Circulating Proteins Influencing COVID-19 Susceptibility and Severity: a Mendelian Randomization Study

Sirui Zhou[1,2],* Guillaume Butler-Laporte[1,2],* Tomoko Nakanishi[1,3,4,5], David Morrison[1], Jonathan Afilalo[1], Marc Afilalo[1], Nofar Kimchi[1], Zaman Afrasiabi[1], Nardin Rezk[1], Meriem Bouab[1], Louis Petitjean,[1] Charlotte Guzman[1], Xioaqing Xue[1], Christos Tselios,[1] Branka Vulesevic[1], Yiheng Chen[1], Vincent Mooser,[3] Vincenzo Forgetta[1], Daniel Kaufmann,[6,7]J Brent Richards[1,2,3,8]

Affiliations:

1) Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Québec, Canada
2) Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada
3) Department of Human Genetics, McGill University Montréal, Québec, Canada
4) Kyoto-McGill International Collaborative School in Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan
5) Research Fellow, Japan Society for the Promotion of Science
6) Division of Infectious Diseases and Research Centre of the Centre Hospitalier de l'Université de Montréal, Montréal, Canada
7) Department of Medicine, Université de Montréal, Montréal, Canada
8) Department of Twin Research, King's College London, London, United Kingdom

***These authors contributed equally to this study**

Corresponding author:
Brent Richards, Professor of Medicine, McGill University
Senior Lecturer, King's College London (Honorary)
Pavilion H-413, Jewish General Hospital
3755 Côte-Ste-Catherine Montréal, Québec, Canada, H3T 1E2
T: +1 514 340 8222 x24362 F: +1 514 340 7529
E: brent.richards@mcgill.ca www.mcgill.ca/genepi

Funding: The Richards research group is supported by the Canadian Institutes of Health Research (CIHR: 365825; 409511), the Lady Davis Institute of the Jewish General Hospital, the Canadian Foundation for Innovation (CFI), the NIH Foundation, Cancer Research UK, Genome Québec, the Public Health Agency of Canada and the Fonds de Recherche Québec Santé (FRQS). TN is supported by Research Fellowships of Japan Society for the Promotion of Science (JSPS) for Young Scientists. JBR is supported by a FRQS Clinical Research Scholarship. SZ is supported by a CIHR fellowship and a FRQS fellowship. GBL is supported by the a CIHR scholarship, and a joint FRQS and Québec Ministry of Health and Social Services scholarship. Support from Calcul Québec and Compute Canada is acknowledged. TwinsUK is funded by the Welcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. These funding agencies had no role in the design, implementation or interpretation of this study. VM is supported by a Canada Excellence Research Chair. The Kaufmann lab is supported by the CIHR, NIH, CFI, AmFAR and FRQS.

Disclosures: JBR has served as an advisor to GlaxoSmithKline and Deerfield Capital.

Key words: Genome-wide association studies, COVID-19, Mendelian randomization, proteomics

Abstract (word count: 231)

Proteins detectable in peripheral blood may influence COVID-19 susceptibility or severity. However, understanding which circulating proteins are etiologically involved is difficult because their levels may be influenced by COVID-19 itself and also subject to confounding factors. To identify circulating proteins influencing COVID-19 susceptibility and severity we undertook a large-scale two-sample Mendelian randomization (MR) study, since this study design can rapidly scan hundreds of circulating proteins and reduces bias due to confounding and reverse causation. We began by identifying the genetic determinants of 955 circulating proteins in up to 10,708 SARS-CoV-2 uninfected individuals, retaining only single nucleotide polymorphisms near the gene encoded by the circulating protein. We then undertook an MR study to estimate the effect of these proteins on COVID-19 susceptibility and severity using the Host Genetics Initiative. We found that a standard deviation increase in OAS1 levels was associated with reduced COVID-19 death or ventilation (N = 2,972 cases / 284,472 controls; OR = 0.48, P = 7×10^{-8}), COVID-19 hospitalization (N = 6,492 / 1,012,809; OR = 0.60, P = 2×10^{-7}) and COVID-19 susceptibility (N = 17,607 / 1,345,334; OR = 0.81, P = 6×10^{-5}). Results were consistent despite multiple sensitivity analyses probing MR assumptions. OAS1 is an interferon-stimulated gene that promotes viral RNA degradation. Other potentially implicated proteins included IL10RB. Available medicines, such as interferon-beta-1b, increase OAS1 and could be explored for their effect on COVID-19 susceptibility and severity.

Introduction

Since its onset in late 2019, the COVID-19 pandemic has caused more than 1 million deaths worldwide, and infected over 34 million individuals.[1] Despite the scale of the epidemic, there are at present no disease-specific therapies that can reduce the morbidity and mortality of SARS-CoV-2 infection, and apart from dexamethasone therapy in oxygen dependent patients,[2] most clinical trials have shown at most mild or inconsistent benefits in disease outcome.[3–5] Further, vaccines are at least several months away, and whether they elicit protective immunity is unclear. Therefore, validated targets are needed for COVID-19 therapeutic development.

One source of such targets is circulating proteins. Recent advances in large-scale proteomics have enabled the measurement of thousands of circulating proteins at once and when combined with evidence from human genetics, such targets greatly improve the probability of drug development success.[6–8] While *de novo* drug development will take time—even in the accelerated arena of COVID-19 therapies—repositioning of currently available molecules targeting those proteins, if they exist, could also provide an accelerated opportunity to deliver new therapies to patients.

Nevertheless, since confounding and reverse causation often bias traditional circulating protein epidemiological studies, disentangling the causal relationship between circulating proteins and COVID-19 susceptibility or severity is challenging. Confounding happens when both the circulating protein and COVID-19 share a common cause, and reverse causation occurs when COVID-19 itself influences the level of a circulating protein. Even with sophisticated statistical methods, both are difficult to control for in observational studies. One way to address these limitations is by using Mendelian randomization (MR), a genetic epidemiology method that uses genetic variants as instrumental variables to test the effect of an exposure (here protein levels) on an outcome (here COVID-19 outcomes). Given that each person's genotype is essentially randomly assigned at conception, due to Mendel's second law, this greatly reduces bias due to confounding. Since genotypes are always assigned prior to disease onset, MR

studies are not influenced by reverse causation. However, MR rests on three core assumptions.[9] First, the genetic variants must be associated with the exposure of interest. Second, they do not affect the outcome, except through the exposure of interest (i.e. a lack of horizontal pleiotropy) and third, they do not associate with the confounders of the exposure-outcome relationship.

Of these, the most problematic is the second assumption—a lack of horizontal pleiotropy. One way to help avoid bias due to horizontal pleiotropy is to use genetic variants that influence circulating protein levels which are adjacent to the gene which encodes the circulating protein through the use of *cis*-protein quantitative trait loci (*cis*-pQTLs).[8] Given their close proximity to the target gene, *cis*-pQTLs are likely to influence the level of the circulating protein by directly influencing its transcription or translation, and therefore less likely to affect the outcome of interest (COVID-19) through pleiotropic pathways. The choice of *cis*-pQTLs also supports the first and third assumptions of MR, as it uses a biologically plausible pathway from the genetic variant to the exposure.

Recently developed MR methods, using two-sample MR,[10] allow for the rapid scanning of hundreds of exposures if genome-wide association studies (GWASs) have been done for both the biomarker and the outcome.[8] Nevertheless, given that MR only requires an association between the genetic instrument and the exposure, a causal association between the two may be confounded by linkage disequilibrium (LD, the non-random association of genetic variants assigned at conception). This would happen if the selected genetic variants associated with both protein levels and COVID-19 are a part of separate causal mechanisms but are also related through LD.[11] To probe this potential problem, colocalization tests can assess for the presence of bias from LD.

