Contents lists available at ScienceDirect

EBioMedicine

Research paper

Evaluating the effects of cardiometabolic exposures on circulating proteins which may contribute to severe SARS-CoV-2

Tom G Richardson[a,*], Si Fang[a], Ruth E Mitchell[a], Michael V Holmes[a,b], George Davey Smith[a]

[a] Medical Research Council Integrative Epidemiology Unit (MRC IEU), Population Health Sciences, Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, United Kingdom
[b] Medical Research Council Population Health Research Unit (MRC PHRU), Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, United Kingdom

A B S T R A C T

Background: Developing insight into the pathogenesis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is of critical importance to overcome the global pandemic caused by coronavirus disease 2019 (covid-19). In this study, we have applied Mendelian randomization (MR) to systematically evaluate the effect of 10 cardiometabolic risk factors and genetic liability to lifetime smoking on 97 circulating host proteins postulated to either interact or contribute to the maladaptive host response of SARS-CoV-2.
Methods: We applied the inverse variance weighted (IVW) approach and several robust MR methods in a two-sample setting to systemically estimate the genetically predicted effect of each risk factor in turn on levels of each circulating protein. Multivariable MR was conducted to simultaneously evaluate the effects of multiple risk factors on the same protein. We also applied MR using cis-regulatory variants at the genomic location responsible for encoding these proteins to estimate whether their circulating levels may influence severe SARS-CoV-2.
Findings: In total, we identified evidence supporting 105 effects between risk factors and circulating proteins which were robust to multiple testing corrections and sensitivity analyzes. For example, body mass index provided evidence of an effect on 23 circulating proteins with a variety of functions, such as inflammatory markers c-reactive protein (IVW Beta=0.34 per standard deviation change, 95% CI=0.26 to 0.41, $P = 2.19 \times 10^{-16}$) and interleukin-1 receptor antagonist (IVW Beta=0.23, 95% CI=0.17 to 0.30, $P = 9.04 \times 10^{-12}$). Further analyzes using multivariable MR provided evidence that the effect of BMI on lowering immunoglobulin G, an antibody class involved in protection from infection, is substantially mediated by raised triglycerides levels (IVW Beta=-0.18, 95% CI=-0.25 to -0.12, $P = 2.32 \times 10^{-08}$, proportion mediated=44.1%). The strongest evidence that any of the circulating proteins highlighted by our initial analysis influence severe SARS-CoV-2 was identified for soluble glyco-protein 130 (odds ratio=1.81, 95% CI=1.25 to 2.62, $P = 0.002$), a signal transductor for interleukin-6 type cytokines which are involved in inflammatory response. However, based on current case samples for severe SARS-CoV-2 we were unable to replicate findings in independent samples.
Interpretation: Our findings highlight several key proteins which are influenced by established exposures for disease. Future research to determine whether these circulating proteins mediate environmental effects onto risk of SARS-CoV-2 infection or covid-19 progression are warranted to help elucidate therapeutic strategies for severe covid-19 disease.
Funding: The Medical Research Council, the Wellcome Trust, the British Heart Foundation and UK Research and Innovation.

Research in context

Evidence before this study

It remains unclear why certain individuals develop more severe symptoms of coronavirus disease 19 (covid-19) compared to others.

However, increasingly findings from the literature suggest that established cardiometabolic risk factors, such as body mass index and smoking, influence the risk of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which is caused by covid-19.

In this study, we used data from a resource involving genome-wide association studies (GWAS) of 97 unique proteins which may play a role in severe SARS-CoV-2 (available at https://omicscience.org/apps/covidpgwas/). This enabled us to use genetic variation to

* Corresponding author.
E-mail address: tom.g.richardson@bristol.ac.uk (T.G. Richardson).

Research in context

Evidence before this study

It remains unclear why certain individuals develop more severe symptoms of coronavirus disease 19 (covid-19) compared to others. However, increasingly findings from the literature suggest that established cardiometabolic risk factors, such as body mass index and smoking, influence the risk of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which is caused by covid-19.

In this study, we used data from a resource involving genome-wide association studies (GWAS) of 97 unique proteins which may play a role in severe SARS-CoV-2 (available at https://omicscience.org/apps/covidpgwas/). This enabled us to use genetic variation to estimate the effects of 10 cardiometabolic risk factors on each of these proteins in turn using an approach known as Mendelian randomization (MR). We additionally used this approach to estimate the effects of these circulating proteins on risk of severe SARS-CoV-2, using studies from the literature who have made GWAS results on this outcome.

Added value of this study

Our study provides a systematic evaluation of the genetically predicted effects of 10 cardiometabolic risk factors on each of the 97 unique proteins. Altogether, we found 105 effects which were robust to multiple testing corrections which may be valuable for future covid-19 research. We also evaluated whether any of these proteins influence risk of SARS-CoV-2, with soluble glycoprotein 130 providing the strongest evidence of a genetically predicted effect using MR. This protein is involved on the interleukin 6 receptor pathway which plays an important role in the body's immune response. However, further data is required to robustly support this gene's putative role in risk of severe SARS-CoV-2.

Implications of all the available evidence

Our findings are important in terms of developing insight into the molecular pathways by which modifiable lifestyle factors influence disease risk. Specifically with respect to severe covid-19, we note that the GWAS datasets of SAR-CoV-2 used in this work will capture genetic effects on increased susceptibility to infection as well as progression to severe symptoms. This is particularly important when considering the implications of therapeutically targeting any of the proteins highlighted by our study.

This protein is involved on the interleukin 6 receptor pathway which plays an important role in the body's immune response. However, further data is required to robustly support this gene's putative role in risk of severe SARS-CoV-2.

Implications of all the available evidence

Our findings are important in terms of developing insight into the molecular pathways by which modifiable lifestyle factors influence disease risk. Specifically with respect to severe covid-19, we note that the GWAS datasets of SAR-CoV-2 used in this work will capture genetic effects on increased susceptibility to infection as well as progression to severe symptoms. This is particularly important when considering the implications of therapeutically targeting any of the proteins highlighted by our study.

