

31 **SUMMARY**

32

33 Long COVID presents a complex and multi-systemic disease that poses a significant
34 global public health challenge. Symptoms can vary widely, ranging from asymptomatic
35 to severe, making the condition challenging to diagnose and manage effectively.
36 Furthermore, identifying appropriate phenotypes in genome-wide association studies
37 of COVID-19 remains unresolved. This study aimed to address these challenges by
38 analyzing 220 deep-phenotype genome-wide association data sets (159 diseases, 38
39 biomarkers and 23 medication usage) from BioBank Japan (BBJ) (n=179,000), UK
40 Biobank and FinnGen (n=628,000) to investigate pleiotropic effects of known COVID-
41 19 risk associated single nucleotide variants. Our findings reveal 32 different
42 phenotypes that share the common genetic risk factors with COVID-19 ($p < 7.6 \times 10^{-11}$),
43 including two diseases (myocardial infarction and type 2 diabetes), 26 biomarkers with
44 seven categories (blood cell, metabolic, liver-related, kidney-related, protein,
45 inflammatory and anthropometric), and four medications (antithrombotic agents, HMG-
46 CoA reductase inhibitors, thyroid preparations and anilides). As long COVID continues
47 to coexist with humans, our results highlight the need for targeted screening to support
48 specific vulnerable populations to improve disease prevention and healthcare delivery.

49 **Keywords:** Long COVID, SARS-CoV-2, SNP, genetic risk, phenotype

50

51

52

53

54

55

56

57

58

59

60 Long COVID, also known as “post-acute sequelae of COVID-19”, has been identified
61 as the next public health concern. This debilitating illness is estimated to affect more
62 than 200 symptoms (1), and may occur in about 12.5% (1/8) people with COVID-19 in
63 the general population (2). The current diagnostic and treatment options are inadequate,
64 and the underlying biological mechanisms need to be further addressed (1). Through
65 genome-wide association studies (GWASs), more than 40 single nucleotide variants
66 (SNPs) involving 35 genes have been identified as potential contributors to
67 susceptibility to SARS-CoV-2 infection or risk of severe COVID-19 disease (3). While
68 translational studies of GWAS findings may provide novel insights into
69 pathophysiology, there remains limited understanding of the pleiotropic effects between
70 genetic predisposition, complications, biomarkers and medications for guiding long-
71 term follow-up and life care of COVID-19 patients.

72 Phenome-wide association studies (PheWASs) based on population have opened
73 up new avenues for formulating effective biomedical hypotheses. In this study, we
74 delved into the pleiotropic loci of known single nucleotide variants (SNPs) (3) that
75 represent either susceptibility to SARS-CoV-2 infection or risk of severe COVID-19
76 disease. To do this, we analyzed 220 deep-phenotype genome-wide association data
77 sets with deep-phenotype information, comprising 159 diseases, 38 biomarkers and 23
78 medication usage from BioBank Japan (BBJ) (n=179,000), UK Biobank and FinnGen
79 (n_{total}=628,000) (4).

80 First, using a comprehensive PheWAS analysis of the genome-wide landscape, we
81 have identified two genetic associations with myocardial infarction and type 2 diabetes
82 that indicate a potential link to COVID-19 infection. Our diagnosis code-focused
83 PheWAS (Figure 1) revealed a significant association between rs912805253 at the
84 *OAS1* locus and myocardial infarction (MI) (ICD-10 code I21/I22/I23/I24/I25) in both
85 the BBJ and European populations [BBJ population: odds ratio (OR) = 1.09, 95%
86 confidence interval (CI) = 1.06-1.12, $p = 3.3 \times 10^{-11}$; European population: OR = 1.04,
87 95% CI = 1.02-1.07, $p = 3.4 \times 10^{-4}$].