Understanding the etiologic role of circulating proteins in infectious diseases can be challenging because the infection itself often leads to large changes in circulating proteins, which may act against the infectious agent. Thus, it may appear that an increase in a circulating protein, such as a cytokine, is

associated with a worsened outcome, when in fact, the cytokine may be the host's response to this infection and help to mitigate this outcome. It is therefore important to identify genetic determinants of the protein levels in the non-infected state, which would reflect a person's baseline predisposition to the level of a protein. Further, circulating proteins may execute their biological roles inside the cell, but then translocate to the peripheral blood, and thus their circulating levels can, in some circumstances, offer insights into their intracellular roles.

In this study, we therefore undertook two-sample MR and colocalization analyses to combine results from large-scale GWASs of circulating protein levels and COVID-19 outcomes[12] in order to prioritize proteins likely influencing COVID-19 outcomes. We began by identifying the genetic determinants of circulating protein levels in large-scale protein level GWASs. Then, after selecting the SNPs associated with circulating proteins, which were proximal to their encoded genes (*cis*-pQTLs), we assessed whether these *cis*-pQTLs were associated with COVID-19 outcomes in the ICDA Host Genetics Initiative COVID-19 outcomes GWASs and then undertook MR analyses to estimate the effect of the circulating protein level of COVID-19 outcomes.

Results

MR using cis-pQTLs, and pleiotropy assessment

We began by obtaining the genetic determinants of circulating protein levels from six large proteomic GWAS of European individuals (Sun *et al*[13] N=3,301; Emilsson *et al*[14] N=3,200; Pietzner *et al*[15] N=10,708; Folkersen *et al*[16] N=3,394; Yao *et al*[17] N=6,861 and Suhre *et al*[18] N=997). A total of 955 proteins from these six studies had *cis*-pQTLs associated at a genome-wide significant level ($P < 5x10^{-8}$) with protein levels, or highly correlated proxies (LD $R^2 > 0.8$), in the COVID-19 Host Genetics Initiative GWASs (https://www.covid19hg.org/results/). We then undertook MR analyses using 1,388 directly matched *cis*-pQTLs and 50 proxies as genetic instruments for their associated circulating proteins on three separate COVID-19 outcomes: 1) Very severe COVID-19 disease, (defined as individuals

experiencing death, mechanical ventilation, non-invasive ventilation, high-flow oxygen, or use of extracorporeal membrane oxygenation) using 2,972 cases and 284,472 controls; 2) COVID-19 disease requiring hospitalization using 6,492 cases and 1,012,809 controls and 3) COVID-19 susceptibility using 17,607 cases and 1,345,334 controls. These case-control phenotype definitions are referred to as A2, B2, and C2 by the COVID-19 Host Genetics Initiative, respectively. In all outcomes, cases required evidence of SARS-CoV-2 infection. For the very severe COVID-19 and hospitalization outcomes, COVID-19 cases were defined as laboratory confirmed SARS-CoV-2 infection based on nucleic acid amplification or serology tests. For the COVID-19 susceptibility outcome, cases were also identified by review of health records (using International Classification of Disease codes or physician notes).

MR analyses revealed that the levels of four circulating proteins, 2'-5'-oligoadenylate synthetase 1 (OAS1), interleukin-10 receptor beta subunit (IL10RB), ABO, and liposaccharide binding protein (LBP) were associated with one or more COVID-19 outcomes after Bonferroni correction for the number of proteins tested ($P<5.2x10^{-5}$) (**Table 1, Table S1-3**). We note that Bonferroni correction is overly conservative given the non-independence of the circulating protein levels. Notably, increased OAS1 levels were strongly associated with protection from all three COVID-19 outcomes. Further, these effect sizes became more pronounced with more severe outcomes, such that each standard deviation increase in OAS1 levels was associated with a decreased odds of COVID-19 susceptibility of 0.81 (95% CI: 0.73-0.90 , $P=5.63x10^{-5}$), decreasing further for hospitalized COVID-19 (OR=0.60; 95% CI: 0.50-0.73, $P=1.69x10^{-7}$) and very severe COVID-19 (OR=0.48; 95% CI: 0.37-0.63, $P=6.77x10^{-8}$) (**Figure 1**).

We next assessed whether the *cis*-pQTL associated with OAS1 levels (rs4767027) was associated with any other phenotypes across more than 5,000 outcomes, as catalogued in PhenoScanner,[19] which catalogues associations of SNPs with outcomes from all available GWASs. We found that the OAS1 *cis*-pQTL was strongly associated with circulating OAS1 levels ($P=6.20x10^{-26}$) in serum, whereas it was not associated with any other traits or protein levels (Bonferroni-corrected $P<5.0x10^{-5}$). These findings reduce the possibility that the MR estimate of the effect of OAS1 on COVID-19 outcomes is due to horizontal

pleiotropy. Finally, except for the susceptibility outcome, the effect of rs4767027 did not demonstrate evidence of heterogeneity across COVID-19 Host Genetics Initiative GWAS meta-analyses (**Table 1).**

Next we identified an independent SNP associated with OAS1 circulating protein levels, which was not at the *OAS1* locus and is thus as *trans*-SNP (rs62143197, P value for association with OAS1 levels =7.10 x 10^{-21}). Due to strong association with many other proteins, such as annexin A2 (P=5.62 x 10^{-237}) and small ubiquitin-related modifier 3 (P=9.12 x 10^{-178}), including this *trans*-SNP could introduce bias from potential horizontal pleiotropy effects and was thus not considered in the MR analyses.

Similarly, using the *cis*-pQTL for IL10RB (rs2834167), we found that a one standard deviation increase in circulating IL10RB level was associated with a decreased odds for hospitalized COVID-19 (OR = 0.51; 95% CI: 0.37-0.70, P=3.61x10^{-5}) and very severe COVID-19 (OR=0.48; 95% CI: 0.31-0.75, P=1.08x10^{-3}). However, circulating IL10RB protein level was not associated with COVID-19 susceptibility. Using PhenoScanner, we could not find evidence of pleiotropic effects of the *cis*-pQTL for IL10RB. The IL10RB *cis*-pQTL also showed a homogeneous effect across cohorts for the three COVID-19 outcomes (**Table 1**).

We found that *cis*-pQTLs for ABO (rs505922) and LBP (rs2232613) were strongly associated with several other proteins, suggesting potential pleiotropic effects (**Table S4**). Given their known involvement in multiple physiological processes, these results were expected, but highlights that these MR analyses may suffer from significant bias from horizontal pleiotropy.

Protein-Altering Variants

cis-pQTLs may be, or may be in LD with, protein altering variants (PAVs)[13] and thus influence MR-prioritized proteins if these PAVs alter the affinity and binding of the molecules used to quantify their level. We thus assessed if the *cis*-pQTLs for the MR-prioritized proteins were PAVs, or in LD (R^2>0.8) with

PAVs, and if so, conditional analyses for *cis*-pQTLs, conditioning on their PAVs in LD, were undertaken to test if the *cis*-pQTL remained associated with the protein level. rs2834167 (IL10RB) and rs2232613 (LBP) are nonsense and missense variants, respectively, therefore both could be subject to potential binding effects. rs4767027 (OAS1) is an intronic variant, which is in LD with missense variants, however, the missense variants in LD with rs4767027 were absent in the original imputed dataset from Sun *et al*, therefore, no assessment for aptamer binding effects could be made. rs505922 (ABO) is not in LD with known missense variants.

Colocalization Studies

To test whether confounding due to LD may have influenced the estimated effect of circulating OAS1 on the three different COVID-19 outcomes, we tested the probability that the genetic determinants of OAS1 circulating protein level were shared with the three COVID-19 outcomes using colocalization analyses. These were performed using *coloc*, a Bayesian statistical test implemented in the *coloc* R package.[11] We found that the posterior probability that OAS1 levels and COVID-19 outcomes shared a causal signal (the posterior probability for hypothesis 4 in *coloc*, PP4) in the 1Mb locus around the cis-pQTL rs4767027 was 0.75 for very severe COVID-19 (**Figure 2**), 0.72 for hospitalization due to COVID-19 (**Figure 3**), and 0.88 for COVID-19 susceptibility (**Figure 4**). This suggests that there is likely one shared causal signal for OAS1 circulating protein levels and COVID-19 outcomes.

