

RESEARCH ARTICLE

COVID-19 susceptibility variants associate with blood clots, thrombophlebitis and circulatory diseases

Areti Papadopoulou^{1,2}, Hanan Musa¹, Mathura Sivaganesan³, David McCoy⁴, Panos Deloukas^{1,2}, Eirini Marouli^{1,2*}

1 William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom, **2** Centre for Genomic Health, Life Sciences, Queen Mary University of London, London, United Kingdom, **3** Barts and the London School of Medicine, Queen Mary University of London, London, United Kingdom, **4** Population Health Sciences, Queen Mary University of London, London, United Kingdom

* e.marouli@qmul.ac.uk



OPEN ACCESS

Citation: Papadopoulou A, Musa H, Sivaganesan M, McCoy D, Deloukas P, Marouli E (2021) COVID-19 susceptibility variants associate with blood clots, thrombophlebitis and circulatory diseases. *PLoS ONE* 16(9): e0256988. <https://doi.org/10.1371/journal.pone.0256988>

Editor: Giuseppe Novelli, Università degli Studi di Roma Tor Vergata, ITALY

Received: June 11, 2021

Accepted: August 19, 2021

Published: September 3, 2021

Copyright: © 2021 Papadopoulou et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: This work was supported by The Barts Biomedical Research Centre funded by the UK National Institute for Health Research (NIHR) in the context of COVIDITY-COHORT (NIHR-BHF flagship project).

Competing interests: The authors have declared that no competing interests exist.

Abstract

Epidemiological studies suggest that individuals with comorbid conditions including diabetes, chronic lung, inflammatory and vascular disease, are at higher risk of adverse COVID-19 outcomes. Genome-wide association studies have identified several loci associated with increased susceptibility and severity for COVID-19. However, it is not clear whether these associations are genetically determined or not. We used a Phenome-Wide Association (PheWAS) approach to investigate the role of genetically determined COVID-19 susceptibility on disease related outcomes. PheWAS analyses were performed in order to identify traits and diseases related to COVID-19 susceptibility and severity, evaluated through a predictive COVID-19 risk score. We utilised phenotypic data in up to 400,000 individuals from the UK Biobank, including Hospital Episode Statistics and General Practice data. We identified a spectrum of associations between both genetically determined COVID-19 susceptibility and severity with a number of traits. COVID-19 risk was associated with increased risk for phlebitis and thrombophlebitis (OR = 1.11, $p = 5.36e^{-08}$). We also identified significant signals between COVID-19 susceptibility with blood clots in the leg (OR = 1.1, $p = 1.66e^{-16}$) and with increased risk for blood clots in the lung (OR = 1.12, $p = 1.45e^{-10}$). Our study identifies significant association of genetically determined COVID-19 with increased blood clot events in leg and lungs. The reported associations between both COVID-19 susceptibility and severity and other diseases adds to the identification and stratification of individuals at increased risk, adverse outcomes and long-term effects.

Introduction

Corona Virus Disease 2019 (COVID-19) pertains to the disease caused by the virus SARS-CoV-2 -severe acute respiratory syndrome coronavirus 2. This RNA virus belongs to a family of coronaviruses that includes Middle East Respiratory Syndrome (MERS) and Severe Acute

Respiratory Syndrome (SARS) [1]. COVID-19 was declared a pandemic by the WHO in March 2020, and as of the 16th of April 2021, there have been 139,802,523 confirmed cases documented in 219 countries [2].

Upon SARS-CoV-2 infection, typical symptoms include dry cough, fever, and loss of smell and/or taste. However, the presentation of infection can range from asymptomatic to critical illness which has led to categorisation of disease based on severity of symptoms. One definition of severe COVID-19 is the need for admission to an intensive care unit and need for a ventilator/supplementary oxygen which is less likely in mild to moderate COVID-19. The need for hospitalisation is mainly due to infection of the lower respiratory tract and therefore inflammation of lung alveoli, also known as pneumonia. This can lead to acute respiratory distress syndrome (ARDS), sepsis and multi-organ failure which are some of the main causes of mortality in COVID patients [3]. Aside from classifying severity of disease based on symptoms, histopathological differences can also be used for classification purposes. For example, it has been suggested that the formation of hyaline membranes, due to lack of surfactant leading to alveolar collapse, are characteristic of pneumonia and therefore of severe COVID-19 [3].

The development of severe COVID is attributed to a dysregulated immune response characterised by a cytokine storm, a severe and sustained activation of the inflammatory pathway mediated by overexpression of cytokines [4]. Cytokines usually act as a signal for innate immune cells, such as neutrophils and macrophages, to move to the site of infection. However, when this is inappropriately sustained, the increased vascular permeability can lead to pulmonary consolidation and other lung tissue injury. In addition to accumulation of pulmonary exudates, the signalling cascade can result in apoptosis of lung epithelial cells which can lead to the complication of ARDS since oxygen transfer is severely impaired. Studies have reported a correlation of higher levels of pro-inflammatory cytokines (e.g. IL-6) and inflammatory markers (e.g. C-reactive protein (CRP)) with more severe clinical presentations [3]. There is also evidence that type 1 interferon, an antiviral defence, is suppressed during the initial phase of the disease, which promotes the over-activity of the inflammatory pathway underlying disease severity [5]. The combination of low interferon levels and high pro-inflammatory cytokines is what indicates the dysregulated innate immune response to COVID-19.

