

INFECTIOUS DISEASES

Understanding COVID-19 through genome-wide association studies

Defining the most appropriate phenotypes in genome-wide association studies of COVID-19 is challenging, and two new publications demonstrate how case-control definitions critically determine outcomes and downstream clinical utility of findings.

Tom H. Karlsen

Exploring self-reported data from more than 700,000 participants in a direct-to-consumer ancestry genetics company, in this issue of *Nature Genetics*, Roberts et al. report how several commonly used phenotype definitions in COVID-19 genetics studies converge to represent either susceptibility to infection by the SARS-CoV-2 virus or risk of severe COVID-19 disease¹. For pragmatic reasons, early genome-wide association studies (GWAS) in COVID-19 focused on hospitalized cases compared with unscreened and often previously genotyped controls^{2,3}. While allowing for rapid assessments during the first and very challenging wave of the pandemic, such study designs are biased towards the biology of complications in COVID-19. The emphasis on patients with mild or no symptoms, including identification

of household COVID-19 exposure as a high-risk measure, allowed the authors to conduct a deep investigation of susceptibility to SARS-CoV-2 infection through comparisons such as exposed individuals who tested positive for COVID-19 versus exposed individuals who tested negative. Not only did these assessments corroborate the controversial *ABO* locus as a bona fide susceptibility gene for SARS-CoV-2 infection^{2,4}, they also suggested the presence of a hitherto unexplored pool of protective variants.

In a dedicated query of rare variants (minor allele frequency (MAF) < 0.005), also reported in this issue of *Nature Genetics*, Horowitz et al. identified an association signal between a non-coding X chromosome variant (rs190509934) upstream of angiotensin-converting enzyme 2 (*ACE2*) and protection against

SARS-CoV-2 infection⁵. The authors go on to substantiate their finding using RNA sequencing - data from liver tissue, showing that the protective allele leads to an almost 40% reduction in *ACE2* expression levels in carriers. The association inherently holds considerable plausibility, with the membrane-bound *ACE2* serving as the binding site for the SARS-CoV-2 spike glycoprotein, initiating virus cell entry⁶. Furthermore, Horowitz et al.⁵ and Roberts et al.¹ utilize rich phenotype data to dissect the chromosome 3p21.31 association into a susceptibility signal and a severity signal, which localize to *SLC6A20* and *LZTFL1*, respectively, as also observed by others⁷. *SLC6A20* encodes the sodium-imino-acid (proline) transporter 1 (*SIT1*), which functionally interacts with *ACE2* (ref. ⁸), and the risk allele has been shown to associate