1. Introduction

On the 11th of March 2020 the World Health Organisation (WHO) declared the coronavirus disease 2019 (covid-19) a global pandemic [1]. Although strict lockdown measures have been enforced in many countries to control the spread of infection, the number of deaths worldwide which have been attributed to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) continues to rise [2]. Furthermore, despite widespread ongoing biomedical research it remains unclear why some individuals develop severe symptoms of SARS-CoV-2 once contracting covid-19, whereas an estimated 80% of individuals display either asymptomatic or mild infections [3]. It is becoming increasingly evident however based on findings from the literature that established cardiometabolic disease risk factors play a role in the severity of symptoms for SARS-CoV-2 [4, 5].

To address this challenge, researchers in the field, led by colleagues from the MRC Epidemiology Unit, have rapidly generated a curated dataset concerning the genetic architecture of 97 unique proteins which may be involved in influencing severe SARS-CoV-2 [6]. These include inflammatory cytokines and antibodies (such as immunoglobulin G) which are involved in immune response to infection, proteins involved in fibrinolysis and blood coagulation and gene products which have been reported to interact with SARS-CoV-2 proteins in human cells [7]. A complete list of these proteins can be found in **Supplementary Table 1**.

This curated resource provides an opportunity to undertake Mendelian randomization (MR) analyzes to develop insight into modifiable risk factors that influence these SARS-CoV-2-related proteins, as well as potential downstream consequences on risk of covid-19. MR can be implemented as a form of instrumental variable analysis which exploits the random assortment of genetic alleles at birth under Mendel's laws of Inheritance [8, 9]. As such genetic variants can be leveraged as instrumental variables to investigate causal relationships between conventional exposures (such as cardiometabolic risk factors) and outcomes (such as circulating proteins) (**Fig. 1A**). As these inherited genetic variants are fixed at conception, MR is typically robust to confounding factors and reverse causation which can bias analyzes in an observational setting which do not make use of human genetics data.

In this study, we systematically applied MR to estimate the effects of 10 cardiometabolic exposures and genetic liability to lifetime smoking in turn on each of the SARS-CoV-2 prioritized proteins. We focused on these types of exposures given findings from the literature providing evidence that they may increase risk of severe covid-19, including observations from cohort studies, healthcare records and previous genetic studies [10-14]. This was followed by a series of sensitivity analyzes as well as applying multivariable MR to evaluate whether exposures independently influence the same circulating protein or act along overlapping causal pathways. We also sought to investigate the potential effects of proteins highlighted by this

estimate the effects of 10 cardiometabolic risk factors on each of these proteins in turn using an approach known as Mendelian randomization (MR). We additionally used this approach to estimate the effects of these circulating proteins on risk of severe SARS-CoV-2, using studies from the literature who have made GWAS results on this outcome.

Added value of this study

Our study provides a systematic evaluation of the genetically predicted effects of 10 cardiometabolic risk factors on each of the 97 unique proteins. Altogether, we found 105 effects which were robust to multiple testing corrections which may be valuable for future covid-19 research. We also evaluated whether any of these proteins influence risk of SARS-CoV-2, with soluble glycoprotein 130 providing the strongest evidence of a genetically predicted effect using MR.
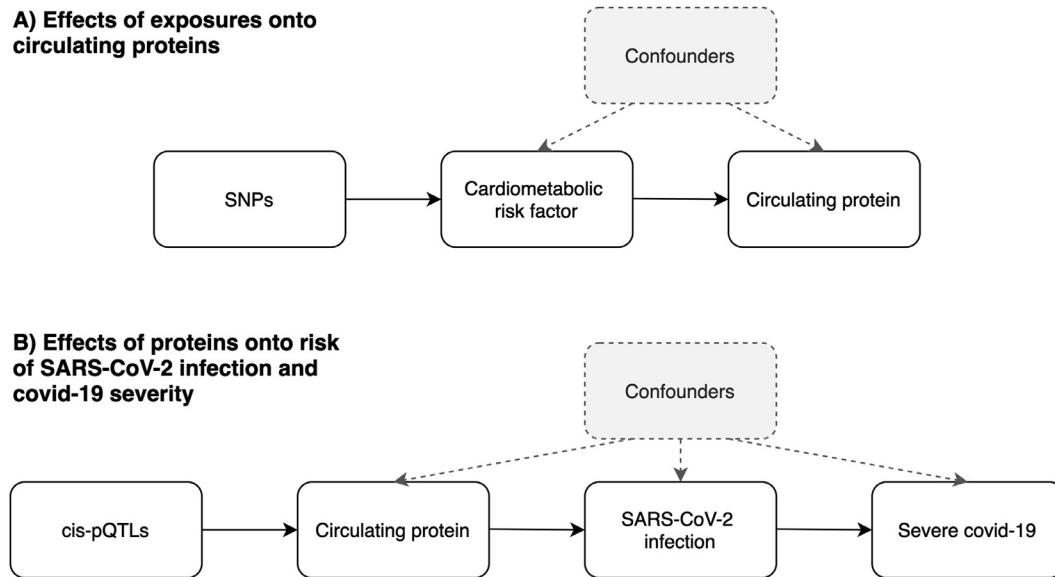
A) Effects of exposures onto circulating proteins

B) Effects of proteins onto risk of SARS-CoV-2 infection and covid-19 severity

Fig. 1. Schematic representation of the analysis undertaken in this study using Mendelian randomization. A) We firstly leveraged genetic variants (referred to as single nucleotide polymorphisms (SNPs)) to systematically estimate the effect of 11 risk factors on 97 circulating proteins related to SARS-CoV-2. B) For proteins highlighted in the initial analysis, we applied MR to estimate their genetically predicted effects on risk of severe SARS-CoV-2. Note that the circulating protein could influence progression from SARS-CoV-2 infection to severe covid-19 without influencing risk of becoming infected (or it could influence the two processes in opposite directions). Instruments for proteins were SNPs robustly associated with their levels and located in the genome at the encoding genes region (commonly referred to as cis-protein quantitative trait loci (cis-pQTL)).

analysis on risk of severe covid-19 using data from recently conducted genome-wide association studies (GWAS).