88 When combining subjects both from BBJ and European population, the OR for
89 rs912805253 with MI was 1.07 (95% CI = 1.05-1.08, $p = 1.3 \times 10^{-12}$). Furthermore,

90 rs912805253 was also significantly associated with type 2 diabetes (T2D) (ICD-10
91 code E11) in both populations [BBJ population: odds ratio (OR) = 1.04, 95%
92 confidence interval (CI) = 1.02-1.06, $p = 3.7 \times 10^{-5}$; European population: OR = 1.05,
93 95% CI = 1.03-1.07, $p = 2.4 \times 10^{-8}$]. When combining subjects from both population, the
94 OR for rs912805253 with T2D was 1.05 (95% CI = 1.03-1.06, $p = 5.6 \times 10^{-12}$). Clinical
95 studies (cohorts) have suggested and increased risk of long-term cardiovascular
96 outcomes (5) and type 2 diabetes (6) after COVID-19 infection, but their genetic risk
97 factors are still unclear. Our findings shed light on the potential genetic link between
98 COVID-19 infection and these diseases, providing a valuable contribution to this field.

99 Second, our findings provide compelling evidence of shared genetic risk factors
100 across 26 clinical lab-measurement, particularly among seven blood cell biomarkers
101 (neutrophil, monocyte, white blood cell, eosinophil, hemoglobin, hematocrit and red
102 blood cell), two metabolic biomarkers (total cholesterol and triglycerides), one liver-
103 related biomarker (aspartate transaminase) and one protein biomarker (total protein).
104 Based on our analysis using the clinical lab-measurement PheWAS, we identified 86
105 significant associations between 12 SNPs and 26 biomarkers (7.6×10^{-11}) (Figure 1).
106 Among these biomarkers, blood cell measurements emerged as one of the top
107 associated categories, followed by metabolic, liver-related, kidney-related, protein,
108 inflammatory, anthropometric. Within the blood cell type category, neutrophil (NEU)
109 (rs912805253, $p = 4.9 \times 10^{-18}$; rs879055593, $p = 1.9 \times 10^{-29}$; rs9411378, $p = 1.1 \times 10^{-22}$;
110 rs1819040, $p = 5.0 \times 10^{-16}$; rs11540822, $p = 3.1 \times 10^{-19}$; rs143334143, $p = 8.7 \times 10^{-20}$),
111 monocyte (MON) (rs912805253, $p = 8.2 \times 10^{-69}$; rs879055593, $p = 9.7 \times 10^{-50}$;
112 rs9411378, $p = 1.3 \times 10^{-45}$; rs17763537, $p = 7.0 \times 10^{-15}$; rs73064425, $p = 2.2 \times 10^{-16}$;
113 rs10490770, $p = 3.8 \times 10^{-15}$), and white blood cell (WBC) (rs912805253, $p = 1.8 \times 10^{-28}$;
114 rs879055593, $p = 9.4 \times 10^{-35}$; rs9411378, $p = 9.7 \times 10^{-24}$; rs11540822, $p = 1.2 \times 10^{-45}$;
115 rs143334143, $p = 1.2 \times 10^{-46}$) were associated with five or more COVID-19 risk
116 associated SNPs. Among metabolic biomarkers, total cholesterol (TC) (rs912805253,
117 $p = 1.1 \times 10^{-47}$; rs879055593, $p = 4.5 \times 10^{-86}$; rs9411378, $p = 3.2 \times 10^{-47}$; rs11540822, $p =$
118 6.2×10^{-11}) and triglycerides (TG) (rs879055593, $p = 6.9 \times 10^{-17}$; rs9411378, $p =$
119 3.4×10^{-11} ; rs11540822, $p = 2.8 \times 10^{-12}$; rs143334143, $p = 1.2 \times 10^{-12}$) were each

120 associated with four SNPs. Finally, aspartate transaminase (AST) in the kidney-related
121 category was enriched for six COVID-19 risk associated SNPs (rs912805253, $p =$
122 3.3×10^{-34} ; rs879055593, $p = 5.4 \times 10^{-49}$; rs9411378, $p = 2.5 \times 10^{-43}$; rs1819040, $p =$
123 4.3×10^{-16} ; rs11540822, $p = 1.3 \times 10^{-14}$; rs143334143, $p = 1.1 \times 10^{-14}$). The identification
124 of shared genetic risk factors across multiple clinical lab-measurements highlights the
125 importance of understanding the interplay between various biological pathways and
126 how they contribute to COVID-19 severity. By analyzing an individual's genetic profile
127 and clinical lab-measurements, healthcare providers may be able to better predict the
128 patient's risk of developing severe symptoms or long-term complications from COVID-
129 19. These findings could help inform treatment decisions and prioritize high-risk
130 patients for early intervention.

