## 125 Annotation

126 For each analysis hosted on the portal, we provide comprehensive annotation for top  
127 results ( $P < 1 \times 10^{-5}$ ) using ANNOVAR<sup>14</sup> and the RESTful API service provided by CADD v1.6.<sup>15</sup> We  
128 also retrieve known phenotype associations extracted from the GRASP<sup>16</sup> and EBI GWAS  
129 catalogs,<sup>17</sup> and known eQTLs extracted from GTEx v8<sup>18</sup> and other eQTL resources compiled  
130 from nearly 150 datasets (built upon the work of Zhang *et al*,<sup>19</sup> detailed in **Table S3**).

131

132 Replication

133           The 6 Covid-19hgi meta-analyses, release version 5, excluding the UKB and 23andMe  
134 datasets, were used for our replication efforts and include results for Covid-19 susceptibility  
135 (labelled C2), hospitalization (B2) and severity (A2). The 3 meta-analyses restricted to European  
136 individuals (EUR) were used to replicate signals observed in our GWAS involving European UKB  
137 participants, while the remaining 3 trans-ancestry meta-analyses (ALL) were used to replicate  
138 signals observed either in our trans-ancestry GWAS or in non-European analyses. A signal was  
139 defined as replicated if the effect direction was concordant in the GWAS and the Covid-19hgi  
140 meta-analysis, and if the p-value in the meta-analysis was below 0.1, corresponding to a one-  
141 sided test threshold of 0.05. In addition, the Bonferroni method was applied to correct for  
142 multiple testing.

143

144 Data Availability

145           Permission was obtained to post UKB summary statistics under an approved application  
146 (ID 28525). The association results are available on the portal, as well as annotated top results.  
147 In addition to UKB summary statistics, results from other efforts are also hosted on the portal.  
148 Authors of Covid-19 GWAS publications have been contacted to seek approval before hosting  
149 the results of their analyses on the GRASP Covid-19 portal. Summary statistics at this time  
150 include multiple releases of the Covid-19hgi group, severe hospitalization results from  
151 Ellinghaus *et al*<sup>6</sup> and the GenOMICC study,<sup>7</sup> with all results are being re-annotated in the  
152 common framework mentioned above.

153

154

155 Results

156

157 UK Biobank Covid-19 demographics

158 Using the latest UKB data release available at this point, analyzed on February 24<sup>th</sup> 2021,  
159 we retrieved 47,413 participants with a Covid-19 diagnostic, of which 15,738 tested positive.  
160 According to inpatient care data, 1,916 positive cases were hospitalized, while 935 patients  
161 with severe Covid-19 diagnostic received respiratory support and/or died from complications  
162 (**Table 1**). Since May 7<sup>th</sup> 2020, we analyzed 15 UKB data releases regarding Covid-19  
163 susceptibility, 7 data releases concerning Covid-19 related deaths, and 4 data releases for both  
164 Covid-19 related hospitalization and Covid-19 severity. Amongst Covid-19 positive participants,  
165 we observed a global increase in the percentage of female cases, starting at 45.3% at the first  
166 release analyzed, and reaching 52.47% in the last, while men were more likely to be infected ( $P$   
167  $= 9.94 \times 10^{-7}$ ), hospitalized ( $P = 1.34 \times 10^{-39}$ ), or develop severe complications ( $P = 3.40 \times 10^{-29}$ )  
168 and die from Covid-19 ( $P < 10^{-300}$ ) in the 02.24.21 data release (**Table S4**). There was also a  
169 decrease in the mean age of positive cases, ranging from 57.02 to 53.77, with a significant drop  
170 after the 2020 summer (**Table 2**), with younger individuals more likely to be infected ( $P < 10^{-300}$ )  
171 while increase in age was associated with hospitalization ( $P = 1.23 \times 10^{-73}$ ), severity ( $P = 1.41 \times$   
172  $10^{-93}$ ) and death ( $P = 3.40 < 10^{-300}$ ) (**Table S4**). Positive cases were mainly of European ancestry,  
173 representing 85.5% of all Covid-19 positive participants in the first analysis and growing to  
174 89.3% in the last. South Asian ancestry, African ancestry, and other ancestry represent 4.4%,  
175 3.2% and 3.1% of positive cases, respectively, in this UKB data release. However, compared to

176 European participants, non-European participants were more likely to be infected ( $P = 9.34 \times$   
177  $10^{-37}$  for AFR,  $P = 6.68 \times 10^{-82}$  for SAS,  $P = 4.31 \times 10^{-8}$  for OTHERS), hospitalized ( $P = 2.78 \times 10^{-28}$   
178 for AFR,  $P = 2.23 \times 10^{-9}$  for SAS,  $P = 1.01 \times 10^{-6}$  for OTHERS), severe ( $P = 2.89 \times 10^{-19}$  for AFR,  $P =$   
179  $9.84 \times 10^{-4}$  for SAS,  $P = 0.02$  for OTHERS) and deceased participants ( $P = 2.72 \times 10^{-10}$  for AFR,  $P =$   
180  $7.04 \times 10^{-3}$  for SAS) (**Table S4**).

181

## 182 Associations with Covid-19 susceptibility

183 The genome-wide significant loci in the last data release analyzed (on 02.24.21) for  
184 Covid-19 susceptibility are presented in **Table 3**, while all associations with  $P < 10^{-5}$  are available  
185 in **Table S5**. In addition, all signals reaching genome-wide significance in any data release  
186 analyzed are presented in **Figures S1-S18**.

187 This project started with the analysis of Covid-19 susceptibility in participants of  
188 European ancestry, as the largest ancestry group in UKB. Consequently, it is the most repeated  
189 analysis with 15 iterations. With untested or negatively tested participants as controls  
190 (Population controls), 8 signals reached genome-wide significance ( $P < 5 \times 10^{-8}$ ) at some point  
191 (**Figure 1**), of which only 3 remained significant in the last data release analyzed: the chr3p21.31  
192 locus encompassing *SLC6A20* and several chemokine receptors (rs73062389-A, MAF = 0.06, OR  
193 = 1.22 [1.16; 1.29],  $P = 7.64 \times 10^{-14}$ ), the *ABO* locus on chromosome 9 (rs9411378-A, MAF =  
194 0.22, OR = 1.10 [1.07; 1.14],  $P = 7.36 \times 10^{-10}$ ) and intronic variants located in *SCRT2* and *SRXN1*  
195 [MIM: 617583] on chromosome 20 (rs749199237-del.G, MAF = 0.35, OR = 1.08 [1.05; 1.11],  $P =$   
196  $1.29 \times 10^{-8}$ ). The chr3p21.31 and *ABO* loci were previously reported to modulate Covid-19  
197 susceptibility,<sup>6-8,10</sup> while the third signal on chromosome 20 was novel.

