Review 2: "Genetic association analysis of SARS-CoV-2 infection in 455,838 UK Biobank participants"

Giuseppe Novelli[1]

[1]**Director Lab, University of Rome Tor Vergata: Universita degli Studi di Roma Tor Vergata, Biomedicine and Prevention, Italy**

[RR:C19 Evidence Scale](#) **rating by reviewer:**

- **Reliable.** The main study claims are generally justified by its methods and data. The results and conclusions are likely to be similar to the hypothetical ideal study. There are some minor caveats or limitations, but they would/do not change the major claims of the study. The study provides sufficient strength of evidence on its own that its main claims should be considered actionable, with some room for future revision.

****

Review:

The authors investigated the human genetic determinants of COVID-19 risk and severity in 455,838 participants in the UK biobank, including 2,003 with COVID-19.

They stratified the test subjects into eight distinct phenotypes (including risks of infection, hospitalization, and severe illness) which were then studied for association with imputed sequencing variants and WES.

The authors then replicated and confirmed the association of COVID-19 with common variants in the loci 3p21.31 (in LZTFL1) and 9q34.2 (in ABO). They also identified due to new loci associated with the risk of infection in the APOE gene (rs429358; OR = 1.29, P = 9 × 10-9).

On the contrary, the authors did not show significant associations of rare coding variants in candidate genes and in the extended exome study.

Considerations:

1.   The relationship between SARS-CoV-2 infection and the ABO phenotype remains controversial and, even when proven, can still be interpreted. Indeed, the often-observed discrepancy can be attributed to the characteristics of the study population in each region. Authors should perform a differentiated analysis not only between comments but also intra-regionally. The distribution of the different phenotypes of the ABO blood group system is a population marker, which is partly associated

with selection pressure relative to infectious agents such as plasmodium in some areas of the world. Even in individual countries, the distribution of allele frequencies of ABO differs from the ethnic origin of the population, as highlighted. **Authors should discuss these aspects.** In addition, it is known that controls recruited in large population studies like UK biobank, are sometimes influenced by socioeconomic factors in certain population groups. **Were these aspects evaluated in the control group?** Socioeconomic factors are frequently associated with comorbidities such as cardiovascular disease and diabetes, which are risk factors for complications and therefore hospitalization. These comorbidities are related to the ABO phenotype: non-O phenotypes are risk factors for ischemic heart disease and thrombosis, while phenotype B is a risk factor for type 2 diabetes.

2.   The authors did not document significant associations with rare variants identified in the sample. **Given the rarity of possible mutant alleles this cannot be ruled out a priori and the authors should discuss this aspect**. **Perhaps an analysis filter aimed at testing disruptive alleles vs. missense alleles, and/or analysis of functional prediction systems of missense alleles, should be considered.** In fact, it is known in this type of studies, the optimal design depends critically on the selection coefficient against deleterious alleles and therefore varies between genes. **Have the authors performed an analysis of the aggregate frequency distribution in cases vs. controls? Is the control of subjects negative for infection right for this type of analysis?** Probably the optimal control for these studies are asymptomatic subjects (without symptoms, but positive for infection). In addition, authors cannot rule out that rare (or private) variants may contribute substantially to the severity of the COVID-19 phenotype. **This should be discussed in the manuscript.**