131 Third, in our PheWAS analysis of medication usage, we found that antithrombotic
132 agents (rs912805253, $p = 3.6 \times 10^{-12}$; rs879055593, $p = 4.1 \times 10^{-11}$), HMG CoA reductase
133 inhibitors (rs912805253, $p = 2.8 \times 10^{-28}$; rs879055593, $p = 2.5 \times 10^{-45}$), thyroid
134 preparations (rs11540822, $p = 3.7 \times 10^{-17}$; rs143334143, $p = 1.3 \times 10^{-16}$), and anilides
135 (rs1819040, $p = 4.6 \times 10^{-12}$) have shared genetic risk factors with COVID-19 (Figure 1).
136 This suggests that individuals with a genetic predisposition to these medications may
137 be more vulnerable to severe COVID-19 symptoms. This findings highlight the
138 importance of personalized treatment and management of COVID-19 based on an
139 individual's genetic background. Furthermore, the discovery of common genetic risk
140 factors between certain medications and COVID-19 opens up new possibilities for
141 developing targeted therapies. By identifying the underlying genetic mechanisms,
142 researchers could develop new medications that could modify these risk factors and
143 potentially improve patient outcomes. In the future, the personalized approach to
144 COVID-19 treatment and the development of new medications may greatly enhance
145 our ability to combat this pandemic.

146 The current diagnostic and treatment protocols for long COVID are inadequate to
147 fully comprehend its overlap with other diseases, the variable onset of symptoms, as
148 well as the effects of vaccinations (1). Definitive diagnostic criteria for long COVID
149 have yet to be established (7). Establishing a comprehensive surveillance system is

150 imperative to effectively manage the ongoing presence of long COVID. Vaccination
151 remains the ultimate protection against SARS-CoV-2 infection. To address these gaps,
152 our study characterized the risk associated phenotypes behind COVID-19 by
153 incorporating PheWASs data sets from BBJ and European biobank (4), and summarized
154 data from published and peer-reviewed GWAS articles on COVID-19 (3). We identified
155 32 significant associated phenotypes across two diseases, 26 biomarkers, and four
156 medications after conservative Bonferroni multiple testing correction ($p < 7.6 \times 10^{-11}$).
157 Our findings provide valuable insights into the pathophysiology of COVID-19 infection,
158 its complications, treatment, genetic risk prediction and lifelong care. This study could
159 aid in the identification and development of new treatment options.

160

161 **Acknowledgments:** The authors wish to thank Dr. Aiko Kurosaka from the University
162 of Tokyo for providing valuable insights during our analysis.

163

164 **Funding:** This work was supported by grants from the Japan China Sasakawa Medical
165 Fellowship.

166

167 **Author contributions:** Z.C. designed the study and wrote the paper; K.M. coordinated
168 the project; all authors reviewed the paper and contributed to scientific content.

169

170 **Competing interests:** The authors have no conflicts of interest to disclose.

171

172 **References**

173

- 174 1. Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings,
175 mechanisms and recommendations. *Nat Rev Microbiol.* 2023;1-14.
- 176 2. Ballering AV, van Zon SKR, Olde Hartman TC, Rosmalen JGM, Lifelines
177 Corona Research I. Persistence of somatic symptoms after COVID-19 in the
178 Netherlands: an observational cohort study. *Lancet.* 2022; 400:452-461.
- 179 3. Karlsen TH. Understanding COVID-19 through genome-wide association
180 studies. *Nat Genet.* 2022; 54:368-369.

- 181 4. Sakaue S, Kanai M, Tanigawa Y, et al. A cross-population atlas of genetic
182 associations for 220 human phenotypes. *Nat Genet.* 2021; 53:1415-1424.
- 183 5. Xie Y, Xu E, Bowe B, Al-Aly Z. Long-term cardiovascular outcomes of
184 COVID-19. *Nat Med.* 2022; 28:583-590.
- 185 6. Xie Y, Al-Aly Z. Risks and burdens of incident diabetes in long COVID: a
186 cohort study. *Lancet Diabetes Endocrinol.* 2022; 10:311-321.
- 187 7. Quinn KL, Razak F, Cheung AM. Diagnosing post-COVID-19 condition (long
188 COVID) in adults. *CMAJ.* 2023; 195:E78-E79.