Following genome-wide association studies (GWAS) of susceptibility to severe COVID-19 infection, Mendelian Randomisation (MR) can then be used for causal inference of the role of a specific variant or risk factor. The outputs of such analyses could serve as guide in clinical trials aiming to repurpose existing medications for effective COVID-19 management. Baricitinib is an example of an anti-inflammatory rheumatoid arthritis drug that has been repurposed to help buffer the cytokine storm that leads to lung injury [6]. Furthermore, evidence of under expression of IFNAR2, which encodes the receptor of interferon, amongst severe COVID-19 patients has led to clinical trials of interferons to reduce mortality rate [4]. In addition, genomic analyses could reveal new therapeutic targets for drug development. 3CLpro, PLpro and RdRp, the main enzymes responsible for SARS-COV-2 replication using host machinery, emerged as potential therapeutic targets aids such as therapy with monoclonal antibodies [7, 8].

GWAS have identified several loci associated with increased susceptibility to COVID-19 infection and severe disease [9]. The underlying genes of these association signals can be broadly divided into underactive genes that enable viral replication and overactive genes that result in pulmonary inflammation and the symptoms of severe COVID-19. The most prominent findings include reduced activity of IFNAR2, increased activity of TYK2 and mutation of OAS1 which means poor activation of an enzyme involved in stopping viral replication [10].

An additional finding is a gene cluster on chromosome 3 that is associated with severe COVID-19 outcomes. COVID-19 individuals with gene variants at this locus had 1.6 odds of

requiring hospitalisation [11]. The same study found that the susceptibility haplotype was most strongly associated with people of Bangladeshi origin which may in part explain the higher COVID-19 related mortality amongst Bangladeshi men compared to White males [12, 13]. It is theorised that historically, gene flow from Neanderthals in this part of the world may have provided protection against certain pathogens through selection pressure amongst people who live in or come from what is now known as Bangladesh, but in the case of COVID-19 it confers a greater risk of severe infection [11]. The COVID-19 susceptibility haplotype showed similarity to Neanderthal Vindija haplotypes [11].

The risk of infection and severity of COVID-19 is modulated by various risk factors. The most frequently reported finding in the literature is that older patients or those with existing respiratory or cardiovascular disease are more susceptible to severe infection.

A study found that a 60–74-year-old individual has a 5-times higher risk of hospitalisation and 90 times higher risk of death from COVID-19 compared to an 18–29-year-old [14]. Additional studies suggest that individuals with chronic inflammatory diseases and vascular disease are also at higher risk of adverse COVID-19 outcomes [12]. The other most significant risk factor is body mass index (BMI) [15], which has a biological basis since obesity is associated with other conditions such as heart disease and diabetes, both of which are independent risk factors. Genetic susceptibility to obesity was found in MR studies to be associated with increased risk of infection [15]. The authors suggest that obesity could be causally linked to infection because of impaired lung function, increased baseline adipokines and cytokines which increase risk to developing ARDs. Additional studies support the causal relationship between obesity as a risk factor, although some findings suggest that this relationship might be mediated by type 2 diabetes [16].

Cardiovascular disease is also a predictor of poor COVID-19 outcomes. The combination of the hyperinflammatory environment due to a SARS-CoV-2 infection also puts the patient into a hypercoagulable state which leads to complications such as multi-organ failure, observed complication of severe COVID-19 [17]. Furthermore, hypertension could be a causal risk factor due to the viral spike protein binding to ACE2 receptors, thereby enhancing the activity of the renin-angiotensin-aldosterone system which leads to inflammation and fibrosis [18]. Similarly, smoking is a risk factor that has been found to increase expression of the ACE2 receptor in the lung which provides some explanation for why smokers are more likely to require hospital admission [19].

In the present study we aim to identify traits and diseases that are related to the genetic susceptibility for COVID-19. We conducted Phenome-Wide Association Analyses to evaluate the association of genetically determined COVID-19 susceptibility and severity using Predictive Risk Scores (PRS) with a comprehensive set of phenotypes including disease in the UK Biobank. COVID-19-related outcomes were derived from the COVID-19 Host Genetics Initiative (<https://www.covid19hg.org/>).

Materials and methods

Phenome-wide association studies (PheWAS) are used to identify the effects of a trait of interest across a potentially large array of available phenotypes, using a hypothesis-free approach [20, 21]. The advantage of the hypothesis-free testing is the possible establishment of novel associations, reporting all the results, and not only the most statistically significant ones [22]. The threshold mainly used for statistical significance is the Bonferroni correction. It is stringent, since the phenotypes under investigation are considered to be independent, which is usually not the case [23]. Despite this, the large sample size used for the analysis counterweights for the multiple testing, leading to the replication of known associations and even the discovery of new ones [22].

We used PHENome Scan Analysis (PHESANT) [21], which is a software package for performing comprehensive phenome scans in UK Biobank [24]. PHESANT tests the association of the exposure (independent variable), in our case the PRS of COVID-19, with quantitative and disease phenotypes in UK Biobank (dependent variables), using regression models. In more detail, PHESANT is implemented in statistical and computing software package, R [25]. A novel rule-based algorithm is used to categorize the phenotypes and the types of regressions, adjusted for the confounders. Based on these, the output of the software is result files depending on each test type: linear, logistic, ordered logistic and multinomial regressions. In addition, the results include a log-file providing information about the processing flow for each field and a model-fit-log file that shows model fit output for categorical unordered models. Lastly, the variable-flow-counts-all file, which includes a set of counts denoting the number of variables reaching each point in the processing flow, for example whether the variable has been analysed or has been excluded, or if some categories from the categorical fields have been excluded, are also provided in the results. In addition, for the continuous data, the inverse normal rank transformation is used to make certain they are normally distributed. To account for multiple testing, Bonferroni corrected threshold is calculated.