2. Methods

2.1. Data resources

2.1.1. Deriving genetic instruments for modifiable exposures

We obtained genetic instruments for 11 exposures using data from large-scale GWAS. These were body mass index (BMI), systolic blood pressure (SBP), diastolic blood pressure (DBP), high density lipoprotein (HDL) cholesterol, low density lipoprotein (LDL) cholesterol, triglycerides, apolipoprotein A-I (Apo A-I), apolipoprotein B (Apo B), genetic liability to lifetime smoking, waist-hip-ratio adjusted for BMI and childhood adiposity based on reported body size at age 10 [15-18]. Details on the study characteristics for the GWAS used to derive these instruments can be found in **Supplementary Table 2.**

We undertook linkage disequilibrium (LD) clumping to identify independent genetic instruments for these 11 exposures assessed using the software PLINK [19]. This process involves removing genetic variants which are correlated with the mostly strongly associated variant with a trait of interested in a region based on pairwise LD (using $r^2 < 0.001$ in this study) using a reference panel of 503 individuals of European descent from phase 3 (version 5) of the 1000 genomes project [20].

2.1.2. Quantitative trait loci data for SARS-CoV-2-related proteins

All pQTL summary statistics for 97 unique proteins were obtained from the https://omicscience.org/apps/covidpgwas webserver [6]. Details on how these pQTL were derived are described in detail in the study by Pietzner et al. and outlined in **Supplementary Fig. 1**. Briefly, plasma samples from 10,708 individuals from the Fenland population-based cohort study were eligible for analysis after exclusions. In total, 409 circulating proteins were prioritized due to any of the following criteria; evidence suggesting that they interact with SARS-CoV-2 ($n = 332$) [7], associated with disease severity ($n = 26$) [21], involved in viral entry ($n = 2$) [22] or that they are clinical biomarkers of adverse, prognosis, complications and disease deterioration ($n = 54$) [23-26]. Of the proteins, SOMAscan proteomic assays were used to derive data on 179 of them.

In the same set of participants, imputed genotype data on 17,652,797 genetic variants were available after imputation using the UK10K+1000 G phase3 reference panel. Summary statistics were available in a total of 97 unique proteins from the webserver as they had at least 1 pQTL acting *in cis*, which is defined here as genetic variants robustly associated with circulating proteins (based on $P<5 \times 10^{-08}$) and located within a 1Mb window around the genes responsible for encoding them.

We undertook LD clumping as before to identify pQTL to be used as instruments in MR analyzes. However, for protein instrumental variables we applied a more lenient LD threshold of $r^2<0.2$ to identify weakly correlated pQTL all within a cis-window of 1Mb either side of the lead cis-pQTL for each protein analyzed. MR was then undertaken for protein targets whilst taking into account LD structure of pQTL. Further details on this method have been described in detail by Burgess et al [27]. In brief, the pairwise correlations between all instruments are incorporated in the standard error of the test statistic for the summary-level weighted generalized linear regression MR when deriving causal estimates.

2.1.3. Covid-19 GWAS datasets

Genetic estimates on SARS-CoV-2 were obtained using data from a GWAS of severe covid-19 [28] based on 1980 patients from intensive care units and wards at seven hospital located in the pandemic epicenters in Italy and Spain (accessed on 11/08/2020). Severe covid-19 was defined as hospitalization with respiratory failure and a confirmed SARS-CoV-2 viral RNA polymerase-chain-reaction (PCR) test using nasopharyngeal swabs or other biological fluids. General population controls analysed in this GWAS had unknown covid-19 status. Effect estimates from these GWAS were mapped to hg19 coordinates using the LiftOver tool (https://genome.sph.umich.edu/wiki/Lift Over). We analyzed GWAS data on severe SARS-CoV-2 data as detection and classification of these is likely to be more complete and less selected than for reported covid-19 symptoms or SARS-CoV-2 test positivity, for which potentially misleading conclusions can be drawn [29]. We also sought out replication of findings using GWAS results from the covid-19 host genetic initiative [30] using data on hospitalized covid-19 cases compared to population controls (https://www.covid19hg.org/results/, accessed on 11/08/2020), as well as a GWAS of mortality attributed to covid-19 in the UK Biobank study compared
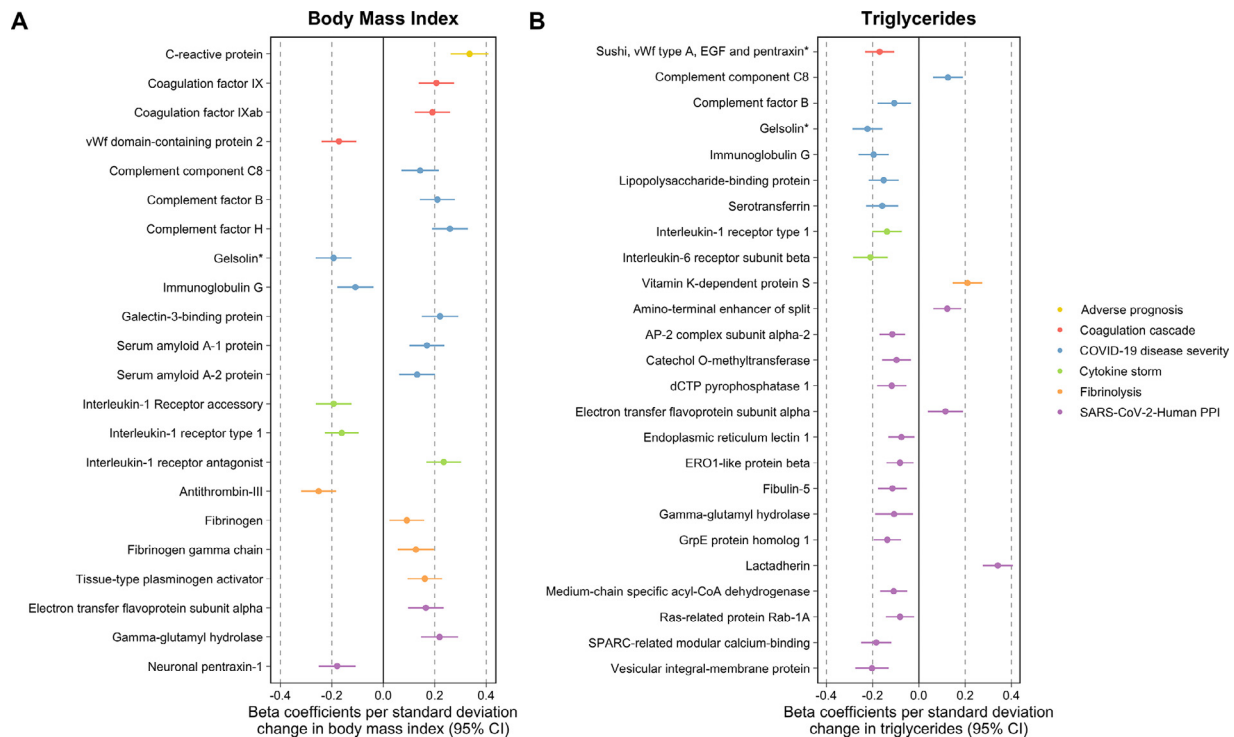
Fig. 2. Forest plots illustrating the Mendelian randomization estimates of genetically predicted A) body mass index and B) triglycerides on circulating proteins related to severe SARS-CoV-2. Proteins have been grouped and colored based on their subcategories. * Gelsolin corresponds to soma ID 16,607−78 as a separate epitope is also available on the SomaLogic assay for this protein.