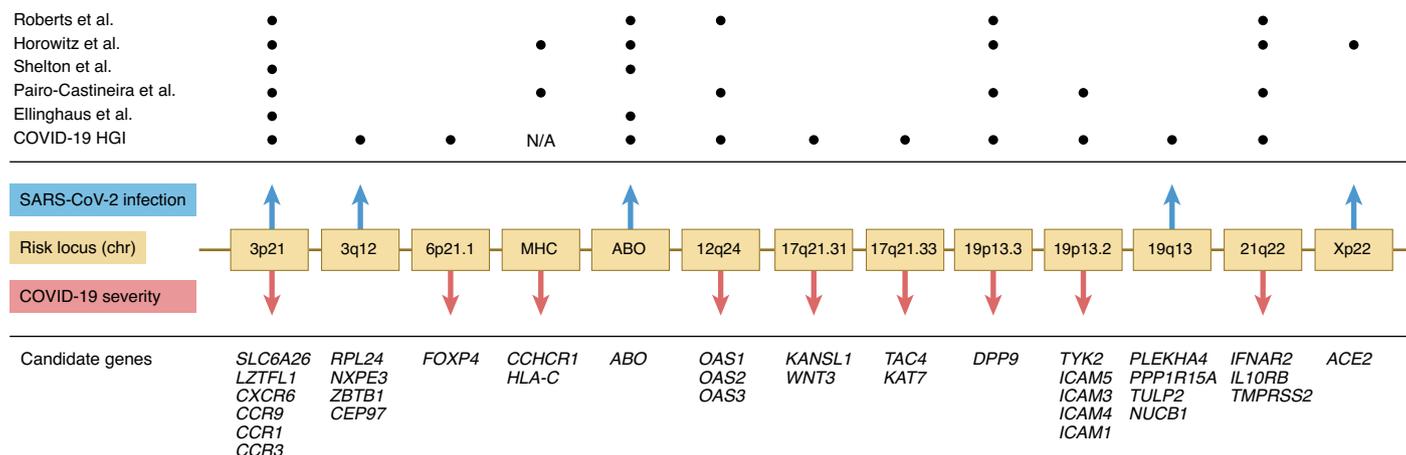


Fig. 1 | Genetic loci from COVID-19 GWAS in peer-reviewed publications to date. The loci represent a mixture of risk variants for SARS-CoV-2 infection (blue upward arrows) and severe COVID-19 with complications (red downward arrows). With increasing sample sizes, further loci are likely to emerge. Nominally significant associations at 3q12 and 6p21.1 in the Horowitz et al.⁵ analysis are not indicated due to substantial sample overlap with the COVID-19 HGI report. N/A indicates that the major histocompatibility complex (MHC) on chromosome 6 was omitted from the reporting in this article due to high heterogeneity of putative associations from the individual studies in the meta-analysis. COVID-19 HGI, COVID-19 Host Genetics Initiative (<https://www.covid19hg.org/>).

with increased expression of *SLC6A20* (ref. ²). Along with data suggesting that the receptor-binding domain of the SARS-CoV-2 spike protein preferentially interacts with blood group A⁹, which is encoded by the risk variant at the *ABO* locus, genetics of the susceptibility to SARS-CoV-2 infection appear to converge on the cell entry apparatus for the virus.

Critical illness in COVID-19 develops in fewer than 10% of individuals infected with SARS-CoV-2 (ref. ¹⁰). Given the window from the first symptoms of COVID-19 to onset of severe disease with respiratory failure (typically about one week)¹⁰, prediction of a severe disease course following SARS-CoV-2 infection is of considerable clinical interest as well as from a therapeutic point of view. Reliable risk stratification may guide therapeutic interventions during this lead-in period, characterized by enhanced viral replication. These interventions potentially include antiviral therapies, convalescent plasma, neutralizing monoclonal antibodies or — possibly more important for hospitalized patients — immunomodulating drugs.

Horowitz et al. found that a high genetic risk score (top 10%) based on six established severity variants was associated with a 1.65-fold and 1.75-fold higher risk of severe disease, in individuals with or without the presence of clinical risk factors such as age and diabetes, respectively⁵. Others have found an odds ratio of 2.0 for the impact of the rs10490770 risk allele at the 3p21.31 locus on the combined end-point of death or severe respiratory failure in an overall COVID-19 patient population¹¹, with almost double the effect size in individuals 60 years or younger (odds ratio of 3.5). These magnitudes are comparable with those associated with clinical risk factors. Findings of lower age in individuals homozygous for the chromosome 3p21.31 risk variant support enhanced utility of genetic risk stratification in the young patient population².