In order to assess the associations of the PRS with Hospital Episode data and data from the General practice (GP), we used the PheWAS library [20] in R [25]. This library is used to test the association of the exposure, the PRS of COVID-19 (independent variable), against each one of the phecodes (dependent variables). The software uses a function to convert the ICD-10 codes to 'PheWAS codes' or phecodes, which represent around 1,600 hierarchical phenotypes formed from grouped ICD-10 codes. In total, the phecodes are divided into 17 distinct categories, such as circulatory system, endocrine/metabolic, mental disorders, neurological, respiratory, plus 1 unlabelled category (denoted as NULL). Next, binary logistic regression models are employed to examine the association of the trait of interest with each phecode, adjusted for the confounders. For our analysis, Bonferroni corrected threshold is applied.

Study population

We examined the UK Biobank, a prospective cohort of 502,504 participants, aged 37–73 years old, recruited between 2006 and 2010. The dataset we used includes a variety of phenotypes including blood measurement, clinical assessments, anthropometry, cognitive function, hearing, arterial stiffness, hand grip strength, spirometry, ECG, data on cancer and death registries, health and lifestyle medical conditions, operations, mental health, sociodemographic factors, lifestyle, family history, psychosocial factors and dietary intake [22]. The analysis was performed using the maximum of 379,655 White British participants, for the variables that had no missing values.

In the present analysis, as trait of interest (exposure) we used the PRS of COVID-19. The PRS is created based on the unweighted sum of increasing risk COVID-19 alleles an individual carries, based on data from the HGI COVID-19 GWAS [26].

For our analysis we used summary statistics from the A2, B2 and C2 analyses. For A2 analysis the phenotype is defined as "Very severe respiratory confirmed COVID-19 vs. population". For analysis B2, the phenotype is defined as "Hospitalized COVID-19 vs. population". For the C2 analysis the phenotype is defined as "COVID-19 vs. population". Our analyses were focused on individuals of European ancestry (White British).

A higher allele score corresponds to more severe COVID-19, depending on the definition used for the GWAS meta-analysis defined above, and it is standardized to have mean of 0 and standard deviation of 1.

SNP selection for the PRS

We used conditional and joint analysis (GCTA-COJO) [27, 28], is employed to select the COVID-19-associated Single Nucleotide Polymorphisms—SNPs for the PRS. A stepwise procedure is used for SNP selection and the joint effects of all selected SNPs are estimated after the model has been optimized. These SNPs are still genome-wide significant, independent and the variance explained by them is larger than considering only the leading SNP at each locus. In more detail, conditional analysis was performed using 50,000 unrelated and randomly sampled European participants from UK Biobank as LD reference. We performed analyses using P-value at $p = 5 \times 10^{-8}$ to declare a genome-wide significant hit. Also, SNPs having allele frequency differences larger than 0.2, were excluded from the analysis along with SNPs having $MAF \leq 0.001$. As adjustments we used age at recruitment, sex, genotype batch, to account for the model variability. We also adjusted for the first 15 genetic principal components PCs to control for confounding via population stratification.

The number of SNPs reported from each analysis are presented in [Table 1](#).

Phenotypes

We used phenotypic data from the UK Biobank including data from Hospital Episode Statistics (HES) and General Practice (GP). HES is a database containing details of all admissions at NHS hospitals in England. All clinical data are coded using International Classification of Disease (ICD) codes and all operations and procedures are coded using the Office of Population Censuses and Surveys (OPCS); Classification of Surgical Operations and Procedures (4th revision). The most recent ones are ICD-10, which are the ones we used, including 11,726 unique ICD-10 codes for 410,321 unique participants.

GP data contain the vast majority of a person's health care information; 40% of GP records are linked to UK Biobank cohort up to now. Multiple clinical coding data are provided using Read codes; a standard vocabulary for clinicians to record patient findings and procedures. The most recent is version 3 CTV3 or Read v3), which is the one we used. Regarding the participants, 999 were excluded because the clinical event or prescription date precedes participant date of birth and 66 were removed because the event date is a date in the future and is presumed to be a place-holder or other system default) [29]. However, the 80,982 unique Read v3 codes, were converted to ICD-10 codes [30], in order to be subsequently used from the PheWAS library [20] in R [25]. As a result, 9,492 unique ICD-10 codes for 161,333 unique participants.

Results and discussion

PheWAS analyses were conducted to evaluate the effect of genetically determined risk for three COVID-19 outcomes, severe, hospitalised and general COVID-19 infection (see Methods), on a comprehensive set of phenotypes including disease in UK Biobank.

Severe respiratory confirmed COVID-19 vs. population (A2) PRS was significantly associated with increased levels of lymphocytes for both counts and percentage ($\beta = 0.013$,

Table 1. Analysis and the corresponding number of SNPs used in the PRS.

Analysis	Number of SNPs from COJO analysis— $p < 5 \times 10^{-8}$
Very severe respiratory confirmed COVID-19 vs. population—A2	11
Hospitalized COVID-19 vs. population—B2	9
COVID-19 vs. population—C2	5

<https://doi.org/10.1371/journal.pone.0256988.t001>

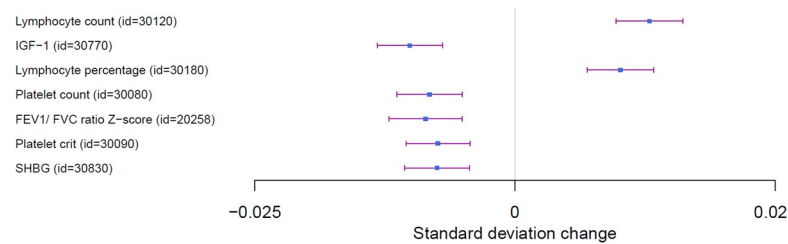


Fig 1. Forest plot of PheWAS analyses of severe COVID-19 risk.