198           Amongst the 5 other signals no longer significant in this data release, the *APOE* [MIM:  
199 107741] variant tagging for the *APOE*- $\epsilon$ 4 haplotype (rs429358-C, MAF = 0.15, OR = 1.38 [1.24;  
200 1.53],  $P = 1.80 \times 10^{-9}$ , on 07.14.20) was the only signal previously described,<sup>20</sup> and was notably  
201 the first report of a genetic determinant for Covid-19 susceptibility. However, this previous  
202 report was based on UKB data and this signal was not replicated in an independent dataset.  
203 This association was greatly attenuated after the summer, when the number of Covid-19 cases  
204 started to rise significantly and the mean age of infected participants decreased. The  
205 interaction between age of participants and the *APOE* variant was significant ( $P = 3.4 \times 10^{-6}$ )  
206 suggesting that a subset of older participant carriers of this variant was more at risk of Covid-19  
207 infection. Remarkably, in the 02.24.21 data release, this variant is still suggestively associated  
208 with Covid-19 severity (OR = 1.42 [1.24; 1.62],  $P = 4.36 \times 10^{-7}$  in EUR) and death (OR = 1.45  
209 [1.25; 1.67],  $P = 4.62 \times 10^{-7}$  in EUR) suggesting a mechanism leading to lethal complications  
210 after infection.

211           Using negatively tested participants as controls (Tested controls), the only signal  
212 reaching genome-wide significance in this last data release was again the chr3p21.31 locus ( $P =$   
213  $1.85 \times 10^{-9}$ ) with a different lead variant (rs73062394-T, MAF = 0.059, OR = 1.2 [1.13; 1.28]).  
214 This variant was in moderate linkage disequilibrium (LD) with rs73062389 ( $r^2 = 0.59$ ), which was  
215 the second lead variant ( $P = 8.46 \times 10^{-9}$ ) for this analysis.

216           In other ancestry-stratified GWAS (AFR, SAS and OTHERS), no signal was found genome-  
217 wide significant in the last data release analyzed. However, in the GWAS combining all non-  
218 European participants, with Population controls, one signal was found significant at the *LRMDA*  
219 [MIM: 614537] locus on chromosome 10 (rs114026383-C, MAF = 0.02, OR = 2.42 [1.78; 3.29],  $P$

220 =  $1.82 \times 10^{-8}$ ). According to gnomAD (v2.1.1),<sup>21</sup> this variant is mostly carried by individuals of  
221 African ancestry (MAF = 0.04) and mainly absent in other ancestries. In the GWAS of African  
222 ancestry participants, this signal is close to genome wide significance ( $P = 3.24 \times 10^{-7}$ ).  
223 Interestingly, *LRMDA* variants have been found associated to lung function<sup>22</sup> [MIM: 608852]  
224 and HIV viral load in an unadjusted GWAS,<sup>23</sup> but there was no evidence of LD between these  
225 variants and rs114026383 ( $r^2 < 0.01$ ).

226 Finally, in the trans-ancestry analyses combining all UKB participants, using the  
227 Population controls, the signals at the chr3.p21.31 and *ABO* loci were also genome-wide  
228 significant (rs73062389-A,  $P = 3.99 \times 10^{-13}$  and rs9411378-A,  $P = 1.12 \times 10^{-9}$ , respectively), and a  
229 third signal was observed at the *MIR3681HG* locus on chromosome 2 (rs112938622-G, MAF =  
230 0.22, OR = 1.08 [1.05; 1.12],  $P = 4.95 \times 10^{-8}$ ). When using Tested controls, only the chr3p21.31  
231 signal was genome-wide significant in the last analysis (rs73062394-T,  $P = 1.99 \times 10^{-8}$ ).

232 In sex-stratified analyses, using Population controls, the chr3p21.31 signal was  
233 significant in women (rs73062389-A,  $P = 1.06 \times 10^{-8}$  in EUR and  $P = 1.35 \times 10^{-8}$  in the trans-  
234 ancestry GWAS) and highly associated in men ( $P = 1.12 \times 10^{-6}$  in EUR and  $P = 3.19 \times 10^{-6}$  in the  
235 trans-ancestry GWAS), whereas the *ABO* signal was significant in men (rs9411378-A,  $P = 2.00 \times$   
236  $10^{-8}$  in EUR and  $P = 2.05 \times 10^{-8}$  in the trans-ancestry GWAS) and moderately associated in  
237 women ( $P = 1.51 \times 10^{-3}$  in EUR and  $P = 1.89 \times 10^{-3}$  in the trans-ancestry GWAS). Notably, the  
238 observed effect of the *ABO* signal was significantly stronger in men ( $\beta = 0.135$ , SE = 0.023 in  
239 EUR) than in women ( $\beta = 0.068$ , SE = 0.021), using the Z-test for the equality of regression  
240 coefficients ( $P = 0.015$ ). In addition, when using Tested controls, an intergenic variant at the  
241 chr15q24.2 locus was significant in the trans-ancestry GWAS of women (rs71401691-G, MAF =

242 0.07, OR = 1.23 [1.14; 1.32],  $P = 4.72 \times 10^{-8}$ ), but not in men ( $P = 0.89$ ). According to GTeX, the  
243 effect allele increases the transcription levels of *MAN2C1* (Mannosidase Alpha Class 2C Member  
244 1, [MIM: 154580]) in several tissues including lungs ( $P = 6.7 \times 10^{-14}$ ) and esophagus-mucosa ( $P =$   
245  $3.4 \times 10^{-10}$ ).

246 Several signals were genome-wide significant in analyses using a specific set of controls,  
247 but were below the significance threshold when using another one. Notably, the *LRMDA* signal  
248 had a stronger effect in the Population analysis ( $\beta = 0.88$ , SE = 0.16) than in the Tested analysis  
249 ( $\beta = 0.62$ , SE = 0.16) and the intergenic variant at the chr15q24.2 locus had a stronger effect in  
250 the Tested analysis ( $\beta = 0.20$ , SE = 0.04) than in the Population analysis ( $\beta = 0.13$ , SE = 0.03), as  
251 represented in **Figure S19**. However, the difference in effects when using 2 different set of  
252 controls was not significant when applying the Z-test for the equality of regression coefficients  
253 ( $P > 0.05$ ).

254

255 Associations with Covid-19 hospitalization, severity, and death

256 The genome-wide significant findings for Covid-19 hospitalization, severity and death  
257 are presented in **Table 4**, while all associations with  $P < 10^{-5}$  are available in **Table S6**. Similarly  
258 to the susceptibility analyses, all signals reaching genome-wide significance in any analysis are  
259 presented in **Figures S20-S42**. Across all analyses of hospitalization and severity, we identified 4  
260 loci reaching genome-wide significance in the last data release analyzed, while no significant  
261 signal was observed in the GWAS of Covid-19 related death.

262 The most recurring signal is again located at the known chr3p21.31 locus. Depending on  
263 the analysis, we observed 3 distinct lead variants at this locus: rs13071258 in the GWAS of

264 hospitalized Europeans, rs72893671 in the trans-ancestry hospitalization GWAS, and  
265 rs13092887 in the severe Covid-19 GWAS of both Europeans and all ancestries. However, all 3  
266 variants were in LD ( $r^2 > 0.84$ ), suggesting a single haplotype modulating Covid-19  
267 hospitalization and severity. As previously reported,<sup>8</sup> this haplotype seems distinct from the  
268 signal of Covid-19 susceptibility (**Figure S43**), as the 2 susceptibility variants (rs73062389 and  
269 rs73062394) are not in LD ( $r^2 < 0.01$ ) with the 3 severity variants.

