Genomic Insights into the Human Susceptibility to SARS-CoV-2 Infection

Perspetivas Genómicas da Suscetibilidade Humana à Infeção por SARS-CoV-2

Ana Cristina MAGALHÃES[1,2,3], Nicole PEDRO[1,2,3], Luísa PEREIRA✉[1,2]
Acta Med Port 2021 Jun;34(6):407-409 ▪ https://doi.org/10.20344/amp.16168

Keywords: Coronavirus Infections; COVID-19; Genetic Predisposition to Disease; Genome-Wide Association Study; Host-Pathogen Interactions

Palavras-chave: COVID-19; Estudo de Associação Genómica Ampla; Infecções por Coronavírus; Predisposição Genética para Doença; SARS-CoV-2

One year after the first reported case of COVID-19 in Portugal, on the 2nd of March 2020, we are acquiring knowledge at an unprecedented pace on this new infectious disease caused by the newly emerged SARS-CoV-2 virus. We are just at the tip of the iceberg, but the breakthroughs have been substantial. Notably, a continuous sequencing characterization of the viral genome is allowing a timely monitoring of new variants, informing public health actions.[1] Machine learning methods and large clinical trials are being conducted for drug repurposing, leading to the identification of widely available treatment drugs, such as tocilizumab and sarilumab.[2] Society's hopes are now centred on the first anti-SARS-CoV-2 vaccines being administered in several countries, after less than one year of development.[3] Complex diseases, such as infectious diseases, result from the interplay between environmental and several human genetic factors. The heritability statistic represents how much of the variation of a trait can be attributed to variation of genetic factors, varying from 0 (no genetic contribution) to 1 (total genetic contribution). International and local scientific efforts are also being dedicated towards improving our understanding about the heritability of COVID-19, and this editorial will focus on these results.

With the transition to the 21st century, technological developments on genomics have improved exponentially. It is now faster, cheaper, and easier to collect big data. The catalogue of human genome variants on several worldwide populations indicates the following mean number of single-nucleotide polymorphisms (SNPs) per human genome in comparison to the human reference genome: 4.31 million in sub-Saharan Africans, 3.53 million in Europeans, and 3.55 million in East Asians.[4] A substantial proportion of these variants are common (> 5% frequency in populations), and they are being used in the design of arrays, which can contain thousands or millions of these variants interspersed regularly throughout the genome. This technology is revolutionising gene mapping on complex diseases[5]: the millions of variants ensure that several variants (tag variants; usually 15 - 20 in non-African populations, and about eight in sub-Saharan African populations) will be linked with an unknown causal/protective variant when comparing two cohorts, e.g. case versus control. These studies are known as genome-wide association studies, abbreviated as GWAS. So instead of testing individually a few usual culprits (such as a few dozen of variants in human leucocyte antigen (HLA), and other immune genes), we now have the means to perform unbiased genome-wide tests. Because the individual contribution of each variant/gene to the complex phenotype is small, the statistical significance, after correcting for the multiple testing of a million of variants, must pass the 10^{-8} *p*-value threshold. This implies that a study only reaches statistical power through the inclusion of thousands of case and control samples. This factor has somehow limited the widespread GWAS application, and currently around 78% of individuals included in GWAS studies are from European ancestry.[6]

The COVID-19 pandemic was an opportunity for the human genetics community to join cross-country efforts and perform well designed GWAS, leading to the COVID-19 Host Genetics Initiative[7]; (https://www.covid19hg.org/). The initiative took advantage of retrospective collections, mostly national biobanks with high-throughput genetic characterization of samples and links to healthcare systems (that enabled the collection of clinical information relevant for COVID-19), such as the Finnish, UK and Icelandic biobanks. But the initiative also includes prospective collections, which obtain direct consenting and sampling from incoming COVID-19 patients. The contributing scientific groups have priority in publishing their results, but they make their data readily available to the initiative group, which regularly performs meta-analyses and makes results publicly available in their website. Most of the individuals included so far in the studies are from North Eurasian ancestry.