to population controls based on analyzes by Johnson and colleagues [31] (available at https://grasp.nhlbi.nih.gov/Covid19GWASResults.aspx, accessed on 11/08/2020).

2.2. Ethics statement

All data analyzed in this study is summary-level data. The relevant ethical approval can be found in the corresponding studies referenced for each dataset.

2.3. Statistical analysis

We firstly applied MR to estimate the effect of each of the 11 exposures in turn on each protein in a two-sample setting [32]. Initially we used the inverse variance weighted (IVW) approach which takes the SNP-outcome estimates and regresses them on those for the SNP-exposure associations. This provides an overall weighted estimate of the causal effect which is based on the inverse of the square of the standard error for the SNP-outcome association [33]. We applied a correction using false discovery rate (FDR)<5% to these results to account for multiple testing. This threshold has been used in this study as a heuristic to highlight findings with the strongest statistical evidence to investigate in further detail.

For effects which survived FDR corrections we applied other MR methods as sensitivity analyzes. This firstly involved applying the weighted median and MR-Egger approaches which are more robust to horizontal pleiotropy in comparison to the IVW method [34, 35]. We also applied the MR directionality test (also referred to as the 'Steiger method') to support evidence that our genetic variant is a valid instrument for the exposure in line with the underlying assumptions of MR [36]. Multivariable MR was undertaken to evaluate the direct effects of exposures on circulating proteins whilst accounting for the effects of other exposures [37, 38]. In doing so, we were able to investigate whether the effect of exposures on proteins were putatively mediated via another exposure. To estimate the

proportion mediated we applied the product method described previously by Burgess and colleagues [39], using genetic instruments from the GIANT consortium for BMI to avoid overlapping samples with the UK Biobank [40].

We next applied MR to investigate the genetically predicted effects of circulating proteins robust to FDR corrections and sensitivity analyzes in the previous analysis on risk of severe SARS-CoV-2 using GWAS results from Ellinghaus et al. This was undertaken by applying the IVW method which uses correlated cis-regulatory variants whilst accounting for their correlation structure [27]. For proteins which provided strong evidence of an effect using the IVW method, we also applied the MR-Egger approach whilst accounting for correlation structure amongst instruments. The MR-Egger is traditionally applied to investigate horizontal pleiotropy in MR studies, although was applied predominantly as a sensitivity analysis in this work given the low number of instrumental variables available for our cis-pQTL analyzes. We also attempted to replicate findings using the GWAS data from the covid-19 HGI and Johnson et al. analyzes [30, 31]. Finally, we applied the cis-correlated IVW approach systematically to all proteins with at least 2 cis-pQTL (as this is the minimum number required for the IVW method) for each covid-19 GWAS dataset. This allowed us to highlight proteins which may play a role in disease but are not strongly under the influence of modifiable risk factors. Importantly, given that these proteins were derived in samples who did not have covid-19, our MR estimates are therefore mainly useful in terms of prioritising candidates for further research rather than implicating them directly in severe covid-19 susceptibility.

All analyzes were undertaken using the 'TwoSampleMR' and 'MendelianRandomization' packages using R (version 3.5.1) [41, 42]. The forest plot in Fig. 2 was generated using 'ggplot2' v2.2.1 [43]. Fig. 3 was generated using the LD link resource [44].