<https://doi.org/10.1371/journal.pone.0256988.g001>

$P = 3.15e^{-15}$; $\beta = 0.01$, $p = 5.6e^{-10}$, respectively. Furthermore, the PRS was associated with decreased levels of platelet count ($\beta = -0.008$, $p = 2.6e^{-07}$). Severe COVID-19 outcome captured by the genetic score was associated with lower sex hormone binding globulin (SHBG) ($\beta = -0.0075$, $p = 2.59e^{-06}$), as well as with decreased FEV1/FVC ratio ($\beta = -0.009$, $p = 1.8e^{-06}$). In addition, severe COVID-19 was associated with lower levels of insulin like growth factor-1 (IGF-1) ($\beta = -0.010$, $P = 3.2e^{-10}$) as shown in Fig 1. For the HES data, none of the phenotypes passed the Bonferroni correction threshold of statistical significance ($p = 3.35e^{-05}$); most significant associations were with hemopoietic traits especially congenital coagulation defects ($OR = 0.85$, $p = 2.68e^{-04}$). Similarly, for the GP data, none of the phenotypes passed the Bonferroni test.

Hospitalized COVID-19 vs. population PRS was associated with decreased levels of platelets counts ($\beta = -0.01$, $P = 1.2e^{-08}$), and with increased lymphocyte counts ($\beta = 0.007$, $p = 8.19e^{-06}$). Furthermore, we observed an association with decreased FEV1/FVC ratio ($\beta = -0.01$, $P = 1.4e^{-08}$) and increased body fat percentage ($\beta = 0.0057$, $p = 2.6e^{-06}$), especially with increased arms and trunk fat mass and percentages ($\beta = 0.003$, $p = 1.56e^{-06}$), and decreased standing height ($\beta = -0.006$, $p = 3.85e^{-07}$), and sitting height ($\beta = -0.006$, $p = 3.38e^{-07}$) (Fig 2). No association signal passed Bonferroni correction in neither the HES nor GP data analysis.

The general COVID-19 infection vs population (C2) PRS was associated with lower levels of alkaline phosphate ($\beta = -0.084$, $p < 2.23e^{-308}$) and aspartate aminotransferase ($\beta = -0.01$, $p = 4.32e^{-09}$). Contrary to aspartate aminotransferase, a positive association was found with alanine aminotransferase ($\beta = 0.01$, $p = 9.41e^{-12}$).

Finally, as shown in Fig 3, several haematological traits showed a significant association including lower levels of monocyte counts ($\beta = -0.02$, $p = 1.81e^{-31}$), RBC cells ($\beta = -0.01$, $p = 1.25e^{-14}$), neutrophils ($\beta = -0.01$, $p = 1.8e^{-10}$) and lymphocytes ($\beta = 0.01$, $p = 6.68e^{-09}$). Moreover, we identified significant signals between COVID-19 susceptibility with blood clots in the leg ($OR = 1.1$, $p = 1.66e^{-16}$) and with increased risk for blood clots in the lung ($OR = 1.12$, $p = 1.45e^{-10}$).

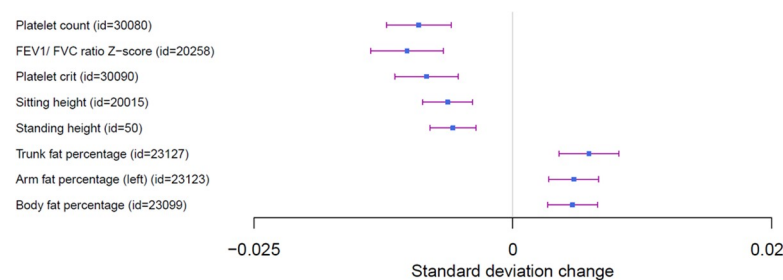


Fig 2. Forest plot of PheWAS analyses of hospitalised COVID-19 risk.

<https://doi.org/10.1371/journal.pone.0256988.g002>

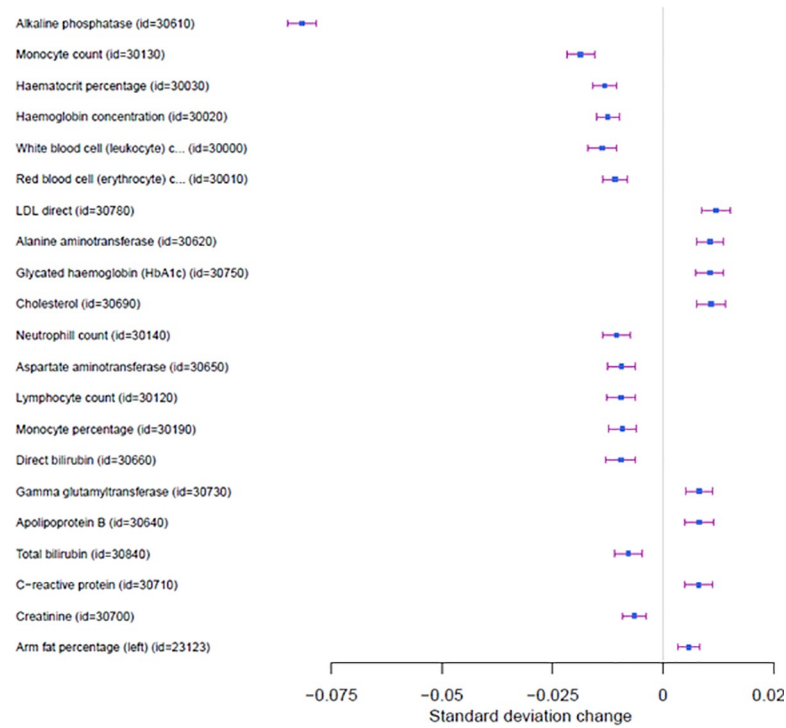


Fig 3. Forest plot of Phewas analyses of general COVID-19 susceptibility.