270 The 3 remaining signals were not previously reported, and were identified in sex-  
271 stratified analyses. First, intronic variants in *ANKS1A* [MIM: 608994] were observed significantly  
272 associated in several analyses of men with Covid-19-related hospitalization, for instance in EUR  
273 using non-hospitalized participants with a positive test as controls (rs112887370-T, MAF = 0.02,  
274 OR = 3.30 [2.18; 5.00],  $P = 1.71 \times 10^{-8}$ ). According to GTEx, this allele increases the transcript  
275 levels of the nearby gene *C6orf106* [MIM: 612217] in some tissues (thyroid,  $P = 1.5 \times 10^{-7}$  and  
276 testis,  $P = 1.7 \times 10^{-6}$ ), while other alleles in LD also reaching genome-wide significance (e.g.  
277 rs2504165-T, OR = 2.84 [1.96; 4.12],  $P = 3.99 \times 10^{-8}$ ,  $r^2 = 0.73$ ) can decrease the transcript levels  
278 of *C6orf106* in skeletal muscles ( $P = 3.2 \times 10^{-5}$ ). *C6orf106* is also known as *ILRUN* (Inflammation  
279 and Lipid Regulator with UBA-like and NBR-1-like domains), a gene involved in the regulation of  
280 antiviral response. Second, an intergenic indel located at the chr12p13.31 locus, between  
281 *CD163* [MIM: 605545] and *APOBEC1* [MIM: 600130], was associated with the hospitalization of  
282 women when using non-hospitalized participants with a positive test (rs370732090-ins.ATTAT,  
283 MAF = 0.14, OR = 0.61 [0.52; 0.73],  $P = 3.37 \times 10^{-8}$ ). Finally, an intronic insertion-deletion  
284 located in *ADGRE1* [MIM: 600493] was associated with Covid-19 severity in European women,

285 using participants without severe Covid-19 as controls (rs770180814-ins.GT, MAF = 0.22, OR =  
286 1.78 [1.45; 2.19],  $P = 4.51 \times 10^{-8}$ ).

287

288 Replication of novel genome-wide significant signals

289 The 7 novel signals identified were further investigated using the Covid-19hgi meta-  
290 analyses results, where the UKB was not included. Unfortunately, variants constituting the  
291 *LRMDA* and *APOBEC1* signals were absent from the meta-analyses, which prevented our effort  
292 to replicate these 2 signals. Out of the 5 remaining signals, 4 did not reach nominal significance  
293 in the replication datasets: at the *SCRT2* locus (rs749199237,  $P = 0.97$  in C2 EUR), at the  
294 *MIR3681HG* locus (rs112938622,  $P = 0.27$  in C2 ALL), at the chr15q24.2 locus (rs71401691,  $P =$   
295 0.82 in C2 ALL) and at the *ADGRE1* locus (rs770180814,  $P = 0.55$  in A2 EUR). Finally, at the  
296 *ANKS1A* locus, the lead variant identified in the analysis of hospitalized Europeans men did not  
297 replicate (rs112887370,  $P = 0.14$  in B2 EUR) but another variant in LD mentioned earlier was  
298 associated at nominal significance (rs2504165-T, OR = 1.13 [1.00; 1.28],  $P = 0.04$  in B2 EUR).  
299 Unfortunately, this association does not survive multiple-testing correction, with a threshold for  
300 significance set at  $P < 0.017$  for 6 one-sided tests. However, the *ANKS1A*, *ADGRE1* and  
301 chr15q24.2 signals were identified in sex-stratified analyses, and the possibility for replication in  
302 an independent sex-stratified dataset is not yet provided by Covid-19hgi.

303

304 Signals with suggestive significance trends

305 The significance trajectory of the most robust signals, at the chr3p21.31 and *ABO* loci,  
306 mostly increased after the surge of new cases following the 09.08.20 data release (**Figure 1**).

307 After this date, the chr3p21.31 signal increased at each of the following 8 data release and  
308 reached genome-wide significance in the 11.03.20 release, while the ABO signal increased 5  
309 times out of 8 and reached significance in the 01.04.21 release. In order to identify signals that  
310 may become significant in future releases, we extracted variants displaying similar positive  
311 trends in significance, meaning variants not yet genome-wide significant, but exhibiting an  
312 increase in significance since the 09.08.20 data release.

313 For each Covid-19 susceptibility GWAS, we extracted variants which had an increase in  
314 significance in at least 6 out of the 8 releases following the 09.08.20 data release, and reached  
315  $P < 10^{-5}$  in the last data release. For other Covid-19 phenotypes, considering that fewer data  
316 releases were available, we extracted variants which increased in significance in all consecutive  
317 releases, reaching the same significance threshold of  $P < 10^{-5}$  in the last data release. After  
318 excluding genome-wide significant signals, 2,409 variants involved in 3,607 associations variants  
319 with suggestive trends were identified across all GWAS (**Table S7**).

320 Next, we sought to identify whether these signals with suggestive trends reached at  
321 least nominal significance ( $P < 0.05$ ) in the Covid-19hgi meta-analyses. Only signals from the  
322 European and trans-ancestry GWAS were considered at this step, to match the corresponding  
323 Covid-19hgi meta-analyses. Signals with suggestive trends from Covid-19 susceptibility GWAS  
324 were sought for replication in the corresponding C2 (susceptibility) meta-analyses, while signals  
325 with trends from the hospitalization and severe Covid-19 GWAS were sought for replication in  
326 the B2 and A2 meta-analyses, respectively. In order to decrease the multiple-testing correction  
327 burden, we only extracted the lead variant at each locus, a locus being defined as a genomic  
328 region containing one or several variants with  $P < 10^{-5}$  separated by less than 1Mb. As a result,

329 329 lead variants involved in 378 associations were extracted (**Table S8**), of which 18 had  
330 concordant effects and reached nominal significance in the Covid-19-hgi meta-analyses (**Table**  
331 **5**). After applying the multiple-testing Bonferroni correction, with one-sided hypothesis, only 2  
332 signals passed the corrected significance threshold ( $0.05 * 2 / 378 = 2.65 \times 10^{-4}$ ), both in trans-  
333 ancestry analyses. The first signal had a positive significance trajectory in the hospitalization  
334 GWAS (**Figure S44**) and is located 5Kb upstream of *ATP5PO* [MIM: 600828], encoding a part of  
335 the ATP synthase complex. According to GTex, the effect allele (rs12482569-A, MAF = 0.17, OR  
336 = 1.22 [1.12; 1.33],  $P = 6.28 \times 10^{-6}$ ) increases the transcript levels of *ATP5PO* in several tissues  
337 including the lungs ( $P = 4.6 \times 10^{-18}$ ) and the esophagus-mucosa ( $P = 1.9 \times 10^{-11}$ ). The second  
338 signal (rs2011959-A, MAF = 0.37, OR = 1.34 [1.18; 1.53],  $P = 6.50 \times 10^{-6}$ ) had a positive trend in  
339 the GWAS of men with severe Covid-19 (**Figure S45**), and is located 200 bp downstream of *CCL3*  
340 [MIM: 182283], a pro-inflammatory cytokine, reported by numerous studies to have unusually  
341 high levels in patients with severe Covid-19.<sup>24-26</sup>

