

# Will GWAS eventually allow the identification of genomic biomarkers for COVID-19 severity and mortality?

Vito Luigi Colona,<sup>1</sup> Michela Biancolella,<sup>2</sup> Antonio Novelli,<sup>3</sup> and Giuseppe Novelli<sup>1,4,5</sup>

<sup>1</sup>Department of Biomedicine and Prevention and <sup>2</sup>Department of Biology, Tor Vergata University of Rome, Rome, Italy. <sup>3</sup>Laboratory of Medical Genetics, IRCCS Bambino Gesù Children's Hospital, Rome, Italy. <sup>4</sup>IRCCS Neuromed, Pozzilli (IS), Italy. <sup>5</sup>Department of Pharmacology, School of Medicine, University of Nevada, Reno, Nevada, USA.

**GWAS involve testing genetic variants across the genomes of many individuals to identify genotype-phenotype associations. GWAS have enabled the identification of numerous genomic biomarkers in various complex human diseases, including infectious ones. However, few of these studies are relevant for clinical practice or at the bedside. In this issue of the *JCI*, Nakanishi et al. characterized the clinical implications of a major genetic risk factor for COVID-19 severity and its age-dependent effect, using individual-level data in a large international multicenter consortium. This study indicates that a common COVID-19 genetic risk factor (rs10490770) associates with increased risks of morbidity and mortality, suggesting potential implications for future clinical risk management. How can the genomic biomarkers identified by GWAS be associated with the clinical outcomes of an infectious disease? In this Commentary, we evaluate the advantages and limitations of this approach.**

*Prediction is very difficult, especially when it comes to the future.*

—Niels Bohr

Numerous research groups around the world have begun to analyze different genomes of COVID-19 patients in order to identify susceptibility loci to SARS-CoV-2 and alleles correlated with the severity of the disease (1–6). Since the beginning of the pandemic, GWAS have represented an indispensable approach to establishing molecular factors and have shed light on the pathological mechanisms of COVID-19.

COVID-19 is a complex disorder in which both genetic and environmental factors, modifiable or not, synergistically influence disease outcome (5, 6). Although the exponential growth of genomic data

has made parallel data sharing and analysis possible during a short time (7), many of the biobanks currently available lack essential complementary patient information, such as ethnic origin, pathological course, pharmacological treatment, and the results of the therapy itself. The collection of samples, for both cases and controls, are influenced by many variables, including sociodemographic factors, which can lead to selection bias (8). Computer algorithms applied to extended genomic studies can lead to interpretive errors if they are based on representative alleles from international reference databases and certain predictive functional bioinformatics characteristics (9, 10). Nonetheless, these predictions represent a valid starting point for studying the impact of different gene loci and are essential for a more in-depth understand-

ing of the cellular pathways involved in the virus-host interaction (5, 6). For example, researchers have recently identified two genomic regions that encompass genes involved in type I IFN antiviral immunity, which are considered the principal targets of germinal mutations of inborn errors of immunity (IEIs) (11). In fact, IEIs compromise type I IFN immunity, either directly or via autoantibodies, favoring the development of potentially lethal forms of COVID-19. In general, the impact of susceptibility alleles identified by GWAS appears to be modest at the individual level, but they can operate as modifiers and, if analyzed as a whole, could reach appreciable ORs. For this reason, genomic analysis is necessary to establish polygenic risk scores used for patient stratification and targeted treatment.