2.4. Role of funding source

The funders had no role in study design, data collection, data analyzes, interpretation, or writing of report.
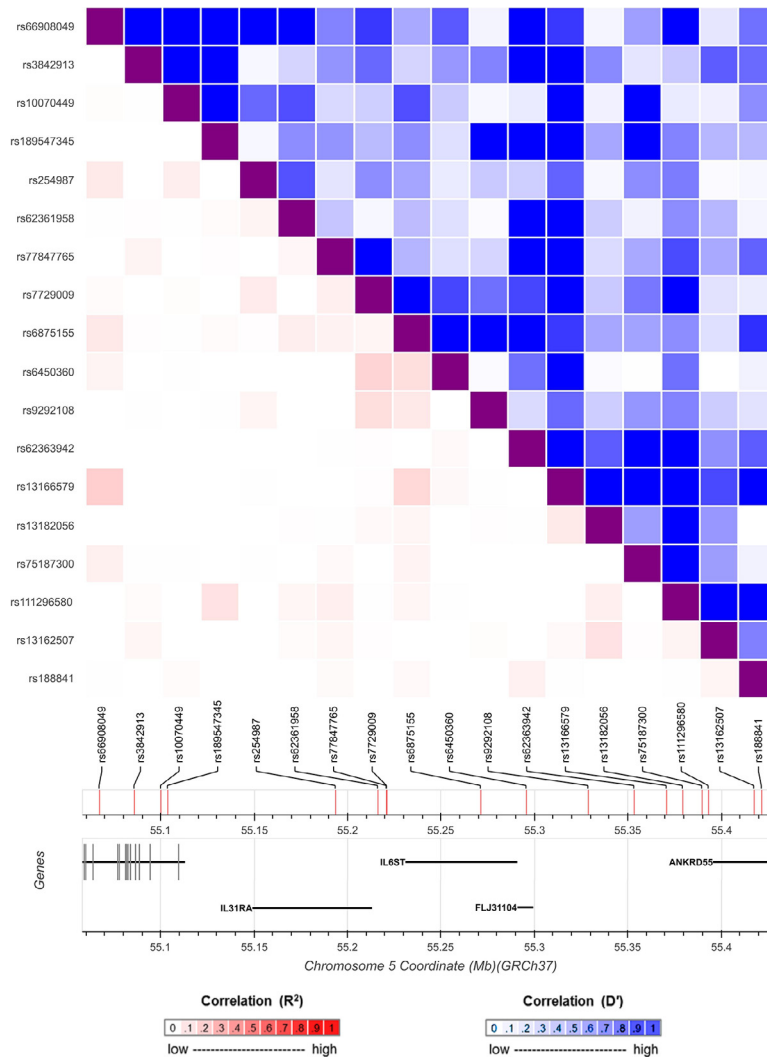
Fig. 3. A diagram of the genetic locus used to identify protein quantitative trait loci (pQTL) as instrumental variables in the Mendelian randomization analysis for glycoprotein 130 (encoded by IL6ST) and severe SARS-CoV-2. The heatmap represents the linkage disequilibrium structure at this region amongst the pQTL used with the bottom-left section representing pairwise r² coefficients (red) and the upper-right section illustrating pairwise D' values (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3. Results

3.1. A systematic Mendelian randomization analysis of circulating proteins

Across the 11 exposures assessed, there were 253 genetically predicted effects on circulating proteins which survived FDR<5% corrections using the IVW method (**Supplementary Table 3**). Amongst top findings was a strong effect of BMI on C-reactive protein (CRP) levels (Beta=0.34 per standard deviation change in BMI, 95% CI=0.26 to 0.41, $P = 2.19 \times 10^{-16}$) which is a well-established marker of chronic inflammation [45]. Elsewhere, there was strong evidence of an effect of HDL cholesterol on elevated levels of serum amyloid A-1 (Beta=0.23, 95% CI=0.16 to 0.29, $P = 1.59 \times 10^{-12}$) and A-2 (Beta=0.24, 95% CI=0.17 to 0.30, $P = 4.38 \times 10^{-13}$) proteins. Undertaking sensitivity analyzes found 106 effects that were robust to FDR<5% corrections using either the weighted median or MR-Egger methods (**Supplementary Tables 4 and 5**). The MR directionality tests provided evidence that assumptions regarding directionality may have been violated for one of these effects, which was between waist-hip-ratio adjusted for BMI and the protein ITIH3 (IVW Beta=−0.31, 95% CI=−0.46 to −0.16, $P = 4.07 \times 10^{-05}$) (**Supplementary Table 6**). IVW estimates for the 105 effects which were robust to sensitivity

analyzes can be found in **Supplementary Table 7**. An overview of the analytical pipeline applied in this section including method and datasets used can be found in **Supplementary Fig. 2**.

Amongst the exposures which contributed most to the remaining 105 effects were BMI (23 effects) and triglycerides (27 effects). As illustrated in **Fig. 2**, the effects driven by BMI were typically spread across the 6 subcategories of circulating proteins. This included effects on coagulation factor IX (IVW Beta=0.21, 95% CI=0.14 to 0.27, $P = 2.42 \times 10^{-07}$), tissue-type plasminogen activator (Beta=0.16, 95% CI=0.09 to 0.23, $P = 2.38 \times 10^{-06}$) and cytokines including interleukin-1 receptor antagonist (Beta=0.23, 95% CI=0.17 to 0.30, $P = 9.04 \times 10^{-12}$). In contrast, the majority of triglycerides effects were found to be on proteins allocated to the SARS-CoV-Human protein-protein interaction (PPI) or severe covid-19 disease subcategories. There were exceptions to this however, such as an effect on clotting factor vitamin K-dependent protein S (Beta=0.21, 95% CI=0.15 to 0.27, $P = 1.25 \times 10^{-10}$) and cytokine signal transducer interleukin-6 receptor subunit beta (Beta=−0.21, 95% CI=−0.28 to −0.13, $P = 4.12 \times 10^{-08}$). We also note the conflicting directions of effect which risk factors have on the proteins assessed even within the same category. For example, within the fibrinolysis category BMI provided evidence of an effect on higher levels of fibrinogen (Beta=0.09, 95% CI=0.02 to 0.16, $P = 0.008$), as well as an inverse effect

on antithrombin III (Beta=−0.25, 95% CI=−0.32 to −0.18, $P = 4.43 \times 10^{-13}$). Findings from the literature supported the direction of effect between BMI and circulating proteins for various effects identified in this analysis (CRP, Factor B and H, the interleukin 1 family of proteins, SAA/2, fibrinogen and antithrombin III), although for others there was no clear prior evidence suggesting that obesity influences their levels (**Supplementary Table 8**).