<https://doi.org/10.1371/journal.pone.0256988.g003>

For the HES data, COVID-19 risk was associated with disorders of circulatory system were (OR = 1.04, $p = 1.31e^{-09}$) (Fig 4). Furthermore, in the GP data, COVID-19 risk was shown to be associated with increased risk for phlebitis and thrombophlebitis (OR = 1.11, $p = 5.36e^{-08}$), phlebitis and thrombophlebitis of lower extremities (OR = 1.15, $p = 3.43e^{-05}$) (Fig 5). Full results for all the analyses are provided in S1 Table.

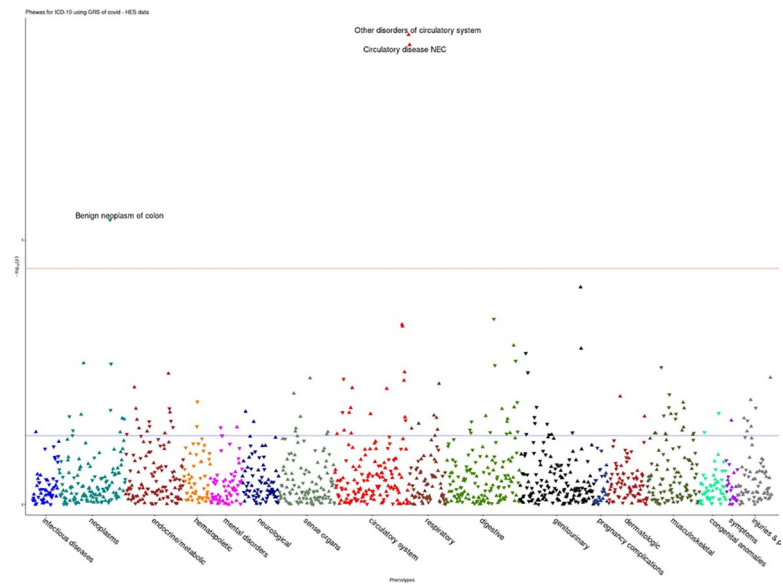


Fig 4. PhewAS associations for COVID-19 susceptibility in HES data.

<https://doi.org/10.1371/journal.pone.0256988.g004>

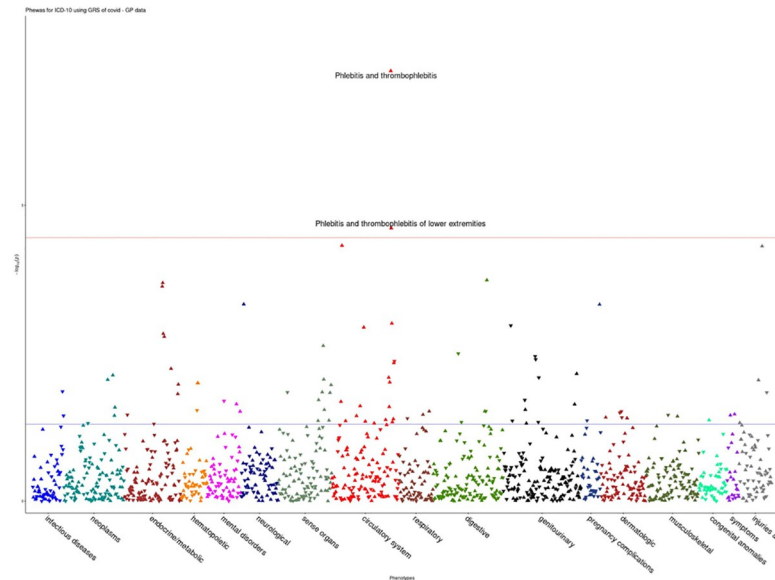


Fig 5. PheWAS associations for COVID-19 susceptibility in GP data.

<https://doi.org/10.1371/journal.pone.0256988.g005>

Our PheWAS analyses in UK Biobank identified a spectrum of associations between both genetically determined COVID-19 susceptibility and severity and a number of traits. Severe COVID-19 outcome was associated with increased counts and percentages of lymphocytes. However, several studies have also reported an association of lymphopenia with COVID-19 severity, in accordance with our findings for general COVID-19 susceptibility [31]. Lymphopenia is rarely observed in children infected with COVID-19, a group in which mortality rate is very low. Lymphocytes are known to play a crucial role in viral infections. Furthermore, atypical reactive lymphocytes are present in COVID-19 patients with better prognosis [32]. However, additional studies are needed to fully delineate the pathophysiology of immunological cells contribution to the disease.