342

343

344 Discussion and conclusions

345

346 This project was initiated with the major aim to share results of Covid-19 host genetics  
347 analyses freely and rapidly on the GRASP portal, during a pandemic where new insights to  
348 improve patient care and develop better treatments were greatly needed. A few months after  
349 the first Covid-19 case was identified in the UK, we started to perform GWAS on each dataset  
350 released by UKB between May 2020 and February 2021, and thus far examined genetic signals

351 associated with Covid-19 phenotypes across 15 data releases. This unique context allowed us to  
352 track the evolution of genetic associations over time, an approach rarely applied. While some  
353 works have examined the statistical properties of phased and nested case-control studies,<sup>27,28</sup>  
354 few studies have proposed using statistical trajectories over time in genetic analyses.

355 As a first major observation, the majority of signals which were observed in the first  
356 stages of the project did not sustain over time. For instance, the first genome-wide significant  
357 association observed in the GWAS of Covid-19 susceptibility in Europeans with rs34338189 as  
358 lead variant is not even nominally significant in the last release ( $P = 0.29$ ). Another genome-  
359 wide significant signal was observed at the *APOE* locus, with a significance that increased in the  
360 first 4 data releases before decreasing continually in the subsequent releases. The lead variant  
361 at the *APOE* locus is coding in part for the *APOE*- $\epsilon 4$  haplotype, known to increase the risk of  
362 Alzheimer's disease [MIM: 607822], dementia, dyslipidemia and cardiovascular diseases [MIM:  
363 617347] and is speculated to cause inflammation and cytokine storms.<sup>29</sup> Notably, the  
364 suggestive association of the *APOE* signal with severe Covid-19 could support this proposition.  
365 The evolution of the *APOE* signal over time could also be due to an initial higher prevalence of  
366 Covid-19 in nursing homes,<sup>30</sup> where dementia patients were at higher risk of being infected and  
367 spreading the virus due to living arrangements, and a poor understanding of transmission  
368 dynamics and appropriate safety guidelines early in the pandemic. Overall, the evolution of  
369 these signals suggests a change in the composition of cases over time, such as the diminution of  
370 age and increase of positively tested women in the later data releases, as well as the  
371 introduction of variant SARS-CoV2 strains. This change was most significant after the summer  
372 when a surge in new Covid-19 infections occurred.

373           The most robust findings from our study are the association of the chr3p21.31 and *ABO*  
374 loci with Covid-19 susceptibility and a distinct signal at the chr3p21.31 locus associated with  
375 Hospitalization and severe Covid-19. These observations corroborate several previous reports,<sup>6–</sup>  
376 <sup>8,10</sup> although we also observed a significant difference in the effect of the *ABO* variant between  
377 men and women. Sex-stratified analyses also allowed us to identify a novel association between  
378 *ANKS1A* variants and hospitalization of men, which reached nominal significance in the Covid-  
379 19hgi hospitalization meta-analysis. Interestingly, the variants comprising this signal can also  
380 influence the expression of *ILRUN*, either as an upregulator or as a repressor, depending on the  
381 tissue. This gene is involved in innate immunity and has been recently shown to act as an  
382 antiviral factor in the context of SARS-CoV-2 infection, notably by downregulating the  
383 expression of *ACE2* [MIM: 300335] and *TMPRSS2* [MIM: 602060], the main SARS-CoV-2 entry  
384 receptors.<sup>31</sup> This role differs from a previous report demonstrating an antiviral effect in cells  
385 infected with Hendra virus,<sup>32</sup> which could hint at diverse roles for this gene in the context of  
386 viral infection.

387           We have developed an original strategy to identify associations displaying increased  
388 significance over time, and identified hundreds of loci with a significance trajectory suggesting  
389 potential future genome-wide significance. Amongst these signals, 2 of them replicated at  
390 nominal p-value in the Covid-19hgi meta-analyses, including a variant downstream of *CCL3*  
391 suggestively associated with Severe Covid-19 in women, a cytokine with a well established high  
392 expression in severe Covid-19 cases.<sup>24–26</sup> The associations we observed changing through the  
393 pandemic could reflect random effects or changes in statistical power, but some of the results  
394 suggest changes due to potential gene-environment interactions such as age, underlying health

395 conditions (*APOE*) or sex makeup of cases exposed to or engaging in risk behavior. This  
396 indicates the general approach of iterative analysis and trends analyses for genetics during  
397 pandemics may have benefits in uncovering pathophysiologic clues. Additionally, other factors  
398 like predominant virus strains and changing treatment strategies through a pandemic might  
399 interact with host genetics, and be better understood by iterative analyses.

400 In summary, our host genomic analyses of Covid-19 have improved the comprehension  
401 of mechanisms involved in the infection and complications due to Covid-19 and we continue to  
402 perform and rapidly share these analyses with the research community. Our study has some  
403 limitations. Most importantly, our data and the work of others support large health disparities  
404 between EUR and non-EUR individuals related to COVID-19 throughout the ongoing pandemic.  
405 Despite an over-representation proportionally among cases and those with severe and fatal  
406 outcomes, the non-EUR component of UKB is a proportionally small sample limiting our  
407 statistical power to address population-specific genetic variants contributing to health  
408 outcomes. Moving forward we feel that having a diverse set of results with different phenotype  
409 definitions, sex-specific, ancestry-specific, and including external group summary statistics, all in  
410 a common genome reference and annotation framework may maximize the chance for new  
411 studies to cross-replicate or meta-analyze results as Covid-19 genetic studies continue to grow.

412

#### 413 Description of Supplemental Data

414 Supplemental Data include 45 figures and 8 tables.

415

#### 416 Declaration of Interests

417 The authors declare no competing interests.

418

419 Acknowledgments

420 All authors were supported by NIH Intramural Research Program funds. The views expressed in  
421 this manuscript are those of the authors and do not necessarily represent the views of the  
422 National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S.  
423 Department of Health and Human Services. This research has been conducted using the UK  
424 Biobank Resource under Application Number 28525. UK Biobank was established by the  
425 Wellcome Trust, Medical Research Council, Department of Health, Scottish government, and  
426 Northwest Regional Development Agency. It has also had funding from the Welsh assembly  
427 government and the British Heart Foundation. All UKB analyses for this manuscript were  
428 conducted on the NIH Biowulf high performance computing cluster (<https://hpc.nih.gov/>). The  
429 Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office  
430 of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and  
431 NINDS. We also thank the NHLBI IT team for their help in keeping the GRASP portal up to date.