Colocalization of circulating LBP levels and the three COVID-19 outcomes showed strong colocalization between LBP levels and hospitalization due to COVID-19 (PP4 = 0.98, **Figure S2**), but lack of colocalization between LBP level and COVID-19 susceptibility (PP4 = 0.48, **Figure S1**) and very severe COVID-19 (PP4 = 0.07, **Figure S3**). Colocalization of ABO levels and different COVID-19 outcomes showed no colocalization between ABO level and different COVID-19 outcomes (posterior probability of single shared signal = 0.03, 0.02 and 0.005 for ABO level and COVID-19 susceptibility, hospitalization due to COVID-19 and very severe COVID-19, respectively). This was likely due to the multiple

independent causal signals at the ABO locus that could not be assessed by *coloc* (**Figures S4-6)**. We were unable to perform colocalization analysis for IL10RB due to a lack of genome-wide summary level data from the original proteomic GWAS.

eQTL studies

We next assessed whether the *cis*-pQTL, rs4767027 for circulating OAS1 levels, influenced expression of OAS1 levels across different tissues using the data from the GTEX consortium.[20] We found that rs4767027 was an eQTL for *OAS1* in whole blood ($P=2.3 \times 10^{-173}$), skin ($P=4.2 \times 10^{-149}$) , adipose ($P=1.4 \times 10^{-147}$) and, critically, in lung ($P=4.5 \times 10^{-139}$). It is also an eQTL for the nearby genes *OAS3* ($P=1.1 \times 10^{-78}$) and *OAS2* ($P=6.9 \times 10^{-21}$) in cultured fibroblasts, spleen, pancreas and whole blood. *OAS1*, *OAS2* and *OAS3* all encode enzymes in the same 2', 5' oligoadenylate synthase family that bind to and activate RNase L, leading to viral RNA degradation and inhibit viral replication.[21] Investigating IL10RB, we found that its *cis*-pQTL is an eQTL for *IL10RB* in the brain ($P=7.9 \times 10^{-34}$), heart ($P=4.9 \times 10^{-22}$) and lung ($P=1.2 \times 10^{-13}$).[20] **Figure 5** summarizes data supporting the role of OAS1 in COVID-19 outcomes.

Discussion:

Disease-specific therapies are needed to reduce the morbidity and mortality associated with COVID-19 outcomes. In this large-scale two-sample MR study of 955 proteins assessed for three COVID-19 outcomes in up to 17,600 cases and 1.3 million controls, we provide evidence that increased OAS1 and IL10RB levels are strongly associated with reduced COVID-19 susceptibility, hospitalization and severe outcomes. For OAS1 the protective effect size was particularly large, such that a 50% decrease in the odds of very severe COVID-19 was observed per standard deviation increase in OAS1 circulating levels. Since therapies exist that activate OAS1, repositioning them as potential COVID-19 treatments should be prioritized.

Each of the four MR-identified proteins has orthogonal evidence implicating their role in viral infections. For example, the OAS gene cluster on chromosome 12 includes *OAS1*, *OAS2* and *OAS3,* which share significant homology and differ only in their number of OAS units. These genes all encode similar enzymes that are essential to host responses to viral infections. They are induced by interferons and activate latent RNase L, resulting in direct viral and endogenous RNA destruction, as demonstrated in *in-vitro* studies.[22] They also increase expression of both IRF3 and IRF7, both genes involved in interferon gene expression. As an interferon stimulated gene[23], *OAS1* polymorphisms have been involved in host immune response to several classes of viral infection including influenza[24], herpes simplex[25], hepatitis C , West Nile[26] Dengue[27], and SARS-CoV[28] viruses. Given that OAS1 is an intracellular enzyme leading to viral RNA degradation, it is probable that the circulating levels of this enzyme reflect intracellular levels of this protein. While further experiments are required to better understand the role of OAS1 in SARS-CoV-2 infection, it is likely that the site of action of this enzyme upon the virus would be in the intracellular compartment.

Molecules currently exist which can increase OAS1 activity. Interferon beta-1b, which is an OAS1 agonist,[29] is currently used to treat multiple sclerosis and has been shown to induce OAS1 expression in

blood.[30] A recent randomized controlled trial showed that interferon beta-1b combined with lopinavir-ritonavir reduced mortality in MERS-CoV infections, another human coronavirus.[31] While lopinavir-ritonavir, rather than interferon beta-1b, may have been responsible for this mortality reduction a recent COVID-19 randomized controlled trial from the RECOVERY group showed no benefit from lopinavir-ritonavir in multiple COVID-19 outcomes.[32] Indeed, according to clinicaltrials.gov, there are multiple on-going phase II randomized controlled trials testing interferon beta-1b for COVID-19 outcomes. *In-vitro* evidence also exists demonstrating that pharmacological inhibition of phosphodiesterase-12, which normally degrades the OAS enzymes, potentiates this OAS-mediated antiviral activity.[21] PDE-12 inhibitors potentiate the action of OAS1, 2 and 3.[33] Interestingly other coronaviruses in the same betacoronavirus family as SARS-CoV-2 have been shown to produce viral proteins that degrade the OAS family of proteins, and antagonize RNase-L activity, leading to evasion of the host immune response.[34,35] Thus classes of medications currently exist that lead to increased OAS1 levels and could be explored for their effect upon COVID-19 outcomes.

Importantly OAS2 and OAS3 circulating levels were not assessed by any of the protein level GWASs and the lead *cis*-pQTL for OAS1 at this locus likely influences OAS2 and OAS3 expression levels. Thus, it remains possible that, in addition to influencing OAS1 levels, the lead *cis*-pQTL may also influence the abundance of OAS2 and OAS3. Nevertheless, all three genes share similar functions, structure and activity, highlighting that increases in the activity, or level of this family of proteins are likely to protect against COVID-19 susceptibility and severity.[36]

IL10RB encodes for the beta subunit of the IL10 receptor, and is part of a cluster of immunologically important genes including *IFNAR1* and *IFNAR2*, both recently implicated in severe COVID-19 pathophysiology.[37] IFNAR1 and 2 encode the interferon alpha/beta receptor subunits 1 and 2, respectively. Interestingly, while there exists a *cis*-pQTL strongly associated with IFNAR1 levels, it was not associated with any of the COVID-19 outcomes (P ~ 0.5). Further, *IFNAR1* had no trans-pQTLs identified, which means that the *IL10RB cis*-pQTL does not likely reflect IFNAR1 levels. However, since

IFNAR2 was not measured in any proteomic studies, we could not test the effect of its circulating levels on COVID-19 outcomes. IL10RB mediates IL10 anti-inflammatory activity through its downstream inhibitory effect on many well-known pro-inflammatory cytokines such as janus kinases and STAT1.[38] While overexpression of IL10 has been involved in the persistence of multiple chronic bacterial infections such as tuberculosis,[39] its role remains poorly understood in acute infections. In sepsis, a disease state characterized by high levels of cytokine activity and a rise in multiple biomarkers associated with inflammation, there is also a well-established paradoxical increase in anti-inflammatory IL10 production by leukocytes, especially in the early stage of the disease.[40] Most importantly, while in a normal physiological state, IL10 is usually only produced at a low level by neutrophils, its production is strongly upregulated by IL4, itself upregulated by lipopolysaccharides (LPS) when they bind LBPs.[41,42] While LPS's are well-known for their role in triggering gram-negative bacterial sepsis, their role in other acute infections and respiratory diseases is likely broader, and involves complex sequences of cytokine signaling.[43–46] Nevertheless, as our MR studies showed that LBP and IL10RB protein levels affected COVID-19 outcome with a concordant effect direction, and given the known role of overt inflammation in COVID-19 morbidity, this pathway likely deserves more investigation.