## A GWAS genomic biomarker without direct biological relevance

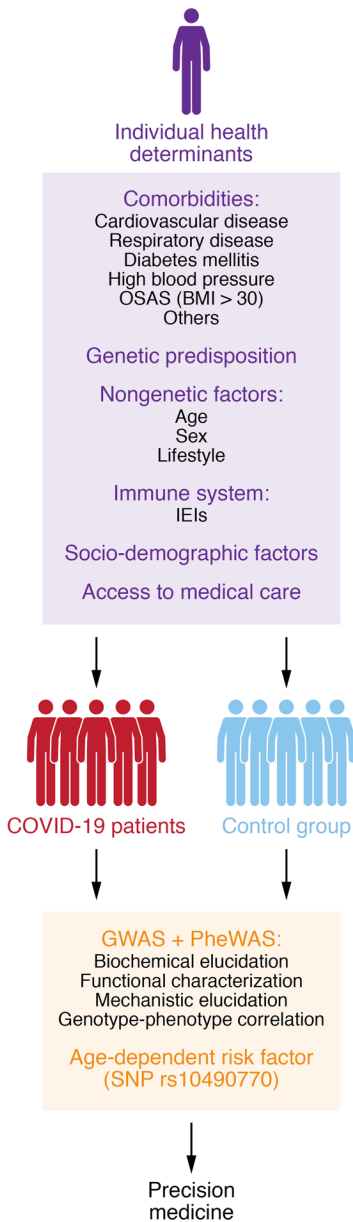
In this issue of the *JCI*, Nakanishi et al. (12) report on their performance of a metaanalysis, providing evidence about the association between a common genetic risk locus (tagged through the SNP rs10490770) at 3p21 and an increase in susceptibility to SARS-CoV-2 infection and COVID-19 severity in subjects who carried the risk allele (C). GWAS related this locus to a greater disease severity, spurring speculation about the clinical implications for stratifying patients in the treatment of COVID-19. The authors analyzed 17 cohorts from 9 different countries, for a total of 13,888 affected patients, of which 7185 were hospitalized. The large-scale reporting better predicted severity in the age group under 60 (OR, 2.7; 95% CI, 1.8–3.9; ref. 12). This haplotype was inherited from Neanderthals and varies in frequency around the world, ranging from less than 1% in East Asia to about 30% in South Asia.

► **Related Article:** <https://doi.org/10.1172/JCI152386>

**Conflict of interest:** The authors have declared that no conflict of interest exists.

**Copyright:** © 2021, American Society for Clinical Investigation.

**Reference information:** *J Clin Invest.* 2021;131(23):e155011. <https://doi.org/10.1172/JCI155011>.



**Figure 1. A model for precision medicine.** In a proactive approach to precision medicine, individual health determinants are systematically complemented with GWAS, PheWAS, and functional studies. Nakanishi et al. found that the common COVID-19 genetic risk factor (rs10490770) associates with increased risks of morbidity and mortality, especially in patients under 60 years (12). Efforts to establish polygenic risk scores, validate candidate genes with mechanistic understanding, and determine additional risk factors will help clinicians stratify patients to provide targeted treatments OSAS, obstructive sleep apnea syndrome.

tion between younger age and genetic condition, the latter acquires a more relevant importance over other known risk factors related to ageing. In a recent study, it has been shown that a recessive X-linked deficiency of *TLR7* showed a high penetrance in 1.8% of male patients who had severe clinical conditions of COVID-19 and were younger than the sixth decade (14).

In addition to elucidating the mechanisms underlying genetic susceptibility to the virus, genomic research, especially when oriented to clinical practice, must complement our understanding of the contribution of nongenetic factors. This combined approach should aim to define the individual risk, not only through a statistical probability algorithm, but also through a holistic approach including functional and biochemical analysis, capable of overcoming traditional biobank cohorts that are often unrepresentative of mixed and complex ethnic groups. For example, the distribution of ABO genotypes correlates to the geographical gradient of *Plasmodium spp* (15), an infective agent that exerts a selective pressure on the alleles. This geographical effect could influence the allelic distribution and subsequently affect correlation with several comorbidities, primarily cardiovascular, that are subject to numerous social and health conditions and cause an increase in hospitalization rates (16, 17).

The interpretation of metaanalysis based on sampling hospitalized patients and the general population as controls for which no information is available about exposure to the virus might lead to an overstatement of the results and collider bias. To prevent or reduce this bias, it is essential to define appropriate inclusion criteria, which in the case of COVID-19 studies,

strictly depends on a correct phenotypic stratification of both samples and controls. To obtain an accurate risk prediction, it is necessary to search and validate risk factors that are independent of other predisposing conditions.