BMI is recognized to causally influence triglycerides and we therefore undertook multivariable MR to evaluate the direct effects of BMI and triglycerides on the 7 proteins which they had in common based on univariable estimates in the previous analysis. The majority of these effects remained robust after accounting for the effects of the other exposure, suggesting that BMI and triglycerides influence these proteins directly via separate causal pathways (Supplementary Table 9). The notable exception to this was the effect of BMI on immunoglobulin G. Effect estimates for BMI on this circulating protein identified in the univariable analysis (Beta=−0.11, 95% CI=−0.18 to 0.04, $P = 0.02$) attenuated to the null when accounting for the effect of triglycerides (Beta=−0.06, 95% CI=−0.13 to 0.02, $P = 0.13$). This suggests that BMI indirectly lowers immunoglobulin G due to its influence on raising triglyceride levels. We estimated that 44% of the BMI effect on immunoglobulin G was mediated via triglycerides using mediation MR.

3.2. Harnessing cis-regulatory variants to evaluate effects of circulating proteins on risk of severe SARS-CoV-2

For each protein highlighted in the previous analysis, we undertook MR using cis-pQTL as instruments to estimate their effects on risk of severe SARS-CoV-2 (**Supplementary Table 10**). Although no results survived multiple testing corrections based on a false-discovery rate threshold of less than 5%, we sought to replicate our lead findings from this analysis. The protein which provided the strongest evidence that it may influence risk of covid-19 was gp130, soluble, also known as glycoprotein 130 (odds ratio (OR)=1.81 increased risk of severe SARS-CoV-2 per doubling of gp130, 95% CI=1.25 to 2.62, $P = 0.002$). This protein is encoded by the *IL6ST* gene and is responsible for signal transduction with all members of the interleukin 6 receptor family [46]. There were 18 weakly correlated pQTL scattered across this locus used as instrumental variables as illustrated in **Fig. 3**. Their pairwise LD correlations can be found in **Supplementary Table 11**. Single variant associations for these 18 pQTL with severe SARS-CoV-2 suggested that the two instruments largely responsible for driving the overall IVW estimate were rs929108 ($P = 0.002$), which is located downstream of *IL6ST* (i.e. the opposite side compared to *IL31RA*) and rs6875155 ($P = 0.006$), located within the gene body of *IL6ST* itself.

Applying the MR-Egger method accounting for correlation structure using the 18 *IL6ST* instruments produced an elevated but imprecisely effect estimates which included the null (OR=1.55, 95% CI=1.00 to 2.39, $P = 0.05$). We were unable to detect robust evidence of replication using the two other covid-19 GWAS datasets (covid-19 host genetics initiative: OR=1.11, 95% CI=0.83 to 1.48, $P = 0.48$ & Risk of death due to covid-19 death in UK Biobank: OR=1.26, 95% CI=0.79 to 2.00, $P = 0.34$). Lastly, we applied the IVW method accounting for correlation structure to all proteins with at least 2 cis-pQTL to evaluate their genetically predicted effects on risk of covid-19. However, using current sample sizes we did not detect strong evidence that these circulating proteins influence risk of severe covid-19 based on multiple testing corrections (**Supplementary Tables 13−15**).

4. Discussion

We have undertaken a comprehensive Mendelian randomization study to systematically evaluate the effect of 11 established risk factors for disease on circulating levels of proteins related to SARS-CoV-2.

Our main findings are that among the modifiable risk factors assessed, BMI and triglycerides showed the widest repertoire of causal effects on these circulating proteins (providing evidence of causation for 23 and 27 effects, respectively). Furthermore, of the circulating proteins investigated by our study, the strongest evidence of an effect on developing severe covid-19 was identified for glycoprotein 130, which is involved in the transmission of molecular signals for inflammatory interleukin cytokines.

It is important to recognise that the case definitions of severe covid-19 in these studies will capture both genetic effects on increased susceptibility to infection as well as increased progression to severe symptoms [47]. It is likely that genetic effects will differ between infection and progression, indeed they could even be in different directions. MR estimates derived using these datasets should therefore take this into account when interpreting the potential implications of therapeutically targeting any proteins highlighted by this (and similar) studies (**Fig. 1B**).

Amongst the 105 effects which were robust to multiple testing and sensitivity analyzes there were several well-established relationships based on the literature. For example, having a high BMI is a known driver of systemic inflammation as indexed by C-reactive protein levels [45] and acute inflammatory markers such as fibrinogen [48]. Other findings fit with the known biology of cardiometabolic risk factors and proteins identified by our analysis, such as the effect of HDL cholesterol on serum amyloid A-1 and A-2 proteins, which have previously been proposed as clinically applicable surrogates of HDL vascular functionality [49]. Whilst our results are therefore of immediate importance for SARS-CoV-2 research, they may also be valuable for future endeavors interested in the therapeutic potential of these proteins with respect to a wide range of disease outcomes.

There were several results from our study which may assist in unraveling the complex pathogenesis of severe SARS-CoV-2. For example, immunoglobulin G (IgG) is a class of antibodies produced by plasma B cells in the immune system in response to a pathogen [50]. Our results indicate that having a high BMI may reduce levels of circulating IgG, suggesting that people with obesity have less of this class of antibody to help protect from infection. That being said, an important consideration when interpreting this finding is that IgG levels were measured in individual's in a healthy state and can therefore only act a proxy for IgG response to infection. Additionally, generic IgG levels were measured rather than the specific adaptive immune response to SARS-CoV-2.

Additionally, our multivariable MR estimates for the effect BMI on IgG attenuated when accounting for the effect of triglycerides on this class of antibodies. This suggests that triglycerides may mediate the lowering effect of BMI on IgG, which we estimated as 41.1% of the total effect of BMI on IgG levels being mediated via triglycerides. Further research into the role of IgG and B cell immunity in response to the covid-19 pathogen is therefore warranted, particularly given that IgG is being measured by tests for antibody responses to SARS-CoV-2 [51]. Along with evaluating the effect of modifiable risk factors on antibody mediated immunity to covid-19, it will be critical to develop insight into how these factors influence cell mediated immunity given the emerging importance of the adaptive immune response to SARS-CoV-2 [52].