Severe COVID-19 risk was associated with lower platelet count, in agreement with meta-analytic studies [33]. Thrombocytopenia has been shown to be common in severe cases of COVID-19 during infection, proposing it as a clinical indicator for illness progression during hospitalisation [34]. Those who are severely affected have shown coagulation abnormalities and clotting complications such as microvascular events, which have been commonly reported in retrieved tissues from severe cases [35]. Furthermore, the presence of platelet aggregates and macro-thrombocytes have been associated with COVID-19 patients admitted to ICU [36]. In a study conducted by Manne et al. [37], they reported an increased expression of platelet genes resulting in an increased activation and aggregation in patients infected by coronavirus reflecting their hyperactive status [37].

SARS-CoV-2 has been associated with abnormal pulmonary function, mainly by causing severe acute respiratory syndrome and being a significant reason for mortality and hospitalisation [38]. In this study, increased risk of severe COVID-19 and hospitalisation were associated with decreased FEV/FVC z-score. Additionally, studies have noted post-recovery changes resembling restrictive pulmonary diseases and variations in resorting to normal range levels [39, 40].

We also observed that genetically determined severe COVID-19 was associated with lower levels of IGF-1. Fan et al. (2021) suggested that a higher concentration of IGF-1 has an association with a lower risk of the mortality of COVID-19 [41]. Although IGF-1 role in severe

respiratory distress has not been fully delineated, further research and understanding of its pathway and contribution shows potential as a therapeutic mean. Severe COVID-19 risk was associated with decreased levels of SHBG. SHBG has a major role in balancing testosterone levels as it has a greater affinity for it more than oestrogen [42]. Interestingly, most male ICU admitted patients with COVID-19 had shown lower testosterone levels, deeming it a screening parameter for early detection of high-risk patients [43, 44]. Nevertheless, current studies lack describing the role of sex hormones and their binding proteins in COVID-19.

The risk of COVID-19 hospitalisation risk outcome was associated with increased trunk and arms fat mass and lower standing and sitting height. Several studies have shown that most hospitalised COVID-19 patients are obese [45]. Furthermore, severe outcome of COVID-19 has been associated with high visceral fat adiposity in European adults [46]. Obesity is identified as abnormal secretions of cytokines and adipokines resulting in increased inflammation; thus, it might contribute to the outcomes requiring hospitalization [47, 48]. Although abnormal liver enzymes were noted in the majority of COVID-19 patients, studies were not conclusive on the significance of liver function enzymes being informative predictors of the severity of the disease progression [49, 50].

For general COVID-19 susceptibility, we observed significant associations with increased phlebitis and thrombophlebitis. This result is in agreement with the available meta-analytic literature as COVID-19 procoagulant activity such as deep vein thrombosis, pulmonary embolism, and superficial phlebitis, are prevalent in moderate to severe cases [51]. Furthermore, we identify significant association of general COVID-19 with increased blood clot events in leg and lungs. In a meta-analytic study blood clots rates of COVID-19 have shown to be a risk of mortality [52].

The reported associations between both genetically determined COVID-19 susceptibility and severity and other diseases adds to the identification and stratification of individuals at increased risk, including those undergoing anti-inflammatory therapy [53]. Further work will be required to better understand the biological and clinical value of these findings including long-term complications after infection.

Conclusions

We conducted Phenome-Wide Association Analyses to evaluate the association between genetically determined COVID-19 susceptibility and severity with a comprehensive set of phenotypes and diseases in the UK Biobank. Our study identifies significant associations of genetically determined COVID-19 susceptibility with increased blood clot events in leg and lungs, thrombophlebitis and circulatory diseases.

Supporting information

S1 Table. Full PheWAs results.
(XLSX)

Acknowledgments

UK Biobank participants (Application 53723).

Author Contributions

Conceptualization: Eirini Marouli.

Data curation: Areti Papadopoulou, Hanan Musa, Eirini Marouli.

Formal analysis: Areti Papadopoulou, Eirini Marouli.

Investigation: Eirini Marouli.

Methodology: Eirini Marouli.

Project administration: Eirini Marouli.

Software: Eirini Marouli.

Supervision: Eirini Marouli.

Visualization: Areti Papadopoulou, Eirini Marouli.

Writing – original draft: Areti Papadopoulou, Hanan Musa, Mathura Sivaganesan, David McCoy, Panos Deloukas, Eirini Marouli.

Writing – review & editing: Areti Papadopoulou, Hanan Musa, Mathura Sivaganesan, David McCoy, Panos Deloukas, Eirini Marouli.