The execution of GWAS in COVID-19 has been remarkably nimble, due in part to robust collaborative networks set up during past GWAS¹², as well as the utilization of previously genotyped study populations such as the UK Biobank, AncestryDNA and 23andme^{1,3–5}. The rapid phenotyping undertaken by several biobanks and direct-to-consumer genetics companies during the COVID-19 pandemic is unprecedented, and the resulting publications deserve acknowledgement as a form of ‘population-level testing’ for genetic clues in emerging diseases. The orchestration of projects by the COVID-19 Host Genetics Initiative has also been an important catalyzer of activities¹³. Figure 1 summarizes published and peer-reviewed GWAS articles on COVID-19. However, even at time of writing, the meta-analysis of the sixth data freeze of the COVID-19 Host Genetics Initiative has been released online, reporting on a total of 23 loci involving in COVID-19 susceptibility (7 loci) and severity (15 loci); adding 10 new loci to the consortium’s own publication only 3 months ago⁷. The 22-month period that has passed since the publication of the first COVID-19 GWAS² appears even more impressive in comparison with the 7 years of Crohn’s disease genetics — spanning from the 2001 nucleotide-binding oligomerization domain 2 (*NOD2*) susceptibility gene discovery to a 2008 meta-analysis^{14,15} — that it took to achieve the same amount of insight. Further exemplified by the 20-year history of genetics of Crohn’s disease, translational studies of GWAS findings take time, but may reveal new and unexpected aspects of pathophysiology. It is in this context that the rapid unravelling of COVID-19 genetics becomes important. Some of the loci hold immediate biological plausibility (for example, *ACE2* and some of the chemokines), whereas the underlying mechanisms of others remain obscure. Following this recent sprint of COVID-19 GWAS to which Horowitz et al.⁵ and Roberts et al.¹ significantly contribute, the

subsequent translational ultramarathon of biological studies can begin — and with this a deeper understanding of the pathophysiology of SARS-CoV-2 infection and its complications will emerge. Vaccination has proven the ultimate protection against SARS-CoV-2 infection. The hope is that the biological insights provided by COVID-19 GWAS will facilitate identification and development of novel treatment options of not only hospitalized and critically ill COVID-19 patients, but also treatment modalities that can prevent hospitalization. □

Tom H. Karlsen ^{1,2} 

¹Department of Transplantation Medicine, Clinic of Surgery, Inflammatory Diseases and Transplantation, Oslo University Hospital Rikshospitalet, Oslo, Norway. ²Research Institute for Internal Medicine, Clinic of Surgery, Inflammatory Diseases and Transplantation, Oslo University Hospital Rikshospitalet and University of Oslo, Oslo, Norway.

 e-mail: t.h.karlsen@medisin.uio.no

Published online: 11 April 2022

<https://doi.org/10.1038/s41588-021-00985-x>

References

1. Roberts, G. H. L. et al. *Nat. Genet.* <https://doi.org/10.1038/s41588-022-01042-x> (2022).
2. Ellinghaus, D. et al. *N. Engl. J. Med.* **383**, 1522–1534 (2020).
3. Pairo-Castineira, E. et al. *Nature* **591**, 92–98 (2021).
4. Shelton, J. F. et al. *Nat. Genet.* **53**, 801–808 (2021).
5. Horowitz, J. E. et al. *Nat. Genet.* (in the press).
6. Yan, R. et al. *Science* **367**, 1444–1448 (2020).
7. COVID-19 Host Genetics Initiative. *Nature* <https://doi.org/10.1038/s41586-021-03767-x> (2021).
8. Kuba, K. et al. *Pharmacol. Ther.* **128**, 119–128 (2010).
9. Wu, S. C. et al. *Blood Adv.* **5**, 1305–1309 (2021).
10. Berlin, D. A., Gulick, R. M. & Martinez, F. J. *N. Engl. J. Med.* **383**, 2451–2460 (2020).
11. Nakanishi, T. et al. *J. Clin. Invest.* <https://doi.org/10.1172/jci152386> (2021).
12. Bulik-Sullivan, B. K. & Sullivan, P. F. *Nat Genet* **44**, 113 (2012).
13. The COVID-19 Host Genetics Initiative. *Eur. J. Hum. Genet.* **28**, 715–718 (2020).
14. Hugot, J. P. et al. *Nature* **411**, 599–603 (2001).
15. Barrett, J. C. et al. *Nat. Genet.* **40**, 955–962 (2008).

Competing interests

The author declares speaker fees from AlfaSigma and Gilead and consultancy fees from Novartis and Intercept. The author has received research grants from Canica A/S.