The first published COVID-19 GWAS[8] included 1610

1. Instituto de Investigação e Inovação em Saúde. Universidade do Porto. Porto. Portugal.
2. Instituto de Patologia e Imunologia Molecular. Universidade do Porto. Porto. Portugal.
3. Instituto de Ciências Biomédicas Abel Salazar. Universidade do Porto. Porto. Portugal.
✉ Autor correspondente: Luísa Pereira. luisap@ipatimup.pt

Recebido: 10 de março de 2021 - **Aceite:** 11 de março de 2021 - ***First published*:** 06 de abril de 2021 - ***Online issue published*:** 01 de junho de 2021

patients with severe disease (respiratory failure) from Italian and Spanish hospitals (the epicentres during the first wave) and 2205 controls, analysed for over eight million SNPs. A 3p21.31 localized gene cluster and the ABO group system were identified as being significantly associated with COVID-19 respiratory failure. For the first gene cluster, it is currently impossible to disentangle which of the *SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6* and *XCR1* genes leads the association, but interestingly, the risk allele GA of rs11385942 is associated with reduced expression of *CXCR6* (which regulates the specific location of lung resident memory CD8 T cells in the immune response to airway pathogens) and increased expression of *SLC6A20* (which functionally interacts with angiotensin-converting enzyme 2, the SARS-CoV-2 cell-surface receptor). Blood group A showed a higher risk (*odds ratio*, OR = 1.45; 95% CI, 1.20 to 1.75) and group O a protective effect (OR = 0.65; 95% CI, 0.53 to 0.79) when compared with other blood groups. The biological mechanisms supporting the ABO group association are not yet clarified.

Pairo-Castineira *et al*[9] studied critically ill COVID-19 patients from UK intensive care units and identified associations with key host antiviral defence mechanisms and with mediators of inflammatory organ damage: gene cluster encoding antiviral restriction enzyme activators (*OAS1*, *OAS2*, *OAS3*) on chromosome 12; regions near the gene encoding tyrosine kinase 2 (*TYK2*) and within the gene encoding dipeptidyl peptidase 9 (*DPP9*) on chromosome 19; and in the interferon receptor gene *IFNAR2* on chromosome 21. Promisingly, both mechanisms can be targeted with existing drugs. Another important result from this work was an initial estimation of the heritability for severe COVID-19 to be 0.065. This value is probably an underestimate, but that may increase with the inclusion of rare variants and larger numbers of cases in future analyses. As a curiosity, the gene cluster encompassing the *OAS1*, *OAS2*, *OAS3* genes, previously associated with other infectious diseases, is known to have the incorporation of genetic material (introgression) from Neanderthal diversity into the European and East Asian genomes. One Neanderthal introgressed haplotype seems to confer risk in humans to severe COVID-19, while another seems to be protective against severe COVID-19.[10]

Horowitz *et al*[11] performed a meta-analysis aggregating four studies and four ancestries (Admixed American, African, European and South Asian), including a total of 11 356 individuals with COVID-19 (19% were hospitalized and 7% had severe disease) and 651 047 individuals with no record of SARS-CoV-2 infection. They identified seven common genetic variants (three novel) that modulate COVID-19 susceptibility and severity, implicating *IFNAR2*, *CCHCR1*, *TCF19*, *SLC6A20* and the hyaluronan pathway as potential therapeutic targets. This work also demonstrated how GWAS information can help identify individuals at high risk of severe disease, who may be prioritized for prophylactic or therapeutic interventions. The authors created a weighted genetic risk score for four variants (near regions *LZTFL1*, MHC, *DPP9* and *IFNAR2*), and compared the risk of hospitalization and severe disease between those with a high score and all other cases, after adjusting for known risk factors (e.g., age, sex, comorbidities). They found that having a high score (top 10%) was associated with a 2.0-fold increased risk of hospitalization and 1.8-fold increased risk of severe disease, and these values were consistent between ancestries. Moreover, the genetic burden for severe disease was higher among individuals with fewer known risk factors: OR = 3.26 (95% CI, 2.03 to 5.23) with one risk factor; OR = 2.19 (95% CI, 1.29 to 3.72) with two; and OR = 1.26 (95% CI, 0.75 to 2.11) with three or more. These results are important as many of these individuals would not be prioritized for prophylactic or therapeutic interventions according to current guidelines.