432

433 Web Resources

434 GTEx : <https://gtexportal.org/home/>

435 CADD : <https://cadd.gs.washington.edu/>

436 EBI GWAS catalog : <https://www.ebi.ac.uk/gwas/>

437 GRASP catalog : <https://grasp.nhlbi.nih.gov/Overview.aspx>

438 Covid-19hgi meta-analyses results : <https://www.covid19hg.org/results/r5/>

439 GnomAD : <https://gnomad.broadinstitute.org/>

440

441 Data and Code Availability

442 The datasets generated during this study are available at the GRASP Covid-19 portal:

443 <https://grasp.nhlbi.nih.gov/Covid19GWASResults.aspx>

444

445 References

446 1. Connors, J.M., and Levy, J.H. (2020). COVID-19 and its implications for thrombosis and  
447 anticoagulation. *Blood* 135, 2033–2040.

448 2. Tang, D., Comish, P., and Kang, R. (2020). The hallmarks of COVID-19 disease. *PLoS Pathog.*  
449 16, e1008536.

450 3. Kwok, A.J., Mentzer, A., and Knight, J.C. (2021). Host genetics and infectious disease: new  
451 tools, insights and translational opportunities. *Nat. Rev. Genet.* 22, 137–153.

452 4. Dragic, T., Litwin, V., Allaway, G.P., Martin, S.R., Huang, Y., Nagashima, K.A., Cayanan, C.,  
453 Maddon, P.J., Koup, R.A., Moore, J.P., et al. (1996). HIV-1 entry into CD4+ cells is mediated by  
454 the chemokine receptor CC-CKR-5. *Nature* 381, 667–673.

455 5. Ciancanelli, M.J., Huang, S.X.L., Luthra, P., Garner, H., Itan, Y., Volpi, S., Lafaille, F.G., Trouillet,  
456 C., Schmolke, M., Albrecht, R.A., et al. (2015). Life-threatening influenza and impaired  
457 interferon amplification in human IRF7 deficiency. *Science* 348, 448–453.

458 6. Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., Fernández, J.,  
459 Prati, D., Baselli, G., Asselta, R., et al. (2020). Genomewide Association Study of Severe Covid-19  
460 with Respiratory Failure. *N. Engl. J. Med.* 0, null.

461 7. Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A.D., Rawlik, K., Pasko, D., Walker, S.,  
462 Parkinson, N., Fourman, M.H., Russell, C.D., et al. (2020). Genetic mechanisms of critical illness  
463 in Covid-19. *Nature* 1–1.

464 8. Shelton, J.F., Shastri, A.J., Ye, C., Weldon, C.H., Filshtein-Sonmez, T., Coker, D., Symons, A.,  
465 Esparza-Gordillo, J., Aslibekyan, S., and Auton, A. (2021). Trans-ancestry analysis reveals genetic  
466 and nongenetic associations with COVID-19 susceptibility and severity. *Nat. Genet.* 1–8.

- 467 9. Kosmicki, J.A., Horowitz, J.E., Banerjee, N., Lanche, R., Marcketta, A., Maxwell, E., Bai, X., Sun,  
468 D., Backman, J.D., Sharma, D., et al. (2021). Pan-ancestry exome-wide association analyses of  
469 COVID-19 outcomes in 586,157 individuals. *Am. J. Hum. Genet.* *0*.
- 470 10. Initiative, T.C.-19 H.G., and Ganna, A. (2021). Mapping the human genetic architecture of  
471 COVID-19 by worldwide meta-analysis. *MedRxiv* 2021.03.10.21252820.
- 472 11. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P.,  
473 Green, J., Landray, M., et al. (2015). UK Biobank: An Open Access Resource for Identifying the  
474 Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med.* *12*, e1001779.
- 475 12. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic,  
476 D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping  
477 and genomic data. *Nature* *562*, 203–209.
- 478 13. Zhou, W., Nielsen, J.B., Fritsche, L.G., Dey, R., Gabrielsen, M.E., Wolford, B.N., LeFaive, J.,  
479 VandeHaar, P., Gagliano, S.A., Gifford, A., et al. (2018). Efficiently controlling for case-control  
480 imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* *50*,  
481 1335–1341.
- 482 14. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic  
483 variants from high-throughput sequencing data. *Nucleic Acids Res.* *38*, e164–e164.
- 484 15. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD:  
485 predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.*  
486 *47*, D886–D894.
- 487 16. Eicher, J.D., Landowski, C., Stackhouse, B., Sloan, A., Chen, W., Jensen, N., Lien, J.-P., Leslie,  
488 R., and Johnson, A.D. (2015). GRASP v2.0: an update on the Genome-Wide Repository of  
489 Associations between SNPs and phenotypes. *Nucleic Acids Res.* *43*, D799-804.
- 490 17. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C.,  
491 McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of  
492 published genome-wide association studies, targeted arrays and summary statistics 2019.  
493 *Nucleic Acids Res.* *47*, D1005–D1012.
- 494 18. Aguet, F., Brown, A.A., Castel, S.E., Davis, J.R., He, Y., Jo, B., Mohammadi, P., Park, Y.,  
495 Parsana, P., Segrè, A.V., et al. (2017). Genetic effects on gene expression across human tissues.  
496 *Nature* *550*, 204–213.
- 497 19. Zhang, X., Gierman, H.J., Levy, D., Plump, A., Dobrin, R., Goring, H.H.H., Curran, J.E.,  
498 Johnson, M.P., Blangero, J., Kim, S.K., et al. (2014). Synthesis of 53 tissue and cell line  
499 expression QTL datasets reveals master eQTLs. *BMC Genomics* *15*, 532.