References

1. Naqvi A. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Molecular Basis of Disease*. 2020;1866(10):165878. <https://doi.org/10.1016/j.bbadis.2020.165878> PMID: 32544429
2. Worldmeter. Countries where Coronavirus has spread 2021 [
3. Zea Zhu. From SARS and MERS to COVID-19: a brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. *Respiratory Research*. 2020; 21(1).
4. The Immune Hallmarks of Severe COVID-19 2021 [updated <https://www.the-scientist.com/news-opinion/the-immune-hallmarks-of-severe-COVID-19-67937>
5. Ye Q, Wang B. and Mao J. The pathogenesis and treatment of the 'Cytokine Storm' in COVID-19. *Journal of Infection*. 2020; 80(6):pp. 607–13. <https://doi.org/10.1016/j.jinf.2020.03.037> PMID: 32283152
6. Kalil ACeA. Baricitinib plus Remdesivir for Hospitalized Adults with COVID-19. *NEJM*. 2021. <https://doi.org/10.1056/NEJMoa2031994> PMID: 33306283
7. Perricone C CP, Ciccacci C, Marcucci E, Cafaro G, Bartoloni E, Perricone R, et al. The differential response to anti IL-6 treatment in COVID-19: the genetic counterpart. *Clin Exp Rheumatol*. 2020; 38(3). PMID: 32452345
8. PLpro inhibitors support immune response to COVID-19 in cell cultures (2021). 2021.
9. Ganna A. Mapping the human genetic architecture of COVID-19 by worldwide meta-analysis. *MedRxiv*. 2021.
10. Eea Pairo-Castineira. Genetic mechanisms of critical illness in COVID-19. *Nature*. 2020.
11. Zeberg HaP S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature*. 2020; 587(7835).
12. von Bartheld CS HM, Butowt R. Prevalence of Chemosensory Dysfunction in COVID-19 Patients: A Systematic Review and Meta-analysis Reveals Significant Ethnic Differences. preprint medRxiv. 2020.
13. England aW-OfNS. Coronavirus (COVID-19) related deaths by ethnic group 2021 [Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/coronaviruscovid19relateddeathsbyethnicgroupenglandandwales/2march2020to15may2020>
14. CDC. Coronavirus Disease 2019 (COVID-19) 2020 [Available from: <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-age.html>
15. Aung N KM, Munroe PB and Petersen SE Causal Inference for Genetic Obesity, Cardiometabolic Profile and COVID-19 Susceptibility: A Mendelian Randomization Study *Front Genet*. 2020; 11(586308).
16. Aea Leong. Cardiometabolic Risk Factors for COVID-19 Susceptibility and Severity: A Mendelian Randomization Analysis. 2020.
17. Cea Petrilli. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ*. 2020:p. m1966. <https://doi.org/10.1136/bmj.m1966> PMID: 32444366

18. Wang M, CR, Zhang L., Yang X., Liu J., Xu M., Shi Z., et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro cell Res. 2020; 30:269–71. <https://doi.org/10.1038/s41422-020-0282-0> PMID: 32020029
19. Jea Smith. Cigarette Smoke Exposure and Inflammatory Signaling Increase the Expression of the SARS-CoV-2 Receptor ACE2 in the Respiratory Tract. *Developmental cell*. 2020; 53(5):514–249. <https://doi.org/10.1016/j.devcel.2020.05.012> PMID: 32425701
20. Carroll RJ, Bastarache L., and Denny J.C.,. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics*. 2014; 30(16):2375–6. <https://doi.org/10.1093/bioinformatics/btu197> PMID: 24733291
21. Millard LAC, Davies NM, Gaunt TR, Davey Smith G, Tilling K. Software Application Profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. *Int J Epidemiol*. 2018; 47(1):29–35. <https://doi.org/10.1093/ije/dyx204> PMID: 29040602
22. Millard LAC, Davies NM, Tilling K, Gaunt TR, Davey Smith G. Searching for the causal effects of body mass index in over 300 000 participants in UK Biobank, using Mendelian randomization. *PLoS Genet*. 2019; 15(2):e1007951. <https://doi.org/10.1371/journal.pgen.1007951> PMID: 30707692
23. Robinson JR, Denny JC, Roden DM, Van Driest SL. Genome-wide and Phenome-wide Approaches to Understand Variable Drug Actions in Electronic Health Records. *Clin Transl Sci*. 2018; 11(2):112–22. <https://doi.org/10.1111/cts.12522> PMID: 29148204
24. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015; 12(3):e1001779. <https://doi.org/10.1371/journal.pmed.1001779> PMID: 25826379
25. Team RC. R: A Language and Environment for Statistical Computing 2019 [September 10, 2020]. Available from: <https://www.R-project.org/>.
26. Initiative C-HG. 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(6):715–8. <https://doi.org/10.1038/s41431-020-0636-6> PMID: 32404885
27. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ATC, Replication DIG, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*. 2012; 44(4):369–75, S1-3. <https://doi.org/10.1038/ng.2213> PMID: 22426310
28. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011; 88(1):76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011> PMID: 21167468
29. Biobank U. First Occurrence of Health Outcomes Defined by 3-character ICD10 code 2019 [cited 2021 04/01/2021]. Available from: https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/first_occurrences_outcomes.pdf.
30. Biobank U. Clinical coding classification systems and maps n.d. [Available from: <https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=592>].
31. Zhao J, Zhu M, Su X, Huang M, Yang Y, Huang J, et al. Clinical features and risk factors for severe-critically ill COVID-19 adult patients in Jiangsu, China: A multiple-centered, retrospective study. *Medicine (Baltimore)*. 2021; 100(5):e24332. <https://doi.org/10.1097/MD.00000000000024332> PMID: 33592878
32. Merino A, Vlasea A, Molina A, Egri N, Laguna J, Barrera K, et al. Atypical lymphoid cells circulating in blood in COVID-19 infection: morphology, immunophenotype and prognosis value. *J Clin Pathol*. 2020. <https://doi.org/10.1136/jclinpath-2020-207087> PMID: 33310786
33. Lanza F, Seghatchian J. Reflection on passive immunotherapy in those who need most: some novel strategic arguments for obtaining safer therapeutic plasma or autologous antibodies from recovered COVID-19 infected patients. *Br J Haematol*. 2020; 190(1):e27–e9. <https://doi.org/10.1111/bjh.16814> PMID: 32407543
34. Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A meta-analysis. *Clin Chim Acta*. 2020; 506:145–8. <https://doi.org/10.1016/j.cca.2020.03.022> PMID: 32178975
35. CMaK Y. COVID-19 associated coagulopathy: An exploration of mechanisms. *Vascular Medicine* 2020; 25(5):471–8. <https://doi.org/10.1177/1358863X20932640> PMID: 32558620
36. Rampotas A, Pavord S. Platelet aggregates, a marker of severe COVID-19 disease. *J Clin Pathol*. 2020. <https://doi.org/10.1136/jclinpath-2020-206933> PMID: 33067181
37. Manne B, DF, Middleton E., Protier I., Rowely J., Stubben C., Petery A., et al. Platelet gene expression and function in patients with COVID-19. *blood* 2020; 136(11).
38. Fan E, Beitler JR, Brochard L, Calfee CS, Ferguson ND, Slutsky AS, et al. COVID-19-associated acute respiratory distress syndrome: is a different approach to management warranted? *The Lancet Respiratory Medicine*. 2020; 8(8):816–21. [https://doi.org/10.1016/S2213-2600\(20\)30304-0](https://doi.org/10.1016/S2213-2600(20)30304-0) PMID: 32645311