Using MR to estimate the genetically predicted effects of circulating proteins on severe covid-19 risk highlighted glycoprotein 130 as the protein with the strongest evidence of an effect on severe SARS-CoV-2 (OR=1.81, 95% CI=1.25 to 2.62, $P = 0.002$), however we were unable to replicate these findings in larger samples. A possible explanation for this could be the heterogeneity between the SARS-CoV-2 GWAS datasets analyzes and their variable case definitions. This requires robust replication that would be necessary before initiating further in-depth analyzes. Glycoprotein 130 is encoded by the *IL6ST* gene and belongs to the interleukin-6 family of cytokines [46]. It's activation is dependent upon the binding of cytokines with their

receptors, such as interleukin-6 (IL6) with interleukin-6 receptor (IL6R) [53]. This is noteworthy due to the extensive interest in repurposing IL6R blockers as a potential therapeutic strategy for SARS-CoV-2 [54-58]. As lowering the levels of circulating IL6R will lead to lower activation of glycoprotein 130, estimates in this study suggest that this might result in reduced risk of severe SARS-CoV-2 symptoms. These findings therefore corroborate results from a recent MR study which used human genetic data to support the efficacy of IL6R inhibition as a potential treatment option for severe SARS-CoV-2 symptoms [59]. However the MR studies to date have not been able to reliably separate influences on risk of becoming infected from risk of progressing to severe disease following infection (Fig. 1B). Thus they do not provide robust evidence as to whether IL6R inhibition would be expected to favourably influence outcome in severe covid-19. Adequately powered randomized controlled trial data are essential for evaluating the clinical value of therapeutic intervention targeting *IL6R* [55].

This study has several limitations which should be taken into account when interpreting its findings. The current sample sizes of the SARS-CoV-2 GWAS are (as one would expect) relatively modest compared to large-scale GWAS data which MR studies are contemporaneously applied to, meaning that our cis-pQTL analysis is likely underpowered. We analyzed severe covid-19 as an outcome to mitigate reported selection bias of cases [29], so larger sample sizes of severe SARS-CoV-2 GWAS in the future should improve the statistical power of our approach. Furthermore, although data from plasma is of unprecedented sample size compared to previous large-scale pQTL analyzes (n = 10,708), it remains comparably modest to the sample sizes of GWAS used to derive instrument for the cardiometabolic exposures in this work. This is exaggerated by the fact that protein MRs are typically conducted using instruments relating to a single gene and therefore these genetic variants will explain a lower proportion of variance in the exposure than if instruments are taken from across the genome [60]. Therefore, although we have undertaken thorough evaluations to interrogate bi-directional relationships between the exposures and proteins in this study, the discrepancies between the samples sizes makes the direction of effect challenging to orientate (the majority of exposure instruments were derived using sample sizes of $n=$~440,000). Finally, although data from plasma pQTL studies provide an exceptional opportunity to leverage instruments for MR studies, it should be noted that serum plasma may not capture signatures confined to disease or cell-type relevant tissues. This is particularly important for a disease with a large autoimmune component such as covid-19 and further emphasis should therefore be noted when interpreting the results of our study on proteins such as IgG. Finally, studies need to be conducted on data that allow MR to separately investigate modifiable influences on acquiring SARS-CoV-2 and on progressing to severe covid-19 or death (Fig. 1B).

In conclusion, our MR study identified many effects between conventional risk factors and circulating proteins which provides a platform for prospective endeavors to dissect related disease pathways. Future research into the pathogenesis of the proteins highlighted by this study are warranted to discern whether they may hold therapeutic potential for severe covid-19.

Funding

Author contributions

TGR undertook data analysis. SF conducted the literature search. All authors contributed to the design of the study and wrote the manuscript. All authors read and approved the final version of the manuscript.

Materials and correspondence

This publication is the work of the authors and TGR will serve as guarantor for the contents of this paper.

Declaration of Interests

Dr Holmes has collaborated with Boehringer Ingelheim in research, and in adherence to the University of Oxford's Clinical Trial Service Unit & Epidemiological Studies Unit (CSTU) staff policy, did not accept personal honoraria or other payments from pharmaceutical companies. All other authors declare no conflicts of interest.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ebiom.2021.103228.

References

[1] The World Health Organisation. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020.
[2] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020;20(5):533–4.
[3] Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese center for disease control and prevention. JAMA 2020.
[4] 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.
[5] Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? Lancet Respir Med 2020;8(4):e21.
[6] Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Genetic architecture of host proteins interacting with SARS-CoV-2. bioRxiv 2020. 2020.07.01.182709.
[7] Gordon DE, Jang GM, Bouhaddou M, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 2020;583(7816):459–68.
[8] Davey Smith G, Ebrahim S. Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003;32(1):1–22.
[9] Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 2014;23(R1):R89–98.
[10] Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020;584(7821):430–6.
[11] Ponsford MJ, Gkatzionis A, Walker VM, et al. Cardiometabolic traits, sepsis, and severe COVID-19: a Mendelian randomization investigation. Circulation 2020; 142(18):1791–3.
[12] Aung N, Khanji MY, Munroe PB, Petersen SE. Causal inference for genetic obesity, cardiometabolic profile and COVID-19 susceptibility: a mendelian randomization study. Front Genet 2020;11:586308.
[13] Liu D, Cui P, Zeng S, et al. Risk factors for developing into critical COVID-19 patients in Wuhan, China: a multicenter, retrospective, cohort study. EClinicalMedicine 2020;25:100471.

[14] Morawietz H, Julius U, Bornstein SR. Cardiovascular diseases, lipid-lowering therapies and European registries in the COVID-19 pandemic. Cardiovasc Res 2020;116(10):e122–5.

[15] O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. Nat Genet 2018;50(12):1728–34.

[16] Wootton RE, Richmond RC, Stuijfzand BG, et al. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. Psychol Med 2019:1–9.