- 500 20. Kuo, C.-L., Pilling, L.C., Atkins, J.L., Masoli, J.A.H., Delgado, J., Kuchel, G.A., and Melzer, D.  
501 (2020). ApoE e4e4 Genotype and Mortality With COVID-19 in UK Biobank. *J. Gerontol. A. Biol.*  
502 *Sci. Med. Sci.* *75*, 1801–1803.
- 503 21. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L.,  
504 Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum  
505 quantified from variation in 141,456 humans. *Nature* *581*, 434–443.
- 506 22. Soler Artigas, M., Loth, D.W., Wain, L.V., Gharib, S.A., Obeidat, M., Tang, W., Zhai, G., Zhao,  
507 J.H., Smith, A.V., Huffman, J.E., et al. (2011). Genome-wide association and large-scale follow up  
508 identifies 16 new loci influencing lung function. *Nat. Genet.* *43*, 1082–1090.
- 509 23. Ekenberg, C., Tang, M.-H., Zucco, A.G., Murray, D.D., MacPherson, C.R., Hu, X., Sherman,  
510 B.T., Losso, M.H., Wood, R., Paredes, R., et al. (2019). Association Between Single-Nucleotide  
511 Polymorphisms in HLA Alleles and Human Immunodeficiency Virus Type 1 Viral Load in  
512 Demographically Diverse, Antiretroviral Therapy-Naive Participants From the Strategic Timing of  
513 AntiRetroviral Treatment Trial. *J. Infect. Dis.* *220*, 1325–1334.
- 514 24. Chua, R.L., Lukassen, S., Trump, S., Hennig, B.P., Wendisch, D., Pott, F., Debnath, O.,  
515 Thürmann, L., Kurth, F., Völker, M.T., et al. (2020). COVID-19 severity correlates with airway  
516 epithelium-immune cell interactions identified by single-cell analysis. *Nat. Biotechnol.* *38*, 970–  
517 979.
- 518 25. Hue, S., Beldi-Ferchiou, A., Bendib, I., Surenaud, M., Fourati, S., Frapard, T., Rivoal, S.,  
519 Razazi, K., Carteaux, G., Delfau-Larue, M.-H., et al. (2020). Uncontrolled Innate and Impaired  
520 Adaptive Immune Responses in Patients with COVID-19 Acute Respiratory Distress Syndrome.  
521 *Am. J. Respir. Crit. Care Med.* *202*, 1509–1519.
- 522 26. Trump, S., Lukassen, S., Anker, M.S., Chua, R.L., Liebig, J., Thürmann, L., Corman, V.M.,  
523 Binder, M., Loske, J., Klasa, C., et al. (2020). Hypertension delays viral clearance and exacerbates  
524 airway hyperinflammation in patients with COVID-19. *Nat. Biotechnol.*
- 525 27. Van der Tweel, I., and van Noord, P.A. (2000). Sequential analysis of matched dichotomous  
526 data from prospective case-control studies. *Stat. Med.* *19*, 3449–3464.
- 527 28. Boessen, R., van der Baan, F., Groenwold, R., Egberts, A., Klungel, O., Grobbee, D., Knol, M.,  
528 and Roes, K. (2013). Optimizing trial design in pharmacogenetics research: comparing a fixed  
529 parallel group, group sequential, and adaptive selection design on sample size requirements.  
530 *Pharm. Stat.* *12*, 366–374.
- 531 29. Numbers, K., and Brodaty, H. (2021). The effects of the COVID-19 pandemic on people with  
532 dementia. *Nat. Rev. Neurol.* *17*, 69–70.
- 533 30. Chen, M.K., Chevalier, J.A., and Long, E.F. (2021). Nursing home staff networks and COVID-  
534 19. *Proc. Natl. Acad. Sci. U. S. A.* *118*,.

535 31. Tribolet, L., Alexander, M.R., Brice, A.M., van Vuren, P.J., Rootes, C.L., Mara, K., McDonald,  
536 M., Bruce, K.L., Gough, T.J., Shi, S., et al. (2021). ILRUN downregulates ACE2 expression and  
537 blocks infection of human cells by SARS-CoV-2. *J. Virol.*

538 32. Deffrasnes, C., Marsh, G.A., Foo, C.H., Rootes, C.L., Gould, C.M., Grusovin, J., Monaghan, P.,  
539 Lo, M.K., Tompkins, S.M., Adams, T.E., et al. (2016). Genome-wide siRNA Screening at Biosafety  
540 Level 4 Reveals a Crucial Role for Fibrillarlin in Henipavirus Infection. *PLoS Pathog.* *12*, e1005478.

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

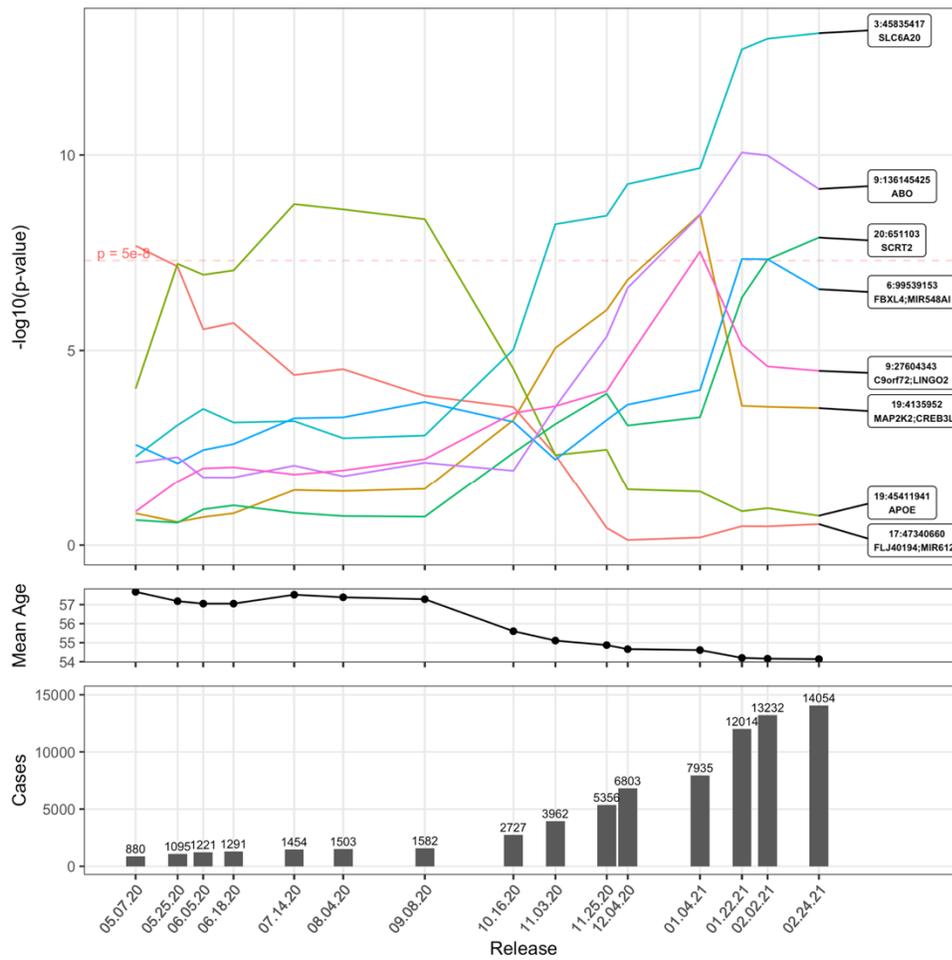
562

563 Figure Titles and Legends

564

565 **Figure 1.** Evolution of significant signals associated with covid-19 susceptibility in UKB

566 participants of European ancestry, using Population as controls.



567

568 The top panel represents the evolution in significance of signals reaching genome wide significance at least  
 569 once across all data releases analyzed in Europeans for Covid-19 susceptibility. The middle panel represent  
 570 the mean age of cases in each data release. The bottom panel represent the number of cases in each data  
 571 release.