This study has limitations. First, we used MR to test the effect of circulating protein levels measured in a non-infected state. This is because the effect of the *cis*-pQTLs upon circulating proteins was estimated in individuals who had not been exposed to SARS-CoV-2. Once a person contracts SARS-CoV-2 infection, levels of circulating proteins could be altered and this may be especially relevant for cytokines such as IL10 (which binds to IL10RB), whose levels may reflect host response to the viral infection. Thus, the results presented in this paper should be interpreted as an estimation of the effect of circulating protein levels, when measured prior to infection. On-going studies will help to clarify if the same *cis*-pQTLs influence circulating protein levels during infection. As emphasized above, these circulating protein levels may also reflect intracellular biological processes, such as may be the case for OAS1. Second, this study suffers a high false-negative rate. Our goal was not to identify *every* circulating protein influencing COVID-19 outcomes, but rather to provide evidence for few proteins with strong *cis*-pQTLs since these

proteins are more likely to be robust to the assumptions of MR studies. Future large-scale proteomic studies with more circulating proteins properly assayed should help to overcome these limitations. Third, the colocalization analyses in this study are also subject to potential bias from mixed ethnicity from the COVID-19 outcome GWAS. Previous studies have already established statistical colocalization methods may be overly sensitive to the accurate estimation of LD structure, and thus fail to colocalize[47] due to small differences in LD, or the presence of more than one causal signal at a locus. Nevertheless, the locus zoom plots (**Figures 2-4**) show similar signals arising from the same genomic regions for the protein and COVID-19 outcomes. Fourth, most MR studies assume a linear relationship between the exposure and the outcome. Thus, our findings would not identify proteins whose effect upon COVID-19 outcomes has a clear threshold effect.

In conclusion, we have used genetic determinants of circulating protein levels and COVID-19 outcomes obtained from large-scale studies and found compelling evidence that OAS1 has a protective effect on COVID-19 susceptibility and severity. Known pharmacological agents that increase OAS1 levels (such as interferon beta-1b and PDE12 inhibitors) could be explored for their effect on COVID-19 outcomes.

Methods:

Cohorts

pQTL GWAS

We systematically identified pQTL associations from six large proteomic GWASs.[13–18] Each of these studies undertook proteomic profiling using either SomaLogic somamer technology, or O-link proximal extension assays.

COVID GWAS and COVID-19 Outcomes

To assess the association of *cis*-pQTLs with COVID-19 outcomes, we used the largest COVID-19 meta-analytic GWAS to date from the COVID-19 Host Genetics Initiative (https://www.covid19hg.org/results/). The participating cohorts provided GWAS summary statistics data from either large-scale biobanks (N = 18), hospital-based cohorts (N = 13), or direct-to-consumer genotyping companies (N = 1), which were meta-analyzed across ancestries and made publicly available. For our study, we used three of these GWAS meta-analyses, based on sample size and clinical relevance. These outcomes were very severe COVID-19, hospitalized patients with COVID-19, and susceptibility to COVID-19 (named A2, B2, and C2, respectively in the COVID-19 Host Genetics Initiative).

Very severe COVID-19 cases were defined as hospitalized individuals with COVID-19 as the primary reason for hospital admission with laboratory confirmed SARS-CoV-2 infection (nucleic acid amplification tests or serology based), and death or respiratory support (intubation, continuous positive airway pressure, Bilevel Positive Airway Pressure, or continuous external negative pressure, Optiflow/very high flow Positive End Expiratory Pressure Oxygen). Simple supplementary oxygen (e.g. 2 liters/minute via nasal cannula) did not qualify for case status. Controls were all individuals in the participating cohorts who did not meet this case definition.

Hospitalized COVID-19 cases were defined as individuals hospitalized with laboratory confirmed SARS-CoV-2 infection (using the same microbiology methods as for the very severe phenotype), where hospitalization was due to COVID-19 related symptoms. Controls were all individuals in the participating cohorts who did not meet this case definition.

Susceptibility to COVID-19 cases were defined as individuals with laboratory confirmed SARS-CoV-2 infection, health record evidence of COVID-10 (international classification of disease coding or physician confirmation), or with self-reported infections (e.g. by questionnaire). Controls were all individuals in the participating cohorts who did not meet this case definition.

Two-sample Mendelian randomization

We used two-sample MR analyses to screen and test potential circulating proteins for their role influencing COVID-19 outcomes. In two-sample MR, the effect of SNPs on the exposure and outcome are taken from separate GWASs. This method often improves statistical power, because it allows for larger sample sizes for the exposure and outcome GWAS.[48]

Exposure definitions: We conducted MR using six large proteomic GWAS studies.[13–18] Circulating proteins from Sun *et al*, Emilsson *et al* and Pietzner *et al* were measured on the SOMAlogic platform, Suhre *et al*, Yao *et al* and Folkersen *et al* used protein measurements on the O-link platform. We selected proteins with only *cis-pQTL*s to test their effects on COVID-19 outcomes, because they are less likely to be affected by potential horizontal pleiotropy. The *cis-pQTL*s were defined as the genome-wide significant SNPs ($P < 5 \times 10^{-8}$) with the lowest P value within 1 Mb of the transcription start site (TSS) of the gene encoding the measured protein.[8] We selected one *cis-pQTL* per protein per each study. We included the same proteins represented by different *cis*-pQTLs from different studies in order to cross examine the findings. For *cis*-pQTLs that were not present in the COVID-19 GWAS, SNPs with LD R^2>0.8 and with minor allele frequency (MAF) < 0.42 were selected as proxies, but MAF > 0.3 was used for palindromic

alignment for proxy SNPs. *cis*-pQTLs with palindromic effects and with minor allele frequency (MAF) > 0.42 were removed prior to MR to prevent allele-mismatches. Bonferroni correction was used to control for the total number of proteins tested using MR. We recognize that this is an overly conservative correction, given the non-independence of the circulating proteins, but such stringency should reduce false positive associations. MR analyses were performed using the TwoSampleMR package in R,[49] using Wald ratio to estimate the effect of each circulating protein on each of the three COVID-19 outcomes. After matching of the *cis*-pQTLs of proteins with COVID-19 GWAS and the removal of palindromic SNPs, a total of 537 proteins (543 directly matched IVs and 22 proxies) from Sun *et al*, 750 proteins (731 directly matched IVs and 20 proxies) from Emilsson *et al*, 94 proteins (85 directly matched IVs and 7 proxies) from Pietzner *et al*, 74 proteins (72 directly matched IVs) from Suhre *et al*, 24 proteins (24 directly matched IVs) from Yao *et al* and 13 proteins (12 directly matched IVs and 1 proxy) from Folkersen *et al* were used as instruments for the MR analyses across the three COVID-19 outcomes (**Table S5-7**).[13–18]

Pleiotropy assessments

A common pitfall of MR is horizontal pleiotropy, which occurs when the genetic variant affects the outcome via pathways independent of circulating proteins. The use of circulating protein *cis-pQTL*s greatly reduces the possibility of pleiotropy, for reasons described above. We also searched in the PhenoScanner database, a large catalogue of observed SNP-outcome relationships involving > 5,000 GWAS done to date to assess potentially pleiotropic effects of the *cis*-pQTLs of MR prioritized proteins, by testing the association of *cis*-pQTLs with other circulating proteins (i.e. if they were *trans*-pQTLs to other proteins or traits). For *cis*-pQTLs of MR prioritized proteins, if they were measured on SOMAlogic platform, we assessed the possibility of potential aptamer-binding effects (where the presence of protein altering variants may affect protein measurements). We also checked if *cis*- pQTLs of MR prioritized proteins had significantly heterogeneous associations across COVID-19 populations in each COVID-19 outcome GWAS.