189

190

191

192

193

194

195

196

197 **Figures and Figure legends**

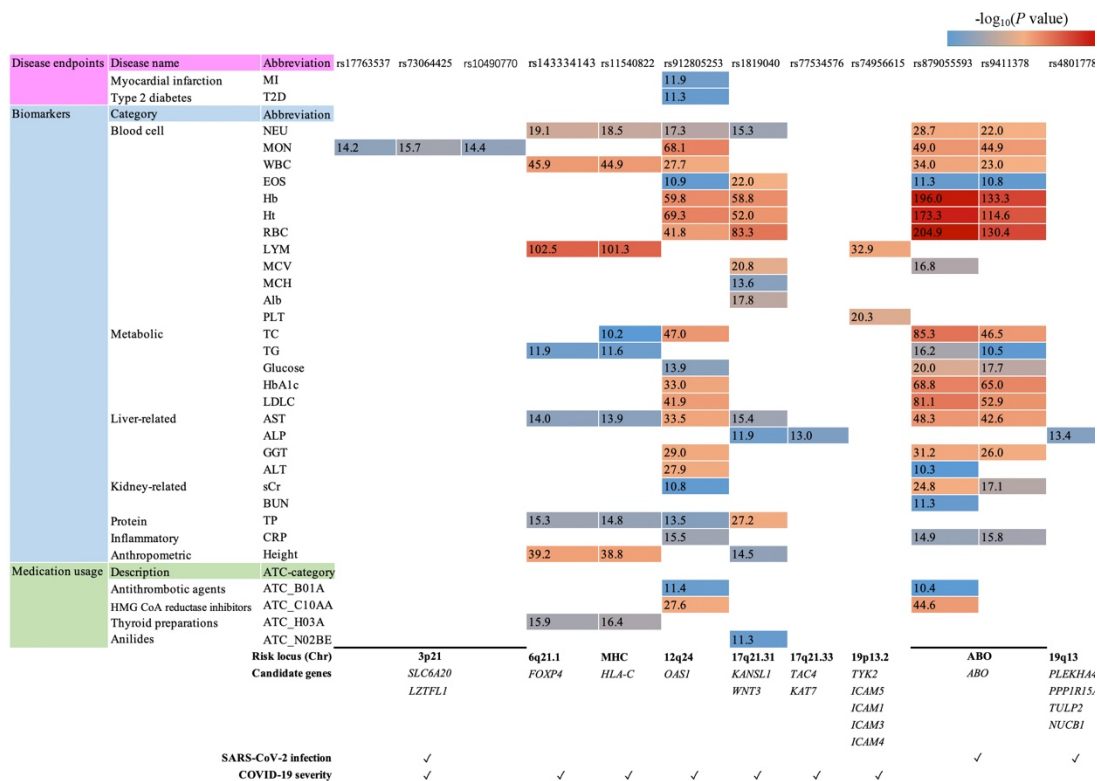


Figure 1. The landscape of associations with diseases, clinical measures (biomarkers) and medication usage for genetic risk loci from COVID-19 GWASs. The association results only considered the SNPs reached the Bonferroni-corrected threshold of $<7.6 \times 10^{-11}$, using $-\log_{10}(p$ value). Risk locus, candidate genes and their representation for SARS-CoV-2 infection and severe COVID-19 with complications from Karlsen’s summary (3). Abbreviations and full names for diseases and biomarkers: MI, myocardial infarction; T2D, type 2 diabetes; NEU, neutrophil; MON, monocyte; WBC, white blood cell; EOS, eosinophil; Hb, hemoglobin; Ht, hematocrit; RBC, red blood cell; LYM, lymphocyte; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; Alb, albumin; PLT, platelet; TC, total cholesterol; TG, triglycerides; LDLC, LDL-cholesterol; AST, aspartate transaminase; ALP, alkaline phosphatase; GGT, γ -glutamyl transpeptidase; ALT, alanine aminotransferase; sCr, serum creatinine; BUN, blood urea nitrogen; TP, total protein; CRP, C-reactive protein.

198

199