39. Huang Y, Tan C, Wu J, Chen M, Wang Z, Luo L, et al. Impact of coronavirus disease 2019 on pulmonary function in early convalescence phase. *Respir Res.* 2020; 21(1):163. <https://doi.org/10.1186/s12931-020-01429-6> PMID: 32600344
40. Torres-Castro R, Vasconcello-Castillo L, Alsina-Restoy X, Solis-Navarro L, Burgos F, Puppo H, et al. Respiratory function in patients post-infection by COVID-19: a systematic review and meta-analysis. *Pulmonology.* 2020. <https://doi.org/10.1016/j.pulmoe.2020.10.013> PMID: 33262076
41. Fan X, Yin C, Wang J, Yang M, Ma H, Jin G, et al. Pre-diagnostic circulating concentrations of insulin-like growth factor-1 and risk of COVID-19 mortality: results from UK Biobank. *Eur J Epidemiol.* 2021. <https://doi.org/10.1007/s10654-020-00709-1> PMID: 33420872
42. Somboonporn W, Davis SR, National H, Medical Research C. Testosterone effects on the breast: implications for testosterone therapy for women. *Endocr Rev.* 2004; 25(3):374–88. <https://doi.org/10.1210/er.2003-0016> PMID: 15180949
43. Rowland SP, O'Brien Bergin E. Screening for low testosterone is needed for early identification and treatment of men at high risk of mortality from Covid-19. *Crit Care.* 2020; 24(1):367. <https://doi.org/10.1186/s13054-020-03086-z> PMID: 32560707
44. Schroeder M, Schaumburg B, Müller Z, Parplys A, Jarczak D, Nierhaus A, et al. The Majority of MALE Patients with COVID-19 present testosterone levels on admission to intensive care in Hamburg, Germany: a retrospective cohort. 2020.
45. Huang Y, Lu Y, Huang YM, Wang M, Ling W, Sui Y, et al. Obesity in patients with COVID-19: a systematic review and meta-analysis. *Metabolism.* 2020; 113:154378. <https://doi.org/10.1016/j.metabol.2020.154378> PMID: 33002478
46. Favre G, Legueult K, Pradier C, Raffaelli C, Ichai C, Iannelli A, et al. Visceral fat is associated to the severity of COVID-19. *Metabolism.* 2021; 115:154440. <https://doi.org/10.1016/j.metabol.2020.154440> PMID: 33246009
47. Ellulu MS, Patimah I, Khaza'ai H, Rahmat A, Abed Y. Obesity and inflammation: the linking mechanism and the complications. *Arch Med Sci.* 2017; 13(4):851–63. <https://doi.org/10.5114/aoms.2016.58928> PMID: 28721154
48. Sattar N, McInnes IB, McMurray JJV. Obesity Is a Risk Factor for Severe COVID-19 Infection: Multiple Potential Mechanisms. *Circulation.* 2020; 142(1):4–6. <https://doi.org/10.1161/CIRCULATIONAHA.120.047659> PMID: 32320270
49. Cai Q, Huang D, Yu H, Zhu Z, Xia Z, Su Y, et al. COVID-19: Abnormal liver function tests. *J Hepatol.* 2020; 73(3):566–74. <https://doi.org/10.1016/j.jhep.2020.04.006> PMID: 32298767
50. Zhang Y, Zheng L, Liu L, Zhao M, Xiao J, Zhao Q. Liver impairment in COVID-19 patients: A retrospective analysis of 115 cases from a single centre in Wuhan city, China. *Liver Int.* 2020; 40(9):2095–103. <https://doi.org/10.1111/liv.14455> PMID: 32239796
51. Tan BK, Mainbourg S, Friggeri A, Bertoletti L, Douplat M, Dargaud Y, et al. Arterial and venous thromboembolism in COVID-19: a study-level meta-analysis. *Thorax.* 2021. <https://doi.org/10.1136/thoraxjnl-2020-215383> PMID: 33622981
52. Malas MB, Naazie IN, Elsayed N, Mathlouthi A, Marmor R, Clary B. Thromboembolism risk of COVID-19 is high and associated with a higher risk of mortality: A systematic review and meta-analysis. *EClinicalMedicine.* 2020; 29:100639. <https://doi.org/10.1016/j.eclinm.2020.100639> PMID: 33251499
53. Pablos JL, Abasolo L, Alvaro-Gracia JM, Blanco FJ, Blanco R, Castrejon I, et al. Prevalence of hospital PCR-confirmed COVID-19 cases in patients with chronic inflammatory and autoimmune rheumatic diseases. *Ann Rheum Dis.* 2020; 79(9):1170–3. <https://doi.org/10.1136/annrheumdis-2020-217763> PMID: 32532753