Colocalization analysis

Finally, we tested colocalization of the genetic signal for the circulating protein and each of the three COVID-19 outcomes using colocalization analyses, which assess potential confounding by LD. Specifically, for each of these MR significant proteins with genome-wide summary data available, for the proteomic GWASs, a stringent Bayesian analysis was implemented in *coloc* R package to analyze all single nucleotide variants (SNV) with MAF > 0.01 in 1MB genomic locus centered on the *cis*-pQTL. Colocalizations with posterior probability for hypothesis 4 (PP4, that there is an association for both protein level and COVID-19 status and they are driven by the same causal variant) > 0.5 were considered likely to colocalize (which means the highest posterior probability for all 5 *coloc* hypotheses), and PP4 > 0.8 was considered to be highly likely to colocalize.

Ethics declarations:

All cohorts contributing data to this study received ethics approval from their respective ethics review board.

Data availability:

Data from proteomics studies are available from the referenced peer-reviewed studies or their corresponding authors, as applicable. Summary statistics for the COVID-19 outcomes are publicly available for download on the COVID-19 Host Genetics Initiative website (www.covid19hg.org).

Author contributions:

Conception and design: SZ, GBL, TN and JBR. Data analyses: SZ and TN. Manuscript writing: SZ, GBL, TN and JBR. Data acquisition: TN, GBL, DM, DEK, JA, MA, NK, ZA, NR, MB, CG, XX, CT, BV, VF and JBR. Interpretation of data: SZ, GBL, TN, YC, DEK, VF and JBR. Intellectual contribution to the manuscript: SZ, GBL, TN, YC, VF, VM, DEK and JBR. All authors were involved in preparation of the further draft of the manuscript and revising it critically for content. All authors gave final approval of the version to be published. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Table 1. MR-Identified Circulating Protein Levels Effecting COVID-19 Outcomes

Protein	*cis*-pQTL	Source	**Very Severe COVID-19** Case N=2,972, Control N =284,472				**COVID-19 Hospitalization** Case N=6,492, Control N =1,012,809				**COVID-19 susceptibility** Case N=17,607, Control N =1,345,334			
			OR	*95%CI*	*P value*	*P het*	*OR*	*95%CI*	*P value*	*P het*	*OR*	*95%CI*	*P value*	*P het*
OAS1	rs4767027	Sun	0.48	0.37-0.63	6.8×10^{-8}	0.63	0.60	0.50- 0.73	1.7×10^{-7}	0.23	0.81	0.73-0.90	5.6×10^{-5}	3.1×10^{-3}
ABO	rs505922	Sun, Emilsson	1.06	1.01-1.11	0.03	0.17	1.10	1.06-1.14	1.3×10^{-6}	0.08	1.07	1.05-1.09	3.4×10^{-10}	0.18
IL10RB	rs2834167	Emilsson	0.48	0.31-0.75	1.0×10^{-3}	0.01	0.51	0.37-0.70	3.6×10^{-5}	0.32	0.88	0.73-1.05	0.15	0.02
LBP	rs2232613	Pietzner, Emilsson	0.85	0.77-0.93	5.5×10^{-4}	0.83	0.86	0.80-0.92	3.5×10^{-5}	0.94	0.94	0.91-0.98	2.4×10^{-3}	0.23

OR: represents the estimated effect of a standard deviation (for Sun *et al* and Pietzner *et al* studies) or one unit (for the Emilsson *et al* study) increase in protein levels on the odds of the three COVID-19 outcomes. P het: P value of heterogeneity for each *cis*-pQTLs across the COVID-19 cohort in the GWAS meta-analysis from the Host Genetics Initiative.