572

573

574 Tables

575

576 **Table 1.** Sample sizes for each analysis, as of 02.24.21

577

Phenotype	Controls	Ancestry	Sex-stratified	Cases	Controls	Female Cases	Female Controls
Susceptibility	Test negative OR not tested (Population)	ALL	yes	15,738	471,582	8,258	255,971
		EUR	yes	14,054	445,196	7,396	241,903
		AFR		496	7,148	277	4,082
		SAS		693	8,724	300	4,031
		OTHERS		495	10,514	285	5,955
		nEUR		1,684	26,386	862	14,068
	Test negative (Tested)	ALL	yes	15,738	47,413	8,258	25,373
		EUR	yes	14,054	44,646	7,396	23,847
		AFR		496	809	277	501
		SAS		693	938	300	430
OTHERS			495	1,020	285	595	
	nEUR		1,684	2,767	862	1,526	
Hospitalization	Not hospitalized (Population)	ALL	yes	1,916	485,404	735	263,494
		EUR	yes	1,712	457,538	655	248,644
		nEUR		204	27,866	80	14,850
	Tested, not hospitalized (Tested)	ALL	yes	1,916	61,235	735	32,896
		EUR	yes	1,712	56,988	655	30,588
		nEUR		204	4,247	80	2,308
	Test positive, not hospitalized (Positive)	ALL	yes	1,916	13,822	735	7,523
		EUR	yes	1,712	12,342	655	6,741
		nEUR		204	1,480	80	782
Severe Covid-19	Not severe (Population)	ALL	yes	935	486,385	320	263,909
		EUR	yes	848	458,402	283	249,016
	Tested, not severe (Tested)	ALL	yes	935	62,216	320	33,311
		EUR	yes	848	57,852	283	30,960
	Test positive, not severe (Positive)	ALL	yes	935	14,803	320	7,938
		EUR	yes	848	13,206	283	7,113
Death	Survivor (Population)	ALL		828	486,492	287	263,942
		EUR		758	458,492	258	249,041
	Tested, survivor (Tested)	ALL		828	62,323	287	33,344
		EUR		758	57,942	258	30,985
	Test positive, survivor (Positive)	ALL		828	14,910	287	7,971
		EUR		758	13,296	258	7,138

578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593

**Table 2.** Composition of the consecutive UK Biobank Covid-19 data releases

Release Date	Cases (tested positive)	Controls	% Tested Positive Females	Age Cases <sup>a</sup>
05.07.20 <sup>b</sup>	1,029	486,291	45.29	57.02 (9.12)
05.25.20 <sup>b</sup>	1,270	486,050	46.93	56.57 (9.25)
06.05.20	1,412	485,908	47.31	56.45 (9.23)
06.18.20	1,485	485,835	47.21	56.47 (9.19)
07.14.20	1,670	485,650	46.59	56.89 (9.14)
08.04.20	1,723	485,597	46.66	56.77 (9.11)
09.08.20	1,821	485,499	46.73	56.66 (9.10)
10.16.20	3,050	484,270	48.69	55.19 (9.09)
11.03.20	4,372	482,948	49.38	54.79 (8.95)
11.25.20	5,868	481,452	49.52	54.56 (8.88)
12.04.20	7,435	479,885	50.46	54.36 (8.80)
01.04.21	8,722	478,598	50.50	54.28 (8.79)
01.22.21	13,401	473,919	52.27	53.85 (8.66)
02.02.21	14,802	472,518	52.44	53.81 (8.64)
02.24.21	15,738	471,582	52.47	53.77 (8.63)

594  
595  
596  
597  
598  
599  
600

<sup>a</sup> Mean baseline Age (standard deviation), at recruitment

<sup>b</sup> For the first 2 releases, the numbers reflect all ancestries, but only cases and controls of European ancestry were considered for analyses

601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616

**Table 3.** Lead variants associated with Covid-19 susceptibility (on 02.24.21)

Ancestry:Controls(:Sex) <sup>a</sup>	CHR:POS:NEA:EA	rsid	EAF	OR	±95CI	P (02.24.21)	Locus
ALL:Population	2:12594718:A:G	rs112938622	0.220	0.92	[0.90; 0.95]	4.95E-08	<i>MIR3681HG</i>
ALL:Population	3:45835417:G:A	rs73062389	0.058	1.21	[1.15; 1.27]	3.99E-13	<i>SLC6A20</i>
ALL:Population:F	3:45835417:G:A	rs73062389	0.058	1.22	[1.14; 1.31]	1.35E-08	<i>SLC6A20</i>
EUR:Population	3:45835417:G:A	rs73062389	0.060	1.22	[1.16; 1.29]	7.64E-14	<i>SLC6A20</i>
EUR:Population:F	3:45835417:G:A	rs73062389	0.060	1.23	[1.15; 1.32]	1.06E-08	<i>SLC6A20</i>
ALL:Tested	3:45839176:A:T	rs73062394	0.057	1.19	[1.12; 1.26]	1.99E-08	<i>SLC6A20;LZTFL1</i>
EUR:Tested	3:45839176:A:T	rs73062394	0.059	1.20	[1.13; 1.28]	1.85E-09	<i>SLC6A20;LZTFL1</i>
ALL:Population	9:136145425:C:A	rs9411378	0.219	1.10	[1.06; 1.13]	1.12E-09	<i>ABO</i>
ALL:Population:M	9:136145425:C:A	rs9411378	0.219	1.14	[1.09; 1.19]	2.00E-09	<i>ABO</i>
EUR:Population	9:136145425:C:A	rs9411378	0.222	1.10	[1.07; 1.14]	7.36E-10	<i>ABO</i>
EUR:Population:M	9:136145425:C:A	rs9411378	0.222	1.14	[1.09; 1.20]	2.05E-09	<i>ABO</i>
nEUR:Population	10:78250184:T:C	rs114026383	0.016	2.42	[1.78; 3.29]	1.82E-08	<i>LRMDA</i>
ALL:Tested:F	15:75602162:A:G	rs71401691	0.069	1.23	[1.14; 1.32]	4.72E-08	<i>GOLGA6D;COMMD4</i>
EUR:Population	20:651103:CG:C	rs749199237	0.350	1.08	[1.05; 1.11]	1.29E-08	<i>SCRT2</i>

617  
618 <sup>a</sup> Indicate the ancestry by label, the control set, and, when it is the case, the corresponding sex-stratified analyses (F: Females,  
619 M: Males)  
620 CHR: Chromosome; POS: Position (hg19 genome build); NEA: Non Effect Allele; EA: Effect Allele; EAF: Effect Allele Frequency;  
621 OR: Odd-ratio; ±95CI: 95% confidence interval [lower bound; upper bound]; P: P-value (from the last data release, analyzed on  
622 02.24.21)