[17] Richardson TG, Sanderson E, Palmer TM, et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: a multivariable Mendelian randomisation analysis. PLoS Med 2020;17(3):e1003062.

[18] Richardson TG, Sanderson E, Elsworth B, Tilling K, Davey Smith G. Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study. BMJ 2020;369:m1203.

[19] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015;4:7.

[20] The 1000 Genomes Project, Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature 2015;526(7571):68–74.

[21] Messner CB, Demichev V, Wendisch D, et al. Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. Cell Syst 2020;11(1):11–24e4.

[22] Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 2020;181(2):271-80e8.

[23] Jose RJ, Manuel A. COVID-19 cytokine storm: the interplay between inflammation and coagulation. Lancet Respir Med 2020;8(6):e46–7.

[24] Merad M, Martin JC. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. Nat Rev Immunol 2020;20(6):355–62.

[25] Zhang L, Yan X, Fan Q, et al. D-dimer levels on admission to predict in-hospital mortality in patients with Covid-19. J Thromb Haemost 2020;18(6):1324–9.

[26] Violi F, Pastori D, Cangemi R, Pignatelli P, Loffredo L. Hypercoagulation and antithrombotic treatment in coronavirus 2019: a new challenge. Thromb Haemost 2020;120(6):949–56.

[27] Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. Stat Med 2016;35(11):1880–906.

[28] Ellinghaus D, Degenhardt F, Bujanda L, et al. Genomewide association study of severe Covid-19 with respiratory failure. N Engl J Med 2020.

[29] Griffith GJ, Morris TT, Tudball MJ, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. Nat Commun 2020;11(1):5749.

[30] The 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(6):715–8.

[31] Johnson A.D. https://grasp.nhlbi.nih.gov/Covid19GWASResults.aspx.

[32] Lawlor DA. Commentary: two-sample Mendelian randomization: opportunities and challenges. Int J Epidemiol 2016;45(3):908–15.

[33] Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013;37(7):658–65.

[34] Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 2015;44(2):512–25.

[35] Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol 2016;40(4):304–14.

[36] Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLoS Genet 2017;13(11):e1007081.

[37] Sanderson E, Smith GD, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. Int J Epidemiol 2020.

[38] Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol 2015;181(4):251–60.

[39] Burgess S, Daniel RM, Butterworth AS, Thompson SG. Consortium EP-I. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. Int J Epidemiol 2015;44(2):484–95.

[40] Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 2015;518(7538):197–206.

[41] Hemani G, Zheng J, Elsworth B, et al. The MR-base platform supports systematic causal inference across the human phenome. Elife 2018;7.

[42] Yavorska OO, Burgess S. Mendelian randomization: an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol 2017;46(6):1734–9.

[43] Ginestet C. ggplot2 elegant graphics for data analysis. J R Stat Soc a Stat 2011;174:245.

[44] Machiela MJ, Chanock SJ. LDassoc: an online tool for interactively exploring genome-wide association study results and prioritizing variants for functional investigation. Bioinformatics 2018;34(5):887–9.

[45] Timpson NJ, Nordestgaard BG, Harbord RM, et al. C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal Mendelian randomization. Int J Obes Lond 2011;35(2):300–8.

[46] Heinrich PC, Behrmann I, Haan S, Hermanns HM, Muller-Newen G, Schaper F. Principles of interleukin (IL)-6-type cytokine signalling and its regulation. Biochem J 2003;374(Pt 1):1–20.

[47] Paternoster L, Tilling K, Davey Smith G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: conceptual and methodological challenges. PLoS Genet 2017;13(10):e1006944.

[48] Fibrinogen Studies Collaboration, Kaptoge S, White IR, et al. Associations of plasma fibrinogen levels with established cardiovascular disease risk factors, inflammatory markers, and other characteristics: individual participant meta-analysis of 154,211 adults in 31 prospective studies: the fibrinogen studies collaboration. Am J Epidemiol 2007;166(8):867–79.

[49] Zewinger S, Drechsler C, Kleber ME, et al. Serum amyloid A: high-density lipoproteins interaction and cardiovascular risk. Eur Heart J 2015;36(43):3007–16.

[50] Reynolds HY. Immunoglobulin G and its function in the human respiratory tract. Mayo Clin Proc 1988;63(2):161–74.

[51] Long QX, Liu BZ, Deng HJ, et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. Nat Med 2020;26(6):845–8.

[52] Le Bert N, Tan AT, Kunasegaran K, et al. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. Nature 2020.

[53] Kurth I, Horsten U, Pflanz S, et al. Activation of the signal transducer glycoprotein 130 by both IL-6 and IL-11 requires two distinct binding epitopes. J Immunol 1999;162(3):1480–7.

[54] Xu X, Han M, Li T, et al. Effective treatment of severe COVID-19 patients with tocilizumab. Proc Natl Acad Sci USA 2020;117(20):10970–5.

[55] Parr JB. Time to reassess tocilizumab's role in COVID-19 pneumonia. JAMA Intern Med 2020.

[56] Gupta S, Wang W, Hayek SS, et al. Association between early treatment with tocilizumab and mortality among critically ill patients with COVID-19. JAMA Intern Med 2020.

[57] Salvarani C, Dolci G, Massari M, et al. Effect of tocilizumab vs standard care on clinical worsening in patients hospitalized with COVID-19 pneumonia: a randomized clinical trial. JAMA Intern Med 2020.

[58] Hermine O, Mariette X, Tharaux PL, et al. Effect of tocilizumab vs usual care in adults hospitalized with covid-19 and moderate or severe pneumonia: a randomized clinical trial. JAMA Intern Med 2020.

[59] Bovijn J, Lindgren CM, Holmes MV. Genetic variants mimicking therapeutic inhibition of IL-6 receptor signaling and risk of COVID-19. Lancet Rheumatol 2020;2(11):e658–9.

[60] Holmes MV, Richardson TG, Ference BA, Davies NM, Davey Smith G. Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug development. Nat Rev Cardiol 2021; in press.