Figure 1. Association of Circulating Protein Levels with COVID-19 Outcomes

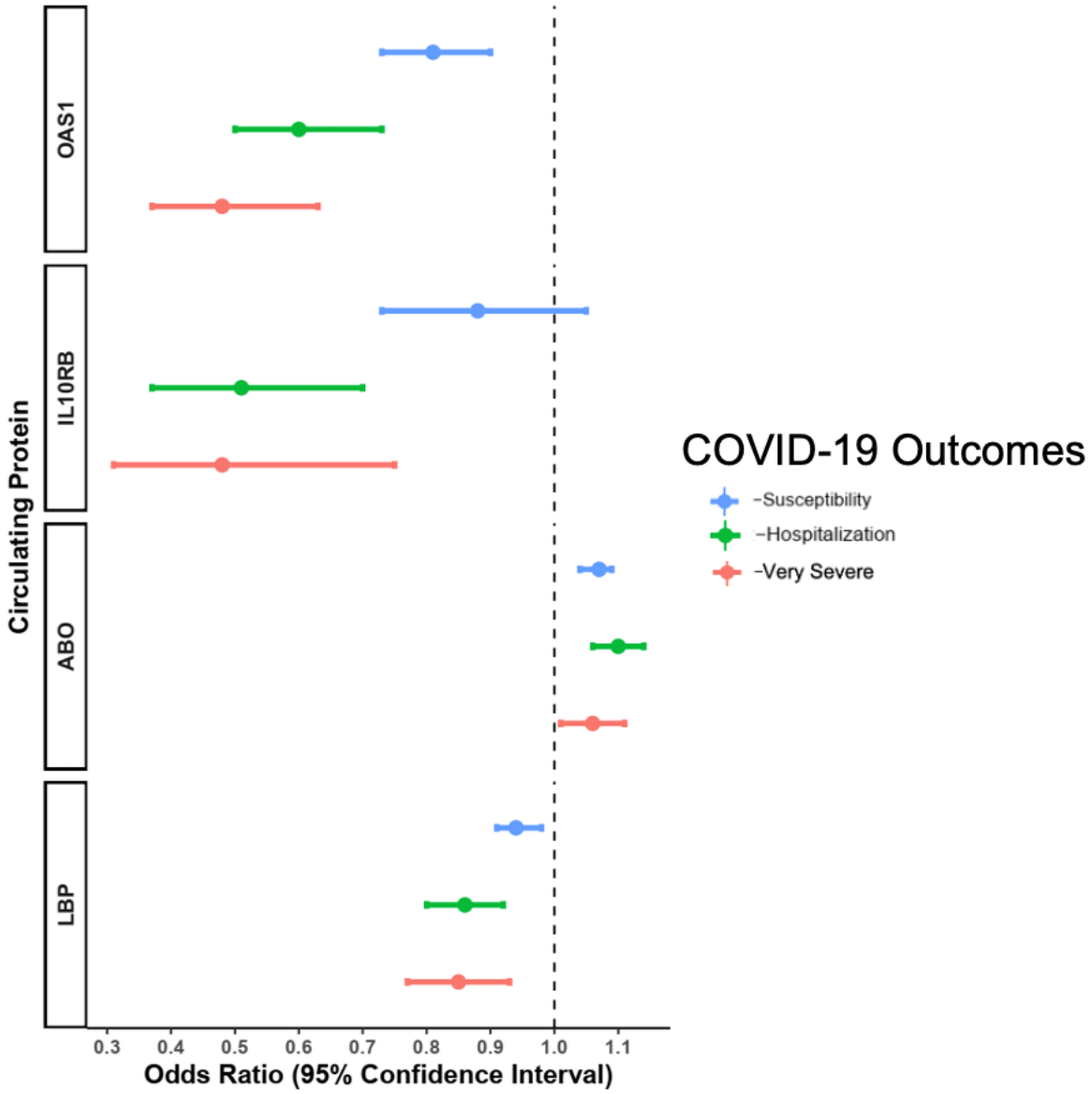
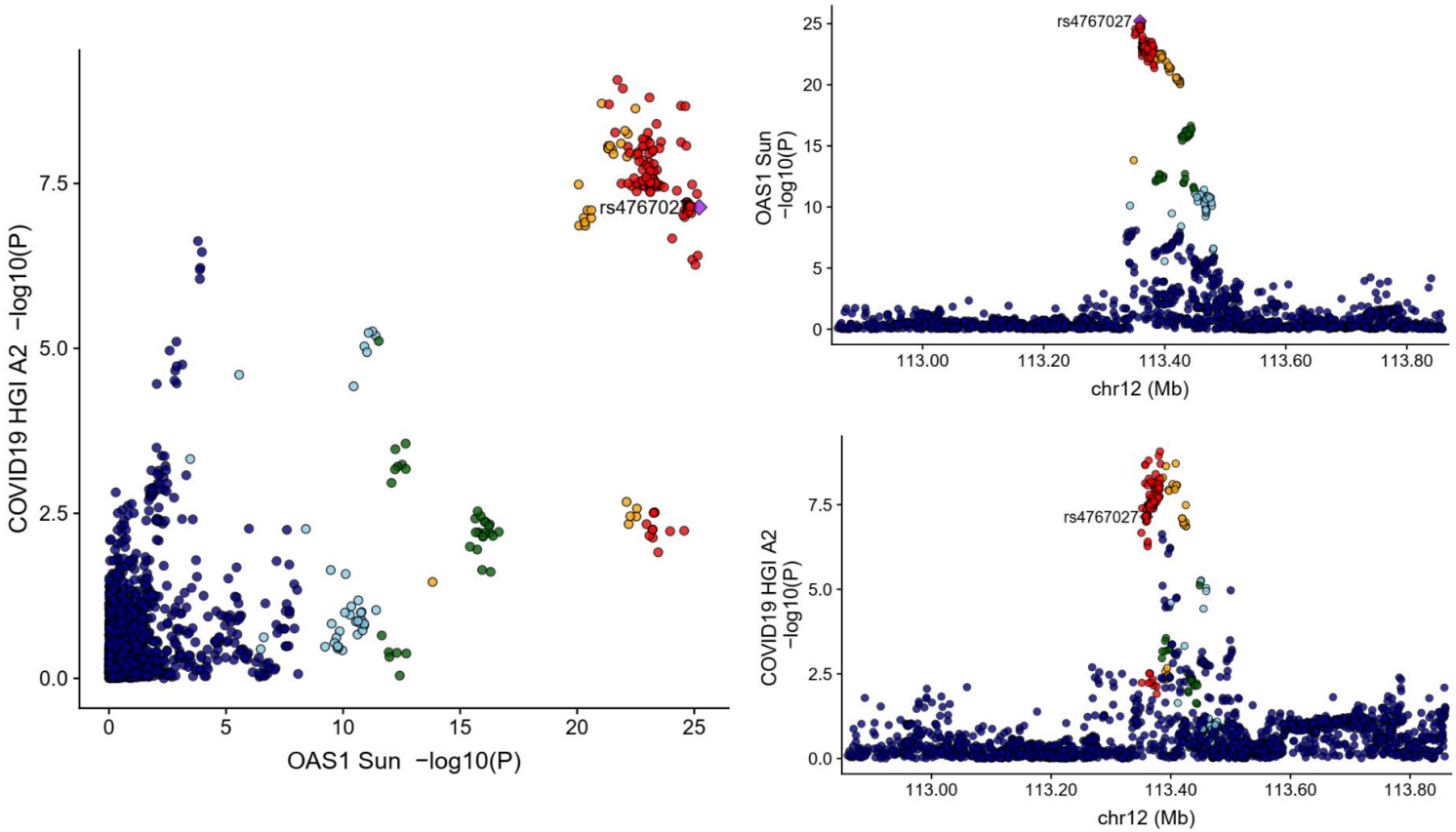
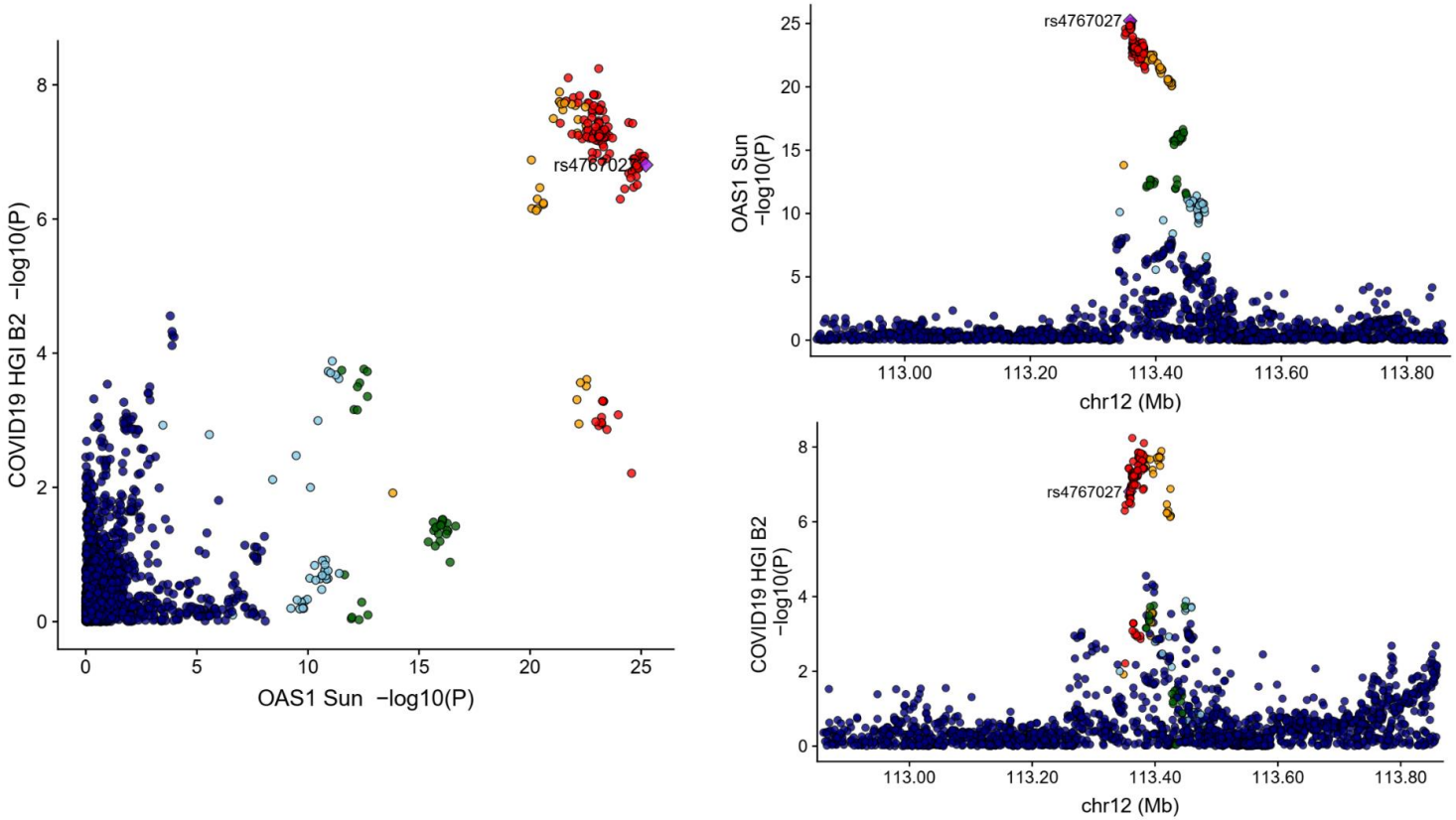


Figure 2. Colocalization of the Genetic Determinants of OAS1 Level and Very Severe COVID-19