We recently[12] took advantage of the available data from the COVID-19 Host Genetics Initiative to calculate a polygenic risk score (PRS) based on significant odds ratios for two phenotypes: severe and hospitalised. We then applied this formula to 198 Portuguese individuals from a population cohort, whom had been characterized by an array containing one million variants. The risk scores for the two-phenotypes followed normal distributions, indicating that the proportion of Portuguese individuals genetically at risk for severe disease or hospitalization is quite low. We then used these results to evaluate between causal hypotheses for the presentation of a severe (six days in intensive care unit) and prolonged (97 days) COVID-19 case in a young (17 years old) patient, with no known risk factors. This young patient had risk scores within the mean values of the Portuguese population, a fact that led us to hypothesize that her severe and prolonged disease may have been due to an early co-infection by two SARS-CoV-2 variants, affiliated in two distinct clades (well defined and genetically distinct variants that have attained significant frequency and geographic spread) diverging by six mutations.

The identified associated variants still need to be functionally tested. The authors are using public transcriptomic databases to perform *in silico* (experiments employing computational approaches) of the associated variants to assess their possible role in controlling the expression of nearby genes. This preliminary evaluation reduces the number of candidates to be tested significantly. It is still urgent to perform large GWAS in non-European (and especially, non-Northern European) populations as trans-ethnic analyses are empowered to identify other significantly associated variants. But it is already clear that GWAS results are already contributing to generate hypotheses for drug repurposing, by revealing the biology of the SARS-CoV-2 infection and should be implemented in guidelines for therapeutic interventions. In the near future, when healthcare systems are not so overwhelmed, GWAS-based predictive tools can help physicians identify unusually high or low risk individuals, and better manage COVID-19 outbreaks.

PROTECTION OF HUMANS AND ANIMALS

The authors declare that the procedures were followed according to the regulations established by the the Helsinki

Declaration of the World Medical Association updated in 2013.

DATA CONFIDENTIALITY

The authors declare having followed the protocols in use at their working center regarding patients' data publication.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

FUNDING SOURCES

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

1. World Health Organization. Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health. Geneva: WHO; 2021.
2. RECOVERY Collaborative Group, Horby PW, Pessoa-Amorim G, Peto L, Brightling CE, Sarkar R, et al. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): preliminary results of a randomised, controlled, open-label, platform trial. medRxiv. 2021.02.11.21249258. doi: https://doi.org/10.1101/2021.02.11.21249258.
3. Dagan N, Barda N, Kepten E, Miron O, Perchik S, Katz M, et al. BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting. N Engl J Med. 2021 (in press). doi: 10.1056/NEJMoa2101765.
4. The 1000 Genomes Project C, Auton A, Brooks L, Durbin R, Garrison E, Kang H, et al. A global reference for human genetic variation. Nature. 2015;526:68-74.
5. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet. 2017;101:5-22.
6. Gurdasani D, Barroso I, Zeggini E, Sandhu M. Genomics of disease risk in globally diverse populations. Nat Rev Genet. 2019;20:520-35.
7. COVID-19 Host Genetics Initiative. The COVID-19 Host genetics initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. Eur J Hum Genet. 2020;28:715-8.
8. Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, et al. Genomewide association study of severe Covid-19 with respiratory failure. N Engl J Med. 2020;383:1522-34.
9. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick A, Rawlik K, Pasko D, et al. Genetic mechanisms of critical illness in COVID-19. Nature. 2021;591:92-8.
10. Zeberg H, Pääbo S. A genomic region associated with protection against severe COVID-19 is inherited from Neandertals. Proc Natl Acad Sci U S A. 2021;118:e2026309118.
11. Horowitz JE, Kosmicki JA, Damask A, Sharma D, Roberts GH, Justice AE, et al. Common genetic variants identify therapeutic targets for COVID-19 and individuals at high risk of severe disease. medRxiv. 2020.12.14.20248176. doi: https://doi.org/10.1101/2020.12.14.20248176
12. Pedro N, Silva CN, Magalhães AC, Cavadas B, Rocha AM, Moreira AC, et al. Dynamics of a dual SARS-CoV-2 lineage co-infection on a prolonged viral shedding COVID-19 case: insights into clinical severity and disease duration. Microorganisms. 2021;9:300.