623  
624  
625  
626  
627  
628

629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644

**Table 4.** Lead variants associated with Covid-19 hospitalization, severity, or death (on 02.24.21)

Phenotype	Ancestry:Controls(:Sex) <sup>a</sup>	CHR:POS:NEA:EA	rsid	EAF	OR	±95CI	P (02.24.21)	Locus
Hospitalization	EUR:Population	3:45843242:G:A	rs13071258	0.071	1.47	[1.29; 1.69]	3.18E-08	<i>SLC6A20;LZTFL1</i>
Hospitalization	EUR:Tested	3:45843242:G:A	rs13071258	0.071	1.49	[1.29; 1.71]	2.03E-08	<i>SLC6A20;LZTFL1</i>
Hospitalization	ALL:Population	3:45850783:T:A	rs72893671	0.081	1.43	[1.27; 1.61]	6.04E-09	<i>SLC6A20;LZTFL1</i>
Hospitalization	ALL:Tested	3:45850783:T:A	rs72893671	0.083	1.45	[1.29; 1.64]	1.99E-09	<i>SLC6A20;LZTFL1</i>
Severe	ALL:Population	3:45909644:C:A	rs13092887	0.075	1.72	[1.43; 2.07]	1.01E-08	<i>LZTFL1</i>
Severe	ALL:Tested	3:45909644:C:A	rs13092887	0.076	1.72	[1.43; 2.08]	1.04E-08	<i>LZTFL1</i>
Severe	EUR:Population	3:45909644:C:A	rs13092887	0.076	1.73	[1.43; 2.10]	2.38E-08	<i>LZTFL1</i>
Severe	EUR:Tested	3:45909644:C:A	rs13092887	0.076	1.73	[1.42; 2.09]	2.72E-08	<i>LZTFL1</i>
Hospitalization	ALL:Positive:M	6:34866730:G:C	rs112018241	0.015	3.10	[2.10; 4.60]	1.56E-08	<i>ANKS1A</i>
Hospitalization	ALL:Tested:M	6:35042053:C:T	rs112887370	0.015	2.89	[1.98; 4.21]	3.35E-08	<i>ANKS1A</i>
Hospitalization	EUR:Positive:M	6:35042053:C:T	rs112887370	0.015	3.30	[2.18; 5.00]	1.71E-08	<i>ANKS1A</i>
Hospitalization	EUR:Tested:M	6:35042053:C:T	rs112887370	0.014	3.16	[2.11; 4.76]	3.19E-08	<i>ANKS1A</i>
Hospitalization	EUR:Positive:F	12:7800075:A:ATTAT	rs370732090	0.857	0.61	[0.52; 0.73]	3.37E-08	<i>CD163;APOBEC1</i>
Severe	EUR:Population:F	19:6921713:G:GT	rs770180814	0.223	1.78	[1.45; 2.19]	4.51E-08	<i>ADGRE1</i>

645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660

<sup>a</sup> Indicate the ancestry by label, the control set, and, when it is the case, the corresponding sex-stratified analyses (F: Females, M: Males)  
CHR: Chromosome; POS: Position (hg19 genome build); NEA: Non Effect Allele; EA: Effect Allele; EAF: Effect Allele Frequency;  
OR: Odd-ratio; ±95CI: ±95% Confidence interval [lower bound; upper bound]; P: P-value (from the last data release, analyzed on 02.24.21)

661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678

**Table 5. Suggestive trends with  $p < 0.05$  in Covid-19hgi meta-analyses**

CHR:POS:NEA:EA	rsid	EAF	UK Biobank GWAS			Covid-19hgi Meta-analysis			Locus
			Phenotype:Ancestry:Controls:(Sex) <sup>a</sup>	OR.022421	P.022421	Meta-analysis	OR	P	
21:35293087:G:A	rs12482569	0.171	Hospitalization:ALL:Population	1.22	6.28E-06	B2_ALL	1.09	1.39E-04	<i>ATP5PO;</i> <i>LINC00649</i>
17:34415356:C:A	rs2011959	0.374	Severe:ALL:Positive:M	1.34	6.50E-06	A2_ALL	1.14	2.63E-04	<i>CCL3</i>
10:134354639:T:C	rs4880413	0.623	Hospitalization:ALL:Positive:M	1.25	7.05E-06	B2_ALL	1.05	0.018	<i>INPP5A</i>
1:63495201:T:C	rs79265284	0.022	Hospitalization:ALL:Positive:M	2.08	9.80E-06	B2_ALL	1.17	0.021	<i>LINC01739;</i> <i>LINC00466</i>
21:34156633:T:C	rs12627142	0.270	Susceptibility:EUR:Tested:M	1.11	5.92E-06	C2_EUR	1.03	0.024	<i>C21orf62-AS1</i>
8:138805432:G:A	rs6577840	0.899	Severe:ALL:Positive:M	1.74	1.10E-07	A2_ALL	1.12	0.025	<i>LOC101927915;</i> <i>LOC401478</i>
8:138803782:T:C	rs7835023	0.904	Severe:EUR:Positive:M	1.72	9.10E-07	A2_EUR	1.15	0.032	<i>LOC101927915;</i> <i>LOC401478</i>
10:134354639:T:C	rs4880413	0.624	Hospitalization:EUR:Positive:M	1.27	2.96E-06	B2_EUR	1.05	0.032	<i>INPP5A</i>
14:34041605:T:C	rs143195932	0.013	Hospitalization:EUR:Positive:M	3.41	6.48E-06	B2_EUR	1.29	0.032	<i>NPAS3</i>
10:37877569:T:C	rs4934590	0.293	Severe:EUR:Population:M	1.38	1.34E-06	A2_EUR	1.10	0.042	<i>LINC00993;</i> <i>MTRNR2L7</i>
5:17189240:G:A	rs188950364	0.019	Severe:ALL:Tested:M	3.21	3.57E-07	A2_ALL	1.22	0.042	<i>BASP1-AS1</i>
9:90051212:T:C	rs17052840	0.086	Severe:EUR:Population:F	2.05	5.72E-06	A2_EUR	1.12	0.043	<i>LINC02872;</i> <i>DAPK1</i>
21:34156633:T:C	rs12627142	0.270	Susceptibility:ALL:Tested:M	1.11	3.26E-06	C2_ALL	1.02	0.044	<i>C21orf62-AS1</i>
1:203193134:T:G	rs2486068	0.185	Severe:ALL:Population:M	1.42	4.01E-06	A2_ALL	1.09	0.046	<i>CHIT1</i>
15:101940428:G:A	rs1552946	0.732	Severe:EUR:Tested:F	0.64	5.64E-06	A2_EUR	0.92	0.048	<i>PCSK6</i>
22:28624425:T:C	rs28701638	0.160	Severe:EUR:Positive	0.73	9.63E-06	A2_EUR	0.92	0.048	<i>TTC28</i>
5:92032219:G:A	rs17082483	0.090	Severe:ALL:Positive:M	1.77	4.26E-07	A2_ALL	1.09	0.048	<i>ARRDC3-AS1;</i> <i>NR2F1-AS1</i>
4:27510227:A:T	rs4377577	0.164	Susceptibility:ALL:Population:F	1.10	8.86E-06	C2_ALL	1.03	0.048	<i>LINC02261;</i> <i>MIR4275</i>

679  
680  
681  
682  
683  
684

<sup>a</sup> Indicate the phenotype, the ancestry, the control set, and, when it is the case, the corresponding sex-stratified analyses (F: Females, M: Males)  
CHR: Chromosome; POS: Position (hg19 genome build); NEA: Non Effect Allele; EA: Effect Allele; EAF: Effect Allele Frequency; P: P-value