**Validating GWAS**

The importance of in vitro, and possibly in vivo, validation of the results obtained through GWAS seems increasingly evident. Recently, Li et al. (18) demonstrated that the rare alleles within the *NFKB1* gene, associated with common variable immunodeficiency (CVID), were found to be neutral on functional testing, despite in silico prediction. Another study from Wickenhagen et al. (19) reports evidence from a cohort of COVID-19 patients that supports the Neanderthal *OAS1* haplotype as having a protective role against disease severity. This study related the gene to function by analyzing *OAS1* transcripts and antiviral activity of prenylated OAS1. Despite the need to strengthen its statistical power, the investigation represents a clear example of the importance of relating disease progression to a hypothetical risk factor (rs10774671; ref. 19). It would be therefore appropriate in the near future to revisit and reanalyze in functional terms results obtained from generalized genomics studies that use GWAS to identify predictive risk variants (8, 20).

A holistic approach requires accurately correlating rare and/or common variants with phenotypic, demographic, and environmental characteristics and with the pleiotropic effects of the characterizing alleles (phenome-wide association studies [PheWAS]; ref. 21). Indeed, the need to correlate a common polymorphism in the European population, such as rs10490770, with clinical phenotype in different ethnic groups can lead to controversial conclusions, as emerged from the analysis by Singh et al. (22). On the other hand, biobank heterogeneity, phenotypic stratification methods, and functional study approaches (i.e., CRISPR interference) can lead to discrepancies in results when some rare variants are taken into consideration (23).

The final purpose of advanced genomics studies is therefore to identify gene-environment interactions (GxE) jointly to identify variants associated with the

Age plays a fundamental role in COVID-19 outcomes and is a well-known risk factor, together with sex, BMI over 30, pulmonary diseases, and cardiovascular diseases. Recently, the presence of autoneutralizing IFN- $\alpha$ 2 or IFN- $\omega$  in relation to age has been demonstrated in the general population, which could explain the substantial proportion of COVID-19 cases and deaths in the elderly. Preexisting autoantibodies could persist silently until SARS-CoV-2 infection occurs. Autoantibodies neutralizing type I IFNs that underlie life-threatening COVID-19 have been found in about 20% of patients with COVID-19 who are over 80 years (13). However, when considering the correla-

variation of gene expression (expression quantitative trait loci [eQTL]; ref. 21). In contrast with the GWAS studies, the eQTL studies require much smaller samples, are sensitive to environmental influence, and allow researchers to identify genetic variants associated with the variation of gene expression in different syndromic and nonsyndromic conditions. These studies also open the way to understanding the molecular mechanisms underlying post-acute (long) COVID-19 (24, 25).

### Benefits of integrated GWAS

GWAS will certainly continue to play a key role in the identification of genes for understanding virus-host interaction. The identification of susceptibility loci represents a fundamental step in scientific research and lays the foundations for the identifying targets relevant for the development of therapeutic approaches and/or the repositioning of drugs in use or disused. The preparation of a polygenic risk score, which can be integrated with the so-called phenome, must be considered as a starting point in continuous evolution, drawing on data from PheWAS.

The complementation of GWAS and PheWAS and the validation of the results obtained through functional studies must be considered as the key objective in a proactive perspective of precision medicine (Figure 1). Achieving this milestone will take time and effort, but it is the only way to integrate genomics into clinical practice.

### Acknowledgments

Studies in the authors' laboratory are supported by Fondazione Roma and Regione

Lazio (Italy, Progetti di Gruppi di Ricerca 2020 A0375-2020-36663, GecoBiomark).

Address correspondence to: Giuseppe Novelli, Viale Oxford, 81 – 00133 Rome (RM), Italy. Phone: 39.06.20900670; Email: novelli@med.uniroma2.it.

1. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19 [published online July 8, 2021]. *Nature*. <https://doi.org/10.1038/s41586-021-03767-x>.
2. Hu J, et al. Genetic variants are identified to increase risk of COVID-19 related mortality from UK Biobank data. *Hum Genomics*. 2021;15(1):10.
3. SeyedAlinaghi S, et al. Genetic susceptibility of COVID-19: a systematic review of current evidence. *Eur J Med Res*. 2021;26(1):46.
4. Pathak GA, et al. Integrative genomic analyses identify susceptibility genes underlying COVID-19 hospitalization. *Nat Commun*. 2021;12(1):4569.
5. Novelli G, et al. COVID-19 one year into the pandemic: from genetics and genomics to therapy, vaccination, and policy. *Hum Genomics*. 2021;15(1):27.
6. Colona VL, et al. Update on human genetic susceptibility to COVID-19: susceptibility to virus and response. *Hum Genomics*. 2021;15(1):57.
7. Jacqueline AL, et al. Workshop proceedings: GWAS summary statistics standards and sharing. *Cell Genomics*. 2021;1(1):100004.
8. Li Y, et al. Genome-wide association study of COVID-19 severity among the Chinese population. *Cell Discov*. 2021;7(1):76.
9. Pairo-Castineira E, et al. Genetic mechanisms of critical illness in COVID-19. *Nature*. 2021;591(7848):92–98.
10. Fricke-Galindo I, Falfán-Valencia R. Genetics insight for COVID-19 susceptibility and severity: a review. *Front Immunol*. 2021;12:622176.
11. Zhang Q, et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science*. 2020;370(6515):eabd4570.
12. Nakanishi T, et al. Age-dependent impact of the major common genetic risk factor for COVID-19 on severity and mortality. *J Clin Invest*. 2021;131(23):e152386.
13. Bastard P, et al. Autoantibodies neutralizing type I IFNs are present in ~4% of uninfected individuals over 70 years old and account for ~20% of COVID-19 deaths. *Sci Immunol*. 2021;6(62):eabl4340.
14. Asano T, et al. X-linked recessive TLR7 deficiency in ~1% of men under 60 years old with life-threatening COVID-19. *Sci Immunol*. 2021;6(62):eabl4348.
15. Cserti CM, Dzik WH. The ABO blood group system and *Plasmodium falciparum* malaria. *Blood*. 2007;110(7):2250–2258.
16. Höglund J, et al. Characterization of the human ABO genotypes and their association to common inflammatory and cardiovascular diseases in the UK Biobank. *Am J Hematol*. 2021;96(11):1350–1362.
17. Dai X. ABO blood group predisposes to COVID-19 severity and cardiovascular diseases. *Eur J Prev Cardiol*. 2020;27(13):1436–1437.
18. Li J, et al. Biochemically deleterious human NFKB1 variants underlie an autosomal dominant form of common variable immunodeficiency. *J Exp Med*. 2021;218(11):e20210566.
19. Wickenhagen A, et al. A prenylated dsRNA sensor protects against severe COVID-19 [published online September 28, 2021]. *Science*. <https://doi.org/10.1126/science.abj3624>.
20. Kasela S, et al. Integrative approach identifies SLC6A20 and CXCR6 as putative causal genes for the COVID-19 GWAS signal in the 3p21.31 locus. *Genome Biol*. 2021;22(1):242.
21. Moon CY, et al. Phenome-wide and expression quantitative trait locus associations of coronavirus disease 2019 genetic risk loci. *iScience*. 2021;24(6):102550.
22. Singh PP, et al. Publisher Correction: The major genetic risk factor for severe COVID-19 does not show any association among South Asian populations. *Sci Rep*. 2021;11(1):12346.
23. Cotsapas C, et al. Do monogenic inborn errors of immunity cause susceptibility to severe COVID-19? *J Clin Invest*. 2021;131(14):e149459.
24. Kerimov N, et al. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat Genet*. 2021;53(9):1290–1299.
25. Marx V. Scientists set out to connect the dots on long COVID. *Nat Methods*. 2021;18(5):449–453.