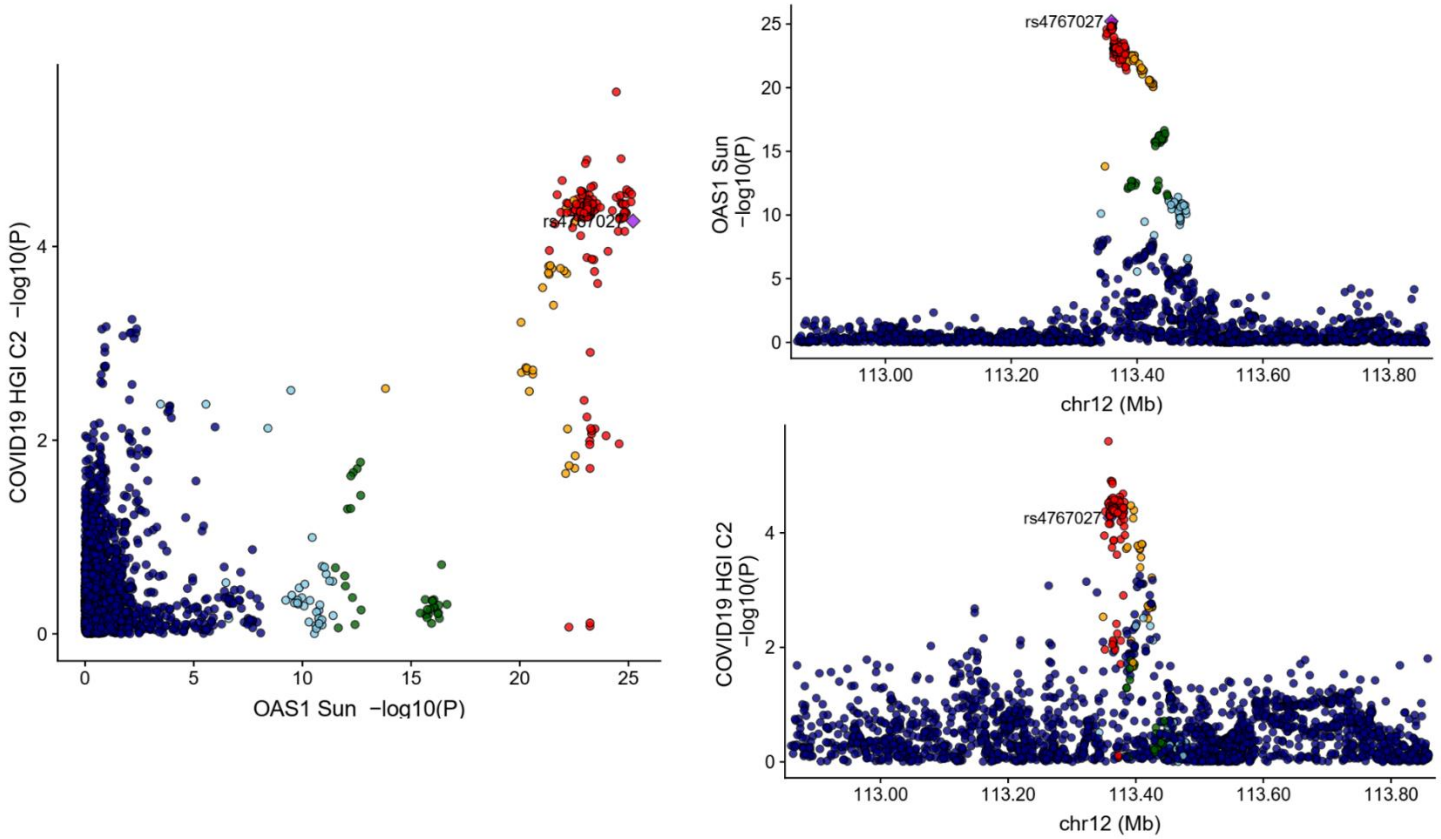
Created using locuscomparer R package,[50] the *cis*-pQTL rs4767027 is shown as a purple diamond. The GWAS used for OAS1 level was Sun et al. LD is based on the 1KG European panel, red dots represent SNPs having LD R2>0.8 with the *cis*-pQTL. Yellow dots represent SNPs having LD R2>0.6 with the *cis*-pQTL. *Coloc* results were PP4=0.75; PP3=0.25.

Figure 3. Colocalization of the Genetic Determinants of OAS1 Level and Hospitalized COVID-19



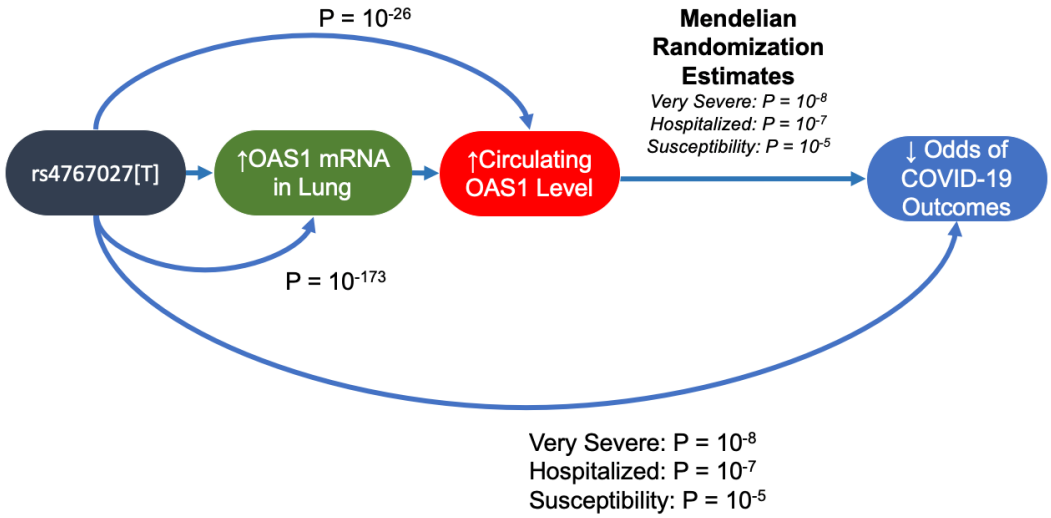
Created using locuscomparer R package, the *cis*-pQTL rs4767027 is shown as a purple diamond. The GWAS used for OAS1 level was Sun et al. LD is based on the 1KG European panel, red dots represent SNPs having LD R2>0.8 with the *cis*-pQTL. Yellow dots represent SNPs having LD R2>0.6 with the *cis*-pQTL. *Coloc* results were PP4=0.72; PP3=0.28.

Figure 4. Colocalization of the Genetic Determinants of OAS1 Level and COVID-19 Susceptibility



Created using locuscomparer R package, the *cis*-pQTL rs4767027 is shown as a purple diamond. LD is based on the 1KG European panel, red dots represent SNPs having LD R2>0.8 with the *cis*-pQTL. Yellow dots represent SNPs having LD R2>0.6 with the *cis*-pQTL. *Coloc* results were PP4=0.88; PP3=0.10.

Figure 5. Summary of Evidence Supporting a role for OAS1 in COVID-19 Outcomes.



Note that this Figure is not intended to be a directed acyclic graph.

References

1. Johns Hopkins. Corona Virus Resource Center. (2020).

2. Dexamethasone in Hospitalized Patients with Covid-19 — Preliminary Report. *N. Engl. J. Med.* 1–11 (2020). doi:10.1056/nejmoa2021436

3. Sterne, J. A. C. *et al.* Association between administration of systemic corticosteroids and mortality among critically ill patients with COVID-19: A meta-analysis. *JAMA - J. Am. Med. Assoc.* (2020). doi:10.1001/jama.2020.17023

4. Beigel, J. H. *et al.* Remdesivir for the Treatment of Covid-19 — Preliminary Report. *N. Engl. J. Med.* 2019–2020 (2020). doi:10.1056/nejmoa2007764

5. Cavalcanti, A. B. *et al.* Hydroxychloroquine with or without Azithromycin in Mild-to-Moderate Covid-19. *N. Engl. J. Med.* (2020). doi:10.1056/nejmoa2019014

6. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).

7. Cook, D. *et al.* Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov.* **13**, 419–31 (2014).

8. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* (2020). doi:10.1038/s41588-020-0682-6

9. Davey Smith, G., Ebrahim, S., Smith, G. D. & Ebrahim, S. *'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? International Journal of Epidemiology* **32**, 1–22 (2003).

10. Dastani, Z. *et al.* Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet.* **8**, e1002607 (2012).

11. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* **10**, (2014).

12. 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.* **28**, 715–718 (2020).

13. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).

14. Emilsson, V. *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science (80-.).* **1327**, 1–12 (2018).

15. Pietzner, M. *et al.* Genetic architecture of host proteins interacting with SARS-CoV-2. *bioRxiv Prepr. Serv. Biol.* (2020). doi:10.1101/2020.07.01.182709

16. Folkersen, L. *et al.* Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genet.* **13**, e1006706 (2017).

17. Yao, C. *et al.* Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nat. Commun.* **9**, 3268 (2018).

18. Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357 (2017).

19. Staley, J. R. *et al.* PhenoScanner: a database of human genotype–phenotype associations. *Bioinformatics* **32**, 3207–3209 (2016).

20. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

21. Wood, E. R. *et al.* The role of phosphodiesterase 12 (PDE12) as a negative regulator of the innate immune response and the discovery of antiviral inhibitors. *J. Biol. Chem.* **290**, 19681–19696 (2015).

22. Kristiansen, H. *et al.* Extracellular 2′-5′ Oligoadenylate Synthetase Stimulates RNase L-Independent Antiviral Activity: a Novel Mechanism of Virus-Induced Innate Immunity. *J. Virol.* **84**, 11898–11904 (2010).

23. Schneider, W. M., Chevillotte, M. D. & Rice, C. M. Interferon-Stimulated Genes: A Complex Web of Host Defenses. *Annu. Rev. Immunol.* **32**, 513–545 (2014).

24. Min, J.-Y. & Krug, R. M. The primary function of RNA binding by the influenza A virus NS1 protein in infected cells: Inhibiting the 2'-5' oligo (A) synthetase/RNase L pathway. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7100–7105 (2006).

25. Hu, B. *et al.* Cellular responses to HSV-1 infection are linked to specific types of alterations in the host transcriptome. *Sci. Rep.* **6**, 28075 (2016).

26. Lim, J. K. *et al.* Genetic Variation in OAS1 Is a Risk Factor for Initial Infection with West Nile Virus in Man. *PLoS Pathog.* **5**, e1000321 (2009).

27. Simon-Loriere, E. *et al.* High Anti–Dengue Virus Activity of the OAS Gene Family Is Associated With Increased Severity of Dengue. *J. Infect. Dis.* **212**, 2011–2020 (2015).

28. Hamano, E. *et al.* Polymorphisms of interferon-inducible genes OAS-1 and MxA associated with SARS in the Vietnamese population. *Biochem. Biophys. Res. Commun.* **329**, 1234–1239 (2005).

29. Cheng, G. *et al.* Pharmacologic Activation of the Innate Immune System to Prevent Respiratory Viral Infections. *Am. J. Respir. Cell Mol. Biol.* **45**, 480–488 (2011).

30. Harari, D., Orr, I., Rotkopf, R., Baranzini, S. E. & Schreiber, G. A robust type I interferon gene signature from blood RNA defines quantitative but not qualitative differences between three major IFNβ drugs in the treatment of multiple sclerosis. *Hum. Mol. Genet.* **24**, 3192–3205 (2014).

31. Arabi, Y. M. *et al.* Interferon Beta-1b and Lopinavir–Ritonavir for Middle East Respiratory Syndrome. *N. Engl. J. Med.* NEJMoa2015294 (2020). doi:10.1056/NEJMoa2015294

32. Horby, P. W. *et al.* Lopinavir–ritonavir in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet* (2020). doi:10.1016/S0140-6736(20)32013-4

33. Poulsen, J. B., Kjær, K. H., Justesen, J. & Martensen, P. M. Enzyme assays for synthesis and degradation of 2-5As and other 2′-5′ oligonucleotides. *BMC Biochem.* **16**, 15 (2015).

34. Zhao, L. *et al.* Antagonism of the interferon-induced OAS-RNase L pathway by murine coronavirus ns2 protein is required for virus replication and liver pathology. *Cell Host*

Microbe (2012). doi:10.1016/j.chom.2012.04.011

35. Zhang, R. *et al.* Homologous 2',5'-phosphodiesterases from disparate RNA viruses antagonize antiviral innate immunity. *Proc. Natl. Acad. Sci. U. S. A.* (2013). doi:10.1073/pnas.1306917110

36. Lee, W.-B. *et al.* OAS1 and OAS3 negatively regulate the expression of chemokines and interferon-responsive genes in human macrophages. *BMB Rep.* **52**, 133–138 (2019).

37. Zhang, Q. *et al.* Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science (80-.).* **5**, eabd4570 (2020).

38. Verma, R. *et al.* A network map of Interleukin-10 signaling pathway. *J. Cell Commun. Signal.* (2016). doi:10.1007/s12079-015-0302-x

39. Moreira-Teixeira, L. *et al.* T Cell–Derived IL-10 Impairs Host Resistance to Mycobacterium tuberculosis Infection . *J. Immunol.* (2017). doi:10.4049/jimmunol.1601340

40. Kasten, K. R., Muenzer, J. T. & Caldwell, C. C. Neutrophils are significant producers of IL-10 during sepsis. *Biochem. Biophys. Res. Commun.* (2010). doi:10.1016/j.bbrc.2010.01.066

41. Lewkowicz, N. *et al.* Induction of human IL-10-producing neutrophils by LPS-stimulated Treg cells and IL-10. *Mucosal Immunol.* (2016). doi:10.1038/mi.2015.66

42. Crepaldi, L. *et al.* Up-Regulation of IL-10R1 Expression Is Required to Render Human Neutrophils Fully Responsive to IL-10. *J. Immunol.* (2001). doi:10.4049/jimmunol.167.4.2312

43. Iyer, S. S., Ghaffari, A. A. & Cheng, G. Lipopolysaccharide-Mediated IL-10 Transcriptional Regulation Requires Sequential Induction of Type I IFNs and IL-27 in Macrophages. *J. Immunol.* (2010). doi:10.4049/jimmunol.1002041

44. Ray, A., Chakraborty, K. & Ray, P. Immunosuppressive MDSCS induced by TLR signaling during infection and role in resolution of inflammation. *Frontiers in Cellular and Infection Microbiology* (2013). doi:10.3389/fcimb.2013.00052

45. Petruk, G. *et al.* SARS-CoV-2 Spike protein binds to bacterial lipopolysaccharide and boosts proinflammatory activity. *bioRxiv* (2020).

46. The crucial role of IL-10 in the anti-inflammatory effects of aerobic exercise in a model LPS-induced ARDS. *Eur. Respir. J.* (2013).

47. Kanduri, C., Bock, C., Gundersen, S., Hovig, E. & Sandve, G. K. Colocalization analyses of genomic elements: approaches, recommendations and challenges. *Bioinformatics* **35**, 1615–1624 (2018).

48. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet. Epidemiol.* 1–12 (2016). doi:10.1002/gepi.21998

49. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, e34408 (2018).

50. Liu, B., Gloudemans, M. J., Rao, A. S., Ingelsson, E. & Montgomery, S. B. Abundant associations with gene expression complicate GWAS follow-up. *Nat. Genet.* **51**, 768–769

(2019).